Governments in the spotlight?

On the use and impacts of freedom of information laws and proactive publication of government data

A thesis submitted by
Mária Žuffová
for the degree
PhD in Politics

School of Government and Public Policy
Humanities & Social Sciences Faculty
University of Strathclyde
Glasgow
2020

Acknowledgements

My interest in government information goes back to 2009 when I started my master's studies at CEU and can be entirely credited to Kristina Irion and Péter Molnár. If I did not have such wonderful professors and mentors, I would be probably doing something else today. Péter and Kristina have both a great talent, each in their own way, to talk about information and data rights and the role of information for democracy in an enthusiastic manner that incites further curiosity. Kristina was also the one who encouraged me to pursue PhD studies. Without her support, I might not have made the decision. She has always trusted in my abilities and pushed me to do a step further. Thank you for that!

Doing a PhD at Strathclyde was a great experience, thanks to many people. First and foremost, yet again, I was fortunate to have excellent and compassionate supervisors. I wish to express my gratitude to Heinz Brandenburg, my primary supervisor for his encouragement, support and a sense of humour when problems (some of them fairly non-standard) occurred. His advice was indispensable in many stages and helped me to produce a more thorough piece of research. I would also like to thank Zac Greene, my second supervisor for having his door always open to advise me with quantitative text analysis or publishing in journals and sending me tons of interesting jobs ads. Heinz and Zac have supported me in many ways and pushed me to be more ambitious. I had fantastic role models in them, and I hope I will once become as kind and approachable mentor to others as they have been to me. I always looked forward to our supervisory meetings, which I guess says it all. I am also grateful for the opportunity to discuss my ideas with Lilian Edwards in the initial stages of my PhD. At this point, I would like to thank my examiners, Gina Yannitell Reinhardt and Scott Cunningham, for their excellent comments and thorough feedback that helped to improve the thesis and Catherine Eschle for being a very supportive chair.

Many other great people were willing to meet and discuss my half-baked ideas. I am especially thankful to Ben Worthy and Dan Berliner for finding the time. At different stages of my PhD, I was lucky to present my work at ECPR and EPSA conferences, Summer doctoral programme at the Oxford Internet Institute and some practitioners events. I would also like to thank the participants at these events for their invaluable feedback.

I am indebted to many at the School of Government and Public Policy – to everyone who provided feedback on my work during school colloquiums and especially to Gemma Traynor, Gabriela Borz and Stefanie Reher for supporting me in different ways. I would also like to thank the library staff who willingly acquired the books I was curious to read – if nothing else, there is a nice collection of books on government transparency at Strathclyde for future students. My gratitude extends to Fabrizio de Francesco, Richard Rose, Melanie McCarry and Anni Donaldson for giving me an opportunity to collaborate on their projects.

"Public bodies, whether national or international, hold information not for themselves but on behalf of the public".

The Joint Declaration on Freedom of Expression, December 2006

In memory of Ján Kuciak and Martina Kušnírová, and all journalists who fearlessly inform us

Abstract

In election times, political parties promise in their manifestos to pass reforms increasing access to government information to root out corruption and improve public service delivery. Scholars have already offered several fascinating explanations of why governments adopt policies that constrain their choices. However, knowledge of their impacts is limited. Does greater access to information deliver on its promises as an anti-corruption policy? While some research has already addressed this question concerning freedom of information (FOI) laws, the emergence of new digital technologies enabled new policies, such as open government data. Its effect on corruption and government accountability remains empirically underexplored due to its novelty and a lack of measurements. The following pages aim to fill this gap. I propose a theoretical framework which specifies conditions necessary for FOI laws and open government data to affect corruption, and I test it on a novel cross-country dataset collated for this thesis. The results suggest that the effect of both FOI laws and open government data on corruption is conditional upon the quality of media freedom. Moreover, other factors, such as free and fair elections, independent and accountable judiciary or economic development, are far more critical for tackling corruption than increasing access to information. These findings have important policy implications. In particular, digital transparency reforms will unlikely yield results in the anti-corruption fight unless robust provisions safeguarding media freedom complement them.

While a cross-country approach has revealed the importance of the media's role as an information intermediary, it does not enable for an in-depth understanding of how media engage with government information. Therefore, in addition to comparative cross-country analysis, two empirical chapters focus on the UK case study. I combine various methods: survey of investigative journalists, qualitative interviews with civic activists and civil servants and quantitative text analysis of FOI requests sent to the UK central government from 2008 to 2017 to investigate how different groups engage with FOI laws and open government data and what their demand for government information is. I find that the use of FOI laws is very heterogeneous. By no means, the proactive publication of open government data could address such a diverse demand, and thus it could not substitute FOI laws. A substantial proportion of topics, which occur in FOI requests covers information in the public interest. However, some FOI uses for private ends could also be linked to the concept of accountability, as they often point to the failure of other government communication channels and poor explanation of newly introduced policies. My work also shows the potential of applying computational social science methods to FOI requests to study the impact of major changes in government policies on people, and rights infringements.

Table of contents

List of Acronyms

ACHPR	African Court of Human and People's Rights
CEE	Central and Eastern Europe
CoC	Control of Corruption
CPI	Corruption Perception Index
CSO	Civil Society Organisation
DExEU	Department for Exiting the EU
ECtHR	European Court of Human Rights
FH	Freedom House
FOI	Freedom of Information
FOIA	Freedom of Information Act 2000 or any FOI act generally
FOISA	Freedom of Information (Scotland) Act 2002
FREX words	Frequent and exclusive words as specified in STM
GCB	Global Corruption Barometer
GODI	Global Open Data Index
IACHR	Inter-American Court of Human Rights
ICO	Information Commissioner's Office
ICTs	Information and Communication Technologies
IMF	International Monetary Fund
INGO	International Non-governmental Organisation
MoD	Ministry of Defence
MP	Member of Parliament
NCTJ	National Council for the Training of Journalists
NGO	Non-governmental organisation
NHS	National Health Service
ODB	Open Data Barometer
OECD	Organisation for Economic Co-Operation and Development
OGL	Open Government License for public information
OGP	Open Government Partnership
ONS	Office for National Statistics
OSA	Official Secrets Act 1911
QTA	Quantitative text analysis
PIS	Participation Information Sheet
RQ	Research question
RTI Rating	Right to Information Rating
SPnnn	Survey participant (number from 1 to 173)
STM	Structural Topic Model
UK	United Kingdom
UN	United Nations

US	United States
USAID	United States Agency for International Development
WB	World Bank
WBGI	World Bank Governance Indicators
WDTK	WhatDoTheyKnow

List of Figures

List of Tables

Introduction

Political parties have different positions on social and economic policies. They disagree on foreign relations or environmental priorities and many other issues. However, they all endorse policies that increase government transparency. Elections after elections, parties on the opposite ends of the political spectrum, have pledged to adopt new and more ambitious transparency policies than those that were adopted by their predecessors. For example, the United Kingdom's (UK) 1997 Labour Party's Manifesto pledged the following:

> "We will clean up politics! ...unnecessary secrecy in government leads to arrogance in government and defective policy decisions... **We are pledged to a Freedom of Information Act** [emphasis mine], leading to more open government, and an independent National Statistical Service" (1997).

> Thirteen years later, the 2010 Conservative Party's Manifesto stated:

> "We will clean up politics: the expenses, the lobbying and problems with party funding... We will publish details of the money the government spends and the people it employs. People will have **a right to government data** [emphasis mine] to make the performance of the state transparent. We will cut the unaccountable quango state and root out waste" (The Conservative Party, 2010: 65, 69).

Despite differences in ideological positions and policy orientations, Labourites and Conservatives used the very same words to express their aspirations to increase government transparency. When they talked about transparency, they portrayed it as a necessary condition for eliminating corruption and cutting spending but took the link between the two at face value. The rhetoric about transparency in other countries is not very different. With the upcoming elections, parties in the government as those in the opposition promise to deliver transparency and point to any available corruption scandals of their competitors despite the evidence that voters do punish politicians who misuse public office for their private gains is inconclusive. Some studies demonstrated that voters base their votes on corruption accusations (Bågenholm, 2013; Remuzat Rennó, 2008; Welch & Hibbing, 1997; Xezonakis, Kosmidis, & Dahlberg, 2016). However, others showed that under certain conditions, politicians get away with their misconducts (Anduiza, Gallego, & Muñoz, 2013; Solaz, De Vries, & de Geus, 2019).

Nonetheless, the pledges in manifestos are not purely symbolic. They delineate serious policy objectives (McMillan, 2018) and indeed, the 1997-2001 Labour

government adopted freedom of information (FOI) legislation and the 2010-2015 Conservative and Liberal Democrat coalition government advanced open government data publication as pledged. The adoption of both policies has undoubtedly produced a dramatic rise in access to government information in the UK. In the pre-FOIA times when the Official Secret Act (OSA) 1911 inhibited any disclosure of government information, an MP who advocated for the adoption of FOIA argued that neither one could "look after one's children in a nuclear emergency nor to know what noxious gases are being emitted from a factory chimney opposite one's house", because everything is an official secret (The House of Commons, 1979). In today's Britain, regularly updated data on levels of on the air pollutant emissions are available at few clicks[1].

There are many reasons why political leaders talk about transparency, and governments adopt policies increasing the opportunities for the public to scrutinise their actions. Transparency scholars have studied them from different perspectives (Berliner, 2012, 2016; Berliner & Erlich, 2015; de Fine Licht & Naurin, 2016; Michener, 2010, 2011; Roberts, 2006). Some research explains the diffusion of FOI laws through international non-governmental organisations (INGOs) advocating for the rights of access to government information and pressuring national governments to pass FOI laws (Banisar, 2004, 2006; Berliner, 2012, 2016). Other studies show that governments pass FOI laws when governing parties face a tight election competition to secure a guarantee that once they are outvoted, they will be able to access information about the actions of their challengers in the office (Berliner & Erlich, 2015). Domestic institutional structures are also recognised as an important factor affecting when and what quality FOI laws are adopted (McClean, 2011; Scrollini, 2015) as well as the extent of media coverage of proposed FOI laws (Michener, 2010). Research on the diffusion of open government data is more modest due to its novelty (Shkabatur & Peled, 2016). The debate about open government data centres mostly on its potential benefits and impact (Noveck, 2009, 2017, 2018; Zuiderwijk & Janssen, 2014) or potential barriers of its adoption (Janssen, Charalabidis, & Zuiderwijk, 2012; Martin, 2014; Moore, 2011; Parycek, Schöllhammer, & Schossböck, 2016; Worthy, 2015b). However, the number of studies that investigate if governments accomplish with transparency policies what they purport to accomplish is limited.

Several empirical studies investigated the causal link between FOI laws and corruption levels and found inconclusive evidence (Costa, 2013; Lindstedt & Naurin, 2010; Peisakhin & Pinto, 2010; Relly & Schwalbe, 2013; Vadlamannati & Cooray, 2017). Vadlamannati and Cooray (2017) and Costa (2013) demonstrated that the adoption of FOI laws increases corruption perception in the short term, in particular

---

[1] See https://www.gov.uk/government/statistics/emissions-of-air-pollutants.

in countries with free media. Lindstedt and Naurin (2010) found contradictory evidence. Similar studies about the effect of open government data are lacking.

This thesis addresses this gap by seeking answers to the following question: **Do FOI laws, and open government data deliver as anti-corruption policies, how do they differ and how are they used?**

In order to answer the first part of the research question and understand the possible causal paths from the access to government information to government accountability defined as an absence of corruption, I develop a theoretical framework in 2. Drawing on Bentham's (1999, 2001) political thought and empirical research in the field (Fox, 2007; Lindstedt & Naurin, 2010; Naurin, 2006; Peisakhin & Pinto, 2010), I argue that having legally guaranteed rights of access to government information or governments proactively publishing their data in open formats is not enough to eliminate corruption. I identify several conditions that need to be met in order for these transparency policies to have an impact on corruption levels.

In line with Lindstedt and Naurin's (2010) arguments, I contend that the fact that government information is accessible does not mean it is also accessed and used. High literacy levels in the population are necessary (Truex, 2011) for available government information to be also used widely. Both requesting information through FOIA and browsing government datasets go beyond basic literacy requirements, and thus, people largely rely on media as their information intermediaries. Hence, another condition necessary for the information to have an effect on corruption levels is media penetration (Besley & Burgess, 2002; Ferraz & Finan, 2008; Reinikka & Svensson, 2005). As the contemporary media are complex, this includes but is not limited to the levels of new circulation, TV and radio transmission, or internet penetration. Also, as Bentham (Schofield, 2006) emphasized, freedom to use the government information and report about it without repercussions is critical for the information to enable the government oversight (Besley & Prat, 2006). In addition to media, civil society can also act as an information intermediary and contribute to the circulation of information revealing corrupt behaviour (Grimes, 2013; van Zyl, 2014). Finally, for such information to have an effect, corrupt behaviour cannot go unpunished. Thus, the independence and accountability of judiciary is another condition that needs to be present on the path from the provision of information to the absence of corruption (Fox, 2007; Smulovitz & Peruzzotti, 2000). I test this theoretical framework on a cross-country dataset collated for the thesis.

This methodological approach enables to address the first part of the main research question, i.e. if FOI laws and open government data affect corruption levels on a global scale (partially) and what facilitates these effects (media freedom). However, it does not tell much about the underlying mechanisms of how these effects

occur in the first place. It also does not shed more light on differences between FOI laws and open government data as a means of information provision. Results from the quantitative cross-country analysis suggest that only if robust provisions safeguarding media freedom are in place and its quality is high, FOI laws and open government data are associated with corruption levels. Nonetheless, the understanding of information uses for government oversight is constrained when relying on quantitative analysis of cross-country data. Moreover, at the moment, no cross-country data on the use of FOI laws and open government data exists.

Therefore, in order to answer the second part of my main research question, which focuses on the differences between FOI laws and open government data and their uses, I selected a case study as a research method. Based on the results of quantitative cross-country analysis, I identify the UK as a crucial case that fulfils the conditions necessary for access to information to affect corruption levels. In the UK as elsewhere, government oversight has been at the heart of any discussion on introducing FOI legislation or launching portal containing open government data (The House of Commons, 1979, 1981, 1984) and any related policy documents (The Cabinet Office, 1997, 2012). Both FOIA and open government data were primarily envisaged to fulfil their government monitoring role. However, empirical research investigating the relevance of FOI requests (Berliner, Bagozzi, & Palmer-Rubin, 2018) or open government data (De Saulles, 2013) for government accountability is scarce.

With the UK case study, I aim to contribute to this literature and provide a useful source for future comparative analyses. Journalists and civic activists represent a group of users who are likely to use FOI laws and open government data for government oversight. Therefore, first, I use a survey method and qualitative interviews to understand how they engage with these transparency policies. The FOI's popularity in the UK[2] and previous research in the field (Dunion, 2011), however, suggested that users represent a very heterogeneous group and seek information to reach different goals. Covering all user groups through surveys and interviews would be extremely time-consuming. A share of FOI requests made to the central government (up to 20%) is publicly available through WDTK, an online participation platform for submitting FOI requests. Using new computational social science methods, in particular, quantitative text analysis to analyse this wealth of unbiased data on public demand for government information can address the above problems and bring interesting, and otherwise difficult to obtain, insights into the uses of FOI laws. I

---

[2] In 2018, almost 50,000 FOI requests were sent to the UK central government bodies, which was an increase of 3 280 requests (+7%) compared to 2017. The majority of requests goes to local governments, so the overall numbers are likely to be in hundreds of thousands. For more information, see FOI statistics at https://www.gov.uk/government/collections/government-foi-statistics#2019.

adopt this approach to explore what people (other than journalists, as journalists self-reported sparse use of the platform) want to know when they ask the government for information through FOI requests if this information is relevant for accountability and if and how it differs from the pro-actively provided data.

The structure of the thesis is as follows. Chapter 1 puts the topic of transparency in a historical perspective, defines the key studied concepts: FOI and open government data and delineates differences between the two. It argues that the call for transparency is anything, but new and vital lessons for contemporary public policy can be drawn from the political thought of the Enlightenment-era philosophers, Jeremy Bentham in particular.

Then, building on his work (Bentham, 1999, 2001), the principal-agent theory (Coase, 1990; Dilulio, 1994) and Lindstedt and Naurin's (2010) empirical research, chapter 2 presents the main theoretical frameworks used. The main argument of this chapter is that for FOI laws and open government data to achieve anti-corruption goals political leaders pledged to achieve, several conditions must be met. Each of them, access to information (news circulation, radio and TV signal coverage, internet penetration and literacy), free media and civil society, electoral democracy, independent and accountable judiciary, is then introduced in greater detail. This chapter also presents the Michener and Worthy's Information-Gathering Matrix (2018), which I use as a theoretical framework for the categorisation of FOI requests studied in Chapter 7.

Chapter 3 details the research design applied in this thesis. First, it briefly explains motivations for researching FOI laws and open government data and why it matters. Then it continues by discussing how others studied these transparency policies and where the gaps lie. The chapter introduces the main research questions and methods that are used to answer them, in particular, a comparative cross-country analysis, which guides the selection of the country case study. The UK was identified as the most likely crucial case as it meets the conditions set by the theoretical framework and thus, is likely to deliver on transparency promises. In this chapter, I also describe the methods used in the case study – the survey and qualitative interviews exploring journalistic uses of FOI laws and open government data and quantitative text analysis estimating the topic prevalence within FOI requests sent to the UK central government bodies from 2008 to 2017.

Chapters 4 to 7 represent each of the separate empirical studies. Although these chapters are linked, they are self-contained, i.e. they entail a brief overview of theoretical assumptions, hypotheses, description of data and used methods and results. Chapter 4 represents an empirical test of the theoretical framework introduced in Chapter 2 and explores the associations between FOI laws, access to government data

and levels of corruption across countries. Following the findings from this comparative cross-country analysis, chapter 5 explains how and why the UK was selected as a country case study. It also provides a brief context on the history of FOIA and open government data in the UK.

Both chapter 6 and 7 are empirical studies of the uses of FOI laws and open government data in the UK. In chapter 6, I explore their uses by journalists and civic activists. Drawing from available literature (Birchall, 2014, 2015; Gurstein, 2011; Worthy, 2015b) I argue that open government data poses high expectations on the skill set of its potential users compared to FOI. I also theorise that given exclusivity is highly valued in journalism; journalists will be reluctant to use FOI in any ways that could publicly reveal their identity and storylines. This chapter also explores journalists' experiences with public authorities.

Chapter 7 examines the topic prevalence in FOI requests sent to central government bodies through WDTK from 2008 to 2017 and compares these topics against proactively published datasets on the national open data portal Data.gov.uk. It also explores the impact of the request outcomes (successful vs. refused) and the party in the government and premiership on the topic prevalence. Finally, Chapter 8 brings together all the findings and discusses their limitations and implications for future transparency research and public policies.

The key thesis contributions are twofold. First, its findings demonstrate that transparency policies lead to better control of corruption and stronger government accountability only when they are accompanied by measures safeguarding media freedom. Therefore, when authoritarian governments adopt transparency policies, we should approach these policies with caution and investigate what aims they follow, whether they shift attention from other substantial domestic policy issues, who is able to use the provided information and for what ends, in particular, whether it can be used to hold the government to account. Second, the UK case study shows that FOI laws and open government data provide a different extent of access to information. While FOI laws are led by the public demand and have in place several mechanisms to guarantee access, the provision of open government data is controlled by governments, which eventually decide what data will be published. Moreover, the benefits of open government data are accessible only to those who have access to digital technologies and have digital and data skills. Inequalities embedded in technologies, thus, translate into inequalities in representation. Information cannot be used to defend someone's interests or to hold the government accountable if it cannot be accessed in the first place. Also, the emphasis on technology and advanced digital and data skills make open government data a policy that has a chance to succeed only in the resourceful countries of the Global North with advanced technological and

human infrastructure. Therefore, FOI laws overall represent a more inclusive transparency policy than open data.

1 The origins and current understanding of transparency

In the past two decades, we have witnessed governments worldwide, be they democratic or authoritarian, to compete in making more ambitious transparency pledges. The post-communist governments passed transparency legislation in the late 1990s and early 2000s at an unprecedented rate. The world's leading governments, such as the UK and US, emphasised that transparency is an essential means to better public service delivery[3] and succeeded in setting it as a priority on the global political agenda[4]. Others, for example, the Brazilian government, has launched the Ministry of Transparency, Supervision and Control[5]. The openness has become a global norm with governments creating international and regional coalitions to join forces in these endeavours. Nonetheless, as such, the quest for more transparent governments is not new.

This chapter provides an overview of the origins of government transparency and its current understanding. It starts by introducing Bentham's political thought on the subject. Bentham's main argument was that for governments to govern well, they need to be exposed to public control. The public ought to have access to the information about the government's activities to be able to assess if it continuously acts for the good of society. "Without publicity, no good is permanent; under the

---

[3] One of the main proclaimed aims of the 2010-2015 Conservative and Liberal Democrat coalition government was to make the British government "the most transparent and accountable government in the world". British Prime Minister David Cameron argued at several occasions that the proactive publication of government data in open formats would create new opportunities for public participation, but also businesses and boost enterprise. Barack Obama, the President of the United States (US) at the time, shared similar views. He argued that his administration is the most transparent in history with everything, be it the list of the White House visitors or legislation, available for the public scrutiny. For more examples of the transparency rhetoric in the United Kingdom (UK), see the following commentaries by the representatives of the Conservative Party: Francis Maude, the then Minister for the Cabinet Office and Paymaster General's editorial in the Guardian: https://www.theguardian.com/commentisfree/2010/nov/19/francis-maude-government-data-published, or David Cameron, the then Prime Minister's opinion in the Daily Telegraph: https://www.telegraph.co.uk/news/politics/david-cameron/8621560/David-Cameron-We-are-creating-a-new-era-of-transparency.html. For more examples of the transparency discourse in the US, please see the following links: The Memorandum for the Heads of Executive Departments and Agencies on Transparency and Open Government: https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

[4] Under the UK presidency, the G8's summit (Group of Eight consisting of leaders from Canada, France, Germany, Italy, Japan, Russia, the UK and US) focus was on government transparency.

[5] The Ministry was created under the provisional measure no 726 in May 2016: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Mpv/mpv726.htm.

auspices of publicity, no evil can continue", he argued (Bentham, 1999: 37). Bentham also emphasised the role of the press, which was presumed to guarantee the information circulates widely. He never used the term transparency, but instead, referred to publicity. I will argue throughout this thesis that the term publicity better captures the underlying dynamics of a range of factors that contribute to good governance and Bentham's ideas are relevant for contemporary policymakers as they were for those in the 18th century. The chapter then continues by defining the two key concepts of this thesis – FOI legislation and open government data as practical instruments, which enable the public to exercise the oversight of the government's activities. I describe the origins of FOI laws and discuss the available research, which has helped to clarify why governments adopt a piece of legislation that restricts them in multiple ways in the first place. I conclude by distinguishing **de jure** and **de facto** right to information and identifying major gaps in the implementation and enforcement of FOI laws and discussing what implications this has for good governance. I then define open government data and formulate its key differences from FOI laws.

1.1 Bentham's principle of publicity

The idea of transparency as a means to achieve good governance has its roots in the 18th century. Many philosophers of the Enlightenment era discussed the necessity of public oversight in their work. Immanuel Kant argued that all actions ought to be compatible with publicity, and thus withstand public scrutiny, to be moral (2015). Utilitarian Bentham wrote along the same lines... "for why should we hide ourselves if we do not dread being seen?" (Bentham, 1999: 30). Bentham proposed perhaps the most detailed account of that time of what transparency means from the perspective of institutional design, and popular oversight (Bruno, 2017). He did so without using the term transparency. In his work, he referred to publicity. The underlying principles and values of both concepts are broadly the same. However, as Lindstedt and Naurin (2010) and Naurin (2006) rightly argued, they have nuances that can lead to different outcomes. Transparency means that the information is made available. Publicity, in addition to that, requires that the information is acted upon, i.e. transparency is a necessary element of publicity, but on its own might not be impactful.

Bentham developed his thoughts on publicity from the premise that people would behave better if they were watched, be they political leaders or ordinary citizens. Thus, publicity ought to be embedded in the architectural and procedural design of institutions. He argued that only when citizens are aware of how their rulers act, they will be able to hold them accountable. No sanctions can be employed in the

absence of knowledge of the violation of the rules. Bentham proposed several measures to increase public access to information and enable the public to monitor its rulers' actions. For example, notetaking in the proceedings, recording the minutes of all speeches, taking questions and answers, allowing strangers to attend the assembly's meetings, and publishing its transactions (Bentham, 2001; Splichal, 2002). The assembly's meetings open to the public would serve as later evidence of what has been said (Bentham, 1999). He argued that the public has to be free to discuss and criticise their rulers' positions to secure their accountability (Schofield, 2006). This presumed that the government would refrain from silencing criticism even when it could incur reputation harm. According to Schofield, Bentham was well aware that individual critical voices may be quickly suppressed or too weak to hold political leaders accountable. He recognised early on the role of the free press in ensuring government's answerability and responsibility. He argued that the invention of the printing press amplified the power of information, and if this information was published in the newspapers, thanks to its "regularity and constancy of attention" its influence increased (2006: 268).

Bentham proposed that publicity would serve both a preventative but also corrective function. Reinforced by the free press, it would help to overcome misbehaviour and misrule and improve society on the whole. With an increased risk of being exposed, political leaders would less incline to break the rules. Those who violate them nevertheless will be forced out from the public life, as owing to the free press, the public will be aware of the extent of the misconduct, and thus will be able to sanction wrongdoers. That said, Bentham recognised that the impact of the press depended on its reach, not only in its geographical sense but also as people's capability to understand the conveyed messages. Other scholars studying Bentham's political thought also stressed his awareness of the publicity being conditional upon literacy (Baume and Papadopoulos 2018: 171). Their interpretation of Bentham's publicity concept is that it requires the engagement with the information to be consequential (ibid). One of the aims of the present thesis is to demonstrate that this condition also holds for contemporary transparency policies.

Initially, Bentham did not attach his publicity concept to any particular political regime. However, in later years, he acknowledged that only liberal representative democracies could fulfil the requirement for publicity, as any other forms of government would inhibit the free exchange of information, control the public image of the government strictly, and avoid accountability (Schofield, 2006). His theory is still valid, and authoritarian countries, such as Belarus, Myanmar, Oman, Saudi Arabia, the United Arab Emirates, or Venezuela, not passing FOI laws is an illustration of thereof. As Schofield notes, Bentham argued that in a representative democracy, publicity is not only beneficial for citizens but also for political

representation, as opening discussions to the public might induce knowledge sharing and overall increase the quality of arguments on both sides (2006). He posited that knowledge of many might outbalance knowledge of few, although, well-educated, which is nothing else than a historical embodiment of the contemporary idea of crowdsourcing.

Schofield suggested that despite Bentham's confidence in the political and social benefits of publicity, he recognised that it ought not to be absolute. The universal preference is for the public system of government, but there might be reasons for secrecy (2006). Bentham argued in his Political Tactics that,

> "It is not proper to make the law of publicity absolute, because it is impossible to foresee all the circumstances in which an assembly may find itself placed. Rules are made for a state of calm and security: they cannot be formed for a state of trouble and peril" (1999: 39).

He identified exemptions from the general rule of publicity. For example, if publicity could deter open deliberations, jeopardise the security or enable offenders to escape justice, secrecy was a better option. In this case, similarly as in others, Bentham's views were progressive and visionary. Contemporary FOI laws specify a set of exemptions from disclosure, for example, information held for criminal investigations or information that could create potential harm. It also creates a safe space for policy discussions.

Bentham's quest for publicity was multidisciplinary and went far beyond the public control of rulers. He envisaged its practical applications in many different areas of life and subjects, in particular, architecture. He argued that not only the procedural design of public institutions might enhance or inhibit accountability, but its architectural design might also noticeably affect it. In this regard, Bentham proposed the Panopticon, a circular building with a centre-dominated inspection tower, which allowed for constant surveillance, as a solution that will address many flaws in public service delivery. The Panopticon, initially envisaged for prisons only, was fuelled again by Bentham's conviction that people behave better when they are supervised (Schofield, 2009). In the prison settings, publicity would defer prisoners from misconduct, but also protect them from harm from their inmates or guards.

In later years, Bentham considered extending the Panopticon to other public institutions whose role encompassed an aspect of supervision, such as hospitals. He foresaw the benefits of the Panopticon's design for political institutions too, in particular, those where assemblies and meetings take place. A circular design with theatre seating on different levels enables that everyone can see and hear each other, which he considered as crucial, as otherwise, members of the assembly are entirely reliant on "borrowed opinions" of those who were able to see and hear (Bentham,

1999). The Panopticon is one of Bentham's most significant legacies, which has implications for contemporary studies of government transparency, but also government surveillance. Although these concepts are intertwined, the present thesis focuses on transparency of governments, not that of people. The government's surveillance is not within its remits, although it will touch upon in several instances when discussing findings.

Bentham's views represent a significant contribution to any contemporary discussion on government transparency. Many of his ideas have come to fruition recently as a result of new inventions of digital technologies. I argue that his thoughts and propositions in the area of government transparency, in particular, the role of the free press in holding the government to account, are still relevant.

1.2 FOI laws: origins, diffusion and implementation

The 20th-century legislation arguably reflects many of Bentham's ideas about publicity, but in the 18th-century government, secrecy was still the prevalent norm. The Tryckfrihetsforordningen, Swedish Freedom of the Press Act, a famous predecessor of contemporary FOI legislation, adopted in 1766, was a deviation from this norm (Underwood & Darch, 2010). It has emerged thanks to individual efforts of a free trade advocate and the diminishing power of the absolutist monarchy. The Freedom of the Press Act abolished press censorship, enabled to write freely about politics but not the religion, and obliged public institutions to provide free access to official documents in their possession. Although the legislation was later abandoned and re-introduced only for shorter periods, it had an impact on political thought in Sweden and abroad (ibid). The Swedish Freedom of the Press Act as stand-alone legislation was an exception in that period, but references to the right to information were made in several documents and treaties. Most of the time, the right to information was implicit, i.e. it could only be inferred from other rights that were mentioned explicitly. For example, the US Declaration of Independence, adopted in 1776, stated:

> "whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such principles and organising its powers in such form, as to them shall seem most likely to affect their Safety and Happiness" (Fray & Spar, 1996b).

It can be argued that to exercise the right to alter or abolish a destructive form of government, citizens must have the ability to recognise when the government is

destructive to "public safety and happiness". Such ability requires access to information about the government's activities in the first place.

The Article 1 of the Bill of Rights of the US Constitution, adopted in 1787, declared the commitment to freedom of speech, assembly and freedom of the press among others (Fray & Spar, 1996a). A point can be made that all these freedoms depend on the freedom to seek and access information. In Europe, Article 11 of the French Declaration of the Rights of Man, adopted in 1789, stipulated that "every citizen may, accordingly, speak, write, and print with freedom" (Fray & Spar, 1996c), and Article 15 noted that "society has the right to require of every public agent an account of his administration" (ibid). In line with the argument made, for citizens to be able to speak, write, and print with freedom, they first ought to have access to information to be able to form an appropriate judgment. Second, as Bentham presumed, they also need to be literate to assess how different kinds of actions might be detrimental or beneficial to their welfare, to form an appropriate judgement.

The right to information was first explicitly mentioned only in the 20th-century, for example, in the Universal Declaration of Human Rights that was adopted in the wake of the Second World War atrocities. Article 19 states:

> "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers" (United Nations, 1948).

With exception to Asia, which does not have a government-endorsed convention or charter to promote or protect human rights, other regional conventions and charters, for example, the European Convention on Human Rights[6], American Convention on Human Rights[7] and African Charter on Human and Peoples' Rights[8] also explicitly mention the right to access information and disseminate it.

---

[6] In article 10, the European Convention on Human Rights states that the right to freedom of expression shall also entail the freedom to receive and further communicate information and ideas without repercussions and regardless of frontiers. The full text of the convention is available at https://www.echr.coe.int/Documents/Convention_ENG.pdf.

[7] The Article 13 of the American Convention on Human Rights establishes, inter alia, that the right to freedom of thought and expression comprise freedom to pursue, obtain, and impart information and ideas of all kinds, and regardless of frontiers. It also specifies that the format of that information can be any of one's choice. The full text of the convention is available at https://www.cidh.oas.org/basicos/english/basic3.american%20convention.htm.

[8] The African Charter on Human and Peoples' Rights states in its Article 9 that everyone shall have the right to receive information, as well as to express and disseminate her views. The full text of the charter is available at http://www.achpr.org/instruments/achpr/.

Nonetheless, some human rights theorists argue, following the Hohfeldian approach[9], that this definition is problematic due to its legal ambiguity. In their view, the Universal Declaration of Human Rights defines access to information as liberty (or privilege) to "seek, receive and impart information" rather than a right, because it does not provide guidance as for the government's duty to give away the sought-for information. Hence, national FOI laws, which address this conceptual gap. In this thesis, I define public access to government information as a right since the FOI laws are usually very specific about the duties of public agencies that are obliged to disclose information under those laws. The definition of FOI in this thesis implies the public has the right to ask the government to provide specific information, and at the same time, the government has a duty to respond to such a request in a manner specified in the particular FOI law.

One of the first countries to enact FOI legislation and specify duties of public agencies was the US after the Second World War, and the press was an essential driving force in the process (Birkinshaw, 1997; Raab, 1994). In the 1960s and 1970s, a few other countries followed and adopted FOI legislation. The highest rate of diffusion of FOI laws occurred in the late 1990s and at the start of the 21st century. The fall of Communist bloc allowed transition countries to adopt new democratic legislation. In some countries, an FOI Act (FOIA) was adopted as a result of efforts of the environmentalist lobby, which sought for government environmental information to become publicly available. These efforts led to the adoption of the Aarhus Convention in 1998[10]. Despite several reasons against adopting FOI legislation, for example, a risk of exposure of incompetence or potential increase in costs and workload caused by high numbers of FOI requests (Ruijer, Grimmelikhuijsen, & Meijer, 2017), governments worldwide have adopted FOI legislation at a relatively quick pace (see Figure 1-1).

---

[9] The jurist Wesley Newcomb Hohfeld (1913) argued that a distinction between liberty and a right needs to be made to avoid conceptual ambiguity. In his view, liberty or privilege means that an entity is free to do a particular act while at the same time, no obligation or duty is derived from this liberty and imposed on another entity. On the contrary, a duty is always attached to the right (1913). Applied to FOI, FOI as liberty would mean that people are free to seek information, but that does not create the obligation for the agency. However, FOI as a right presumes that the agency has to respond to the requester.

[10] The full text of the Aarhus Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters, is available at http://www.unece.org/fileadmin/DAM/env/pp/documents/cep43e.pdf.

Figure 1-1: FOIA adoption rate from 1950 to 2016
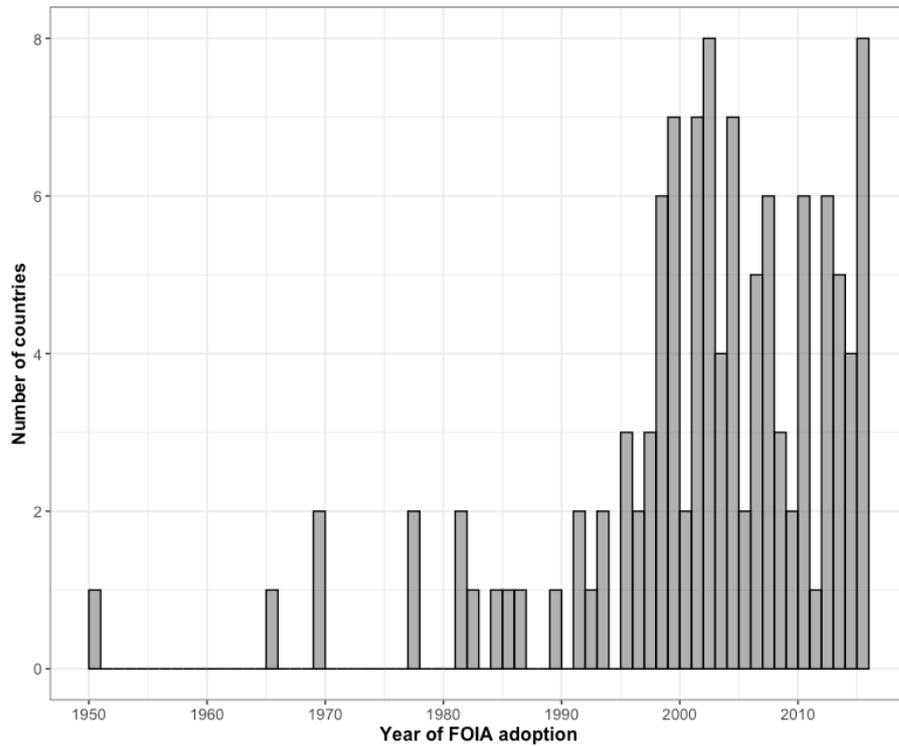
Note: For effective visualisation of the figure, the first FOIA adopted in Sweden in 1766 was removed. The figure starts with the Finnish Act on the Openness of Government Activities, which was passed in 1951.

Scholars propose different reasoning behind the rapid diffusion of FOI laws. Banisar (2004) and Berliner (2012, 2016) argued that its passage in some countries owes mainly to INGOs advocating for human rights causes. Others too proposed that the presence of INGOs is required for any issue to get on the national or global agenda, as among other things, they provide stakeholders from the public and private sector the platforms for sustained interactions and deliberations (Eccleston & Woodward, 2014; Wang & Rosenau, 2001). Banisar (2004) posited that a significant pressure for adopting FOI legislation has indeed come from the World Bank and the International Monetary Fund (IMF), which make their financial support to beneficiary countries conditional upon them implementing anti-corruption legislation, FOI legislation included. Another example of an INGO, influencing transparency policies, is Open Government Partnership (OGP), an international initiative which encourages national and local governments to make pledges to advance open government policies. OGP scores its potential members on the existence of FOI legislation (2018). Also, Berliner (2012, 2016) provided evidence that not only INGOs have an impact on the passage of FOI legislation; they do influence its design. Using the example of Article 19, British human rights organisation established to promote freedom of expression and information, Berliner (2012, 2016) demonstrated that countries, where Article 19 carried out an expert legal assessment, passed stronger FOI legislation subsequently

(ibid). Overall, INGOs contribute to the creation of international norms in the area of access to information.

Despite its positive normative goals, some scholars do not see the diffusion of human rights legislation, right to information included, fuelled by the INGOs and Western governments, in particular, the US and the UK, as positive. Mutua (1996) made an argument that countries that are not based on Western individualism must have a chance to "negotiate the normative content of their human rights law". The one-size-fits-all approach offering an exemplary piece of legislation regardless of specific contexts does not allow for that. Posner (2014) also proposed a thesis that the global human rights norm is "too expansive in scope and feeble in enforcement to have any real impact", and thus is preordained to fail, in particular in developing countries. He further argued that while human rights do not require affluent western liberal democracies to change their institutional practices and behaviours substantially, the extent of change for governments in the Global South with limited resources and fragile security and order is enormous.

The Joint Declaration on Freedom of Opinion and Expression in 2004[11] stated that "steps should be taken, including through the allocation of necessary resources and attention, to ensure effective implementation of access to information legislation". However, the amount of resources that could be freed up for this purpose radically differs across countries. As a result, although every country can afford to pass FOI legislation, not every country can afford to implement and enforce it.

Posner (2014) argued that for governments with scarce resources, it is always a choice between rival human rights, as they cannot afford to implement and enforce all of them. Often, they have to choose between the right to government information and the right to health care or the right to clean water, and the choice is obvious. As can be seen in Figure 1-2, the first countries to adopt FOI legislation were high-income and upper-middle-income countries. Lower-middle income and low-income countries have caught up and adopted FOI legislation only recently, mostly in the past decade.

---

[11] The Joint Declaration by the UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the OAS Special Rapporteur on Freedom of Expression is available at https://www.osce.org/fom/38632?download=true.

Figure 1-2: FOIA adoption rate by WBGI Income status from 1950 to 2016
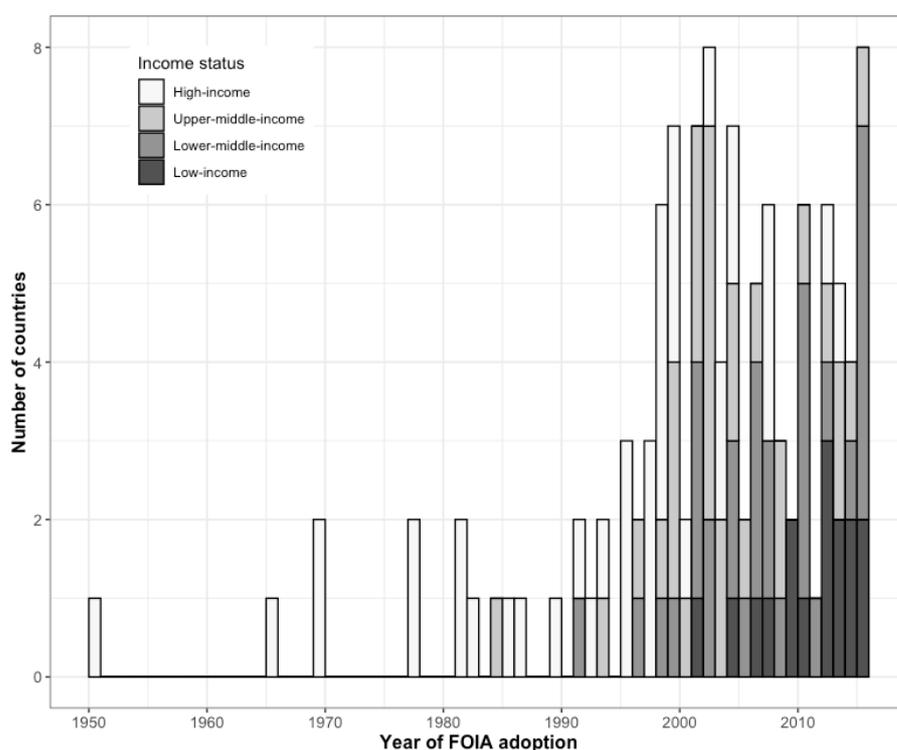
Note: For effective visualisation of the figure, the first FOIA adopted in Sweden in 1766 was removed. The figure starts with the Finnish Act on the Openness of Government Activities, which was passed in 1951.

Moreover, in low-income countries, low literacy levels exclude a large part of the population from being able to exercise its rights and engage politically and enable several political dysfunctions to continue (Fukuyama 2001). Human rights legislation in these contexts is often purely symbolic (Posner, 2014). However, despite the criticism of the INGOs fuelled diffusion of human rights laws, their pressure and the funding they provide have a positive spillover effect on local civil society.

The appeals for the FOIA adoption can also come from the government itself. For instance, in transition countries, governments decide to adopt FOI legislation to secure access to government information for the case of future electoral loss and make the democratic shift less likely to be reversed. Most of Central and Eastern European (CEE) post-communist regimes hastened to pass FOI laws not only to meet the conditions set by the European Union to increase their chances to join the union but also to set the standards for access to government information for any other government to come in the future. Indeed, there is a reason why Underwood and Darch (2010) called the post-Cold War Europe the golden age of FOI. For illustration, Slovakia was on the brink of an authoritarian crisis in the 1990s under Vladimír Mečiar's government. In 1998, when Mečiar's Movement for a Democratic Slovakia lost power, one of the first pieces of the legislation that new government prepared and

passed was an FOIA. Besides that, in transition countries, FOI laws have also had significant importance for reconciling with the past. For example, with FOI legislation in place, victims of the communist and other oppressive regimes and their families are given a chance to study archival materials of state secret services to understand their injustices better.

Berliner and Erlich (2015) empirically tested and demonstrated that tight political competition triggers faster adoption of FOI laws. Mexican states where a margin between an incumbent party and challenger party was small were more likely to pass FOI law prior to the election compared to states where an incumbent party was expected to win the election by a large margin. Nonetheless, while some governments might pass FOI laws tactically, others might approach them as administrative governance reforms and genuinely pass them to improve their performance, public service delivery, and their records management. The benefits of access to government information are well documented (Peisakhin & Pinto, 2010; Relly & Schwalbe, 2013; Vadlamannati & Cooray, 2017). None of them, however, comes from the mere existence of FOI laws, but their diligent implementation and enforcement.

Berliner (2012) pointed out, FOI legislation **de jure** and **de facto** can be markedly different. In other words, an excellent piece of legislation can easily fail in practice. Policy diffusion scholars Holzinger and Knill (2005) argued that adopting a policy predicts poorly its implementation, as several intervening variables influence it. As for FOI laws, the gap between rhetoric and reality is easy to spot when the rating of their quality is compared with the indices assessing the state of civil liberties and political rights. The Global Right to Information Rating[12] (RTI Rating) provides a comprehensive assessment of the quality of FOI legislation worldwide. Access Info, Madrid-based civil society organisation (CSO) and the Canadian Centre for Law and Democracy have been compiling the rating since 2011 in 124 countries and counting, as almost every year there is a state that adopts an FOIA[13]. The RTI Rating examines several aspects of FOI legislation, for example, recognition of a right of access, the scope of the right, clarity and simplicity of requesting procedures and availability of assistance if needed. The assessment also looks at reasonable timelines and fees, unrestricted reuse of disclosed information, exemptions consistent with international standards, public interest override, right to appeal, the presence of oversight bodies and sanctions for not complying with FOI legislation. It also explores if training

---

[12] The Global Right to Information Rating is available at https://www.rti-rating.org/.

[13] This information was from 12 June 2019.

programmes for civil servants and awareness campaigns for the public are available (Access Info Europe & Centre for Law and Democracy, 2017a).

Although democratic countries were first to adopt FOI laws (see Figure 1-3), a glance at the latest RTI Rating shows that no traditional Western democracy is present among the top twenty-five countries with strongest FOI laws (see Table 1-1).

Figure 1-3: FOIA adoption rate by FH Freedom status from 1950 to 2016
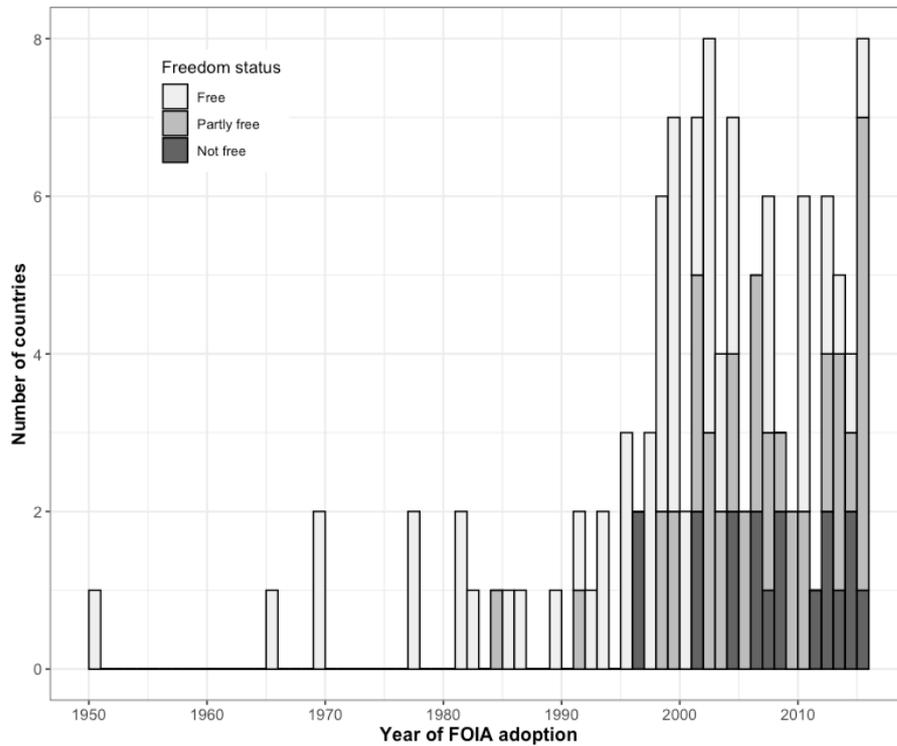
Note: For effective visualisation of the figure, the first FOIA adopted in Sweden in 1766 was removed. The figure starts with the Finnish Act on the Openness of Government Activities, which was passed in 1951.

Many countries have a robust legal framework, but its application and enforcement are questionable. In 2019, with Afghanistan, Mexico and Serbia respectively leading the rating, no Western European democracy made it to the top ten countries. According to the Freedom House's (FH) annual study on the state of freedom in the world only a half of countries represented in the top ten – Serbia, Slovenia, India, Croatia and El Salvador – can be classified as free countries. Afghanistan, with the most comprehensive FOI legal framework, is classified as a not free country. With the elections accompanied by frauds, frequent violent assaults on journalists and civilians, and high levels of corruption, it is unlikely that common bureaucratic application of the new FOI law holds to its standards written in the

law[14]. FH ranks the remaining countries - Mexico, Sri Lanka, Albania, and Liberia as partly free countries (Aghekyan et al., 2018). A lack of tolerance for dissenting views is present in all of these countries. I argue that in a hostile environment where citizens and civil society cannot voice concerns freely, the institutionalised right to seek information is of minimal use. Accessing government information has value for the public only if it can be further used. If citizens cannot use the information to hold the government to account without fearing repercussions, the information cannot serve its intended corrective function.

Table 1-1: Top twenty-five RTI Rating countries by their FH status

	RTI Rating	FH Status
1. Afghanistan	139	Not free
2. Mexico	136	Partly free
3. Serbia	**135**	**Free**
4. Sri Lanka	131	Partly free
5. Slovenia	**129**	**Free**
6. Albania	127	Partly free
7. India	**127**	**Free**
8. Croatia	**126**	**Free**
9. Liberia	124	Partly free
10. El Salvador	**122**	**Free**
11. Sierra Leone	122	Partly free
12. South Sudan	120	Not free
13. Tunisia	**120**	**Free**
14. South Africa	**119**	**Free**
15. Vanuatu	**119**	**Free**
16. Maldives	116	Partly free
17. Azerbaijan	115	Not free
18. Saint Kitts and Nevis	**115**	**Free**
19. Antigua	**113**	**Free**
20. Kenya	113	Partly free
21. Nepal	113	Partly free
22. Macedonia	112	Partly free
23. Ethiopia	111	Not free
24. Nicaragua	111	Partly free
25. Moldova	110	Partly free

Note: RTI – Right to Information, FH – Freedom House, the countries in bold are considered *free* by Freedom House.

---

[14] According to an Afghani researcher who wished to remain anonymous, journalists face several challenges when requesting information under FOI law.

All in all, the letter of the law is not enough. Examples of the mismatch between the letter of FOI laws and their implementation and enforcement are many. For instance, Azerbaijan adopted the Law on the Right to Obtain Information in 2005. Since then, it went through several amendments and in 2019, Access Info Europe and Centre for Law and Democracy assessed it as the 17th most comprehensive in the world (2017b). Nonetheless, according to the reports by INGOs and Azerbaijani CSOs, at the time, most public agencies did not have designated civil servants to deal with FOI requests. Civil servants had a limited understanding of the law due to a lack of training, and they often worked in the absence of any internal guidance. On top of that, enforcement of the law did not have any oversight. The Office of the Information Commissioner was not established despite initial plans and promises, and FOI-related tasks were delegated to the Office of the Commissioner for Human Rights who was not provided with additional human or financial resources towards performing these new tasks (Bayramov, Ibadoghlu, & Imanov, 2016; IRM Staff, 2016; Transparency Azerbaijan, 2016). Lastly, obtaining politically sensitive information in Azerbaijan often comes with the risk of being harassed or threatened (Article 19 et al., 2017). In such contexts, journalists might decide not to use FOI requests as their information source at all.

There are other reasons which might discourage journalists from using FOIA. The advancement of new digital technologies has not only created new opportunities for governments to interact with their constituents and share the information with them but also track what information they are interested in. Michener et al. (2019) conducted an experiment with FOI requests submitted to over 700 Brazilian municipalities and found out that requesters with institutional affiliation are more likely to receive the requested information than non-institutional requesters. They propose a compelling argument that this is a demonstration of "identity-questioning" when governments are googling FOI requesters to get cues about their potential abilities to inform their decision on information disclosure or non-disclosure. These practices, which have been revealed in different countries, go against the spirit of equality in public service provision. Roberts (2005) offered a similar case from Canada. In his article, he describes the revelations of a Canadian investigative journalist Ann Rees[15] who uncovered that Canadian government treats politically sensitive requests differently than regular requests and provide civil servants with detailed guidance on how to spot politically sensitive requests and whom to inform about them.

---

[15] For more information, see Red File alert: Public access at risk available at http://classic.austlii.edu.au/au/journals/FoIRw/2004/13.html.

The surveillance of requesters' interests might not only have an impact on the outcome of their requests in terms of disclosure but might also have repercussions for their safety (Michener & Rodrigues, 2018). In 2018, when an investigative journalist Ján Kuciak was murdered in Slovakia, speculations emerged about how those who ordered the murder knew on what he was working. Kuciak was an avid and experienced FOI user, and his requests were very detailed to maximise chances for information disclosure. Slovak FOIA, like many others, requires the requester to reveal his full name and address. Some suggested that a civil servant in charge of handling FOI requests might have blown a whistle about Kuciak's cues[16]. While this remains a subject of an on-going investigation, journalists have shared multiple stories about civil servants revealing their identity to the subjects about whom they requested the information.

These are maybe extreme examples of how an FOI legislation, robust on paper, can be watered down in practice, but there are manifold ways how to discourage requesters from submitting FOI requests. Some examples are delays in responding to FOI requests, increasing fees for information disclosure or transferring some of the public services to private actors who are commonly excluded from the obligations under FOI legislation in most jurisdictions. The evidence of these practices was pointed out, for example in the US under Obama administration by journalists (Elliott, 2016; Gerstein, 2012; Nation, 2015; Thacker, 2013) and human rights organisations (Downie Jr. and Rafsky, 2013; Susman, 2016). They claimed to have experienced years-long administrative silence, unjustifiably excessive fees for disclosure, incorrectly applied exemptions from the law to avoid scrutiny or privacy breaches. The Electronic Frontier Foundation argued that in terms of compliance with the FOI legislation, Obama's administration was "as secretive – if not more so – as his predecessors and the FOIA has become the prime example of his administration's lack of progress" (Timm, 2012). These malpractices have been present since the inception of the US FOI legislation in 1966. Katz (1969) argued that the exemptions from the US legislation allowed for very vague interpretations, making it possible for the agencies to curtail access to information against the spirit of the law.

Nevertheless, de Smaele (2004) suggests that despite all their imperfections, the adoption of FOI laws is a necessary precondition for any government to be responsive and trusted by their constituents. Despite its poor implementation and

---

[16] For a more detailed debate on this, see the article "Freedom of Information Law: Reporter's Best Friend or Killer?" published on the Organized Crime and Corruption Reporting Project website (Lavrov and Kubaniova 2018) available at
https://www.occrp.org/en/amurderedjournalistslastinvestigation/freedom-of-information-law-reporters-best-friend-or-killer.

enforcement, FOI legislation gives at least some leverage to citizens, journalists and civil society to hold government accountable. Once there is a law in place, the local and international community can at least monitor its implementation and enforcement and require improvements (Michener 2011).

Moreover, the European Court of Human Rights (ECtHR), the Inter-American Court of Human Rights (IACHR), and the African Court of Human and People's Rights (ACHPR), which are charged with supervising the enforcement of related regional human rights conventions and charters have generated in the past decades a vast jurisprudence on access to information. The courts have repeatedly recognised access to government information as a fundamental human right and thus provided further legal certainty for requesters. For instance, the IACHR decided in the case of Claude-Reyes et al. vs. Chile that complainants who requested information from the Foreign Investment Committee on the forestry company Trillium and a planned deforestation Río Condor Project were refused this information unlawfully (Inter-American Court of Human Rights, 2006). The court decided that the agency "did not provide any valid justification under Chilean law" to conceal the information. The judgement was notable for several reasons. First, it confirmed that the requesters do not need to prove why they are interested in obtaining a specific piece of information. Second, it acknowledged that the right to freedom of thought and expression could only be entirely exercised if an individual can obtain the information requested to form an informed opinion. At the same time, the judicial decision acknowledged that the right to seek government information is a stand-alone right.

Similarly, the jurisdiction of the ECtHR has been instrumental in enforcing the right to information in Europe, e.g. rulings in the case of Kenedi vs. Hungary (ECtHR, 2009) or the case of the Youth Initiative for Human Rights vs. the Republic of Serbia (ECtHR, 2013). One of the major judicial decisions that have advanced access to public information was the recent judgment in the case of Magyar Helsinki Bizottsag vs. Hungary. The claimant, Hungarian Helsinki Committee, requested the list of lawyers appointed by the government to represent suspects in criminal proceedings to conduct research in its capacity of an NGO. While some public authorities disclosed the information, some withheld it justifying their decision as a measure against the potential privacy breach. The national court supported the decision of the authority not to disclose the information. However, the ECtHR recognised, referring to Article 10 of the European Convention of Human Rights, that if the information is vital for public debate, the refusal to disclose it violates the right to freedom of expression and information. This reference to Article 10 has been crucial, as, until this judicial decision, the ECtHR was reluctant to employ Article 10 when access to government information was rejected. With this judgment, it confirmed that if access to specific information is a condition for an individual to exercise her freedom

of expression, and is in the public interest, then this information has to be provided. This decision is of utmost importance for journalists, CSOs or anyone who seeks the information to trigger an informed public deliberation.

All in all, while FOI laws have flaws in their design or are weakly enforced, the jurisprudence of national constitutional courts and supranational human rights courts makes them hard accountability measures. Compared to other transparency policies, FOI laws provide some certainty that governments cannot decide at their full discretion what information they will share with their citizens or not. Similarly, having an independent oversight agency such as the Information Commissioner is of utmost importance, as it also ensures that the public has tools to appeal if the information is arbitrarily withheld from them (Birkinshaw, 2010).

1.3 Government transparency in the digital era: open data

Bentham acknowledged that for transparency to have an impact on government accountability, the free press is needed (Schofield, 2006). However, the effects of the press are not conditional only upon its freedom, which guarantees that the government does not interfere with editorial processes. The scale of the press' reach is also crucial. Thus, technologies have always played an essential role in enabling government transparency, as they facilitate the circulation of information. In Bentham's era, the invention of the printing press made it possible. In the 20th and 21st century, the internet has been replacing it and made seeking and disseminating information in real-time as easy as just a few clicks.

With increasing internet penetration and levels of digital literacy, governments also gained new opportunities to communicate with their constituents and provide them with information. The advancements of digital technologies have increased the efficiency of data collection and dissemination to an unprecedented scale. "Data, data everywhere", said the headline of the Economist's special report in 2010 (The Economist, 2010). High expectations about the political, economic and social benefits arising from the greater access to data made it one of the key topics for the next decade. The discourse in the media and that of civil society has been full of superlatives, referring to data as a new oil or gold and claiming it will be a new fix for a vast array of long-standing political and social problems, such as corruption (Granickas, 2014; OECD, 2016; Santiso & Roseth, 2017; Vrushi & Hodess, 2017), climate change (Millner, 2014), and famine (Laperrière, 2019; Vark, 2013).

Therefore, it comes as no surprise that data-driven agenda has also permeated the party manifestos and election campaign speeches. Delivering transparency through new digital technologies was central to both Barack Obama's presidential campaign in the US in 2008 and the Conservative Party's general election campaign in the UK in 2010. Both Obama and The Conservatives won the elections and set transparency and anti-corruption as high-priority topics on their national, but also global political agenda. In the following years, the US and UK governments hosted key anti-corruption and open government summits. They adopted several reforms increasing access to government data; in particular, they launched national open data portals and published hundreds of thousands of datasets in open formats there – open government data.

Open government data has often been presented as a perfected way of providing government information (Noveck, 2009, 2017). While FOI legislation assumes a proactive individual, open data relies on the proactiveness of the government. It is one of the aims of this thesis to explore the theoretical and empirical differences between the two. The governments positioned themselves as advocates of proactive information disclosure. For instance, Obama, in his several campaign speeches, contended that his administration "will put government data online in universally accessible formats and let citizens track federal grants, contracts, earmarks and lobbying contracts" (2007, 00:00:21). David Cameron, the Prime Minister of the UK Conservative and Liberal Democrat coalition government from 2010 until 2015, argued at several occasions that the proactive publication of government data in open formats would bring a radical change to the use of public resources and services. He foresaw a range of benefits coming from this new digitally-enabled transparency, in particular, substantial savings in public spending, and the elimination of duplication in processes. Francis Maude, the then Minister for the Cabinet Office and Paymaster General, claimed that the government "should be proactively making public everything that is appropriate... and make redundant the need for people to ask for access to information" (The Cabinet Office, 2014). Nonetheless, as Paterson (2015) argued, for media and civil society to exercise their watchdog role, they need to be in a position to be able to request government information of their interest.

Public access to government data in open formats has been advocated on the basis that the collection, manipulation and use of government's data have been happening at citizens' expense. It is the taxpayer's money that funds the government, so people shall have a right (subject to reasonable exemptions) to benefit from anything that has been created with this money, be it more or less physically tangible (Kitchin, 2014). Also, most of the data, which the government collects, is about its citizens. Thus, they shall have a right to access it. Overall, releasing the government's data has been viewed as the government's obligation (Kitchin, 2014) but also as an

opportunity (Mayer-Schonberger & Cukier, 2013). Once the data is in the public domain, it creates prospects for multidisciplinary mass collaboration and the creation of new services and products (Surowiecki, 2005) and new opportunities for civic participation (Noveck, 2009).

In technical terms, internet activists and scholars define open data as any data which is "complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary, and license-free" (No author, 2007). In more layman's terms, this means that anyone can use, modify and share data for any purpose without a necessity to ask for approval from its owner, which in the case of open government data is the government. However, there is a limited agreement on the definition of open data in political and social terms. McGee and Edwards (2016) argued that open government data had become one of the catchwords that donors like, but its arbitrary use in many different contexts has put it at risk of becoming an ambiguous concept devoid of any meaning (ibid). Mayernik (2017) also agreed that both terms "open" and "data" are very vague and invite several, and often competing, explanations.

While the technicalities of open data are of a little interest to a political scientist, its informational value, quality and implications for government accountability are. Yu and Robinson (2012) were the first to flag the distinction between data that meets the technical definition of open data and advance public service delivery but is politically unimportant and data that opens the government to scrutiny. They argued that the openness of government data does not necessarily mean government openness. The conflation of the two might, in their view, decrease the credibility of open data as a transparency measure for achieving government accountability. They warned that governments might exploit this conceptual stretching to use open data as a substitution for "hard political change" while still positioning themselves as accountable (ibid).

Weinstein and Goldstein (2012) disagreed and posited that conceptual purity is of limited usefulness in this debate since public service delivery and government accountability are intertwined. In their view, fulfilling acute needs and increasing the quality of public services is a sign of responsible governments, in particular in the Global South. In some contexts, providing data on clear water resources might be of greater importance than data on government officials' meetings with lobbyists. Nonetheless, as Peixoto (2013) argued, a line needs to be drawn between data, which allows government oversight and those which do not, to prevent governments from claiming laurels for accountability when there is none. Some governments might deliver high-quality public services, and at the same time, they might be extremely opaque to reveal their inner workings to the public. In particular, authoritarian governments which restrict political rights and civil liberties are unlikely to publish

data that could potentially limit their choices and expose them to public scrutiny (ibid).

For instance, Singapore provides many useful and interesting data but does not publish any data on government spending. The distinction between politically consequential and non-consequential government data is vital because governments have been selling open data as a new powerful means to fight corruption and INGOs have also appropriated the open data agenda based on this anti-corruption discourse.

As with FOI laws, there are several reasons why governments might resist increasing access to government data. Governments and bureaucracies share a risk-averse culture (Barry & Bannister, 2014; Kornberger, Meyer, Brandtner, & Höllerer, 2017; Martin, 2014; Ruijer et al., 2017; T.-M. Yang, Lo, & Shiang, 2015). Exclusive possession of information has always represented a vital source of bargaining power and income for them (Galnoor, 1977; Grimmelikhuijsen & Welch, 2012; Martin, 2014). Giving this information away means for governments giving up control over it. In examining government information, the public might reveal mismanagement of public resources, but also poor management of data, its inaccuracy, incompleteness, or outdatedness. Exposed lack of competence might have consequences for agencies and civil servants in charge of the data.

Barry and Bannister (2014) argued that duplicities in the work of multiple agencies might also get exposed and lead to budget cuts or even the closure of some agencies whose work might be deemed as redundant. Giving government data away for free also means having to make up for a financial loss from data trade (Martin, 2014). However, Kitchin (2014a) argued that the main barrier to publishing government data in open formats is a lack of financial resources. Building technical infrastructure for publishing datasets, maintaining it, producing and cleaning datasets comes at a cost. The price tags attached to building necessary human infrastructure are also not marginal. Some studies emphasised a critical need for additional training for civil servants charged with publishing government data (Yang, Lo and Shiang 2015). In developing countries with poorly equipped statistical offices, these costs might make open data agenda an unachievable goal (Kitchin, 2014a).

Despite these barriers to make data available to the public, governments have been committing to greater transparency, launching open data programmes, and publishing datasets. No prior transparency policy was adopted at such high rates as open government data, and by such a diverse range of governments, from democracies to autocracies. Similarly, as in the case of diffusion of FOI laws, several INGOs have played an essential role of agents of change. For example, the Organisation for Economic Co-Operation and Development (OECD) and World Bank's Digital Development and Governance units have supported governments' endeavours in

releasing their data in open formats. Since 2012 the World Bank has also been funding open data initiatives in developing countries. From 2012 to 2017, it assisted to more than 50 low and middle-income economies with their open data projects and redistributed more than €40 million[17]. The availability of technical support and funding might serve as an incentive for some governments, but also domestic CSOs to adopt the issue (Schnell 2014). Other important INGOs in the field are Transparency International (TI)[18], Open Knowledge Foundation (OKF) [19] and OGP[20]. Owing to the available funding and pressure from international and domestic civil society, governments worldwide launched open data portals where they aim to publish data they collect or produce.

All in all, open government data represents new digitally-enabled transparency, which empowers the digitally skilled public in accessing and using data in different areas. Compared to traditional transparency policies, such as FOI legislation, the role of government in open data is more proactive as the government publishes the data without waiting to be prompted by the public. Many perceive affirmative disclosure as an answer to the flaws of FOIA detailed previously (Kwoka, 2018b; Noveck, 2017, 2018; Pozen, 2017). However, the proactive publication also comes with disadvantages. For instance, the decision what data will be published lies entirely with the government. Hence, if the government wishes to curtail access to politically sensitive data, it can easily do so. More in-depth comparison of government information and data provided proactively as open data and reactively through FOI requests follows in the next section.

---

[17] For more information about the funds distributed by the Trust Fund for Statistical Capacity Building see https://blogs.worldbank.org/opendata/five-years-investments-open-data

[18] TI was established in 1993 with a mission to curb corruption and encourage governments to adopt anti-corruption legislation and policies. Thus, it comes as no surprise that it appropriated open data agenda early on, as it was promoted as a potential anti-corruption policy. Although its national chapters are independent and self-governed, they liaise very closely with the headquarters and pursue many goals in a coordinated manner. Wang and Rosenau (2001) pointed out that their agenda is often diffused from established and well-resourced chapters in high-income countries to branches in low-income countries.

[19] OKF, similarly, like TI, has national chapters. Although they operate independently, they are supported by the headquarters in the UK. Since OKF is a single-issue organisation and focuses solely on greater access to information and data, it also stands a good chance to affect the diffusion of open data policies.

[20] OGP was created in 2011 as a multilateral initiative to promote transparency, accountability and civic participation, and the use of new digital technologies for achieving these. OGP requires its members, now more than 70 national and 15 local governments, to produce actionable commitments, the fulfilment of which is closely followed by the Independent reporting mechanism. Open government data is one of the central topics in governments' national action plans. So far, more than 400 governments' commitments were related to open data, either in a specific area, such as an increase in access to government budget and spending data, education data.

1.4 FOI vs. open data: Complements or substitutes?

Whether the government information is obtained through an FOI request or published on the government's data portal, there is an implicit expectation that this information is accurate. This thesis takes it at face value, and while it examines the content of available information, it does not question its accuracy. The information made available by governments might be defective in manifold ways, intentionally or unintentionally, but it is not within the remits of this thesis to address this limitation. That said, this limitation is repeatedly mentioned throughout the thesis[21]. Further research into the credibility of government information more generally, and the debate about the right of access to **accurate** government information is needed.

There are several critical distinctions between information provided through FOI requests and made available as open data on governments' dedicated platforms. First, who makes the decision what information and when will be disclosed is different for the two. Why does it matter? This distinction has obvious consequences for what information is eventually made available and whether it is relevant for government accountability. As for FOI requests, the public authority is usually obliged to disclose within legally set time limits any requested information which is of public interest and not a subject of exemptions from FOIA. If the authority withholds information unlawfully, a requester has several legal means to challenge this decision. Proactive publication of open government data relies on the authority's goodwill.

Second, the responsibility for credibility of the information disclosed through FOI requests as well as that of open data rests entirely with the government. However, anyone can modify open data, and thus the credibility of its derivatives cannot be guaranteed.

Third, FOI and open data require a different set of skills and impose different expectations on individuals interested in exercising their right to information and data. Some other distinctions between the access to government information under the FOI regime and proactive publication of open data are outlined in Table 1-2.

---

[21] For examples where governments misled the public see empirical chapter 4.

Table 1-2: Key distinctions between FOI and open data

	FOI	open data
Accessibility in terms of time	statutory time limits	immediate once published
Availability of an offline equivalent	yes	no
Legal framework	FOIA/FOISA	RPSI*
Legal safeguard of information provision	public interest test	none
Availability of appeal mechanisms	yes	no
Oversight body	yes	no

Note: * The Reuse of Public Sector Information Regulations 2015

The concept of open government data relies on a proactive government which presumes what data might be in the public interest and publishes it without waiting for people to ask for it. Fung (2013) defines this new digitally-enabled transparency as "the information on tap" in contrast to "the information on-demand". Shkabatur (2012) refers to the publication of government datasets as discretionary transparency. Both these definitions are eloquently accurate. Indeed, it is at the government's discretion to decide what datasets it will make publicly available. Noveck (2017) defines the information provided through FOI requests as "ex-post" and as open data as "ex-ante". She argues that this shift from ex-post to ex-ante puts more emphasis on collaboration and innovation rather than monitoring and accountability. She proposes that the ex-ante publication makes open data less adversarial tactics than FOI. Therefore, it is not surprising that governments have generally welcomed the shift towards greater proactive disclosure of information as a more cooperative and less oppositional approach.

However, collaborations require reciprocity and equality. O'Neill (2006) argues that open data represents a heavily one-sided way of interaction where the government directs the information flow. The question then stands, if governments are in a position of information suppliers and decide what data will be published, what reason do they have to release any information that will put them into a bad light? Samaha (2006: 918), drawing from Stiglitz, argued that governments will always prefer to disclose "information that makes the administration look public-spirited, effective, and efficient, but withhold information to the contrary". With open data, governments can easily do it as there is no legal requirement for them to respond to public requests for specific datasets (unless, they are filed under FOIA, and these datasets already exist, as FOIA does not oblige agencies to create new information).

FOIA is different in the extent to which it reflects the public demand. It provides information that people explicitly want to know. Shkabatur (2012) refers to

FOIA as mandatory transparency because government agencies have a legal obligation to provide information. In theory, they cannot arbitrarily decide what information they will make available and what they will withhold from a requester. In the FOI regime, the government has limited ways to avoid disclosing information lawfully. The law sets specific conditions in terms of timeframes to respond, means used to respond and exemptions from disclosure. Anything beyond what is specified in the law represents an unlawful withholding of information. Moreover, as mentioned, if government agencies withhold the information from requesters unlawfully, requesters have several legal means to appeal the decision, taking their case to supranational courts if all lower-instance courts decide in their disadvantage. FOI regime works against the arbitrariness of governments' decisions.

I argue that this distinction between FOI legislation and open government data is crucial, as it produces different outcomes for government accountability. So far, only a few scholars have explored empirically what kind of transparency the public gets with the provision of government datasets in open formats, and how this transparency is different from that delivered through FOIA. In one of a few studies, De Saulles (2013) investigated the nature of data published on the UK open data portal Data.gov.uk. Examining a random sample of 100 datasets on Data.gov.uk, he found that the majority of datasets was oriented on public service delivery. There were significantly fewer datasets that revealed the inner workings of government, decision-making processes or public spending. His findings add to an argument that open government data and FOI might be complementing each other but are no substitutes.

In conclusion, the FOI regime is driven by public interest and publishing open government data relies on the government's goodwill. I follow Worthy's and Hazell's argument (2014) that open government data cannot match precisely the expectations put on FOI, as FOI requests are very specific and diverse at the same time and follow individual interests. Thus, due to a large volume of FOI requests, it would be difficult to provide all information that is requested through FOI as open data. It would also be impossible to predict future demand and publish proactively information in which the public might be interested. I will also empirically test this argument later in the thesis (see Chapter 7).

Another important distinction between government information provided as open data and through FOI requests is how this information can be further used. In both cases, the government or public authority, which provides the information to a requester or publishes a dataset, is ultimately responsible for its accuracy, credibility and overall quality. As for open government data, the opportunity to reuse the data is included in its very definition. However, guidance on whether the information

provided in response to an FOI request can be further used is conflicting. In general, requesters might usually reuse this information, but only in a way fully compliant with other regulations in place. For example, the UK Information Commissioner's Office (ICO) further explains this matter:

> "Although the disclosure under FOIA does not carry any restrictions, the restrictions imposed on the further use of that information by the Copyright, Designs and Patents Act 1988 still apply" (2016).

So, if the disclosed information is protected by copyright, it cannot be freely reused. And vice versa, if it is not protected by copyright, it can be generally copied, distributed, edited or built upon for both non-commercial and commercial purposes freely with a small number of conditions. In such cases, public authorities disclose this information under Open Government License (OGL), which was developed by the National Archives as the default license for government information (ibid). At the same time, ICO's position endorses openness. ICO argues that if a disclosure is public, then it can inform any public debate and even when the information is disclosed just to one requester, any other requester asking for the same information, will be able to get it (ibid). Hence, ICO places the responsibility to judge if disclosure can infringe someone's copyright on the public authority. On the contrary, the National Archives state in their guidance that it is ultimately the responsibility of the requester to make sure she does not violate copyright (The National Archives, 2016).

One of the fundamental properties of open data, expected to generate various benefits, is a possibility of using and reusing it freely without any restrictions. The primary idea behind open data use and reuse draws from a concept of crowdsourcing (Surowiecki, 2005). He argued that making (government) data available to the public creates a unique opportunity for multidisciplinary mass collaboration, which could result in the creation of new useful services (ibid). Howe (2006) who coined the term "crowdsourcing", defined it as:

> "...The act of taking a job performed traditionally by a designated agent (usually an employee), and outsourcing it to an undefined, generally large group of people in the form of an open call".

Some scholars argued that crowdsourcing is anything but new (Morozov, 2013). However, while the concept has been routinely applied to businesses, governments started to use it more extensively only recently along with developing their open government data policies. In the recent past, governments preferred to keep their data and control over designing and providing data-based services. Now, when they have been making data available to the public, they shift competencies to the "crowd".

Kornberger et al. (2017) studied how the Vienna city administration responded to public demand for its datasets and concluded that by putting them in the public

domain, the city administration lost control over these datasets. The interviewees from the city administration had mentioned few instances when a third party, e.g. creator of an application drawing on municipal data provided erroneous information or a user did not update an application, and thus had information, which was not timely. However, it was eventually the administration who suffered reputation harm and had to resolve the dispute with a citizen. Kornberger et al. (ibid) illustrated it using an example of a mobile parking application user who did not update it and unknowingly parked her car in a lot, which was not free of charge, got fined and refused to pay. While this error did not have a severe impact, as it concerned a minor civil penalty, it is easy to imagine a number of similar scenarios with far more damaging consequences, which would compromise the reliability of public services or even put lives in danger. The loss of autonomy and control over published data could also diminish a role of the state in providing services that were traditionally viewed as public (Schillemans, Van Twist, & Vanhommerig, 2013; Sieber & Johnson, 2015).

Responsibilisation, the shift of responsibility for tasks that have traditionally been managed by the government to the third parties, is problematic. An example of that would be the concept of citizens as armchair auditors that David Cameron, the former UK Prime Minister, linked to the concept of open government data (2009). He suggested that once his government publishes spending data, people will act as armchair auditors and actively monitor the data to spot the irregularities. Birchall (2015) argued that such vision construes people as knowledgeable and engaged citizens who are able to act as a watchdog of the local and central government. It expects them to have time, interest, and skills to browse governments' accounts at their leisure. She concluded that not only this is unlikely to be the case, but this rhetoric also has serious democratic consequences. Morozov (2013) provided an example of the Conservative Party using the rhetoric about armchair auditors to justify the cuts in funding for BBC investigative journalism and thus, weaken the media's watchdog function to crowdsource it among ordinary citizens[22].

In this debate, the capacities of FOI and open data user are largely overlooked. Open government data is, by definition, available to anyone. In technical terms, anyone can access it, download it, use it and adapt it. However, what is open and non-discriminatory in technical terms is not necessarily open and non-discriminatory in universal terms. The fundamental fact that transparency delivered as open data requires the use of digital technologies makes it a discriminatory practice in the first place. It furthers inequalities and exclusion. First, not everyone has physical access to

---

[22] For more information on proposed cuts, see the article "Chris Bryant warns over BBC cuts" available at https://www.theguardian.com/media/2015/jul/01/chris-bryant-bbc-cuts-conservative-government.

digital technologies and can afford using them. Second, the awareness of available government open data platforms cannot and should not be taken for granted. Also, perhaps most importantly, the skills necessary to use open data are very specific and advanced, including statistical and programming skills. They are available only to a small fraction of the population (Davies, 2014; Gurstein, 2011). Thus, the benefits of open data will reach only a few (Birchall, 2015; Dai & Li, 2016; R. P. Lourenço, 2015; Worthy, 2015a). Gurstein (2011) warned that a "data divide" is parallel with a "digital divide" and affects the same disadvantaged population, and data inclusion is as necessary as digital inclusion. Those who do not have access to digital technologies or skills to use them would be likely excluded from this new digitally-enabled transparency. The key role of technology in open data also make it a policy that is feasible in the Global North but puts a lot of pressure on capacities of the Global South governments.

At the same time, although the information provided through FOI requests comes with stronger guarantees in terms of government responsibility, FOIA is not easy to navigate for citizens either. Having a legal right to access government information does not equate to being able to exercise it. By law, everyone can submit an FOI request. In some jurisdictions, FOI guarantees access to government information only to its citizens or permanent residents[23]. Compared to open data, FOIA does not rely on technology. Requesters can ask for information using different channels, offline channels included. FOIA is less restrictive and discriminatory in this sense. However, the legal jargon in FOI requests might act as a discouragement to submit a request. The time frames for responding also affect the size and composition of the audience and result in non-egalitarian distribution of information provision (Pozen, 2017). For investigative journalists who are expected to produce stories in a relatively short time, the legal time limits to respond to an FOI request might not be feasible, even more so, if their requests are met with administrative silence or responses are significantly delayed.

Overall, it can be concluded that both FOIA and open government data "empower empowered" (Gurstein, 2011). Be it in India (Saxena & Janssen, 2017) or the USA (Schrock & Shaffer, 2017), the open data users are more likely to be male, young and educated. In the US, the use of open data is also determined by race, with a prevalence of white people benefiting from it (Schrock & Shaffer, 2017). Neuman

---

[23] For example, Canadian FOIA requires a requester to either be a Canadian citizen or a permanent resident. Alternatively, who is neither and would like to submit FOI request can mandate a citizen or resident to submit on her behalf. On the contrary, in the US, there are no restrictions for requesters. Anyone from anywhere in the world can submit an FOI request to the US government authorities.

(2016) found that as a result of poorer access to education, women are more likely to be disadvantaged at accessing government information even in more traditional ways, such as through FOIA.

Both in theory and practice, FOI and open government data are distinct transparency policies. I argue that these differences have implications for the kind of transparency they deliver (the type of information they provide) and subsequently its leverage to hold governments to account.

2 Exploring conditions for effective transparency policies

> "And finally, we will unleash an army of 'armchair auditors' to crawl over the government's accounts - members of the public who can see for themselves whether their government is really delivering value for money for them"[24]
> (Maude, 2009).

> "Digital government can only be transparent to a digital society" (Margetts, 2006: 203).

Despite the limitations outlined in the previous chapter, both FOI laws and open government data have the potential to bring the benefits the governments promise. However, their extent depends on the use. As was pointed out, neither making sense of FOIA nor navigating through government data published on government platforms is an easy and straightforward task. It requires access to digital technologies, digital skills and legal awareness. As Margetts (2006) wrote, the digital government will make a difference for those digitally literate only. However, arguably, only a small fraction of the population has necessary skills, leaving Cameron and Maude's army of armchair auditors too few in numbers or poorly equipped to win any major battle. I argue that if FOI laws and open government data are a minority specialism, then their impact as anti-corruption policies heavily relies on intermediaries who master this specialism and are able to extract the key information and communicate it to the public. Below I propose a contingency framework of the path from access to government information to government accountability, defined here as an absence of corruption. I identify access to information and literacy, free media, free civil society, electoral democracy, the independent and accountable judiciary as necessary conditions for government information to affect corruption levels and discuss these conditions in detail. I then make a case for researching who uses FOI laws and open government data and for what purposes.

---

[24] The full text of Francis Maude's speech at the Conservative Party Conference in 2009 is available at https://conservative-speeches.sayit.mysociety.org/speech/601301.

2.1 FOI laws, open data and corruption: Identifying the missing links

The US Supreme Court judge Brandeis famously stated in his appeal for greater transparency that "publicity is justly commended as a remedy for social and industrial diseases, sunshine is said to be the best disinfectant, electric light the best policeman" (1914). The perception of transparency as a means to the accountable government goes back to the Enlightenment era and Bentham's political thought. Information flows are at the heart of citizen-government relations and if unrestricted, can serve as an essential government's corrective mechanism.

From a principal-agent perspective, increasing access to government information decreases information asymmetry between the government (agent) and the public (principal). As a result, it creates more favourable conditions for citizens (principals) to hold their elected representatives accountable (Dilulio, 1994) and make well-informed choices and decisions (Heald, 2006; Stiglitz, 1999b). Many scholars already demonstrated how important the access to information is for public service delivery, accountability, but also a political competition (Berliner, 2014; Berliner & Erlich, 2015; Besley & Burgess, 2002; Reinikka & Svensson, 2005). Hence, Stiglitz's (1999a) claim that information is a public good benefiting all members of society is not in the slightest an exaggeration.

While research on the role of FOI legislation as a means for fighting corruption has been relatively abundant and diverse, the role of open government data has not been explored much due to its novelty. I modify Lindstedt and Naurin's (2010) theoretical framework to explain the effects of both FOI and open government data on corruption.

For the purposes of this thesis, corruption is perceived as a proxy for government accountability. The definition of corruption for purposes of this thesis is not limited to the abuse of a public office for private gain (Mauro, 1995). Its sense is broad and also includes petty corruption among citizens. Lindstedt and Naurin (2010) proposed that increasing transparency makes undertaking fraudulent behaviour more complicated and dangerous. For instance, if the FOI legislation is robust, well implemented and enforced, then political leaders engaged in fraudulent behaviour have to make an additional effort to cover it up in a more sophisticated way so that an FOI requester will not be able to phrase her question precisely enough to get a revealing answer. Strong transparency policies might altogether discourage from engaging in corrupt activities. In such a case, access to government information serves to prevent corruption. However, when investigative journalists are able to get the controversial information, it can also serve as a corrective measure pointing out to committed

infringements and calling for justice. As a result, access to information leads to better, more efficient and more equitably distributed public services. If public service delivery is efficient and equitable, citizens do not have to rely on paying bribes for getting access to these services in the long run.

The absence of corruption is a sign of accountable governments that govern public resources well. The presence of corruption is a departure from good governance. As Rose & Peiffer (2019) argue, both corruption and governance are socially constructed and context-dependent phenomena. However, formal Weberian bureaucratic and informal normative standards exist that prescribe how government and public administration should behave. Transparency, enabling the public scrutiny of adherence to these standards, is often articulated as a means of controlling corruption and promoting good governance and accountability. In literature, the absence of corruption and good governance are often used interchangeably, as a function of each other. In this thesis, control of corruption is a proxy for government accountability too.

Lindstedt and Naurin (2010) distinguished between agent-controlled and non-agent-controlled transparency. Under conditions of agent-controlled transparency, it is the agent who determines what information will be made available. They argued that this kind of transparency is less likely to affect corruption because agents will have the discretion to disclose the information which puts them in a favourable position and conceal the opposite. I categorise both FOI laws and open government data as agent-controlled transparency.

However, I recognise that the level of the agent's control substantially differs between the two. As for open government data, there is no legal obligation for the government to react to the public demand for specific datasets. It is ultimately the government who decides what data will be published on an open data portal or elsewhere. Some might argue that access under FOI legislation is different in this respect because public agencies are legally bound to respond to FOI requests and disclose demanded information unless it is a subject to an exemption from disclosure under FOI law. While ***de jure***, this is true, as argued in the previous chapter, practical application of the legislation often does not meet its own standards stipulated on paper. In the end, requesters are often left with the deliberateness of the government's decisions. However, due to precisely specified legal obligations and available appeal mechanisms, FOI appears to be more powerful accountability mechanism than proactively published government datasets.

Scholars and practitioners have documented a number of cases in different countries where grand corruption was revealed with the help of FOI requests made by

journalists and civic activists. The US Investigative Reporters and Editors (IRE), a CSO dedicated to investigative journalism, has created a database of investigative news stories, many of which used FOIA as one of their informational sources. Majority of these stories dealt with government corruption, frauds, or financial mismanagement (Lee 2001: 383). News Media for Open Government, a coalition of news media and journalism organizations in the US, runs a similar database Without FOIA Tumblr (Kwoka, 2018a). British journalist Matt Burgess launched a website and twitter feed FOIDirectory[25], which highlights stories from British news media that used FOI requests. Berliner and Erlich (2015) described how Mexican journalists used the then newly introduced FOI legislation to expose misappropriation of funds by the Estado Mayor Presidencial, the presidential guard, and Wal-Mart corrupt practices used for expanding to the Mexican market. Thanks to the journalist Heather Brooke's use of FOIA, the British public learnt about excessive expenses claimed by their members of parliament (MPs) (2009). FOIA was also critical to Kuciak's revelations about potential ties between the globally operating mafia 'Ndrangheta and Slovak high-ranking public officials (Davies & Harding, 2018).

Not only is information obtained through FOI requests instrumental in uncovering the misuse of public funds, but it also helps to identify serious gaps in the provision of different public services. For illustration, journalists in the UK learnt through FOI requests about unanswered police emergency calls (Miller, 2015), negligent home care visits (Donnelly, 2015), and incorrect classification of the patients with degenerative and progressive conditions, which resulted in cutting their benefits (Dugan, 2014) among others. In Canada, FOI requests pointed to hours-long ambulance wait times in remote areas (McKenna, 2017).

These, although anecdotal, examples suggest that without intermediaries, access to government information might not be sufficient to affect corruption and a lack of accountability. The previous chapter, which introduced the main features of FOI laws and open government data suggested that they both are far from straightforward to use for an ordinary citizen – hence, the important role of intermediaries. Lindstedt and Naurin (2010) argued that the information needs to be widely circulated (publicity condition), and subsequently acted upon (accountability condition). I modify their framework and identify the following conditions as necessary for the government information to have an impact on corruption control and government accountability: access to information, access to media, news media freedom, electoral democracy, free civil society and independent and accountable judiciary.

---

[25] http://www.foi.directory/

I propose a contingency framework of the effects of FOI laws and open data policies on corruption control as a proxy for government accountability. I argue that unless the judiciary is independent and accountable and the media or civil society actors are free to convey the information obtained through FOI requests or from government data portals to the public, and they regularly do so, FOI laws and open data policies will not deliver as anti-corruption policies. If these conditions are not met, these transparency policies will not necessarily contribute to stronger government accountability either. The proposed framework, thus, assumes that journalists and civic activists use both FOI laws and open government data. While I argue that their use is crucial for greater transparency to have an impact on corruption control, I also posit that how they use it matter no less.

I apply Michener's and Worthy's (2018) Information-Gathering Matrix (see Figure 2-1) to conceptualise the use of FOI laws and open government data. They developed this matrix to explore different uses of FOI laws in response to the overly political narrative about FOI use. Michener and Worthy proposed four main utilisations of FOI laws, i.e. seeking political information for public interest reasons (top left quadrant) or individual/private reasons (top right quadrant) and seeking non-political information in the public interest (bottom left quadrant) and the private (bottom right quadrant).

Figure 2-1: Michener's and Worthy's FOIA Information-Gathering Matrix (2018)
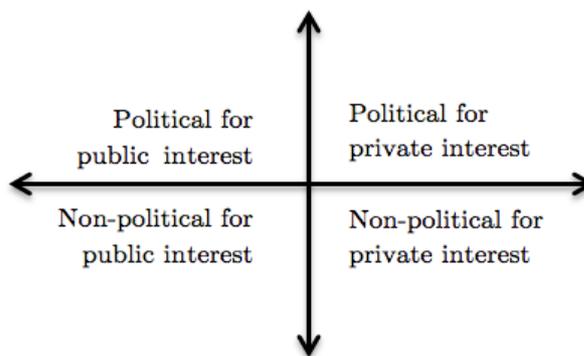
Thus, the proposed contingency framework of the path from access to government information to accountability (see Figure 2-2) also reflects for what purposes the information is being gathered. I argue that, normally, only information sought in the public interest would lead to better corruption control and stronger government accountability. Moreover, the mere availability of this information is insufficient. It has to be acted upon. I propose that news media, institutionalised civil society or engaged individuals with a specific set of skills who seek information in the public interest represent important intermediaries who distil the important messages from the abundant information and data and convey these messages to the wider

public. On the contrary, if the information is predominantly used for private interests, democratic and accountability functions of FOI laws and open government data are weakened.

Figure 2-2: A contingency framework of the path from access to government information to corruption control and greater government accountability
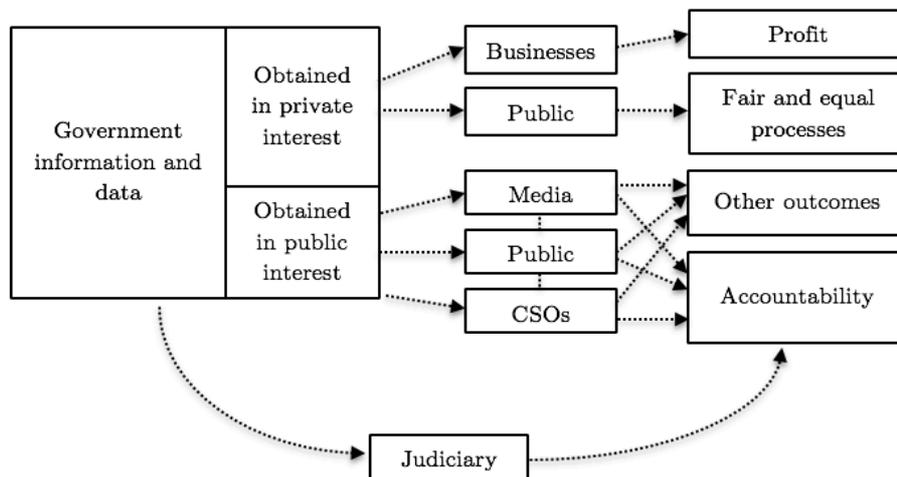
2.1.1 Condition 1: Access to information

My definition of access to information includes several stages. First, the information has to be available in the public domain. This is guaranteed through FOI laws or facilitated by the proactive publication of government data. Second, this information has to be accessed, i.e. people either have to have means themselves to access this information or this information has to be circulated via intermediaries. In both cases, access to information has to be matched with literacy. I describe what I mean by these steps in more detail below.

The available research suggests that very few people would be able to access, and analyse, and thus, benefit from published government datasets (Birchall, 2015; Dai & Li, 2016; R. P. Lourenço, 2015; Worthy, 2015a). Also, research on FOI shows that several groups are excluded from exercising their rights of access to information as a result of structural inequalities (Neuman, 2016). Lindstedt and Naurin (2010) posited that intermediaries, such as news media or CSOs, are required to transmit important messages to the public. The importance of news media for government accountability has also been acknowledged in the jurisprudence of supranational human rights courts. For example, ECtHR argued in its decision in the case of Voskuil vs. the Netherlands that the standards of protection of freedom of information and expression for media should generally be high, and even more so when the information

it disseminates uncovers the corruption of political representation. The judgment reads that "the press is one of the means by which politicians and public opinion can verify that public money is spent according to the principles of accounting and not used to enrich certain individuals" (ECtHR, 2008).

Bentham emphasised the role of the unrestricted flow of information for compliance with moral and social norms and pointed out the amplifying effects of the press already in the 18th century (Bentham, 1999, 2001; Splichal, 2002). His ideas are still valid. First, the news media acts as intermediaries between government and citizens whom they help to interpret the information obtained through FOI requests or data scraped from open data portals. Second, the news media performs the role of a watchdog. If political leaders have to operate in the free and pluralistic media environment, and they want to retain their positions, they cannot afford to be identified with fraud, nepotism or any other serious misconduct. If news media disseminates any reputation-harming information, their chances for re-election are likely to drop. Many scholars have already demonstrated that governments are more responsive to their constituents in countries where citizens have access to news and media are free and pluralistic (Besley & Burgess, 2002; Besley, Burgess, & Prat, 2002; Besley & Prat, 2006; Brunetti & Weder, 2003; Chowdhury, 2004; Djankov, McLiesh, Nenova, & Shleifer, 2003; Ferraz & Finan, 2008; Norris, 2008; Reinikka & Svensson, 2005; Solis & Antenangeli, 2017; Svensson, 2005; Yazaki, 2017).

For instance, Ferraz and Finan (2008) provided evidence in this direction. They explored the effects of the disclosure of audit reports on the electoral performance of incumbents in selected Brazilian municipalities. They found that in the municipalities where inhabitants had access to the local radio, the corrupted mayors stood a significantly lower chance to get re-elected than in those without access. Also, non-corrupt incumbents experienced the same effects, just in the other direction. Their chances to get re-elected rose significantly in the municipalities with access to the local radio (ibid).

Reinikka and Svensson (2005) came to similar conclusions in their research, i.e. that corruption and state capture stand better chances to thrive where citizens lack information and capabilities to obtain and evaluate it. They looked at the Ugandan capitation grant programme for primary schools, which was supposed to cover primary schools' nonwage expenses. However, only 20% of central government spending on the programme on average reached the schools with a large variation across them. There were cues that the remaining 80% ended up in the hands of the local government officials in charge of the programme. The central government ran an experimental newspaper campaign to familiarise the local population with the

programme and transfers made to their local primary schools. The schools in the areas with higher newspaper circulation reported lower misuse rates.

Besley and Burgess (2002) demonstrated that news media also serves vulnerable citizens as protection. Using longitudinal data from 16 major Indian states, they explored how newspaper circulation affects the extent of government accountability. They observed that in states with higher newspaper circulation, governments were more responsive and spent more on public food distribution and calamity relief. The impact was significant. Controlling for other variables, they found out that "a one per cent increase in newspaper circulation was associated with a 2.4% increase in public food distribution and a 5.5% increase in calamity relief expenditures" (Besley and Burgess, 2002: 1435). All these empirical studies point to the importance of media penetration for access to information and subsequently, access to information for government accountability.

2.1.2 Condition 2: Media freedom

Besley and Burgess (2002) also found that newspaper circulation was higher in states where political parties did not own any prominent news media outlets. At the same time, these were the more responsive states that spent more on public food distribution and calamity relief. This finding suggests that media penetration, be it newspaper circulation, radio or TV signal coverage, mobile or internet penetration, might not be sufficient for achieving government accountability. Although the news media helps to disseminate important messages, the critical question is whether every message stands the same chance to make it to the public sphere. Media freedom is, thus, no less important than media penetration. Media freedom plays a critical role. Besley's and Burgess' (2002) findings also suggested that it is closely linked to ownership.

Brunetti and Weder (2003) demonstrated on a large cross-country analysis that high levels of press freedom are associated with significantly lower levels of corruption. They argued that in a pluralistic media environment, a politician might buy some journalists to keep politically important information at bay but cannot buy all of them. Where the free news media exists, efforts to make such arrangements are risky, as other journalists might uncover them and report about them (ibid).

Zhu's, Lu's, and Shi's (2012) findings also confirmed the importance of diverse media ownership. They found that coverage of corruption by government-controlled news media reduced perceived corruption as they are forbidden to report on it. However, that does not mean that corruption does not exist where it is not recognised and openly discussed. According to the public choice theory, governments are motivated by self-interest. Therefore, the government-owned news media have low

incentives to provide its audience with independent and balanced investigative journalism, as that is likely to harm governments' interests. As Stiglitz (1999b) and Samaha (2006) argued, governments have no rational reason to enable information, that would make them look bad, to circulate. Thus, being in control of the media market and being responsive and accountable to their constituents at the same time is an irreconcilable task for the government. The availability of government information or data indicating misuse of public resources is likely to have a limited impact when journalists are unable to make use of it. Only media that can freely serve its watchdog function without political or business interference and pressures are able to strengthen government responsiveness (Norris, 2008). Djankov et al. (2003) observed that countries that lack the diversity of media ownership exhibit similar characteristics. They are more likely to represent low-income non-democratic regimes, widely perceived as corrupt, with low literacy levels in the population.

However, as Besley and Prat (2006) rightly argued, the absence of government censorship does not mean other types of suppression, such as corporate censorship or self-censorship are absent too. They suggested that the governments or corporations that bribe journalists are an extreme and less frequent example these days. That said, subtler and more sophisticated ways of influencing media coverage, such as passing regulations favourable for media owners, enabling collusion between powerful corporations and state, are more wide-spread (Besley & Prat, 2006; Whitten-Woodring & James, 2012). Hallin and Papathanassopoulos (2002) provided several examples from Italy where major corporations were in control of key news media outlets for decades and used them either to make their way to the government or to gain other benefits. For instance, the daily newspapers La Stampa and Corriere della Sera used to be controlled by the automobile company Fiat. Besley and Prat (2006) noted that during those times, the government passed restrictive regulations on importing Asian cars and financed building road infrastructure in exchange for favourable reporting by Fiat-owned outlets (both La Stampa and Corriere della Sera were considered as pro-government).

While these are examples of corporate censorship, self-censorship can occur, in particular, in countries where governments make extensive use of libel laws to silence critical voices. Levitsky and Way (2002) cited the late 1990s Croatia as an example, where, according to the Open Society Institute reports, major news media outlets faced more than 200 libel suits per year. The Philippines can serve as a more contemporary example. Libel and defamation are criminalised and penalised severely by imprisonment up to four years and two months if occurred in the offline environment and up to eight years if occurred online. The libel and defamation laws discourage Philippine journalists from investigating public interest cases, which involve government officials or powerful corporations. The Freedom on the Net report

from 2017 stated that almost 500 online libel cases were filed and followed by imprisonment in criminal trials, which happened out of the public sight (Freedom House, 2017a). According to the Committee to Protect Journalists, the Philippines are one of the deadliest countries for journalists. In the past decade, 45 journalists were murdered with very few cases properly investigated (2018). In these working conditions, even if the government increases access to information and data, news media might be reluctant to use it fearing potential repercussions. All in all, the above empirical research demonstrates that media freedom matters for corruption control and government accountability. Also, it is a complex issue and entails freedom from direct or indirect government and corporate interference and pressure.

2.1.3 Condition 3: Free civil society

It is not only news media that performs a watchdog role. CSOs also act as intermediaries between the government and the public and monitor the government's actions. The importance of civil society in disseminating information in the public interest was also confirmed by the jurisprudence of supranational human rights courts. ECtHR argued in its several judgements, for example, in the case of Steel and Morris vs. the UK that there is "a strong public interest in enabling such groups (as CSOs) and individuals outside the mainstream to contribute to the public debate by disseminating information and ideas on matters of general public interest". The judges adopted a position that watchdog function is not limited to the news media or professional journalists and CSOs should equally enjoy the same levels of protection of freedom of information and expression (ECtHR, 2005).

Scholars recognised civil society's diverse roles and their importance for democracy early on (de Tocqueville, 2002; Habermas, 1984, 1998). Gellner (1994) argued that without civil society, there is no democracy and modernity. INGOs also adopted this position in the late 1990s and the early 2000s. Hansen (1996), from the US Agency for International Development (USAID), posited in his reports that civil society played a crucial role in the USAID mission, in particular, for countries transitioning towards democracy. The acknowledgement of the CSOs role also translated into the availability of further funding for local anti-corruption CSOs operating in developing or transitioning countries. The OECD stated "civil society plays a key role in fighting corruption. Today, this statement is unchallenged: it has become a leitmotiv of anti-corruption discourses" (2003: 7).

There is vast empirical evidence supporting the importance of civil society for government accountability (Ackerman, 2004; Blair, 2000; Grimes, 2013; Ostrom, 1996; Smulovitz & Peruzzotti, 2000; van Zyl, 2014). However, the majority of comparative

cross-country studies demonstrated that its impact is significant only if other factors introduced above, such as institutionalised right to information, free and pluralistic media are present (Grimes, 2013; van Zyl, 2014).

Grimes (2013) studied the link between civil society and the prevalence of corruption. Using data from 133 countries, she tested the impact of the civil society density, measured as the number of CSOs relative to the size of the population and found support for her argument that dense civil society is associated with lower corruption. This argument builds on the expectation that CSOs contribute to distributing and interpreting politically important information; for instance, they expose wrongdoings and subsequently act upon this information. They encourage political engagement by mobilising the public and other important stakeholders to join their endeavours. Lastly, but no less importantly, they compel the formal oversight bodies, such as audit offices or judiciary to act and use their legal competencies to sanction wrongdoings (van Zyl, 2014).

Grimes (2013) provided an example of an Indonesian CSO that mobilised the public to protest and created a pressure leading to the prosecution of 43 public officials in the Padang province. Smulovitz and Peruzzotti (2000) mentioned the impeachment of Brazilian President Fernando Collor de Mello in 1992 as an outcome of coordinated actions of civil society, among other factors. Ackerman (2004) described that in India, the government-led participatory processes and institutions are often highly politicised and do not serve their purpose, as they lack external control mechanisms. However, local communities have created their own successful grassroots control mechanisms that exist in parallel with the formal ones.

Ostrom (1996) demonstrated that the inclusion of the civil society in decision-making processes eventually improves public service provision, although it requires significant initial time and effort investment from both public officials and citizens. Using the example from Brazil of co-producing water and sanitation projects with the public, Ostrom (1996) showed that the involvement of civil society is associated with lower levels of corruption and higher efficiency of service delivery. She also offered another case from Nigeria, where the decision-making processes in local primary education happened without civil society and led to poor service delivery and high political capture of the funds designated for schools, teachers and students. Ostrom (1996) concluded that government accountability without the involvement of civil society does not exist.

Therefore, despite its lack of hard legal sanction mechanisms, civil society is not toothless in holding governments to account (Smulovitz and Peruzzotti 2000). Not only might uncovering corrupt practices and mobilising the public to demand accountability damage the reputation of involved public officials but eventually, it

also might force them to leave the public arena for good. Hence, although civil society is unable to impose legal sanctions, it can mobilise the public and encourage the formal oversight bodies, which have this competence, to use it.

2.1.4 Condition 4: Independent and accountable judiciary

The accountability condition means that some redress mechanisms are in place, and they enable that corrupt agents are held accountable. A free and fair election is the most basic accountability mechanism, but many scholars argued that although it is a necessary one, it is not sufficient for different reasons. First, the elections happen once in a few years only, leaving political leaders to choose to be unresponsive in between the elections. Second, voters, given they possess all critical information, have still only one vote to hold accountable several public officials for their performance in many different fields. Last but not least, Smulovitz and Peruzzotti (2000) rightly concluded that voting is an individual and private act, and voters do not coordinate with each other to punish a specific politician by not voting for him.

While electoral democracy serves as an accountability mechanism despite its limitations, constitutional equality before the law represents a stronger accountability mechanism. Greater access to government information and data might help to uncover irregularities in awarding contracts, favouring certain suppliers, but it can serve as an effective anti-corruption measure only if those responsible are sanctioned. Otherwise, knowledge about corrupt behaviour and impunity might lead to a higher perception of corruption, greater social acceptance of it and higher public distrust (O'Neill, 2006).

The assumption that greater access to information will enhance government accountability also relies on the power of shame, i.e. those named and shamed for various misconducts would react in a self-reflective way. However, contrary to this common expectation, those named and shamed are not necessarily ashamed (Fox, 2007). Greater access to information alone, thus, might not be efficient in achieving accountability (ibid). Data published on government portals might reveal corrupt practices, but if those who were engaged in them, retain their positions and avoid any repercussions, it might contribute to further erosion of public trust. Other hard accountability mechanisms able to apply legal penalties are required to foster accountability - for example, independent national audit offices, police, prosecutors and courts (Fox, 2000). Exposure of wrongdoings is a necessary but not sufficient condition to curb corruption. As an example, Fox offered truth commissions, which revealed past wrongdoings but did not lead to any implications for perpetrators (ibid). Many scholars agreed with him that the government has to justify its actions in front of the public but also in front of the courts if there is a suspicion that these actions

were against the law (Ackerman, 2004; Bovens, 2007; Bovens, Schillemans, & Hart, 2008; Steffek, 2010).

Prosecutors and judges play a significant role in pursuing the rule of law. Nonetheless, to be able to play this role well, they must be independent. According to the US legal scholar Rosenn, judicial independence can be defined as

> "the degree to which judges actually decide cases in accordance with their own determinations of the evidence, the law and justice, free from the coercion, blandishments, interference, or threats from governmental authorities or private citizens" (1987: 7).

Larkins (1996) emphasised that judicial independence guarantees equality before the law for everyone, members of the government included. While there is a universal agreement among scholars and practitioners about the importance of judicial independence for democracy, not so much clarity is about measuring it (Larkins, 1996: 607). Several formal arrangements have been widely recognised as signals of the judicial independence from the government, its opposition or the parties to the trial. The judges should be guaranteed life tenure as opposed to being reappointed in order to decrease motivation to satisfy those in charge of the selection procedures (Feld & Voigt, 2003; Larkins, 1996). In addition, their compensation should be comparable with the salaries of lawyers working in the private sector and academia. These measures should prevent judges from taking bribes from the parties to the trial. Judges should also strictly devote themselves to their judicial role. To avoid a conflict of interest, they should not be involved in commercial or political activities.

Independence from the government can be ensured through other measures. For instance, Sberna and Vannucci (2013) argued that selection and re-appointment procedures are crucial. In their view, judges should be selected on merit by their colleagues, not by the representatives of the executive branch of government. They also should not be transferred to other courts against their will and should enjoy judicial impunity. Rosenn (1987) added that the publicity of judicial decisions could also act as a preventative mechanism against deciding arbitrarily or favouring one party to the trial, even more so, if a detailed justification is required in judicial decisions. The principles of judicial independence from inferences by the executive branch of government are often described in the specific laws on the judiciary or anchored in the constitution (ibid). Nonetheless, relevant legislation and constitution represent **de jure** judicial independence. Similarly, as with FOI laws, the text of the law is often distant from its practical application. Hayo and Voigt (2014) argued that yet it is an important predictor of **de facto** judicial independence.

However, it is not only legislation on the judiciary that needs to be assessed when evaluating judicial independence in a country. The government can also adopt

policies with negative implications on the courts' work, such as decriminalising certain corrupt practices or shortening the statute of limitations – statute prescribing maximum periods within which legal proceedings must be initiated, leading to impunity. As an example of hindering the courts' functions, Sberna and Vannucci (2013) mentioned the circumstances of the Mani Pulite ("clean hands") judicial investigation into political corruption in Italy in the 1990s. They cited Gherardo Colombo, one of the judges involved in the Mani Pulite, who stated that due to the short statute of limitations only 1,000 out of 3,200 offenders were convicted. The trend has not improved since then. In 2005, more than one in ten court trials ended due to the statute of limitations (ibid). They further suggested that this low number of sentences might eventually decrease public trust in the courts (Sberna and Vannucci, 2013). Last but not least, judicial decisions should not be politicised in a way that their impartiality is questioned. Nonetheless, scholars agree that while independence is necessary, the judiciary cannot be autonomous in a way that its work is not subjected to any control (Rios-Fígueroa, 2012; Stephenson, 2004).

I argue that if all these conditions are in place, greater access to government information through FOI legislation or pro-active publication of government datasets could improve corruption control and government accountability. Meeting these conditions means that, in theory, government information can be used to hold the government to account. However, it still matters what information and data are sought and how they are used in practice. In the next subsection, I will present Michener's and Worthy's (2018) Information-Gathering Matrix and explain how it relates to the present research and fit into the main contingency framework.

2.2 Conceptualising the use of FOI laws and open government data

When governments present statistics on the use of FOI laws or government datasets, the high numbers of requests or datasets downloads are portrayed as a demonstration of the FOI laws and open data successes. However, according to the model proposed above, only particular uses of government information obtained through FOI requests or from open data portals are likely to translate into better control of corruption and government accountability. The contingency framework presented above assumes that the end-users who facilitate these effects are primarily journalists, civic activists affiliated with institutionalised CSOs or acting in their private capacities. What they have in common is that their use of FOI laws and open data is in the public interest, e.g. calling public attention to the mismanagement of public resources or unfair

treatment. Such use of FOI laws and open data is, by its nature, expected to create tensions.

Michener and Worthy (2018) argued that in some countries, legitimate questions about the government's actions and policies might be seen as adversarial. They refer to several how-to FOI guides, which describe the legislation as a means of forcing recalcitrant public authorities to disclose information of public interest (ibid). Politicians, too, often see FOI as overly hostile. Surprisingly, this is also true of politicians who helped to pass the law in their countries. The former UK Prime Minister Tony Blair, in his memoirs, labelled passing FOIA as the biggest mistake of his government:

> "Freedom of Information. Three harmless words. I look at those words as I write them and feel like shaking my head till it drops off my shoulders. You idiot. You naive, foolish, irresponsible nincompoop. There is really no description of stupidity, no matter how vivid, that is adequate. I quake at the imbecility of it. Some people might find it shocking. Oh, he wants secret government; he wants to hide the foul misdeeds of the politicians and keep from "the people" their right to know what is being done in their name" (Blair 2010: 516 - 517).

He further defends his statement by arguing that FOIA has never been used by "the people", instead journalists use it as a weapon against the government, not to "bestow knowledge on the public" (ibid).

Nonetheless, his view does not agree with the historical developments of FOI laws, as these were legislated precisely for the journalistic use in the first place. In the previous chapter, I documented that the first FOIA, adopted in Sweden in 1766, was limited solely to the press' use. The US FOIA, the third oldest in the world, was adopted in 1966, thanks to the efforts of journalists and their associations (Fenster, 2006). They were in the centre of legislating processes, and once the FOIA was adopted, they were perceived as its key end-users. Kwoka (2016) cited Congress' archival materials which demonstrate that some members of Congress believed that for FOIA to have intended impact, news media must engage with it. In the UK, news media has also been an important actor engaged in the debates about FOIA.

As argued in the previous chapter and supported by the ECtHR judgements, institutionalised civil society or engaged individuals can also fulfil government watchdog role. In Germany and the UK, CSOs were the main lobbyists for FOI legislation and were perceived as their future users once it is adopted and enforced. In the UK, civil society campaigned for FOIA since the 1970s in response to the domestic legislation, which enhanced government secrecy and surveillance for decades (McClean, 2011). In Germany, the primary expected users of FOIA were environmental activists (ibid).

Kwoka (2018a, 2018b, 2016) meticulously documented the contemporary use of FOIA in the US and found that, at least in the US, FOIA is used for purposes far from initially intended. Her findings suggest that journalists or civic activists do not file the majority of requests in the US. In reality, most requests she reviewed were non-political and did not follow public interests (ibid).

However, the proportion of FOI requests covered in her research is small. Moreover, similar studies from other countries, researching public demand for government information, are few. In this thesis, I study the public demand for government information in the UK, and I adapt Michener's and Worthy's (2018) Information-Gathering Matrix (see Figure 2-1) to conceptualise how FOI laws and open government data are used.

The division between political and non-political information represents to what extent a requester seeks to influence policy, hold government accountable or exercise her other rights. If an interest in influencing policy is absent, then sought-for-information is non-political. For example, if someone is interested in the map of cycling paths in a particular area, this is non-political information. However, the boundaries are blurry and conditional upon the further use of disclosed information. What initially was a non-political and private interest request, can be turned into a political public interest issue if the information is used, for instance, in the local election campaign against the mayor in the office or following a cyclist's death in a road accident in an area with no cycling paths.

The division between public and private purposes reflects whose interests are followed when seeking information. If someone requests the statistics on air pollution, it is clearly in the public interest to have access to it, as poor air quality might pose a health risk to many. However, if someone asks for her data, e.g. own medical or criminal records, the purpose is private. Again, the division between privately and publicly motivated FOI requests might also become indistinct. For example, a public grant participant might request the names of evaluators who assessed her grant application to satisfy her curiosity. At that point, it is a privately driven FOI request. However, if she forwards the obtained information to an investigative journalist who spots that some of the evaluators are in the conflict of interests, and reports on it, from that point on, it becomes an FOI request driven by public interest.

While the common narratives about FOI heavily emphasise its use for political purposes, the available research suggests that political use in the public interest represents, in fact, a minority of requests. Dunion (2011) argued that available research on the uses of FOIA suggests that the population of users is heterogeneous with different levels of skills and varied interests. Kwoka (2018a, 2018b, 2016) was one of the first researchers to conduct an in-depth study on the use of FOIA in the

US. She reviewed commercial FOI requests submitted to the Securities and Exchange Commission, Food and Drug Administration, Environmental Protection Agency, Defence Logistics Agency, Federal Trade Commission, and National Institutes of Health (Kwoka, 2016) and "first-person FOIA requests" submitted to a large number of authorities (Kwoka, 2018a). She defines first-person FOIA requests as requests through which requesters seek information and data about themselves. They often use FOIA as their last resort when other government information channels intended for this purpose failed. In the next paragraphs, I locate hers and other available research on the use of FOIA within the categories introduced by the Information-Gathering Matrix. I also propose how the matrix can be adapted and applied to the use of open data, and what uses matter for government accountability.

2.2.1   Use 1: Non-political information in the private interest

The available research suggests that most frequent users of FOI laws seek information for private purposes. Relly and Schwalbe (2016) argued that despite the accountability rhetoric, commercial interests in the US FOIA have been strong from its inception and have substantially shaped its content. Kwoka (2016) documented that in six selected public authorities, the share of FOI requests for commercial purposes was the largest. She found that commercial companies submit FOI requests for various reasons, such as obtaining information about other companies in their field to get a competitive advantage, uncovering regulatory risks or selling obtained information for profit to the third parties.

The first type of use can be illustrated by an example of a catering company which seeks information from public authorities about the costs of catering services by current providers only to be able to offer their services cheaper when the next procurement takes place. However, it, of course, can be argued that eventually, if public resources are saved, such FOI requests was in the public interest. Commercial companies also submit FOI requests to explore potential regulatory risks, requesting materials about the course of various inspections. Some commercial companies use FOIA to submit identical FOI requests to several public agencies, so-called round-robin requests, to collate obtained information and resell it to third parties which do not want to file FOI requests on their own behalf. FOIA is an entire business model of these commercial companies, according to Kwoka (2018a, 2016).

Such use of FOIA has serious democratic and political-economy implications. Pozen (2017) argued that the use of FOI laws for commercial ends is not wrong per se. Going back to the example of a catering company, if, through an FOI request, requester obtains the information about a particular ministry's excessive spending on

old and rare spirits and forwards this information to the news media, FOIA fulfils its initially intended function, i.e. holding government to account, as the ministry will have to justify its purchases. Following the Information-Gathering Matrix, the request will then move from quadrant with non-political requests for private purposes to political requests in the public interest. However, this is not a commonly observed shift, and the consequences of FOIA use for commercial purposes are rarely as those above. Most of the time, commercial companies use FOIA for their own ends.

Pozen argued that, as a result, FOI departments, which are often underfunded, have fewer capacities to deal with the requests from journalists and watchdogs who were the intended users of FOIA. Public resources are, thus, reduced by private interests, often unrelated to government oversight (2017). Administrative management of FOI requests becomes even more inefficient, leading to delayed responses and making the FOIA an inapt tool for journalists and watchdogs, as by the time they receive the requested information it is likely to become irrelevant. Kwoka (2016) illustrated this with several examples of journalistic use of FOIA when responses to requests took from several months to several years. Such long delays are incompatible with FOIA's initially intended purpose of holding the government to account (Kwoka, 2016, 2018a, 2018b; Pozen, 2017). One way how this could be addressed is exploring how demand for non-political information in the private interest could be answered via pro-active publication of open government data.

2.2.2 Use 2: Political information in the private interest

Similarly to commercial requests, "the first-person FOI requests" where people seek information about themselves also shift public resources away from the intended FOIA use (Kwoka 2018a, 2018b). While they, as she argues, might promote equality and justice, as they often help the requesters to exercise their other rights, they do not advance the oversight of the government primarily. She documented several examples in the US where FOIA was the last resort for people to get information about themselves to put the record straight in pending or expected administrative disputes. She found the most common recipients of first-person FOI requests were Department of Homeland Security, its enforcement component - Immigration and Customs Enforcement, Veterans Health Administration or Equal Employment Opportunity Commission. The examples of requests were, for example, getting access to the so-called "Alien File", which the US government holds on every noncitizen. As Kwoka (2018a, 2018b) argued, the information in this file is often critical for achieving justice in removal proceedings of a noncitizen. More generally, having a means of accessing one's own records, for example, criminal, medical records in civil litigation or criminal prosecution is a necessary condition for due process.

However, Kwoka (2018a, 2018b) and Pozen (2017) disputed that FOIA is well-suited for these purposes. As in the case of commercial requests, first-person requests put a significant burden on public authorities, which, as a consequence, have fewer resources to deal with requests which seek information in the public interest. They argued that if the prevalent use of FOIA is unintended, FOIA cannot fully serve its democratic functions. Moreover, since these unintended uses make FOIA less efficient, seeking one's own data through FOIA might take too long to make a difference in disputes. The requester might get information far later than it would have been useful. In the noncitizen removal proceedings, FOIA delays might have disastrous repercussions, as Kwoka (2018b) illustrated. However, given that these requests demand information that falls under categories safeguarded by data protection legislation, it cannot be substituted by pro-active publication or "the information on tap" as Fung (2013) defines it.

Two above categories – commercial and first-person FOI requests represent requests made to follow private ends. Nonetheless, requests filed in the public interest are those that fuelled the adoption and diffusion of FOI laws worldwide, and I argue that public interest-driven requests are also those that primarily contribute to government accountability and corruption control.

2.2.3 Use 3: Non-political information in the public interest

In the US, legislators perceived journalists as watchdogs and primary users of FOIA (Kwoka 2016). However, many CSOs have watchdog function too and thus, can equally advance FOIA's democratic function. They both file FOI requests that ask for information which is in the public interest. This information though, does not necessarily need to be political. For illustration, a health charity, such as Cancer Research UK, requesting statistics on attendance rate of cervical screening and using this information for raising awareness is an example of a non-political request in the public interest. Another example of non-political use of the FOIA for public purposes are requests submitted by researchers and scientists.

Walby and Luscombe (2017) made an argument that FOIA is a fantastic resource for researchers in hardly accessible sites where ethnographic methods would be difficult to negotiate. Similarly, it might be a useful tool for shedding light into the issues which are in the public interest, but researchers are often discouraged from pursuing it (Lee 2001). Fuller (1988: 99) defines these issues as "forbidden research terrains". For example, the management of universities might be resistant to provide the statistics of violent incidents on the campus due to potential reputation risks.

However, in many countries, FOI laws oblige them to disclose these statistics if requested.

If researchers approach FOIA in a systematic manner, and requests are well planned; they can yield valuable and rigorous findings (Walby & Luscombe, 2017). Additionally, data on the variation of the FOI requests' management by different public authorities are generated as a by-product (ibid). Thus, FOIA can offer rich comparative data on a research topic of interest, e.g. education, health (Savage and Hyde 2014), and at the same provide interesting procedural and performance data. Researchers are also in a comparative advantage to journalists and civic activists whose work has an important temporal aspect. While the news lives perhaps days, causes weeks, even smaller research projects are planned over months and years, enabling researchers to better cope with the delays in responses to their FOI requests. Much of this information could possibly technically be made available as open data. However, much of it might also represent information that institutions would like to keep a low profile, not attracting attention towards these facts.

2.2.4 Use 4: Political information in the public interest

Whenever introducing a new policy, politicians rarely miss a chance to make a point that the policy is for "the people". Blair's quote cited earlier is a good illustration. He argued that he passed FOIA for "the people" to open the government and parliament to scrutiny, but complained that who is using it, in the end, are mostly journalists to scandalise private lives of politicians (Blair 2010). While he was wrong about journalists' use of FOIA, as they seem to represent a minority user, he was right about people. The available research suggests that only a small proportion of ordinary citizens use FOIA. Many are unaware of its existence and how it can be used for their benefits. Data on the use of FOI laws is rarely collected in a centralised systematic manner, and thus, it is difficult to provide any flawless statistics. Therefore, different measures, such as the ratio of requests to the total population, are used as a proxy (McClean, 2011; Vleugels, 2008). Except for the US, the ratio is lower than 0.1% in many countries, for example in France, Germany and the UK.

McClean (2011) argued the assumption that "the people" will use FOIA contradicts what we know about public attitudes. Many scholars concluded that public interest in politics is overestimated (McCarthy & Fluck, 2016). Schumpeter (2010) was particularly sceptical about the engagement of ordinary citizens. He argued that the majority is uninformed and uninterested. In uneventful times, the public will not provide sufficient oversight of the government. In his view, political élites are those who matter in this respect, as they have to compete for the office (ibid). To win it, if

they are not able to prove that they are the most competent for the role, they at least have to put their opponents in the comparatively worse off position. Exposing their wrongdoings is a safe bet. Therefore, FOI laws can be a powerful political weapon in the hands of MPs as Worthy (2014) proposed. He argued that, in many respects, FOIA is a stronger accountability measure than other means MPs have at their disposals, such as parliamentary questions. While questions can be skilfully evaded, FOI requests require a response by law, and if the requested information is in the public interest, it has to be disclosed (ibid).

Despite these advantages, Worthy (2014) found that British MPs are very sporadic users of FOI laws, which supported the previous evidence from other countries. He argued that their low use is a combination of different reasons. First, British MPs are among the oldest in Europe. The mean of their age was 60.2 years after the 2017 elections. It is a well-known fact that older people are more likely to be conservative users and prefer familiar tools over new ones, such as Parliamentary Questions that also require written questions and responses that are published and archived (Worthy, 2014). Second, the basic principle of FOIA is the equitable applicant-blind treatment of all requests. However, research from Canada (Roberts, 2005) and Brazil (Michener et al., 2019) suggested that FOI officers were instructed to identify potentially politically sensitive requests. These were, then, often treated differently, for example, the responses were being drafted with the press departments to keep official lines (Worthy 2014). In such cases, FOIA is of limited use to requesters.

As for the initially envisaged use of FOIA by journalists, in the US, their FOI requests represents only a marginal portion of the whole population. Galka (2017), the founder of FOIA Mapper, quantified the use on the federal level. In 2017, he analysed over 200,000 FOI requests to 85 federal government agencies and estimated that journalists account only for 7.6% of all FOI requesters. The three most frequent users were commercial companies (39%), individuals (20.1%) and law firms (16.7%) (ibid). The journalistic use of FOIA has been on the further decline mostly due to the financial constraints, and in particular in local newspapers (Hamilton, 2016, 2018). With shrinking revenues from advertising, which have shifted to large internet companies, such as Google and Facebook, newsrooms are working on tight budgets, which have resulted in layoffs (Cornford, Wilson, Baines, & Richardson, 2013; Mullin, 2016; Schmidt, 2017) or the shrinkage of the provided investigative content.

Investigative reporting that involves submitting FOI requests or analysing patterns in data requires time. Hamilton (2016) found that an average investigative story entered in the IRE contests took six months to produce if it was a story that appeared in print, five months if it was aired on radio or television, and almost eight months if the story was published online. The news media, which is interested in

conducting investigations, has to have resources to cover the salaries of investigative and data journalists for several months and at the same time accept that the months-long work might not translate into an influential story. Given the delays in responding to FOI requests and resolving FOIA lawsuits, using FOIA is an expensive endeavour for news media who conduct investigative stories. Another reason that might discourage journalists from using FOIA as their key informational source is similar to that mentioned in relation to FOI requests submitted by MPs, i.e. differential treatment. Public agencies tend to be more vigilant towards journalistic FOI requests and might obstruct or refuse them (Kwoka, 2018b). While, again, perhaps much of information sought by journalists and civic activists via FOI requests could be released to the public domain pro-actively, governments would only disclose it if it were not in their interest. Therefore, the possibility to demand this information and a legal guarantee that such a demand will be met or at least the request responded is extremely important for this category of information in the matrix.

That said, the journalistic use of FOIA has had a great impact on government accountability and contributed to uncovering misuses of public resources or positions. Michael Morisy, the cofounder of MuckRock, a web-based platform for submitting FOI requests, gathered several FOI-driven stories. Many uncovered connections and communication between government officials and lobbyists or negligence and ignorance of important information which could have prevented severe emergencies, such as the Flint Water poising[26]. These stories often had political consequences, followed by resignations or prosecutions of corrupt public officials or by the adoption of fairer policies.

Hamilton (2016, 2018) argued that investments in investigative journalism have high returns in public money savings. The benefits of FOIA used by investigative journalists who seek information in the public interest are widespread if their stories become influential. He (2016) illustrated them with an example of Pat Stith, an investigative journalist from North Carolina who conducted over 300 investigations in his career. Almost half of them led to substantive changes on an institutional and individual level, such as adoptions of new laws and policies or personnel changes in public agencies.

---

[26] See the article by Chad Livengood from Detroit News Lansing Bureau: "Flint water warnings reached gov's inner circle" available at https://eu.detroitnews.com/story/news/michigan/flint-water-crisis/2016/02/26/snyder-aides-urged-switching-flints-water-oct/80967048/.

2.2.5 Applying Information-Gathering Matrix to open government data

The Information-Gathering Matrix can also be applied to data provided on governments' open data portals as the type of published data populates each of quadrants except for political information sought in private interest. People can submit FOI requests asking for their own data even though such requests are formally dealt under data protection legislation in most jurisdictions. However, because one's own data is a protected category, subject to data protection laws, and the publication of personal data constitutes a serious breach in most cases, the public agencies cannot publicly respond to such FOI requests. Also, personal data, thus, cannot be included in any proactive publication.

That said, other types of information from the Information-Gathering Matrix can be found on open data portals, although the boundaries between them are less prominent as once the data is published it can be used by anyone. For example, contracting data and spending data fits clearly into the category of political information in the public interest if this data is used by investigative journalists or civic activists who inform the public how public money is spent. At the same time, contracting and spending data might also be used by commercial companies interested in public procurement bids. Nonetheless, because this data is published proactively, its use by commercial companies does not strain the resources of public agencies at the expense of data uses in the public interest, as is the case with FOI requests. This example suggests that proactive publication of government data that is in high demand might be a solution to some of FOI flaws as was suggested above.

Not all uses of open government data matter for government accountability similarly as not all uses of FOI laws do. However, the use of government datasets in open format by journalists and civic activists shows a similar trend as the use of FOIA. Although there is a country variation, in general, the use is low. The US exhibits the highest journalistic uses of both FOI laws and open data. A glance at the recent Pulitzer Prize awarded stories, such as Washington's Post story on police shootings and its victims[27] or the Wall Street Journal's reporting on the US healthcare providers' motivations and practices[28], reveals the use of FOI legislation, public records and government data. In the UK, journalists' interest in open government data is low. Worthy (2015) found less than 200 journalistic stories on the local government spending over £500 in the span of four years (2010 – 2014). Stoneman's (2015) interviewees revealed that they prefer submitting an FOI request to mining data from

---

[27]The articles from the series are available at
https://www.washingtonpost.com/?utm_term=.3ba81e0529cd

[28] For a list of related Pulitzer awarded stories, see https://www.pulitzer.org/winners/wall-street-journal-staff.

governments' portals. In developing countries, the take up of open government data by journalists has been even lower (ibid). Reasons behind it vary. First, the use of open data requires statistical and programming skills. Some of these skills have become a part of the curriculum at the schools of journalism only lately. Journalists, thus, have to either rely on the help of others in their teams (if they are a part of a multidisciplinary investigative team), prequalify to address new job demands or drop the story which requires the use of data. Schrock (2016) argued that due to high demands on skills, open government data remains a minority specialism of civic hackers, who have time and interest in issues of public interest.

Stoneman (2015) argued that public availability, one of the critical characteristics of open data, does not correspond with how newsrooms work. Often, journalists compete for their stories, and hence, confidentiality is crucial. Anyone with the skills can scrutinise government datasets on the open data portal, but specific FOI requests are known only to the requester and public agency, and thus, in theory, are confidential. Precisely for these reasons, I argue, journalists prefer to use their emails to submit FOI requests over the public platforms, such as WDTK in the UK or MuckRock in the US. In conclusion, journalists use open government data only if it fits their story, the competition cannot trace it, and working with data does not require a significant time investment (Brugger et al., 2016; Stoneman, 2015). On the other hand, getting information through FOI requests might take weeks and months, already published data is available instantly.

Last reason why journalists might decide not to use data on national open government data portals is potential bias, which I discussed when I introduced the difference between "information on-demand" and "information on tap" in the previous chapter. Open government data as "information on tap" represents data that is carefully selected by the government as data that can be presented to the public. It is unlikely that government departments would pro-actively publish data that puts them in a bad light. However, data published by government departments on the dedicated national open data portal is not the only government data available in open formats.

Administrative data, such as data produced by the Office for National Statistics on business and economy, employment, and society more broadly on national, regional and local levels is an important source of information for journalists and civic activists. As Allen & Blinder (2018) demonstrated, journalists prefer administrative data over the information provided by individual government departments, in particular when they report about topics that are highly politicised, such as immigration. Allen & Blinder (2018) showed that in an effort to guards its autonomy, the media routinely interacts with the non-political Office for National Statistics and uses its data, which is perceived as independent. Based on Allen's and

Blinder's findings (2018), I argue that administrative data has a special role within the "information on tap" concept and its use contributes to government accountability far more than data published by individual government departments on the national open government data portal. By using data generated by the public institution that maintains bureaucratic independence, media manages to keep its autonomy too and frame important political and social problems through a different lens and propose a picture that is different from that offered by political leaders. Administrative data is "independent of partisan politics" (W. L. Allen & Blinder, 2018: 206)

I argue that the uses of FOI laws and open government data (administrative data included) for public interest purposes are those that have the potential to strengthen government accountability and corruption control. As Michener and Worthy (2018: 482) posited "a relatively small cadre of journalistic requesters produce an asymmetrically large impact (precisely) due to the public nature of their work and what they drag into the light in substantive and procedural terms". Thus, these uses of government information and data can be far-reaching.

3 Research Design: A mixed-methods approach to studying the use and impacts of FOI laws and open government data

In this chapter, I first offer several motivations why studying FOI laws and open government data is relevant for political science. I briefly discuss how others studied this subject, and outline what the available research has not satisfactorily addressed so far and how the present thesis aims to fill this gap. I further present the research design, starting from research questions, which guided the selection of particular methods. I adopt a mixed-methods approach to studying the use and impacts of FOI laws and open government data for various reasons. First and foremost, I considered the availability of quantitative and qualitative sources and their limitations, and to what extent the inclusion of these sources can offer some general conclusions but also some in-depth insights. Second, a mixed-methods approach makes findings more accessible to wider audiences, not restricted to the scholarly audience only, which I believe is vital, given the study of both FOI laws and open government data is very relevant to policymakers and practitioners. In this thesis, I combine a quantitative comparative cross-country study (n = 110 countries) with a case study, which focuses on the use of FOIA and open government data in the UK. The methods used in the case study include expert survey, qualitative interviews and quantitative text analysis of FOI requests sent to the UK central government from 2008 to 2017. Each of these methods, its strength and weaknesses, and mainly the justification of its choice, are described in detail below.

3.1 Why study FOI laws and open government data

Popper (2001), drawing from the Greek philosophical tradition, argued that problems are central to both natural and social sciences. Without problems, there is no scientific advancement. Scientists want to solve these problems and to do so, they locate them in broader theoretical contexts, formulate plausible but falsifiable hypotheses and conjectures and test them often only to learn they were wrong. Popper made a case (ibid: 13) that refuting a theory creates new knowledge as "we learn not only that a theory is wrong; we learn why it is wrong". He posited that this perseverant trial-and-error approach generates "new and more sharply focused" scientific problems that demand solutions, or in other words, new research questions that require answers. It makes the scientific inquiry an unfinished and unfinishable endeavour with no clear

boundaries (ibid: 15). This piece of research has also been motivated by a real-word puzzle. Secretive governments and bureaucracies eagerly adopt FOI laws and release their data in open formats in the public domain, in their own words, to combat corruption, promote good governance and strengthen accountability.

One of the aims of this thesis is to explore whether governments achieve what they purport to accomplish with their transparency policies and if they do, how access to information translates into better corruption control, i.e. what underlying causal mechanisms are at work. Governments have subscribed to new transparency policies at a quick pace, but not much detail has been shared about their inner workings. What information and data exactly governments share? What information and data do people want from the governments? Is it the same information and data? Does the government information supply meet the demand? Who has the skills to access the information?

Transparency has become a common denominator of discussions at global forums and summits. Political leaders have pledged to increase access to government information and routinely argued it would reinforce government accountability and restore public trust (see the former UK Prime Minister David Cameron's and the former US President Barack Obama's quotes mentioned in previous chapters). However, these rather vague arguments have not been typical of the election campaign period only, but they remained vague later in the government manifestos and other strategic government documents. A rich source of governments' pledges in the field of transparency policies is national action plans submitted to OGP. For illustration, the Government of Canada states in its OGP national action plan that,

> "Maximising the release of data and information will enable Canadians to better
> engage with their government, and hold it accountable, creating an environment
> that supports meaningful civic engagement and drives social and economic
> benefits through the innovative reuse of data and information" (2014: 8).

In the action plan, there is no reference as to how government accountability or any other set goals would be achieved through the release of government data. However, the example of the Canadian OGP national action plan does not represent an exemption from the rule, but the rule. The Albanian government in its latest action plan argues that opening government datasets will deliver "efficient and effective governance, economic growth and innovation, transparency and accountability" (Government of Albania 2018: 12) "Data is powerful", the action plan claims but does not explain how it exerts its power. The plan then goes on that "...many countries have launched [open data] programs" (ibid: 12). This particular statement suggests that the genuine reason why the Albanian government has decided to adopt open data policy might lie elsewhere. For instance, committing to the release of government data

might signal to the EU, which of Albania seeks to become a member, that it takes anti-corruption seriously. Open data is to some candidate countries' governments now what FOI laws were to CEE candidate countries in the 1990s.

Leaving the motivations aside, it is not very clear from most of these national action plans how the governments are going to achieve accountability, responsiveness and eradication of corruption with increasing access to government information. These potential benefits are taken up at face value. Transparency and corruption control or government accountability are used interchangeably and lead to a definitional confusion where source and outcome are hard to disentangle. Some governments assume that the release of data will be followed up by public interest. With many more eyes on them, contractors and public officials might become more attentive of their actions and transactions. Obama and Cameron also suggested that the main aim of transparency policies was to allow citizens to monitor their governments and help to root out corruption.

Nonetheless, the relationship between transparency and corruption control as a proxy for government accountability is unlikely to be straightforward. While some evidence of the effect of FOI laws on corruption is already available, the investigations of the effect of open government data on corruption have been limited, mostly because open data is a new concept. It is the mix of this novelty, a high level of political attention devoted to open data, and ambitious (but often unsubstantiated) political claims about its impact that makes it an interesting and timely research subject.

3.2 How others studied FOI laws and open government data

The available research on transparency policies has primarily focused on FOI laws. The global spread of FOI laws in the past three-four decades has inspired research questions about governments' motivations to adopt policies that subject them to scrutiny. Berliner (2012, 2016), McClean (2011), and Michener (2010) provided several valuable explanations. The influence of INGOs (Banisar, 2004, 2006; Berliner, 2012, 2016), differences in institutional arrangements (McClean, 2011), uncertain political future and political competition (Berliner & Erlich, 2015), or the extent of media coverage of proposed FOI laws and presidential powers (Michener, 2010, 2011) affect the passage of FOI laws and the strength of their design.

Once in place, FOI laws have various effects. Scholars studied their impact on public service delivery (Meijer, 't Hart, & Worthy, 2018; Prat, 2005) and corruption

(Costa, 2013; Lindstedt & Naurin, 2010; Peisakhin & Pinto, 2010; Relly & Schwalbe, 2013; Vadlamannati & Cooray, 2017). The results of these studies are mixed. Escaleras, Lin, and Register (2010) found no associations between FOI laws and corruption levels except for countries with weak institutions where FOI laws were linked with rising levels of corruption. Vadlamannati and Cooray (2017) and Costa (2013) found that the passage of FOI laws increases corruption perception in the short term. The effect is larger for countries with a higher degree of media and civil society's freedom and political corruption, suggesting that media and CSOs amplify the messages about corruption. Lindstedt and Naurin (2010) observed that government transparency is associated with lower levels of corruption under the condition of media freedom.

Others (Prat, 2005) discussed the unintended consequences of government transparency, in particular, chilling effects on internal deliberations. Sunstein (2018) also argued that transparency of inputs could discourage openness and honesty and erode trust. A large share of FOI-related research falls within a legal research category, in particular with a focus on human rights law (Banisar, 2006; Birkinshaw, 1997, 2006b, 2006a; Kwoka, 2016, 2018a, 2018b; Pozen, 2017).

While the research on FOI laws has been rich across disciplines, from journalism to political science to human rights and administrative law, open government data is both under-theorised and lacking in empirical studies in social sciences. The systematic literature review conducted for this thesis covered articles, written in English, published between years 2009 and 2017 and revealed a prevalence of articles in the field of community informatics with a strong emphasis on technical aspects of the phenomenon.

Afful-Dadzie and Afful-Dadzie (2017) examined the differences between research on FOI laws and open government data. They found that research on FOI laws occurred more frequently in political science and law journals and focused on access to government information as a human right necessary to hold the government accountable. Research on open government data was published mostly in journals focused on information science and discusses the quality of data platforms, barriers to data publication and use, and presented new theoretical frameworks and concepts. When political scientists write about open government data, they avoid the term, but rather focus on the particular source of data, e.g. crime statistics, education data. To account for that, I searched for a combination of the following different key terms: open government data, government data, open government, open data, data, transparency, accountability, and politics. Adding terms from political science and public policy also helped to exclude articles in computer science, which were not the

subject of the review. Altogether, I reviewed 126 journal peer-reviewed articles and grey literature (reports by World Bank, OECD, OKF, TI and many others).

I found that in terms of geographical representation, high-income Anglo-Saxon countries were overrepresented (see Table 3-1 for more detailed breakdown). The majority of articles discussed open data initiatives in Europe and North America with a particular focus on the UK and the US. Given they were frontrunners in open government data, this is not surprising. The studies assessed open data portals, Data.gov in the US and Data.gov.uk in the UK. Recently, more studies have explored emerging initiatives in Asia and Africa (Bello, Akinwande, Jolayemi, & Ibrahim, 2016; Kassen, 2017; Yang et al., 2015). Majority of articles reviewed the central government's open data initiatives.

Table 3-1: Geographical representation of open government data research (by regions)

Region	Number of articles	Percentage
Europe	40	31.7
North America	16	12.7
Asia	10	7.9
South America	9	7.1
Africa	4	3.2
Australia and Oceania	1	0.8
Several regions covered	22	17.5
Not applicable	24	19
Total	**126**	**100**

There is a clear predominance of qualitative studies, drawing mainly from semi-structured interviews and case studies. A few quantitative studies occurred after 2014, but their proportion has been low. A couple of mixed-methods articles and one experimental study were also among reviewed articles.

The reviewed articles covered a variety of topics. Many authors discussed barriers to the publication and use of open government data on both sides: from agencies and citizens' perspective (Gonzalez-Zapata and Heeks, 2015; Wang and Lo, 2016; Wirtz et al., 2016; Kornberger et al., 2017; Ruijer, Grimmelikhuijsen and Meijer, 2017). They explored a conflict between the traditional Weberian understanding of bureaucracy and a loss of control over information related to open government data (Kornberger et al., 2017) and how bureaucracies battle it (Heimstädt, 2017). A significant part of the open government data research explored specific open government data projects and platforms, for instance, open data portals in Africa (Bello et al., 2016), Italy (Origlia et al., 2016; Vetrò et al., 2016), Mexico, Russia and the USA (Styrin, Luna-Reyes, & Harrison, 2017) or Taiwan (Yang & Wu, 2016), and

their potential benefits, from better service delivery, be it better access to the public toilets (Bichard & Knight, 2012), or smarter public transportation (K. Kuhn, 2011) to government accountability (P. R. Lourenço, Piotrowski, & Ingrams, 2017). Others explored the use of open government data (Afful-Dadzie & Afful-Dadzie, 2017; Brugger et al., 2016; da Silva Craveiro & Albano, 2017; Hellberg & Hedström, 2015; Saxena & Janssen, 2017; Schrock & Shaffer, 2017; Stoneman, 2015).

Nonetheless, while the open government data research is diverse, it rarely addresses the big political questions. Open government data were adopted worldwide as part of ambitious anti-corruption programmes, at least they were presented as such, but little is known if they achieve these goals and how. This thesis aims to fill this gap.

3.3 Research questions

Sandberg and Alvesson (2011) argued that compelling and interesting research questions go beyond spotting gaps in the existing literature but challenge existing assumptions. They do not focus on what is already known, but how what is already known can be thought differently. De Vaus (2001) categorised research questions into two types: descriptive questions, which describe what phenomena we observe and explanatory questions, which explain why these phenomena happen. He posited that the observation always precedes explanation. Research questions can also be categorised as theory-driven or policy-oriented. White (2017) argued that policy debates could be a helpful basis for generating research ideas, as they motivate the questions that are important and relevant in the real world. King et al. (1994) considered such relevance as one of the key criteria of good social science research projects.

Nonetheless, answering policy-oriented questions also requires a theoretical background. Theories are invaluable for establishing potential causal links and interpreting them, and eventually refuting or conforming them (Popper, 2001). Many scholars agreed that theories are a cornerstone of any useful research (Mason 2002; Goertz 2008). Another important criterion is that the research question contributes to the knowledge by formulating testable hypotheses (King et al., 1994). Following the theoretical framework proposed in the previous chapter, I formulate the central question the present thesis seeks to answer as follows:

Do FOI laws and open government data work as anti-corruption policies, and how are they used?

Drawing from the above definitions (de Vaus, 2001; Popper, 2001; White, 2017), the first part of this research question can be classified as an explanatory question. It aims to explain how FOI laws and open government data affect levels of corruption across countries and what facilitates these effects. As was previously mentioned, the absence of corruption is perceived as a proxy of good governance and accountable governments in this thesis. The second part of the question is more descriptive and exploratory, as it seeks to investigate how FOI laws and open government data are used. As for the origin of the research question, it comes from the intersection of theory and practice. Although the central question comes from ongoing policy debates, it is also fed by well-established theories in political science and communication studies, from Bentham's political thought on publicity and its link to government accountability to principal-agent theories.

To answer the overarching question, I formulate several research sub-questions. White (2017) argued that breaking the central research question into several sub-questions contributes to creating a more nuanced description or explanation depending on what the question seeks to do. He differentiated between subsidiary questions that directly relate to the central question and ancillary questions that are linked to the central question more loosely in a sense that they "follow from it" but do not necessarily answer it (ibid). He proposed that these sub-questions also help to clarify the final research design, as they often ask about specificities, e.g. what specific population will be studied, where and when. In his view, having only a central question might later restrict or confuse data collection and analysis choices (ibid). On the contrary, if several research questions are well framed, they might indicate what data needs to be collected for them to be thoroughly answered. In line with this, I formulate the following sub-questions:

RQ1. Do FOI laws and the availability of open government data affect corruption control across countries?

RQ2. If so, what does facilitate these effects?

RQ3. Who does use FOI laws and open government data, and for what ends?

RQ4. What information do FOI requesters seek, and how is that different from pro-actively provided open government data?

In the past decade, the governments have shown a great interest in open government data as a new digitally-enabled government policy that could increase government transparency and further the oversight by the public who would act as armchair auditors. The growth of interest in the topic, ambitious claims about the

potential benefits of open government data and a lack of available empirical evidence to support these claims all, among other things, motivated this research.

As I argued in the previous chapter, the concept of citizens as armchair auditors is simplistic and very unlikely to materialise, i.e. intermediaries, such as news media and CSOs are necessary for the information to have an impact. Following the first two questions (RQ1 and RQ2), I aim to explore the associations between greater availability of open government data and corruption control. One of the key objectives is to establish if there are causal mechanisms that facilitate the greater availability of information to have anti-corruption effects. De Vaus (2001) and Yin (2014) argued that good research design should indicate what sources of evidence will be needed to collect to provide convincing answers to the research question. Yin (2014) compared designing research to solving a logical puzzle, finding the right tools to answer the question. These tools are unambiguous definitions of central concepts and valid and reliable measurements (White, 2017). Research design stage also includes evaluating the feasibility of the proposed approach and identifying limitations and trade-offs (Rog & Bickman, 2008). Converting the research questions into a research project, I first identified the measures of open government data availability, FOI laws, and corruption and what other controls and intervening variables are known to affect corruption and how to measure them. Given the availability of different measurements across countries, the first two questions are answerable using quantitative cross-country analysis (n=110). However, as open government data is a relatively new phenomenon and data measuring its availability across countries dates back to 2013, a longitudinal analysis was not possible. Instead, a cross-sectional analysis was selected as more appropriate and feasible.

The scope of the remaining two questions (RQ3 and RQ4) and a current lack of measurement across countries, however, do not allow for large-N analyses. Nonetheless, both RQ3 and RQ4 are instrumental in responding to the central research question aiming to explain whether FOI laws and open government data work as anti-corruption policies and contribute to government accountability. As the theory presented in chapter 2 proposes, news media and civil society are important intermediaries who transmit information about the corrupt behaviour of political representation to the public. Unrestricted access to news media and favourable conditions for its free operation are instrumental for the released information to have consequences for the political representation. Therefore, it is crucial to explore how journalists and civic activists engage with the information obtained through FOI requests and released as datasets to understand the extent of this impact better. No less importantly, we need to know what kind of information and data is being disclosed. However, at the moment, this is not feasible to do from a comparative perspective due to various measurement constraints. Therefore, at the moment, there

are only isolated studies that explored how journalists and civic activists used FOI laws and open data in different national contexts, e.g., the US (Camaj, 2018; Dyer, 2016; A. R. Schrock, 2016; A. Schrock & Shaffer, 2017) and Europe (Davies, 2010; Janssen et al., 2012; Susha, Zuiderwijk, Janssen, & Grönlund, 2015; Worthy, 2015a; Zuiderwijk & Janssen, 2014).

Conducting systematic comparative research on the use of FOI laws and open government data on a global scale is difficult for a number of reasons. Some, not all governments publish only the overall statistics on FOI compliance, i.e. the number of accepted and rejected requests categorised into very broad topics. The full text of requests and agencies' responses are largely unavailable. Few governments, for example, the Mexican federal government, have a centralised government platform for submitting FOI requests[29]. The absence of tools for systematic receiving and archiving FOI requests makes investigating the whole population of FOI requests practically impossible. It also largely impedes any meaningful cross-country comparative analyses.

As Michener and Worthy (2018) argued, new platforms launched and run by CSOs offer a partial solution as they are public, and thus provide a publicly available archive of requests and responses from public agencies. However, FOI requests submitted through these platforms do not constitute a representative sample, as they are tied to a particular communication channel and some user groups, for instance, journalists, might purposefully avoid due to a lack of confidentiality. Still, these platforms provide a wealthy and at the moment, the only source of data on the public demand for information. The uses of open government data are even more difficult to trace since the data is public, and anyone can download it. Often, the only information available (and only to the platform owner) is the number of page visits and views. If government platforms require users to register, then the government has some basic statistics on the type of users and data they downloaded, but it does not know how they used the data.

FOI and open government data platforms vary across jurisdictions significantly, and this variation does not allow for a large-N analysis. Therefore, to answer the central research question, and, in particular, sub-questions RQ3 and RQ4, I adopt a country case study approach, which in itself combines both quantitative and qualitative methods. The central research question as well as sub-questions contribute to the scientific understanding of the effects of FOI laws and open government data and are also of substantial normative importance to democratic theory as the right to information is one of the fundamental democratic values. From the perspective of

---

[29] See the platform at https://www.infomex.org.mx/.

political representation theory, it is imperative who can exercise this right and how. Pitkin (1967), who provides one of the most comprehensive accounts of the concept of political representation, argues that in its simplest terms, it is about making citizens **present** in policymaking processes. Being able to ask questions and receive answers is its basic prerequisite.

3.4 Case selection and methods

This thesis employs a mixed-methods approach to answer the above research questions. The mixed-methods approach aims to generate a better and more in-depth understanding of the researched phenomena (Greene, 2007). It originates from the assumption that different methods are equally legitimate for the analysis. At the same time, none of them separately is sufficient to answer the research question fully (ibid). Therefore, it is necessary to use them in combination to increase validity and reliability, and address the complexity of the research phenomenon (ibid). The proponents of mixed-methods approach argued that the division between quantitative and qualitative scholars is counterproductive, as instead of focusing what method or methods answer the research questions most appropriately and accurately, it makes false assumptions that some methods are superior to others (Lieberman, 2005). King et al. (1994) made a case that methods should, first and foremost, serve its original purpose of helping to answer research questions. If one method suffices to provide a convincing answer, then a researcher should go with that one. If several methods are needed, then the mixed-methods approach is preferred. Coppedge (1999) posited that the combination of large-N and small-N analysis counterbalances the weaknesses of either. The findings of quantitative analysis can yield generalizable results, and qualitative inquiry can assist in developing thicker concepts and theories (ibid).

There are several rationales to use the mixed-methods approach, i.e. triangulation, complementarity, development, initiation and expansion (Greene, 2007; Greene, Caracelli, & Graham, 1989). Triangulation seeks to strengthen the validity of the results by using different methods simultaneously to explain the same phenomena. Using a mixed-methods approach for purposes of complementarity means to acquire broader, multi-faceted and comprehensive knowledge of various aspects of the researched phenomena through the use of multiple methods. The goal is to provide the findings with additional layers of explanations.

The mixed-methods approach is also used for development when findings obtained through a specific method prompts using an additional method to produce more nuanced findings. The first method helps to shine a light on a particular

phenomenon and inform how it can be further explored. The mixed-methods approach for development is sequential. Usually, the quantitative analysis provides the insights into general trends of the researched phenomenon and is followed by an in-depth qualitative inquiry that focuses on a specific aspect. The main aim of qualitative findings here is to support or further explain the findings of quantitative analysis. An example of a mixed-methods approach used for development is a nested analysis (Lieberman, 2005). The nested analysis uses the results of a large-N analysis to guide the small-N case selection to increase the validity of the findings and provide in-depth insights. If the preliminary results from large-N analysis yield robust and satisfactory results, then the small-N analysis is used to test the theoretical model. If the results are not satisfactory, then the small-N analysis is used to build a new theoretical model.

Scholars also employ a mixed-methods approach to get novel insights and perspectives (initiation). Lastly, a mixed-methods approach can also be used with the aim to expand the scope of the research project. In such a case, different methods are used to study different phenomena (Greene, 2007). Greene et al. (1989) argued that the research design of a mixed-methods study is either guided by the priority assigned to specific methods or by the data collection process.

This thesis is located within the varieties of mixed-method approach as a sequential combination of quantitative and qualitative methods (mixed methods for development) to achieve in-depth knowledge of the effects of FOI laws and open government data on corruption control and government accountability and their use (mixed methods for complementarity).

I employ the comparative cross-country analysis to test the theoretical framework presented in the previous chapter. Subsequently, the results of the comparative cross-country analysis guide the selection of the case study, which explores how different user groups engage with FOI laws and open government data. The country case study combines an expert survey and qualitative interviews on the uses of FOI laws and open government data, and quantitative text analysis of FOI requests submitted through WDTK between 2008 and 2017. The overview of these methods and their linkage to the research questions are offered in Table 3-2.

Denscombe (2009) argued that both quantitative and qualitative data have their pros and cons with regard to their initial objectives. Quantitative comparative cross-country analysis provides insights into the effects of the factors that contribute to lower levels of corruption. Nonetheless, as Ragin (2014) posited, in the quantitative comparative analysis, it is often difficult to identify the causal combinations that lead to the outcome. It provides with a breadth of knowledge and helps to identify patterns, but it does not offer in-depth knowledge.

Table 3-2: The overview of employed research design

Research question	Method	Data source
RQ1. Do FOI laws and the availability of open government data affect corruption control across countries?	comparative cross-country analysis + IV	WGI + others
RQ2. If so, what does facilitate these effects?	comparative cross-country analysis + IV	WGI + others
RQ3. Who does use FOI laws and open government data, and for what ends?	expert survey + qualitative interviews	Own collection
RQ4. What information do FOI requesters seek, and how is that different from pro-actively provided open government data?	QTA – structural topic models	WDTK + Data.gov.uk + own compilation

Note: IV stands for instrumental variable, QTA for quantitative text analysis, WGI for Worldwide Governance Indicators and WDTK for online participation platform WhatDoTheyKnow.

Therefore, it is often combined with case studies that compensate for that (George and Bennet, 2005; Flyvbjerg, 2006; Yin, 2009). I decided to complement the quantitative cross-country analysis with a single country case study. I identified the UK as the most likely crucial case as it both ranks high on open government data and FOIA, as well as on Control of corruption measure. The rationale of the most likely case is that if the expected outcomes do not occur in this case, they will hardly occur in other cases with less favourable conditions (Levy, 2008). As for methods use, Gerring (2007) argued that case studies could use both quantitative and qualitative methods.

3.4.1 <u>Comparative cross-country analysis and addressing endogeneity concerns with instrumental variables</u>

Comparative politics has long played an important role as a subfield of the study of politics. Cross-country empirical analyses proliferated in this political science subfield in the second half of the 20th century as a result of the availability of national-level data in different areas[30]. International institutions established after the World War II with the goal to contribute to reconstruction efforts, such as the World Bank, started to collect data in the area of agriculture, economy, education, environment, health, or public infrastructure, which allowed for quantitative comparisons across countries (Jackman, 1985; Schmitter, 2009). Since then, the popularity of global ratings and rankings has soared. As Cooley and Snyder (2015) argued in their book Ranking the

---

[30] See, for example, the comparative research by Dahl in Europe (1966) or Lewis in West Africa (1965).

World: Grading States as a Tool of Global Governance, most of the political and socio-economic issues are now quantified, and governments are usually very responsive to how they are being assessed through these ratings and rankings.

The quantification of complex political and social phenomena, such as corruption, accelerated empirical comparative cross-country research. Since many of the available measurements have already existed for decades, they have generated abundance of longitudinal (panel) data, which are, from the methodological point of view, generally preferred over cross-sectional data, as they can substantially improve causal inferences. However, estimating longitudinal models is not always possible, which is also the case of the present thesis. Since the measures of open data are new and available only for limited years (2013 - 2016) across which there is almost no variation in the data for outcome variable (corruption control), the cross-sectional approach had to be adopted to estimate the effect of open data and FOI laws on corruption control. Cross-sectional studies analyse observational data at a specific point in time as opposed to longitudinal data that captures dynamic changes over time. They can be quantitative (applied to large samples) as well as qualitative (applied to small samples) (Gerring, 2012). Due to the availability of different measurement across countries, I adopt a large-N cross-sectional analysis (n = 110 countries as units of analysis). The estimated models are specified in detail in the relevant empirical chapter (chapter 4).

While cross-sectional studies are instrumental in revealing correlations and patterns, they have severe limitations in providing causal inferences (Jackman, 1985). Since they observe researched phenomena at a certain point in time, they make it extremely difficult to disentangle a cause from an outcome. This is particularly true in the area I study in this thesis – the relationship between transparency and corruption as a proxy of government accountability. Previous research on transparency and corruption often mentions potential endogeneity problems. Scholars emphasised that many corruption measures, in this analysis CPI as an element of CoC, in particular, are constructed through eliciting opinions of experts, which is an imperfect measure. Although experts are knowledgeable of country contexts, Peisakhin and Pinto (2010) argued they might fail to untie the nature of the causal relationship between policy and changes in behaviour. Endogeneity problems may occur, i.e. corruption may well be caused by factors other than an identified predictor, but which are correlated to that predictor. This might lead to biased ordinary least squares (OLS) estimators for the effect of the predictor, as they will absorb the effect of omitted variables. Reverse causality and measurement errors may also cause endogeneity.

For example, one may argue that less corrupt countries, might self-select themselves to pass FOI laws and open up government data. Thus, the causal relationship between FOI laws and open government data on one hand and corruption levels on the other might well run in the opposite direction. While this is plausible, and less corrupt countries might opt-in for more transparency, more corrupt countries might be motivated equally if not even more, to pass but not properly implement FOI laws or increase access to government data to signal they are too transparent and accountable. Schnell (2018) demonstrated that this was the case in Romania, where levels of corruption are high. Scholars emphasised that, in particular, open data can easily be misused for these purposes as they do not require a hard-accountability commitment (Peixoto, 2013; Yu & Robinson, 2012). On the empirical level, the issue of endogeneity can be addressed through instrumental variables that account for the effect of omitted variables, potential reverse causality and measurement error (Angrist & Pischke, 2015; Bascle, 2008; Semadeni, Withers, & Trevis Certo, 2014). The adopted strategies to identify an instrumental variable and its application are discussed in detail in the relevant empirical chapter (chapter 4).

3.4.2 Expert survey and semi-structured interviews

Different methods, for example, interviews or surveys, might be employed to explore how British investigative journalists engage with FOI legislation and open government data. Gillham (2000) argued that both have its advantages and disadvantages. Interviews provide the researcher with rich materials, as interviewees have enough time to express their views fully. If prompting questions are well-timed, interviewees might reveal information that would not be accessible for the researcher using other methods. However, at the same time, interviewing is a resource-intensive method (ibid). It comes with high costs, both in financial and human resources terms. Setting up interviews, holding them at different locations convenient to interviewees, and transcribing them requires resources in terms of time and finances. Besides, interviews expect a certain time-commitment from interviewees, which might affect their decision to engage with the research in the first place.

Precisely for these reasons, I eventually decided to conduct an online survey on the use of FOI legislation and open government data among journalists instead of interviewing them. Initially, I tried to recruit journalists for interviews, but that proved to be difficult. I had to adapt and adjust the methods to capture the views and experiences of journalists as their input was crucial. I complemented the survey with a small number of interviews with CSO representatives who were more accessible.

The survey does not provide with the same richness of information and insights as interviews for different reasons. For instance, it does not allow for picking up non-verbal cues. If the survey is administered online, respondents do not have a chance to consult questions and seek clarifications, which might lead to omitting some questions or guess-answering. However, it has also its advantages over interviews. It has the potential to reach wider audiences and thus, provide more generalizable findings. The main aim of the survey is to explore how journalists in the UK engage with FOI legislation and proactively published datasets. It is important to note that there are two FOIAs in the UK: FOIA 2000 that governs access to information held by all public authorities in England, Northern Ireland and Wales and a separate FOIA 2002 that regulates access to information held by all Scottish public authorities. Therefore, the emphasis in the survey was put on the central UK government and the Scottish government. The full text of the survey is provided in Appendix 2.

Fowler Jr. and Cosenza (2008) argued that every survey is only as good as its design is. If surveys are poorly designed, neither they will yield any meaningful results nor will they help to explain a phenomenon in question. I followed standard procedures when designing the survey in order to increase reliability and validity of the results and minimise potential sources of errors, which often occur mainly at the design stage (Sudman & Bradburn, 1974).

Non-response represents one of the most common sources of survey error. It can be influenced by different factors, from the lack of knowledge to answer a particular question to lack of time. The greater time commitment from respondents is expected, the higher the chance they will decide not to participate in the research. First and foremost, to address this issue, the actual survey was preceded by a brief but informative Participation information sheet (PIS), provided in Appendix 2. PIS explained the main aims of the survey and located it in the wider context of my research project and explained the character of the collected data and how it will be managed.

It is also essential to keep the questionnaire short and include a minimum of open-ended questions to minimise non-response (Reja, Manfreda, Hlebec, & Vehovar, 2003). Prior to designing the survey, I established that it must be possible to complete the survey within five minutes. For the sake of time, the survey also followed a certain displaying logic that allowed respondents to skip questions that were irrelevant to them. I also tried to avoid open-ended questions. Overall, I included only three, one at the end of each survey section and one at the end of the survey. Two questions prompted respondents to describe in their own words the benefits of FOIA and open government data for their journalistic work. The last question invited them to provide any comments or insights on the overall topic of the survey or the survey itself.

However, non-response might also occur in closed-ended questions when provided answers do not reflect respondents' true views and experiences (Reja et al., 2003). There are several ways how to address this issue, e.g. by providing an "other" option and making responding a voluntary act.

Gillham (2000) argued that asking respondents for very precise recollections of their past experiences might leave the researcher with inaccurate and uninformative data. Relying on memory can be partially mitigated by offering a range of options. For instance, in the present survey, the journalists were asked to state how many FOI requests they submit per year, but they were not required to come up with a number. Instead, they were offered several options (less than 10, 10 − 50, 51 − 100, more than 100). Alternatively, an option "don't know" and "don't remember" can be included to avoid that respondents are forced to give an ultimate answer even when they are uncertain.

The use of jargon is undesirable, as it might hamper consistency of understanding, and again lead to errors. If the respondents do not have a common understanding of the concepts in question, their insights will not be able to generate valid results (Fowler Jr. & Cosenza, 2008; Gillham, 2000). They might try to guess answers, skip questions or give the survey up entirely. The language of the survey, thus, needs to consider its audience. In order to strengthen the understanding across respondents, any word or concept that might be ambiguous can be explained in brackets. Since the present survey is targeted at expert respondents who are expected to understand the terminology in the FOI and open government data field, the use of jargon was not a major concern here. The survey assumes that a journalist will be aware of government websites and databases, such as Data.gov.uk or platforms for journalists run by CSOs, such as WDTK, which might not be true of every journalist. However, any questions that used an overly legalistic language, in particular, questions on the exemptions from FOIA, included an explanation of each term. For instance, class-based and prejudice-based exemptions were described in more detail providing illustrative examples.

Non-response is also more likely to occur when respondents are unwilling to share their views and experiences, e.g. when researching sensitive topics (Fowler Jr. & Cosenza, 2008). Rasinski et al. (1999) argued that ensuring high privacy standards might help to overcome a lack of willingness to participate. Although the survey does not explore sensitive issues as such, the piloting process revealed that journalists might not feel comfortable about sharing what their sources are, even in very general terms. For these reasons, I decided not to collect any personal data. The survey is entirely anonymous. Although it asks some socio-demographic questions, such as gender and age, it does not collect any additional information, such as IP addresses, that would

allow for identifying the respondents. In addition, there are no required questions in the survey, so if respondents feel uncomfortable about sharing their age, they can skip the question.

Apart from non-response, errors in surveys might come from social desirability bias when respondents "like to appear to be other than they are" (Brace 2008: 195). Although at first sight, social desirability bias might not be an issue in the present survey, as it does not aim to capture behaviour or attitudes where a standard desirable norm has been established. As Bradburn, Sudman, and Wansink (2004) argued, we have a common understanding about what it means to be a good citizen, well-informed, cultured and socially responsible person and we would like to match these definitions. Generally, there is no such thing as "the right use" of FOI and government data. However, some journalists might have underestimated the number of vexatious requests they made, as these are perceived as a clear misuse of FOIA because they lack a serious purpose or put an undue burden on authorities. Social desirability bias, in this case, might also come from what the respondents see as a socially desirable outcome. This phenomenon is called instrumentation (Brace, 2008). For illustration, they might think that if the survey results reveal substantial use of FOIA, the government could not discard FOIA as irrelevant.

These are just a few issues that could have affected the reliability and validity of the survey results and had to be considered at the design stage. Similarly, the best sampling strategies had to be considered. The survey respondents were recruited using homogenous purposive sampling. The sampling and recruiting strategies are described in greater detail in the relevant empirical chapter (chapter 6).

Undoubtedly, the survey on the journalistic use of FOIA and open government data can provide interesting insights about what their preferred source of information is, what information they are interested in and what challenges they face in obtaining it. However, the survey has its limitations. As flagged up above, even if the response rate is satisfactory, the survey findings are not representative of the population of journalists in the UK. However, alternative ways of researching the journalistic use of FOIA and open government data also do not come without their own flaws.

Some scholars (Hamilton, 2016) examined the full texts of news investigations in the US, which were submitted for the annual prize competition of IRE and searched for the key terms that would suggest the use of FOIA. Nonetheless, as he acknowledged, this approach also comes with limitations. First, journalists might use terms other than those envisaged by Hamilton. Second, the journalistic use of FOIA is not straightforward. FOIA can be used in very different stages of investigations and might end up not being mentioned at all in the final published piece. For example, some investigations start with FOI requests fishing for information in different areas.

Others' main source of information might be a confidential source or a whistle-blower or an anonymous leaker, and FOIA is used only to confirm their cues.

Also, the survey on the journalistic use of FOIA and open government data disregards potential other uses. Journalists are not the only major group of FOI requesters, quite the opposite. As Kwoka's (2016, 2018b, 2018a) research suggests, in the US, other FOI user groups prevail, such as private companies and individuals interested in their own data which they failed to obtain using other means and see submitting an FOI request as a last resort. There is no reason to assume that FOI users are significantly different in the UK.

Thus, while the expert survey will help to shine a light on the journalistic uses of FOI and open government data, other purposes and their extent will remain concealed. Second, as was previously mentioned, for a variety of reasons, survey respondents might not always provide their true views. Sometimes, they might skew reality to follow their ends or look in a better light. Last, survey respondents have to rely on their memory to respond how many FOI requests they submitted, in what topics they were interested, and more importantly, they might decide not to engage with the survey at all. Laver, Benoit, and Garry (2003) argued that quantitative text analysis might address these problems and provide additional evidence to corroborate the findings from other methods, such as surveys or interviews.

3.4.3 Quantitative text analysis

Not only the advancements in digital and computational technologies have contributed to the emergence of open government data – a research subject of this thesis, they have also transformed research practices. New computational research methods have enabled to study an unprecedented amount of information and data. As for this thesis, in addition to surveying journalists and CSO representatives who might engage with FOI laws and open government data, the content of FOI requests and government datasets published on open data portals could also be fruitfully explored. Decades ago, analysing the texts of FOI requests would have been either impossible or a very costly and years-long endeavour due to their large volume. However, the rapid development of new methods and software, enabled by the growth of computing processing power, significantly decreased the costs associated with this kind of research.

I decided to employ a quantitative text analysis (QTA) to study the content of FOI requests. QTA is different from traditional content analysis as it approaches the text as "collections of word data" (Laver et al., 2003). The first aim of this analysis is to corroborate the survey evidence and explore other uses of FOI legislation that

the survey did not cover. The second aim is to explore what transparency FOI delivers. In other words, what the public wants to know, in what topics they are interested in, and how the prevalence of topics is linked to the request outcome or particular government and prime minister. Is the central government more responsive when it comes to some topics than others? Do topics reflect the major political events? Lastly, are topics of FOI requests also represented in datasets published on open data portals? The quantitative text analysis makes researching these questions feasible and less labour-intense (Grimmer & Stewart, 2013; Laver et al., 2003).

Grimmer and Stewart (2013) pointed out that QTA might be useful but never fully replaces thorough and careful readings of the texts. Rigorous validation is necessary to arrive at meaningful, reliable and valid conclusions. The automation of the analysis makes the process more feasible but does not remove the requirement to engage with the analysed texts. The researcher's role in using QTA is critical for guiding the process, making decisions and interpreting the results of the analysis (ibid). There are two main types of QTA: supervised and unsupervised. When employing supervised QTA, a researcher has already assumptions about categories that will appear within the text, for example, based on previous research. An unsupervised method is usually selected when it is difficult to derive these categories beforehand. The main aim of the unsupervised method is, thus, to estimate the categories within the text and assign individual texts that are linked to these categories (ibid). Both methods are complementary. For instance, unsupervised QTA can yield categories for the subsequent supervised QTA.

Since its inception, FOI legislation has attracted the attention of scholars in different fields, from political economy to human rights law. The available research covers various aspects of FOI laws – their diffusion, their impact on democracy and economic development, and their role for investigative journalism. However, the content of FOI requests is a largely unexplored research terrain. There are several studies examining the topic prevalence within FOI requests sent to Mexican federal government agencies (Berliner et al., 2018), and which topics are related to the government's unresponsiveness (Bagozzi, Berliner, & Almquist, 2019).

However, FOI is country and context-specific. Also, FOI legislation is universal in the sense that anyone (subject to the legal definition provided in different jurisdictions) can ask the government for any information that it has at its disposal. This universality of FOI laws makes predicting the variation in FOI requests' topics difficult, and the employment of the unsupervised method more appropriate and useful. Grimmer and King (2011) also made a case for the use of unsupervised methods in understudied research areas. I decided to estimate a structural topic model, as it enables using the metadata about the texts as covariates. Topic models as an

unsupervised method require extensive validation (Quinn, Monroe, Colaresi, Crespin, & Radev, 2010). Some of the validation techniques include, for example, the assessment of semantic and predictive validity (Quinn et al., 2010). Semantic validity reflects the extent of homogeneity within the topic and its distinctiveness from other topics. Predictive validity means that major external events will be reflected in the topics as an increase in interest in a specific topic (Grimmer, 2010).

Altogether, the combination of selected methods, discussed above, should generate new knowledge about the impact of FOI laws and open government data on corruption control and government accountability and shed light on how these transparency policies are used by different user groups.

4 Do FOI laws and open government data work as anti-corruption policies? Evidence from a comparative cross-country analysis

As was outlined in previous chapters, access to government information has long been perceived as critical for democracy and development. New digital technologies have created new opportunities for government-citizens interactions in the online environment. Many governments used this opportunity to demonstrate greater transparency and in addition to existing FOI laws, started proactively publishing its data in open formats for people to use and reuse it. In the past decade, open government data has become central to the global anti-corruption agenda and political campaigns. While research on the effect of FOI laws on corruption is available (Costa, 2013; Vadlamannati & Cooray, 2017), the impact of open government data on corruption has only seen a preliminary investigation due to its novelty and a lack of measurements.

In this chapter, I aim to fill in this gap. First, I present the hypotheses. I argue that both FOI laws and open government data have limited effect on corruption control unless accompanied by robust provisions safeguarding media freedom and an independent and accountable judiciary. Then, I detail what data I use to test these propositions and how I measure it. I also discuss potential measurement problems and how I address them. I then present my estimation methods and the results, which, after instrumenting for the availability of open government data to address its potential endogeneity, corroborate some of my expectations. At last, I discuss the limitations of the study and what implications my findings have for policies advancing access to government information in different political and economic contexts.

4.1 Aims and hypotheses

Drawing from the theoretical framework introduced in chapter 2, I propose the following hypotheses:

> **h4-1a.** Countries with greater availability of open government data are more likely to exhibit lower levels of corruption than countries with limited availability.

h4-1b. The relationship between open government data and corruption is conditional upon the quality of media freedom.

h4-1c. The relationship between open government data availability and corruption is conditional upon the degree of judicial accountability and independence.

h4-2a. Countries with a long tradition of FOI laws are more likely to exhibit lower levels of corruption than those where FOI legislation was introduced just recently.

h4-2b. The relationship between FOI laws and corruption is conditional upon the quality of media freedom.

h4-2c. The relationship between FOI laws and corruption is conditional upon the degree of judicial accountability and independence.

4.2 Data and measurement

Since data on the availability of open government data exists only for years from 2013 to 2016[31] and the variation of data on corruption levels is minimal within this period; longitudinal analysis was excluded as a possible estimation method. Instead, cross-sectional analysis using the latest available data from 2016 was employed.

4.2.1 <u>Dependent variable: Control of corruption</u>

Low levels of corruption are a sign of governments which are responsive to their constituents and responsibly governs public resources. In this analysis, I use the Control of corruption score[32] from the Worldwide Governance Indicators as a measure of corruption. It is an aggregate and robust measure compiled from a variety of indices, such as Afrobarometer, Latinobarometro, Transparency International's Global Corruption Barometer (GCB) and Corruption Perception Index (CPI), World Economic Forum Global Competitiveness Report and others. The estimate ranges from -2.5 to 2.5. A negative score indicates high, and a positive score indicates low levels of corruption. In other words, countries that are able to control corruption well have positive score and countries that are unsuccessful in controlling corruption have

---

[31] See Open Data Barometer at https://opendatabarometer.org/?_year=2017&indicator=ODB and Global Open Data Index at https://index.okfn.org/

[32] See https://info.worldbank.org/governance/wgi/pdf/cc.pdf.

a negative score. In 2016, corruption was highest in Equatorial Guinea with the -1.81 estimate and lowest in New Zealand with the 2.28 estimate.

Control of corruption score captures both petty and grand forms of corruption, as well as *state capture* by elites and private interests. It is a robust measure that can withstand the criticism that is often raised of the components of Control of corruption individually (Andersson & Heywood, 2009; Heywood & Rose, 2013). Measures of corruption perception, in particular, CPI, are criticised mostly for its expert nature. Scholars argued that expert evaluators might only repeat common knowledge about particular countries and perpetuate labelling them *corrupt* or *non-corrupt* even when the situation has already changed (Heywood & Rose, 2013; Rose & Mishler, 2010).

Other scholars (Mills, 2017; Sharafutdinova, 2010) made an important point that in hybrid regimes, (false) accusations of corruption are often used to gain an advantage or eliminate political competition, which results in driving the overall public perception of corruption in a country up. CPI also captures grand corruption mostly, i.e. cases of overpriced tenders in the state administration, but does not reflect petty corruption among citizens, for example, to access essential public services. However, as available research in the field demonstrated, although grand and petty corruption is connected more broadly, discrepancies between them might be substantial in some cases. While citizens might have a high perception of corruption if they are commonly informed about poor management of public resources, they might still have good access to public services without relying on bribing (Heywood & Rose, 2013; Rose & Mishler, 2010).

At the same time, however, indices measuring actual corruption suffer from flaws of other types. For instance, GCB, the largest world-wide public opinion survey on corruption, is likely to suffer from the compromising effect of social desirability bias. Respondents might not tell about the bribes they had to pay. They might be reluctant to speak the truth, fearing potential repercussions for their dishonest conduct or simply not wanting to be associated with socially unacceptable behaviour (Arnold & Feldman, 1981; Chung & Monroe, 2003). Using Control of Corruption for the analysis is, thus, a reasonable choice, as aggregating several measures of corruption can compensate for their individual imperfections to some extent.

4.2.2 Independent variables

The availability of open government data is measured as the Implementation sub-index of Open Data Barometer, an index designed by the World Wide Web Foundation in 2013. It is a continuous variable measuring the availability of key

government datasets in open formats. The score scale ranges from 0 to 100. In 2016, the lowest achiever was Mali with an 8.67 score, and the highest achiever was the UK with a 91 score. Open Data Barometer's sub-index examines the availability of detailed budget data, government spending data, data on public contracts, company registration data, international trade data, map data, land ownership data, national environmental statistics, public transport timetable data, crime statistics, detailed census data, legislation data, election results data, primary and secondary education performance data and health sector performance data (World Wide Web Foundation, 2016).

It is important to note that differences in the open government data availability score between individual datasets are substantial. The highest mean of data availability score for all countries is for census data (μ = 53.91, Md = 65, n = 115), the lowest mean is for spending data (μ = 11.87, Md = 5, n = 115). As can be observed in Figure A-1 in Appendix 1, which provides supplementary materials for this chapter, countries that are considered free and democratic release more data on average. However, all countries are reluctant to release data on contracts, company registration data, and land ownership data, i.e. data that are of utmost importance for monitoring the use of public resources and exposing patterns of corrupt and fraudulent behaviour.

Open Data Barometer offers broad geographical coverage, and from available open data indices, has the most stable methodology over time. Nonetheless, it has its limitations too. While the measure evaluates against the open data criteria, i.e. whether data is machine-readable and reusable, openly licensed, regularly updated, provided free of charge and many other indicators, Open Data Barometer does not assess reliability, accuracy or quality of the data. Therefore, taking Open Data Barometer at face value assumes trust that the government has been collecting and managing its data rigorously and also providing it in its entirety in the public domain.

However, this might not always be the case. Some governments might be tempted to censor or doctor the data that could reveal their poor performance in certain areas. In summer 2019, Brazilian President Jair Bolsonaro suggested that the data published by the National Institute for Space Research, the government agency in charge of monitoring deforestation, will require prior approval from the administration following the publication of figures that showed 88% rise in deforestation in a year[33]. Similarly, the example of Tanzania which recently amended its Statistics Act to restrict disseminating any statistical data distorting or discrediting

---

[33] The article on the rise of deforestation in Brazil is available here. The government platform, monitoring deforestation and publishing satellite data, is available here.

the official statistics, and thus limited their citizens' ability to question official government data, may suggest why cautiousness about data quality is substantiated (World Bank, 2018). Other infamous cases in the past also suggest that there is a reason for scepticism about reliability, quality or entirety of data provided by the government. For example, the Hungarian government led by Ferenc Gyurcsány from 2004 to 2009 concealed essential data about the state budget from the public (Dempsey, 2006).

The tradition of FOI laws is measured as the <u>number of years since FOI legislation has been adopted</u> and enforced. It is a continuous variable, ranging from 0 for countries that do not have FOIA in place to 250 for the Swedish FOIA that has been adopted in 1766. The length of the FOI law tradition in a country is a useful proxy for measuring the openness of public authorities and the implementation of FOI laws. For instance, Scandinavian governments that were among the firsts to adopt FOI laws, are well-known for their positive approach to transparency, which they perceive primarily as public access to government information (Grønbech-Jensen, 1998). Having decades-long experience with FOI laws is also likely to affect their implementation and enforcement positively. Measuring the impact of FOI laws through the length of their existence has, of course, limitations. For instance, it can be conflated with the overall democratic tradition in a country. A cross-country measure of the quality of the implementation of FOI laws would be desirable for the analysis. However, since it does not exist, I argue, the tradition of FOI laws is the best available measure. Although there is a measure of the quality of the FOI laws' design (RTI Rating), as I argued in chapter 2, the practice can be miles away from the letter of the law. In other words, *de jure* FOIA does not necessarily reflect *de facto* FOIA. For instance, as can be observed from Table 1-1, according to the RTI Rating, Afghanistan has the most comprehensive FOI legal framework. However, it is classified as a not free country. The implementation and enforcement of FOIA in such conditions might be problematic.

For measuring <u>media freedom,</u> I use the Freedom House's Freedom of the Press Index and the number of murdered journalists in a given country in a given year. The Freedom of the Press Index is based on experts' survey responses. Experts, usually media scholars assess legal, political and economic media environment in a particular country. The range of the score goes from 0, indicating that a country has none or close to none media censorship and enjoys the highest levels of media freedom to 100, which suggest heavy restrictions and censorship. In 2016, the country with the highest media freedom was Norway, and country with the most severe censorship was North Korea (Freedom House, 2017b).

The Committee to Protect Journalists have collected and published data on assaults on journalists and the press for more than 30 years. They keep evidence of killed and imprisoned journalists, as well as journalists who went missing. As for journalists' killings, the Committee collects data on the causes of the death, whether it was a murder or journalist died in crossfire or combat, or on a dangerous assignment. Since the death of journalists in conflict zones is unlikely to be related to corruption in the first place, I included in the analysis murders only. This variable has additional explanatory power, as Pearson's correlation coefficient between Freedom of the Press Index and murders of journalists was very low (0.15).

Civil society variable is measured as an additive score created from three variables from the Varieties of Democracy (V-Dem) dataset: v2cseeorgs, v2csreprss and v2csprtcpt (Pemstein, 2015). The range is from 0, indicating severely restricted civil society to 11 indicating free civil society. The first variable v2cseeorgs measures CSOs entry and exit conditions, i.e. to what extent they are controlled by the government, and its range is from 0 which represents monopolistic control to 4 which indicates that the government does not hinder formation and operation of CSOs except when they are involved in violent activities, such as planning a coup. The second variable v2csreprss measures the extent of the government's repression of CSOs. It is an ordinal variable with five levels where 0 indicates severe repression of civil society aiming for its elimination and 4 means CSOs are free to organise without fearing repercussions. The last variable v2csprtcpt quantifies the extent of people's engagement in CSOs, and its range is from 0 where associations are government-sponsored, and engagement is often compulsory to 3 where CSOs are numerous and active in diverse agenda, and people are involved at least in one of them.

Judicial accountability and independence are also measured as an additive score created from two variables from the Varieties of Democracy (V-Dem) dataset: v2juaccnt and v2jucorrdc (Pemstein, 2015). The range is from 0, indicating low judicial accountability and independence to 8 showing the opposite. The first variable v2juaccnt measures judicial accountability, i.e. how often are judges removed from their posts or otherwise disciplined if they are found responsible for a serious misdemeanour. It is an ordinal variable with five levels (never = 0, seldom = 1, about half of the time = 2, usually = 3, always = 4) and for the analysis, it was converted to a continuous variable. The second variable v2jucorrdc assesses judicial independence, i.e. how often do individuals or businesses pay bribes to speed up or delay the process, or to obtain a favourable judicial decision. Again, it is an ordinal variable with five levels (always = 0, usually = 1, about half of the time = 2, not usually = 3, never = 4) ordered in reverse order. These two variables were merged for the analysis.

4.2.3 Control variables

Free and fair elections are also measured using a variable from the V-Dem dataset, which defines free and fair elections as the elections free from registration fraud, systematic irregularities, government intimidation of the opposition, vote-buying, and election violence (Pemstein, 2015). The variable is an aggregate measure consisting of other V-Dem indicators. In the original dataset, it is an interval variable ranging from 0 to 1, converted to 0 to 100 scale for better interpretability of the results. In the literature, the free and fair election is described as the essential government accountability mechanism (Smulovitz & Peruzzotti, 2000).

GDP per capita data (in 1,000 USD) from the World Bank is used as a control variable as previous research in the field has demonstrated that the level of economic development and growth matters for an absence of corruption. GDP can also serve as a proxy for assessing wider living standards and literacy, as it is associated with other measures, such as the Human Development Index (HDI).

Table 4-1: Descriptive statistics

Variable	N	Mean	St. Dev.	Min	Max
Control of Corruption	193	-0.1	1	-1.8	2.3
Corruption Perception Index	172	42.8	19.3	10	90
Bribery rates	110	22.9	17.1	0	77
Open government data	115	32.4	18.7	6	91
Budget data	115	40.4	32.2	5	95
Spending data	115	11.9	21.8	0	95
Data on contracts	115	23.7	25.6	0	95
Data on companies	115	20.5	26.4	0	100
Land ownership data	115	13.6	19.3	0	90
Years of FOIA	196	9.2	20.5	0	250
Press freedom	195	48.5	23.8	9	97
Murdered journalists	196	0.1	0.3	0	2
Civil society	171	7.8	2.9	0	11
Judiciary	171	4.3	1.8	0	8
Free and fair elections	171	59.6	30.1	0	98
GDP (in 1,000 USD)	190	13.3	20.5	0.25	165

4.2.4 Estimation method and dealing with endogeneity

Linear models were used to test for a significant effect of open government data, relative to FOI laws, and other factors on corruption levels. I controlled for the presence of free and fair elections, and gross domestic product per capita. I estimated

the model using an ordinary least squares (OLS) regression as based on the proposed theory I expect the relationship between access to information and corruption as a proxy for government accountability to be linear with greater transparency leading to better control of corruption. All final models for estimating the effect of FOI laws and open government data on corruption levels were visually checked (diagnostic plots) for normal distribution of residuals, constant variation (homoskedasticity), influential observations and independence of residuals. In addition, Shapiro-Wilk (a formal test of the normal distribution of residuals) and Breusch-Pagan (a formal test of heteroskedasticity) tests were conducted. These formal tests, together with a visual examination of diagnostic plots, suggested that the models meet the assumptions of general linear models. For a detailed description of these tests and diagnostic plots, see Appendix 1. All analyses were conducted in R (v3.5.1).

In Chapter 3 that discusses the research design and methodological issues in more detail, I outlined potential endogeneity problems that are associated with OLS models applied to observational data. The estimators produced by OLS can be biased for a number of reasons (some of them described in Chapter 3) and lead to false causal inferences (Wooldridge, 2002). The decisions to adopt FOI laws and open data policies are not random, and as researchers, we cannot obviously affect them. The ethical possibilities for conducting randomised control trials to study the effect of transparency policies (as other government policies) are extremely limited if any at all (Bascle, 2008).

On the empirical level, the issue of endogeneity can be addressed by identifying the instruments and thus accounting for the effect of omitted variables, potential reverse causality and measurement error (Angrist & Pischke, 2015; Bascle, 2008; Semadeni et al., 2014). If an exogenous instrument for open government data measure can be identified and open government data maintains a significant coefficient when instrumented in the second stage of the instrumental variable (IV) estimation, it can be concluded that open government data affects corruption rather than vice versa. Sovey and Green (2011) argued that if instrument variables are not formed through random assignment, as is the case in the experiments, the good practice for observational studies is to provide substantial theoretical justification for why the selected instrument is uncorrelated with the error term, and hence with the dependent variable. At the same time, it should be a good predictor of the endogenous independent variable in question.

I identified an instrumental variable that fulfils criteria – membership in the Open Government Partnership (OGP). The OGP's role is to secure actionable commitments from national governments to increase transparency, accountability, and public participation. Governments draft national action plans to formulate their

commitments. So far, more than one-sixth of them were related to open data. I argue that the membership in OGP affects domestic politics and contribute to the global diffusion of open data policies since OGP has monitoring mechanisms in place to assess how governments fulfil their pledges and implement policies to which they committed. As a result, open data commitments in most countries eventually translate into higher availability of government datasets. Nonetheless, at the same time, the membership in OGP does not make countries free from corruption. The CoC estimates for OGP member countries ranged in 2016 from -1.03 score in Nigeria to 2.28 score in New Zealand. In fact, out of 75 OGP member countries in 2016, the majority (42) had a negative CoC estimate, i.e. were considered rather corrupt than non-corrupt. Also, aspiring countries can formally meet the eligibility criteria for OGP membership relatively easily, which is why I argue that the OGP membership is, to a great extent, exogenous and detached from other variables. In 2016, some OGP members did not represent democratic regimes, infringed political rights and civil liberties, and were known for restricting journalists in their work. In 2016, some did not have FOI laws.

4.3 Results

The results are presented as follows. The associations between FOI laws, open government data and corruption levels measured as Control of corruption score are discussed below (see Table 4-2, Table 4-3, Table 4-5 and below). All models are estimated using ordinary least square and two-stage least square regression to address potential endogeneity problems.

Based on the theory proposed in 2, the following models were estimated:

> <u>Model 1:</u> Corruption = α + βopen data + βFOI years + βpress freedom + βmurdered journalists + βcivil society + βjudiciary + βfree election + βGDP + ε

> <u>Model 2:</u> Corruption = α + βopen data + βFOI years + βpress freedom + βmurdered journalists + βcivil society + βjudiciary + βfree election + βGDP + βopen data*press freedom + ε

> <u>Model 3:</u> Corruption = α + βopen data + βFOI years + βpress freedom + βmurdered journalists + βcivil society + βjudiciary + βfree election + βGDP + βopen data*judiciary + ε

> <u>Model 4:</u> Corruption = α + βopen data + βFOI years + βpress freedom + βmurdered journalists + βcivil society + βjudiciary + βfree election + βGDP + βFOI years*press freedom + ε

Model 5: Corruption = α + βopen data + βFOI years + βpress freedom + βmurdered journalists + βcivil society + βjudiciary + βfree election + βGDP + βFOI years*judiciary + ε

4.3.1 Open data and the moderating effect of press freedom

First, I will describe the results from models estimating the effect of open government data on corruption levels. As can be observed from Table 4-2, except for model 3, the findings are consistent with the **h4-1a** hypothesis, i.e. greater availability of government data is significantly associated with higher corruption control (lower levels of corruption). In model 1, holding all other predictors constant, for every score increase in open government data score, CoC score increases by 0.007 scores, i.e. control of corruption is stronger, and levels of corruption are lower.

Interestingly, freedom of the press was not significantly associated with corruption in any model. However, in model 1 and 3, the importance of media freedom for uncovering and tackling bribe and corruption is demonstrated through the variable measuring the numbers of murdered journalists. Every additional murder of a journalist is significantly associated with an increase in corruption levels (decrease in CoC score) by almost 6% in both models (p < 0.05). In a plain language, the risk of death linked to investigative reporting might put off other journalists from investigating and informing about corruption scandals and help corruption to flourish further.

Also, all models (1-3) revealed the importance of the independent and accountable judiciary and the presence of free and fair elections and economic growth for tackling corruption. Holding all other predictors constant, for every score increase in the independence of judicial system CoC score increases by 0.15-0.22 scores, which on the scale from -2.5 to 2.5 represents a substantive increase (4%). GDP per capita is also a statistically significant predictor of corruption with high-income countries being associated with lower corruption. Every additional 1,000 USD of GDP per capita is associated with approximately 0.02 increase on the CoC scale. An increase in free and fair elections score is also significantly associated with higher corruption control (lower corruption levels). The relationship between free civil society and corruption levels goes surprisingly in the opposite direction, i.e. freer civil society is significantly associated with lower corruption control (higher levels of corruption). This finding is against my central argument that dense and active civil societies are a necessary presence in controlling corruption. Therefore, it deserves further exploration.

Table 4-2: The effect of open government data on corruption (OLS)

Control of corruption	Model 1	Model 2	Model 3
Open data	0.007*	0.022***	-0.004
	(0.003)	(0.006)	(0.007)
Years of FOIA	0.002	0.001	0.002
	(0.002)	(0.002)	(0.002)
Press freedom	-0.004	0.008	-0.003
	(0.004)	(0.005)	(0.004)
Murdered journalists	-0.269*	-0.204·	-0.266*
	(0.113)	(0.110)	(0.112)
Civil society	-0.073**	-0.070**	-0.072**
	(0.025)	(0.024)	(0.025)
Judiciary	0.216***	0.212***	0.153**
	(0.032)	(0.031)	(0.051)
Free and fair elections	0.008**	0.010***	0.008**
	(0.002)	(0.002)	(0.002)
GDP	0.020***	0.017***	0.019***
	(0.003)	(0.003)	(0.004)
Open data*press freedom		-0.0004**	
		(0.0001)	
Open data*judiciary			0.002
			(0.001)
Constant	-1.132**	-1.833***	-0.936*
	(0.400)	(0.443)	(0.416)
N	110	110	110
R^2	0.879	0.890	0.882
Adjusted R^2	0.870	0.880	0.872
Residual Std. Error	0.364 (df = 101)	0.349 (df = 100)	0.362 (df = 100)
F Statistic	92.019*** (df = 8; 101)	90.216*** (df = 9; 100)	83.295*** (df = 9; 100)

Note: * p < .05; ** p < .01; *** p < .001, the results in bold are statistically significant.

As I mentioned above, when I described the estimation method used for the analysis, I conducted a number of formal tests. I also visually examined diagnostic plots for all models (available in Appendix 1). Based on their results, I can conclude that the models meet the assumptions of general linear models. I explored a relationship between civil society measure and CoC measure of corruption control and estimated a univariable OLS model to understand better what might lie behind this finding. In this univariable OLS model, the relationship between civil society and corruption control is also significant but has a different direction, i.e. for every score increase in freedom of civil society, CoC score increases by 0.167 scores (p < 0.001), i.e. control of corruption is stronger, and levels of corruption are lower. On the CoC

scale from -2.5 to 2.5, this coefficient represents a substantive increase (3-4%) in the expected direction.

However, although I would normally expect to observe this direction in the multivariable OLS model above, too, the direction is opposite. Literature studying problems that occur in general linear models with many independent variables in different fields point to the potential multicollinearity (Dohoo, Ducrot, Fourichon, Donald, & Hurnik, 1997; Kalnins, 2018). Multicollinearity arises when independent variables relate to each other. I used VIF statistics in all models. Based on the results, I was able to reject potential multicollinearity problems. However, Kalnins (2018) demonstrated that VIF statistics might be misleading if independent variables are correlated via a latent variable, which is common to multiple of them.

One of the approaches to deal with multicollinearity is reducing the number of independent variables (Dohoo et al., 1997). In this case, it would mean omitting civil society variable from the multivariable OLS model. However, there are a number of reasons why this might not be the best approach in the present case. First and foremost, it goes against the central theory presented in this thesis. I argue that dense and active civil society is indispensable on the path from government transparency to better corruption control and government accountability. Therefore, I cannot deliberately remove it from the model only because it fails to confirm one of the proposed hypotheses.

Second, from the statistical point of view, keeping civil society variable in the multivariable OLS model improves the model. I estimated both models – the more complex model including civil society variable and the model without it. Then, I explored the Akaike's Information Criterion for both models to investigate their quality (Bozdogan, 1987). Following the Occam's Razor (principle of parsimony which prefers simplicity), Akaike developed a statistical criterion - Akaike's Information Criterion (AIC), which helps to identify "an optimal and parsimonious model in data analysis from a class of competing models" (Bozdogan, 1987: 346). In practical application, a model that offers the highest information gain with the least complexity has minimum AIC compared to other models. The more complex model with civil society variable had a lower AIC than the model without civil society variable. Therefore, keeping civil society variable in the models is preferable. In addition to AIC, I also performed a likelihood ratio test to test for the significance of civil society variable. The results again confirmed that the more complex model with civil society variable should be retained. That said, the coefficient for civil society measure should be taken with caution.

Models with added interaction effect (model 2 and 3) yield further interesting results. The moderating effect of press freedom (model 2) modifies the size and

strength of the association between open government data and corruption and is consistent with the proposed theory. The association of open data and corruption is greater in the model with press freedom as a moderator. In model 2, for every score increase in open data availability, CoC score increases by 0.02 scores (robust standard error = 0.006, p < 0.001), i.e. levels of corruption go down and control of corruption strengthens. Substantively speaking, this is a 0.4% increase on the scale of CoC (from -2.5 to 2.5). As can be observed from the interaction effect in Table 4-2, for every decrease in the press freedom score, the effect of open data on corruption diminishes by 0.0004. In other words, as the level of media censorship increases the slope of the relationship between open data and corruption becomes more negative.

To unpack the nature of moderating effect, I also conducted a simple slopes analysis, which helps to determine whether the relationship between open data availability and corruption is significant for specific values of the press freedom (Hayes, 2018). I set the values of the press freedom score at 10 (free press, e.g. Norway, Sweden), 30 (moderate government or corporate influence, e.g. Italy), 50 (high levels of violence against journalists or state control, e.g. Albania and Bosnia and Herzegovina) and 90 (not free, e.g. Saudi Arabia). For the press freedom score of value 10, the coefficient is 0.02 (p < 0.001), for the score of value 30, it decreased to 0.01 (p < 0.001). However, for the press freedom score of value 50 as well as 90, the relationship between open data and corruption is no longer significant.

As the Johnson-Neyman plot illustrates more precisely in Figure 4-1, open data availability has no effect on corruption when the press freedom score is higher than 43.08. This finding is consistent with the proposed theory as well as Lindstedt's and Naurin's (2010) results. It means that greater open data availability positively affects corruption and contributes to its greater control in countries which enjoy a free press, i.e. news coverage is vigorous, the safety of journalists is assured, the government interference in media is marginal, and the press is not imperilled by the interests of powerful economic groups. At a level of diminishing press freedom score and increasing censorship beyond a score of 43.08, open data is no longer significantly associated with corruption levels.

This finding is also interesting in light of Schnell's (2014, 2018) work. Using the case study of Romania, she demonstrated that even when the government adopts transparency and anti-corruption policies, in particular FOI laws and asset disclosures for high-level officials, as *cheap talk*, they might still "generate costs for decision-makers" and they did in Romania (2018, p. 420). She argued that while the pressure from the EU to adopt and retain adopted transparency and anti-corruption policies was an important factor, domestic actors were at least equally vital, and every attempt to weaken or retract these policies was followed by public outrage. The findings above

aptly complement Schnell's research and propose an argument that this public outrage owes to a great extent to the free press. In countries with heavily state-controlled media or otherwise violated journalistic rights, journalists might not be able to fully benefit from adopted transparency and anti-corruption policies and their reporting might be heavily censored, leaving the public uninformed or mislead about corruption cases.

Figure 4-1: Conditional effect of open data on corruption as function of press freedom
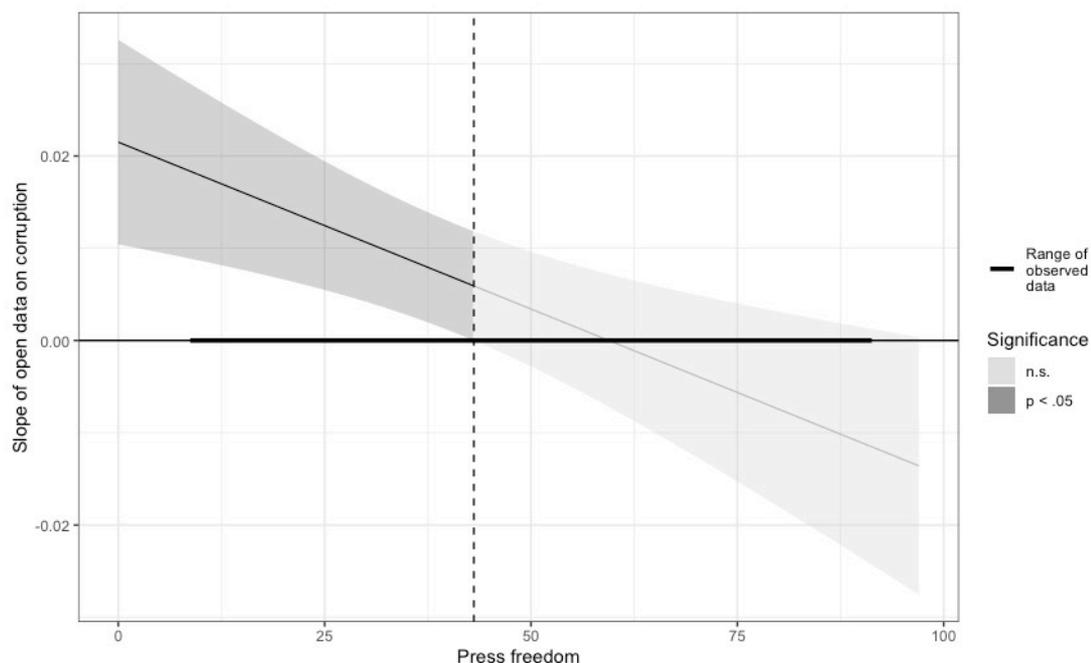
As can be seen in model 3 in Table 4-2, the presence of judiciary does not modify the association between open government data and corruption, and thus the findings cannot confirm the **h4-1c** hypothesis. Also, in model 3, the individual effect of open data on corruption vanishes.

I also included membership in OGP as an instrumental variable and used two-stage least squares regression to estimate the effect of open government data on corruption and address potential endogeneity problems. By using the instruments, I get a variance of open government data measure that is not correlated with the confounding variable. As can be observed from Table 4-3 below, the F-test statistic on instruments is 12.993 (p < 0.001), interacted with press freedom is 6.284 (p < 0.001), and interacted with judiciary is 15.761 (p < 0.001). Therefore, the null hypothesis that instruments are weak cannot be confirmed.

However, the Wu-Hausman test is statistically significant only in model 1 and 3 (p < 0.05), suggesting that IV estimates are more consistent than OLS estimates in

these models. In model 2, the Wu-Hausman model was not significant, and thus OLS estimates are consistent for this model. In conclusion, the models confirm that the associations between both open government data and corruption levels are significant only in the interactions with press freedom.

Table 4-3: The effect of FOI laws and open data on corruption (IV)

Control of corruption	Model 1	Model 2	Model 3
Open data	-0.013	0.023·	-0.031
	(0.011)	(0.014)	(0.028)
Years of FOIA	0.004·	0.001	0.003·
	(0.002)	(0.002)	(0.002)
Press freedom	-0.003	0.023·	-0.001
	(0.004)	(0.012)	(0.005)
Murdered journalists	-0.110	-0.012	-0.116
	(0.157)	(0.167)	(0.153)
Civil society	**-0.073***	**-0.066***	**-0.071***
	(0.029)	(0.028)	(0.029)
Judiciary	**0.201*****	**0.198*****	0.089
	(0.039)	(0.037)	(0.133)
Free and fair elections	**0.014****	**0.017*****	**0.015****
	(0.004)	(0.005)	(0.005)
GDP	**0.030*****	**0.022*****	**0.026*****
	(0.007)	(0.006)	(0.006)
Open data*press freedom		**-0.001***	
		(0.0003)	
Open data*judiciary			0.004
			(0.004)
Constant	-1.074*	-2.564**	-0.721
	(0.471)	(0.780)	(0.617)
N	110	110	110
R^2	0.833	0.850	0.840
Adjusted R^2	0.820	0.837	0.826
Residual Std. Error	0.428 (df = 101)	0.408 (df = 100)	0.422 (df = 100)
Weak instruments: open data	**12.993*****	**8.634*****	**10.467*****
Weak instruments: open data*press freedom		**6.284****	
Weak instruments: open data*judiciary			**15.761*****
Wu-Hausman	**5.162***	2.339	**3.175***
Sargan	NA	NA	NA

Note: *p < .05; **p < .01; ***p < .001, the results in bold are statistically significant.

This analysis also explores the effect of different types of government datasets on corruption. The score of data availability is an aggregate score and consists of scores for different government datasets. Therefore, it is possible to examine the effect of datasets that are particularly crucial for monitoring how the government manages

public resources, such as data on government spending, contracts or land ownership data. I estimate the following model:

Corruption = α + βbudget data + βspend data + βcontracting data + βcompany data + βland ownership data + βFOI years + βpress freedom + βmurdered journalists + βcivil society + βjudiciary + βfree election + βGDP + ε

The results in Table 4-4 show that some datasets essential for monitoring public resources; in particular, land ownership data are significantly associated with corruption levels.

Prior to this analysis, I refrained from proposing directional hypotheses, as the literature on the link between data on the government's management of public resources and corruption has been inconclusive in this regard. Birchall (2015) and O'Neill (2006) argued that publishing more data will bring more revelations on mismanagement of public resources to the light, and if these are left unpunished, the public will become even more sceptic and mistrustful of the government, perceiving it as corrupt.

In particular, data on contracts may reveal cases of suspicious use of public resources, overspending, illegitimate expenses etc. Such revelations might turn the public to become disillusioned with politics in a broader sense and see corruption in a country as pervasive. Bac (2001) suggested that access to contracting data might also improve outsiders' ability to detect the patterns of corrupt behaviour and increase their incentives to replicate these behaviours for personal gain. He proposed that this might eventually lead to an increase in corruption, which contradicts prevalent theories of transparency. A practical example of this would be if increased access to data on contracts would allow detecting which public agencies award procurement contracts repeatedly to the same companies, and how their value differs from comparable contracts in other public agencies. Recognising such agencies may encourage establishing connections at these agencies to offer bribes in exchange for awarded procurement bids. However, I find no evidence to confirm these speculations.

I find that an increase in land ownership data is associated with lower levels of corruption. However, this effect is small, with a coefficient of 0.005 (robust standard error = 0.002) statistically significant at $p < 0.05$. Similarly, as with the data on government contracts, the evidence about the direction of the relationship between the availability of data on land ownership and corruption is inconclusive (Benjamin, Bhuvaneswari, Rajan, & Manjunatha, 2007). Lastly, the effect of free and fair elections, journalists' murders, civil society, independent and accountable judiciary and GDP on corruption levels remained statistically significant as in the majority of models.

Table 4-4: The effect of the availability of different government datasets on corruption (OLS)

	Control of corruption
Data on government budget	0.002
	(0.001)
Data on government spending	0.0002
	(0.002)
Government contracting data	-0.003
	(0.002)
Company register data	0.002
	(0.001)
Land ownership data	**0.005***
	(0.002)
Years of FOIA	0.001
	(0.002)
Press freedom	-0.004
	(0.004)
Murdered journalists	**-0.235***
	(0.115)
Civil society	**-0.063***
	(0.026)
Judiciary	**0.209*****
	(0.033)
Free and fair elections	**0.008****
	(0.002)
GDP (in 1,000 USD)	**0.023*****
	(0.003)
Constant	-1.141***
	(0.397)
N	110
R^2	0.888
Adjusted R^2	0.874
Residual Std. Error	0.359 (df = 97)
F Statistic	63.974*** (df = 12; 97)

Note: *p < .05; **p < .01; ***p < .001, the results in bold are statistically significant.

4.3.2 FOI laws and the moderating effect of press freedom

Table 4-5 presents the model without interaction effect and models where interaction effect between the number of FOI years and press freedom was added (model 4a, 4b and 5). In the first model, the tradition of FOI legislation, measured as the number of years since FOIA was adopted in a country, was not significantly associated with

corruption levels. Therefore, the results are inconsistent with the ***h4-2a*** hypothesis that countries with a long tradition of FOI laws are more likely to exhibit lower levels of corruption than those where FOI legislation was introduced just recently. This is an interesting non-finding, contrary to the intuitive assumptions. These results might be indicative of the problems with the measurement discussed above. While FOI law tradition might be associated with the quality of the law implementation, a more accurate measure of FOI implementation might be more appropriate.

The diagnostic plots of residuals for model 4a[34], in particular, the Residuals vs. Leverage plot reveals that Sweden, with the longest tradition of FOI legislation in the world, is an influential observation because it falls beyond Cook's distance of 1 and thus, might affect the slope of the regression line and have an effect on the parameter estimates (see Figure A-8 in Appendix). Therefore, in addition to model 4a where Sweden was included, a model without Sweden was estimated (4b). As can be observed from Table 4-5, the inclusion of Sweden in the dataset influences the parameter estimates. The results of model 4a are consistent with some of my hypotheses (***h4-2a*** and ***h4-2b***), but the moderating effect of press freedom is significantly stronger when Sweden is excluded from the analysis (4b).

The association between FOI tradition and corruption is significant only when interactions between FOI years and press freedom are added. In model 4a, for every additional year of FOI legislation in place, CoC score increases by 0.005 scores (robust standard error = 0.002, $p < 0.05$), i.e. levels of corruption go down. In model 4b, these associations are even stronger. For every additional year of FOI law in place, CoC score increases by 0.017 scores (robust standard error = 0.006, $p < 0.01$), i.e. levels of corruption lower. As can be observed from the interaction effect in Table 4-5, for every decrease in the press freedom score, the effect of FOI tradition on corruption diminishes by 0.0005. In other words, as the level of media censorship increases the slope of the relationship between FOI tradition and corruption becomes more negative.

---

[34] Model 4a: Corruption = α + βopen data + βFOI years + βpress freedom + βmurdered journalists + βcivil society + βjudiciary + βfree election + βGDP + βFOI years*press freedom + ε

Table 4-5: The effect of FOI laws on corruption (OLS)

Control of corruption

	Model 1	Model 4a	Model 4b	Model 5
Open data	0.007*	0.008**	0.007*	0.007*
	(0.003)	(0.003)	(0.003)	(0.003)
Years of FOIA	0.002	0.005*	0.017**	-0.007
	(0.002)	(0.002)	(0.006)	(0.012)
Press freedom	-0.004	-0.003	0.00004	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)
Murdered journalists	-0.269**	-0.253*	-0.245*	-0.267*
	(0.113)	(0.112)	(0.110)	(0.114)
Civil society	-0.073**	-0.070**	-0.067**	-0.073**
	(0.025)	(0.025)	(0.024)	(0.025)
Judiciary	0.216***	0.208***	0.214***	0.203***
	(0.032)	(0.032)	(0.032)	(0.036)
Free and fair elections	0.008**	0.008**	0.008***	0.008**
	(0.002)	(0.002)	(0.002)	(0.002)
GDP	0.020***	0.019***	0.017***	0.020***
	(0.003)	(0.003)	(0.004)	(0.003)
Years of FOIA*press freedom		-0.0003*	-0.0005**	
		(0.0001)	(0.0002)	
Years of FOIA*judiciary				0.001
				(0.002)
Constant	-1.132**	-1.191**	-1.379***	-1.083**
	(0.400)	(0.395)	(0.399)	(0.407)
N	110	110	109	110
R^2	0.879	0.884	0.885	0.880
Adjusted R^2	0.870	0.874	0.874	0.869
Residual Std. Error	0.364 (df = 101)	0.358 (df = 100)	0.353 (df = 99)	0.365 (df = 100)
F Statistic	92.015*** (df = 8; 101)	85.058*** (df = 9; 100)	84.244*** (df = 9; 99)	81.480*** (df = 9; 100)

Note: *p < .05; **p < .01; ***p < .001, the results in bold are statistically significant.

A simple slopes analysis was again conducted to understand better the complexion of the moderating effect of press freedom on the relationship between FOI tradition and corruption. I followed the same procedures as in the case of exploring the moderating effect of press freedom on the slope of the relationship between open data and corruption. I set the values of press freedom score at 10, 30, 50 and 90 scores. The results were different. The FOI tradition is significantly associated with corruption levels only in countries with excellent press freedom score, such as Canada, Norway or Switzerland or opposite, in countries with an extremely poor score, such as Iraq, Malaysia or Qatar. Else, the relationship was not significant. For example, for the press freedom score of value 10, the coefficient is 0.01 (p < 0.001), but for the

score of value 30, it is insignificant. The relationship between FOI tradition and corruption is also insignificant at press freedom score of value 50. The Johnson-Neyman plot shows that the effect of press freedom on the FOI tradition-corruption relationship is not significant when press freedom score is outside of the interval 18.81 – 65.72. Once the press freedom score increases and the censorship strengthens beyond a score of 65.72, the association between FOI tradition and corruption becomes significant again.

Figure 4-2: Conditional effect of FOI tradition on corruption as function of press freedom
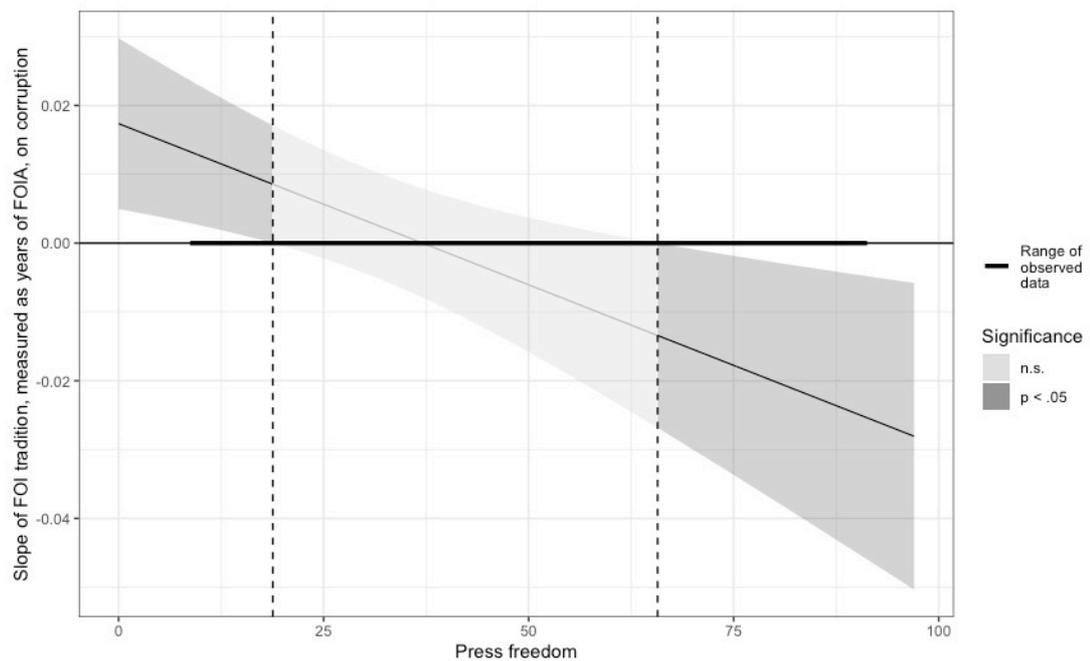
FOI tradition is significantly associated with decreasing corruption and stronger control of corruption (higher CoC score) in countries with press freedom score lower than 18.81, i.e. countries with unrestricted conditions for media operation. However, in contrast with open data, the tradition of FOI legislation becomes significantly and negatively associated with corruption levels (CoC score) once press freedom score is higher than 65.72 (very high level of censorship). The interaction effect between the number of years with a FOI law in place and independence and accountability of judiciary was also not significant (model 5). Therefore, hypothesis *h4-2c* cannot be confirmed.

Overall, the findings mostly support previous research on transparency policies and their link to corruption, i.e. the effect is conditional upon media freedom (Besley & Burgess, 2002; Besley et al., 2002; Besley & Prat, 2006; Brunetti & Weder, 2003;

Chowdhury, 2004; Djankov et al., 2003; Ferraz & Finan, 2008; Norris, 2008; Reinikka & Svensson, 2005; Solis & Antenangeli, 2017; Svensson, 2005; Yazaki, 2017). The results add to the previously accumulated knowledge in the field and demonstrate that access to free and pluralistic media might be crucial for any transparency measure to be effective also as an anti-corruption measure. It might not be enough for the information and data to be out in the public domain; the information needs to be acted upon. Investigative journalists and civic activists who use FOI laws or government datasets to obtain information, act as intermediaries, who are able to extract crucial pieces from the abundance of the information and interpret them to the public. If their working conditions worsen, become dangerous or compromised by different types of censorships, it might allow corruption to flourish, as it will go unnoticed.

4.4 Conclusions and discussion

Political leaders have always pledged transparency. In the past, many built their election campaigns on the promise to pass FOI laws. Recently, they made a case for open government data as a new digitally-enabled transparency policy that will help to eradicate mismanagement of public funds. Since then, governments worldwide have launched national open data portals and published thousands of datasets. This study is one of the first to investigate the relationship between open data, relative to FOI laws, and corruption levels. The results suggest that both FOI laws and open data are significantly associated with corruption levels only in the interaction with press freedom.

These results are interesting for theory as well as policies. First, from the theoretical perspective, they validate Lindstedt's and Naurin's framework and demonstrate its versatility for estimating the effect of different measures of transparency on corruption. Second, the findings suggest that enthusiasm for open data as a powerful anti-corruption means, shared by political leaders and anti-corruption activists should be approached with caution and the relationship between open data and corruption should be further investigated. The present results suggest that media freedom modifies the relationship between transparency policies and corruption levels. While in countries that enjoy high degrees of media freedom, the increase in open data availability or longer tradition of FOI laws is significantly associated with lower corruption levels and stronger control of corruption, these associations become insignificant once media freedom deteriorates beyond a specific threshold.

In view of these findings, transparency policies might be less effective for addressing corruption in conditions where journalists or discontented citizens are censored. For instance, Mexico has adopted several transparency policies and is one of the leaders in open data. However, at the same time, its rate of killings of investigative journalists (Committee to Protect Journalists 2018) makes the profession extremely dangerous. It might discourage journalists from using government information and data indicating that powerful politicians or entrepreneurs are involved in corruption. In countries with restricted conditions for media operation, open government data and other transparency policies might serve as smokescreen boosting the international image of governments as being transparent while shifting attention from substantial domestic policy issues. The analysis has also shown that not all types of government datasets might matter for the anti-corruption fight equally. I find an association between land ownership data and corruption levels only.

4.5 Limitations

There are several limitations to this analysis. It offers insights into the relationship between open data, relative to FOI laws, and corruption in a certain point of time (the year 2016). Due to the unavailability of measures of open data availability over time, conducting a longitudinal study was not possible. Such a study could have provided a clearer picture of whether the moderating effect of media freedom on the relationship between government transparency and corruption is temporary and how it changes as other factors change over time. The previous studies in the field suggest what findings could emerge if panel data were available.

For example, Vadlamannati and Cooray (2017) and Costa (2013) found that the adoption of FOI legislation drives the corruption perception up shortly after their passage. They find that this is more the case for countries that have free media and dense civil society. Once an FOI law is passed, journalists and civic activists can get hold of information that was previously inaccessible to them. Some of this information will inevitably put the government and public administration in a bad light. In countries where media penetration is high, and journalists are able to write freely, the messages about mismanagement of public resources will travel fast and might decrease CoC score in the short term. However, I posit that the long-term effect will depend on the quality and independence of institutions that are formally charged with punishing the instances of corrupt or fraudulent behaviour. If criminal justice system, i.e. police, prosecutors and courts, are able to do their work without interference and checks and balances are in place to ensure they continue to do so, it is likely, that any behaviour that represents a criminal offence will eventually be penalised. The sense of

justice and fairness can, thus, contribute to lower levels of corruption perception in the long term. On the contrary, if criminal justice institutions are weak and grand corruption goes with impunity, the public might become cynical about politics, and mistrust in government and its perception as inherently corrupt might grow in the long run as O'Neill (2006) argued. Longitudinal studies could offer interesting insights towards this end.

Similarly, qualitative case studies using, for example, process tracing techniques, could also add inferential leverage, and explain the impact of specific transparency policies and the role of media and both offline and online civic activism for their success. To some extent, the following chapters address this by providing an analysis of demand for government information and data by different user groups in the UK. The survey of journalists and interviews with civic activists in chapter 6 sheds light on how these groups engage with government datasets and FOI laws, what ways of obtaining government information they prefer and how digital technologies have changed it. Chapter 7 is an investigation into the public demand for government information in the UK and how the central government meets this demand. Using quantitative text analysis (structural topic modelling), I explore the prevalence of topics within FOI requests submitted to the UK central government through the online participation platform between 2008 and 2017.

Another drawback of this study that should not be omitted is that it takes at face value that government-held information and data is accurate. However, the information made available by governments might be defective in manifold ways, intentionally or unintentionally. While I provided several examples when the accuracy of government information was questionable, it is not within the remits of this study to address this limitation. That said, further research into the credibility of government information more generally, and the debate about the right of access to accurate government information and data is needed.

Another caveat of the analysis is that the used measures are imperfect. While measures are only proxies of researched phenomena, some reflect them better than others. For instance, as already raised, the FOI laws tradition can be conflated with democratic tradition to some extent as non-democratic countries started to pass FOI laws only recently. This measure does not tell us much about the implementation of the law. While an excellent piece of legislation might be implemented poorly, imperfect legislation might well deliver good levels of openness. Nonetheless, other than to acknowledge this limitation, a little can be done to address it, as there are no cross-country measures of the quality of FOIA implementation available because they are notoriously difficult to develop.

5 The UK at the forefront of transparency and anti-corruption initiatives

In this chapter, I will briefly explain why the UK has been chosen as a case study to investigate the use and effects of FOI laws and open government data in more detail. First, following the comparative cross-country analysis introduced in the previous chapter, I will describe the steps that led to the selection of the UK as a crucial case. I will also discuss the media environment in the UK, given that the results of the analysis demonstrated the importance of media freedom for transparency policies, in particular, open government data to affect corruption (as a proxy for government accountability). Second, I will describe the path to the UK FOI legislation that despite decades-long national and international pressures did not get adopted until 2000 and how, in quite a contrast, open government data made a steep rise to the top of government agenda from one year to another. I also document the major legislative changes governing access to government information and analyse the key policy papers related to open government data.

5.1 The UK as a country case study

The selection of the UK as a country case study is primarily based on the results from quantitative comparative cross-country analysis presented in the previous chapter. One of the main models, which confirmed the effect of open government data on corruption only in the interactions with press freedom, was used to guide the selection of a country case. Figure 5-1 shows the actual Control of corruption score plotted against the model-fitted values of Control of corruption. As can be observed from the figure, although the UK is not on the line, and the model-fitted values are somewhat lower than actual values, the UK belongs among the cases where corruption levels (both actual and model-fitted) are among the lowest. In other words, the Control of corruption score is among the highest in the UK.

I labelled all country cases where actual and model-fitted values of Control of corruption were equal or higher than one, i.e. corruption is well controlled. These are New Zealand, Finland, Denmark, Norway, Sweden, Singapore, Canada, Switzerland, Iceland, the Netherlands, the UK, Germany, Australia, Belgium, Ireland, Austria, Japan, France, United States, Uruguay and Israel (ordered by descending Control of corruption score). As can be observed from the figure, these country cases do not only generally share low levels of corruption, but they also exhibit other similarities. Most

of them enjoy high levels of media freedom, have FOI legislation in place and have relatively high levels of the availability of open government data. The only exception is Singapore. Singapore's model-fitted values of Control of corruption almost mirror the actual values. However, in Singapore, low corruption levels are not a consequence of government transparency policies or media freedom as in other labelled countries. Singapore does not have FOI legislation, and also, media freedom is limited (see Figure 5-1). The low corruption levels in the country owe mostly to its independent and accountable judiciary (Haider Malik, 2007) and open economy, in particular, absence of barriers to international trade, which translates to high levels of economic growth (Gatti, 2004). Thus, although Singapore represents a fascinating case for studying factors contributing to combating corruption, these are outside of the thesis' theoretical framework, which focuses on government transparency.

Figure 5-1: Model-fitted CoC score vs. CoC score in 2016
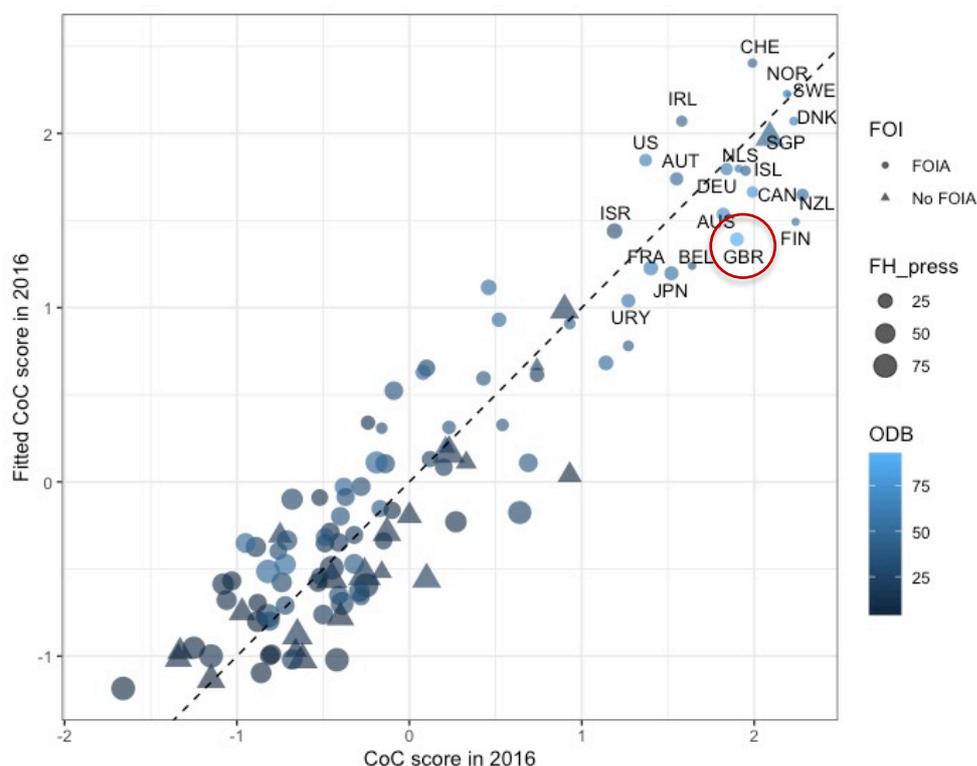
Note: FH_press stands for press freedom measured as a reverse score taken from Freedom House (0 – a free press, 100 – not free press) and is visualised as the diameter of the circle or the size of the triangle. ODB stands for the Open Data Barometer's sub-score, which measures the availability of government datasets (0 – no availability, 100 – full availability) and is visualised as the scale of blue. FOI shows the existence of the FOI law (circle) or its absence (triangle). The UK is circled in red and is labelled as GBR (ISO country code).

More importantly, another story emerges from the figure above. Labelled country cases where government transparency policies are in place, and which succeed in controlling corruption, all, except Singapore, represent industrialised democracies

of the Global North, i.e. high-income countries that value civil rights and liberties and participatory processes. No Global South country in the figure exhibits low levels of actual and model-fitted Control of corruption values. The previous chapter showed that democracy, in particular, media freedom matters for transparency policies to achieve its intended goals, such as good governance and accountability. The figure above shows that wealth is relevant too. Uneven distribution of resources, lack of necessary technological and human infrastructure and unequal access to opportunities to obtain digital and data skills causes that many countries in the Global South are unable to convert the advantages new digital technologies bring for their benefit.

Therefore, if another democratic but developing country was considered for the case selection, I argue that the results would have been very different. First, developing countries have limited state capacities. The infrastructure necessary for the successful implementation of FOI laws and open data policies might be limited or entirely missing. For instance, efficient management of FOI requests requires digital management practices. The publication of open government data has to be preceded by systematic collection of the data. Due to lack of resources, the capacities of developing countries to produce and disseminate data of high quality are constrained (Devarajan, 2013). Thus, their efforts to adopt open data policies and publish government data would be largely dependent on international aid and how it could be used. While the availability of international aid might substantially improve both the quantity of available data and its quality, it might also create perverse incentives and lead to the creation of open data portals resembling Potemkin villages. Researchers have already documented that some governments (in this particular case it was Moldovan government) where open data publication was funded through international aid, did not sustain a regular publication. They published the data in bulk shortly before an audit visit by the donor (Shkabatur & Peled, 2016). For the reasons stated above, most developing countries would not have the capacity to sustain open data publication.

Therefore, the case study was selected from industrialised democracies that have favourable conditions for transparency policies to achieve their intended aims. While all other countries labelled in Figure 5-1 could serve as interesting case studies, my limitations as a researcher also had to be taken into account. Since the case study included expert survey and qualitative interviews, and the analysis of texts of FOI requests, language competence had to be factored into the final decision. This limited the selection to English-speaking countries: Australia, New Zealand, the UK and the US. With new digital technologies, the question of geographical proximity does not represent any major concern. While it did not influence the selection of the UK as a case study, other factors did. As can be observed from the figure, the UK stands out on one variable, in particular, the availability of open government data with a score

of 100 out of 100. Second Canada has a 10 points lower score than the UK. The UK has been one of the innovators in open government data and repeatedly occupied the top place in the international open data rankings[35]. The UK has been an early adopter of new digital technologies. Internet penetration and use are almost universal, creating excellent conditions for the use of published data. The UK was also one of the founding members of OGP and is involved in other major global transparency partnerships, such as the Extractive Industries Transparency Initiative. Therefore, the assumption that the shift towards greater transparency will bring substantial outcomes and improve governance and strengthen accountability in the UK is reasonable.

Also, even though the design of FOIA in the UK is not one of the strongest according to RTI Rating[36], it has a wide scope and covers a range of public authorities. The availability of appeal procedures and mechanisms for the enforcement of sanctions still make it a robust piece of legislation safeguarding the requesters' rights of access to government information. RTI Rating also favourably assesses the promotional measures in the UK's FOIA. The Information Commissioner has an important role in promoting the law, encouraging its use and providing guidance. While the late adoption of FOIA in the UK might have negatively affected public administration's attitudes towards the right of access to government information compared to countries where FOIA was introduced early, FOIA as a tool has become very accessible and also widely used in the UK. Even though the users of FOIA are likely to represent a tiny fraction of the UK population, the number of requests they send is very high in the UK compared to other European countries (as far as it can be judged from publicly available online participation platforms and available government statistics).

Last but not least, the UK was selected because of its long tradition of the strongly professionalised, and formally autonomous practice of journalism, and public broadcasting oriented on the information (Hallin & Mancini, 2004), which assumes critical engagement with government sources. Notwithstanding the deteriorating conditions for journalists both in the UK and world, the UK still remains to be in the group of countries with free media[37]. Journalists are still able to use government information freely without facing harassment or other repercussions. However, the UK performs in media freedom significantly more poorly than other Western European

---

[35] For more information, see the Global Open Data Index by Open Knowledge Foundation at https://index.okfn.org/ and Open Data Barometer by Worldwide Web Foundation at https://opendatabarometer.org/.

[36] For more information, see the methodology of RTI Rating at https://www.rti-rating.org/country-data/scoring/.

[37] For more information, see, for example, the Freedom House Freedom of the Press index at https://freedomhouse.org/report/freedom-media/freedom-media-2019 or the Reporters without borders index at https://rsf.org/en/ranking.

countries. Few major outlets dominate the media market accounting for more than 70% of the audience in each media segment, including TV, radio or print media market (Smith & Cavaliere, 2017). Media are protected from the direct political pressure by prohibiting political bodies to hold licences in the broadcasting sector. The Office of Communications, commonly known as Ofcom, media regulating authority, also guarantees the protection of editorial autonomy through the Ofcom Broadcasting Code. Its section five, on due impartiality and due accuracy, requires these principles are upheld (Ofcom, 2019). However, the print sector does not have any equivalent of the code (Smith & Cavaliere, 2017). While the instances of violence against journalists are rare, the online harassment and hate speech is on the rise and might still be underestimated as it is likely to be formally unreported (Clark & Grech, 2017). One of the main recent impediments to media freedom in the UK was the adoption of the Investigatory Powers Act in 2016 which has expanded the government's surveillance of communications and weakened the protection of journalists and whistle-blowers (Reporters without borders, 2019).

Since the UK is in many respects similar to other industrialised democracies depicted in Figure 5-1, I do not expect to find substantial differences in the impact of transparency policies in the UK compared to these countries. However, I assume that the different approaches to FOI legislation and the extent of media freedom might slightly alter to what extent FOI laws and open data policies can contribute to good governance and accountability. For instance, as can be observed from Figure 5-1, Scandinavian governments, in particular, Norwegian, Swedish and Danish governments are likely to achieve the best results. This might be due to a long experience with FOI legislation. Sweden was the first country in the world to adopt FOI legislation in 1766, Denmark and Norway passed FOI laws, among the first in Europe, in 1970 (see Table 5-1). At the same time, Scandinavian countries enjoy the highest levels of media freedom in the world and thus create favourable conditions for engagement with government information and data.

5.2 A long journey to FOI legislation

Despite its long democratic tradition, the UK adopted FOIA notably later than other Western European democracies (see Table 5-1). The way to legislating the right to access government information was far from straightforward. As FOI scholar Worthy (2017) argued, to understand the lukewarm relation to FOIA fully, one has to know about the pervasive culture of secrecy that has burgeoned in the UK government in the 20th century first. Discussing openness without looking at the opposite side of it will offer a partial response only.

Table 5-1: Year of FOIA adoption in Western European democracies

Country	Year of FOIA adoption
Austria	1987
Belgium	1994
Denmark	1970
Finland	1951
France	1978
Germany	**2005**
Greece	1986
Ireland	**2003**
Italy	1990
Netherlands	1978
Norway	1970
Portugal	1993
Spain	**2013**
Sweden	1766
Switzerland	**2004**
UK	**2000**

Note: Countries in bold adopted FOIA only recently, after 2000.

The first Official Secrets Acts (OSA) in the UK was adopted in 1889 as a response to the leaks from the government officials to the press (White, 1980). Using the sanctions under the civil law, the government was inefficient in preventing the officials from selling government documents, and thus decided to draft new legislation – OSA that would include criminal sanctions as well. This legislation was rarely used in practice and was soon repealed by a newer one (Ponting, 1990).

The pre-war government, fuelled by unfounded fears of the presence of foreign spies, hastily passed the OSA 1911, which had far more wide-raging scope and powers than its predecessor (French, 1978). The OSA's section 2 was relevant for FOIA, as it made it an offence to disclose official information to an unauthorised person as well as to receive any official document knowing its disclosure violated the act. Many scholars argued that the new law had a ***catch-all*** quality due to its ambiguity and a lack of legal definitions (Bartlett & Everett, 2017; Ponting, 1990; White, 1980). They reasoned that if the letter of this act was strictly followed, ordinary civil servants were unable to share any (no matter how banal) details from their professional life with their families. The law did not make a difference between a document related to national security and information about a lunch served at the government canteen (ibid). The powers and the scope of the act were further strengthened in 1920 when it was amended. Ponting (1990) mentioned that the opposition to the act pointed to potential ramifications for the press, who is likely to work with leaks and the

infringement of civil liberties. At the same time, the government started to develop its secret intelligence services, including MI5, MI6 and Government Communications Headquarters (GCHQ).

The secretive environment kept FOIA at bay for decades despite several major attempts by both Labour and Conservative governments to repeal the OSA's controversial section 2 and legislate access to government information. The October 1974 Labour Party Manifesto stated:

> "Labour believes that the process of government should be more open to the public. We shall: replace the Official Secrets Act by a measure to put the burden on the public authorities to justify withholding information; and establish compulsory registers of interest for all MPs, councillors, peers, senior civil servants, senior council officials, and others in the upper reaches of the public service" (The Labour Party, 1974).

Following these pledges, in 1978, the 1976-1979 Labour government published the White Paper on the Official Secrets Act. It made a case for greater government openness, arguing that it is essential for democracy, as it enables people to make informed decisions and oversee the workings of their government.

In January 1979, Clement Freud, MP for the Liberal Party presented the House of Commons with the Official Information Bill, which proposed to repeal the section 2 of the OSA 1911 but also introduce a statutory right of access to government information for the public. As the bill was in line with the official position of the 1976-1979 Labour government under PM James Callaghan, the chances of its passing were expected to be good. Freud's speech at the House of Commons emphasised that under the then conditions, it was impossible to get much essential information about the state of the environment, public health and safety risks or new taxpayer-funded research findings. He demonstrated the absurdity of the OSA 1911 that had then been in force for 68 years with an example of an applicant for a gardener job at the Hampton Court Palace who was required to sign the declaration form that anything acquired during the job he discloses to unauthorised persons makes him liable to prosecution (The House of Commons, 1979). Despite broad parliamentary support for the bill, the government did not proceed further due to the lost vote of no confidence and following early elections in which the Labour Party lost.

Soon after, in 1981, another effort to legislate access to government information came from the Labour MP Frank Hooley (The House of Commons, 1981). However, the 1979-1983 Conservative government was not in favour of the proposition and defeated it at second reading. As the Commons debates from those years suggest, there was also a mounting international demand for the UK to legislate the access to government information, in particular after the CoE published the recommendation

on the FOIA introduction (Council of Europe, 1976). Often, the inspiration for the UK reform was drawn from the US FOIA. Parliamentarians referred in their speeches to overseas practice, including not only the US but New Zealand and Canada to support their arguments both for and against FOIA. The next move to legislate FOIA came from the Liberal MP David Steel in 1984. The draft was informed by Campaign for Freedom of Information (CFOI). Nonetheless, the bill was opposed. The main argument against FOIA was similar to those commonly used in the present: "Once this Pandora's box is opened, it will take some closing" (The House of Commons, 1984).

Along with the legislative efforts to introduce FOIA, several Private Members' Bills were proposed to repeal the OSA's section 2 but without success. Finally, in 1989, the 1987-1990 Conservative government adopted the 1989 OSA, which came into force in 1990, repealed the controversial section 2 and introduced category-based offences related to the unauthorised information disclosure. It consisted of information related to security and intelligence, defence, international relations, foreign confidences, information potentially leading to crime, and the special investigation powers under the Interception of Communications Act 1985 and the Security Services Act 1989 (Bartlett & Everett, 2017).

As for FOIA, until its adoption in 2000 (and 2002 in Scotland), the access to government information had only seen partial advancements. Worthy argues that FOIA was "simply too radical for a succession of governments wedded to the orthodoxies of information control" (2017a: 22). Prior to the passage of the 1989 OSA, several pieces of legislation increased access to government information. For example, the 1987 Access to Personal Files Act allowed people to access some of the records about themselves held by the government (Access to Personal Files Act, 1987). Other private members' bills increased access to medical reports for employers or insurers and the environmental information (Access to Medical Reports Act, 1988; Environment and Safety Information Act, 1988).

Also, the regulation that brought advancements in the access to government information continued to be adopted in the early 1990s, e.g. EC Directive 2003/4/CE on public access to environmental information was transposed into the UK legal system.

Worthy (2017a) noted that both Conservative and Labour governments in turns spoke in favour of the open government. Some of the Labour manifestos that endorsed government openness were mentioned above. The Conservative Party also wanted to be part of this conversation as the topic clearly was one that attracted political capital. A section of its 1992 manifesto was devoted to open government (The Conservative Party, 1992).

However, the promise for greater openness was limited to widening the access to personal records held by the government, disclosure of the members of the Cabinet's Committees and ministerial guidance. While the party pledged to review existing restrictions to the access to information, it did not explicitly mention FOIA. The Conservatives won the elections, and in line with its manifesto, they adopted only non-statutory regulation, the White Paper on Open Government (The Cabinet Office, 1993). The paper's emphasis was on transparency and accountability but also on public service delivery (ibid). Nonetheless, the paper made it clear that there are groups of information that require confidentiality, such as national security, defence, public safety or personal data. The white paper was followed by a voluntary Code of Practice on Access to Government Information. While the code was undoubtedly useful, with other EU directives governing the use of and the access to personal data, it became increasingly clear that the full statutory FOI regime will be necessary (Gay, 1997). Not only was the UK lagging behind its counterparts in Western Europe but also some new democracies that emerged after the fall of Communist bloc had drafted FOI laws by then.

When the Labour Party formed the government in 1997, the combination of the international diffusion of FOI laws, the emergence of new information technologies and 18 years *out of power* created momentum for legislating FOIA (Worthy, 2017a). Worthy (ibid) argued that the last factor played a particularly important role, as FOIA helped the Labourites to distinguish themselves from Conservatives as having a new radical approach to government-citizen interactions.

In 1997, the government published the White Paper Your right to know – the government's proposal for FOIA. The tone of the paper continued to present FOIA as a radical policy and the fulfilment of a significant election campaign pledge. The introduction of the paper discussed the values behind the proposition – government transparency and accountability, and modernisation of public services (The Cabinet Office, 1997). The paper also described the rights and obligations of both requesters and public authorities, and the technicalities of the planned legislation, e.g. appeal procedures. It introduced the position of independent commissioner with wide-ranging powers who will oversee the implementation. It also set the exemptions from the law and principles that require protection - national security, defence and international relations, law enforcement, personal privacy, commercial confidentiality, public safety, the information supplied in confidence and the integrity of the decision-making and policy advice processes in the government. The white paper suggested broad coverage of public authorities except for intelligence services – MI5, MI6 and GCHQ, and included both government records and information as the subject to disclosure (ibid). It also explained the principles of harm and public interest tests. The former was aimed to prevent harm by disclosure. The latter should ensure there is a balance

between disclosing or withholding information. The paper also admitted that the successful implementation of the law would require the government to prepare strategies for public engagement with the law, manuals for authorities, training for civil servants in charge of dealing with requests and monitoring mechanisms (ibid).

However, the final version of the law that made it to the parliament was criticised by CSOs and media for broad definitions, blanket exemptions and for generally falling short of its initial ambitions (The Guardian, 1999). Some of the criticism was addressed, but the passed bill remained somewhat weaker than the white paper envisaged. The act came into force in 2005 together with The Freedom of Information Scotland's Act 2002 (FOISA), which was assessed as more liberal, providing greater rights of access than FOIA 2000 (Nelson, 1999).

Since then, FOIA has been widely used. In its first five years, requesters sent to public authorities over half a million requests (Goodall & Gay, 2010). The Constitutional Reform and Governance Act 2010 introduced some changes to FOIA, some increasing, others shrinking the access to information. For instance, the period for disclosing historical government records in the National Archives reduced, but exemptions were widened to include the Royal family (ibid).

Once the law was adopted and came into force, the political interest in the topic faded away. Neither Labour nor Conservative Party had mentioned FOIA in their 2005 election manifestos. The Labourites pledged to provide better information about health services and children's services and parental entitlements. A section was devoted to new digital challenges (The Labour Party, 2005). Also, the Conservatives neither mentioned FOIA nor did they promise further access to information. Their manifestos in 2010 also did not touch upon FOIA (The Conservative Party, 2010). The closest to FOIA and its values was the Labour Party's pledge to propose new libel legislation to protect "the right of defendants to speak freely" (The Labour Party, 2010). The focus of both manifestos rather shifted to the proactive publication of government data, which will be discussed in more detail in the next section.

Already the 2007-2010 Labour government introduced reforms to increase access to data held by the government. The 2010-2015 Conservative and Liberal Democrat coalition government continued to advance open government data agenda and created a narrative that framed it as a modern substitution to FOIA. Minister Francis Maude stated in one of his speeches that:

> "... [his aim] is to make Freedom of Information redundant. We should be proactively making public everything that is appropriate. You should make redundant the need for people to ask for access to information" (The Cabinet Office, 2014).

Shortly after that, the government announced that it would set up the Independent Commission to review FOIA and its implementation. This motion raised concerns among CSOs and media, who argued that the requesters' rights should not be restricted (Broad, 2015; Freeguard, 2015; The Guardian, 2015). The Commission led by Lord Burns collected over 30,000 written responses from organisations and individuals and oral evidence from 21 individuals representing various sectors in response to the call for evidence (The Cabinet Office & Independent Commission on Freedom of Information, 2016). The Commission's final report concluded there is no evidence for restricting the right of access to government information. The report admitted that FOIA furthered openness and transparency and changed the culture of the public sector for better (The Cabinet Office & Independent Commission on Freedom of Information, 2016). It provided several recommendations for advancing the access to information, some of which did not require the amendments of the law. The report strongly endorsed wider proactive publication but as a parallel complementing mechanism, not as a substitution of FOIA. Interviewees argued that it was mostly the powerful resistance from the press and CSOs that eventually prevented changes to FOIA[38].

However, despite the government's strong emphasis on transparency[39], FOIA fell off the radar as also many interviewees noticed[40]. For instance, the UK government presided the G8 summit in 2013, and one of its main topics was the promotion of greater transparency[41]. In 2016, it hosted the international Anti-Corruption Summit[42]. Nonetheless, the statistics on the FOIA implementation, in particular, the share of refused FOI requests do not reflect this strong pro-transparency rhetoric (Institute for Government, 2019; The Cabinet Office, 2019). The recently created ministries, such as the Department for Exiting the EU (DExEU) and Department of International Trade (DfIT), have continuously been among the opaquest. Also, the UK government has not signed the CoE Convention on Access to Official Documents yet. The Convention has been open since 18 June 2009. So far, 17 CoE member states have signed it, and nine have also ratified it. It is the first binding international legal instrument to acknowledge a general right of access to the government's official documents.

---

[38] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[39] See https://www.gov.uk/government/news/pm-outlines-plans-for-transparent-government.

[40] INT6_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[41] See https://www.gov.uk/government/topical-events/g8-2013.

[42] See https://www.gov.uk/government/topical-events/anti-corruption-summit-london-2016.

5.3 A new way of the government's operation: open data

The Labour government's efforts in open data in the mid-2000s can be seen as a continuation of the narrative they created about the modern government that will revolutionise government-citizens relations. In 2007, two years after the FOIA has become effective, the government commissioned an independent review to look into new ways of how government information could be used. The review was carried out by Ed Mayo, the then Chief Executive of the National Consumer Council and Tom Steinberg, the then director of mySociety, a CSO running a number of online participation platforms. Its main topics could be grouped into three categories, in particular, (1) economic and (2) social value that government information generates and (3) the emphasis on enabling to reuse this information (Steinberg & Mayo, 2007). The word "reuse" occurred in the text 71 times. Other words that frequently occurred in the text were words "help" (69), "value" (52), "economic" (50), "commercial" (46), "innovate" or "innovation" (43), and "social" (42). The review was written in a positive tone, illustrating on several examples and case studies how better access to government information in open formats could bring improvements to different areas of life, from health care to education.

Steinberg and Mayo (2007) made a case that with increasing internet use, getting information online will become mainstream, and the government has to respond to that with providing more information online and creating new web-based services. They offered the government several recommendations, for example, engaging more with citizens online, encouraging excluded groups to participate as well, sharing raw data free of charge and introducing open licensing to government information to enable its further use. The government published a response to the review, and partially or fully accepted proposed recommendations as valid and useful (The Cabinet Office, 2007). The measures that followed in the next three years demonstrated that the government engaged with these recommendations to a great extent.

For instance, soon after the review, the government started preparations to launch an open data portal that would serve as a central repository of government data. They managed to launch it in February 2010, shortly before the general election, in which the Labour Party lost to Conservatives who eventually created a coalition with Liberal Democrats. It is important to mention that a major scandal preceded the election, which revealed that MPs widely misused their allowances and expenses (Telegraph Reporters, 2009). The investigation known as the Expense files cost several MPs their jobs. Also, criminal charges were brought against eight MPs in total; some of them were later jailed. This scandal had also determined what topics were discussed in the upcoming election campaign. Access to government information remained to be on the top of the agenda, but election campaign narratives narrowed the agenda down

to government transparency. For instance, the Conservative Party's Manifesto stated that,

> "Drawing inspiration from administrations around the world which have shown that being transparent can transform the effectiveness of government... [we] will bring in new measures to enable the public to scrutinise the government's accounts to see whether it is providing value for money. All data will be published in an open and standardised format" (The Conservative Party, 2010: 69).

This also showed that the UK political leaders did not want to fall behind what had been done in the area abroad. In particular, they followed closely transparency reforms that were being adopted by the US government under Obama, from the adoption of the Open Government Directive to the launch of open data portal Data.gov.

The narrative about transparency remained strong also after the election, and the 2010-2015 Conservative and Liberal Democrat Coalition government continued to profile itself as the most transparent government in the world, which will release an unprecedented amount of government data. In one of his speeches during the first month as a prime minister, Cameron said,

> "With a whole army of effective armchair auditors looking over the books, ministers in this government are not going to be able to get away with all the waste, the expensive vanity projects and pointless schemes that we have had in the past" (Prime Minister's Office, 2010).

Many were sceptic about broader take-up by citizens and perceived open data as a subject of interest primarily to data journalists, CSOs and private sector (Stoneman, 2015). One interviewee indeed argued that there was a strong push from the civil society sector for the government to publish open data. A part of this push was also the creation of new non-partisan organisations, such as the Open Data Institute in 2012[43].

Also, in 2012, the Cabinet Office presented to the Parliament the major policy paper on open data - Open Data White Paper[44], which to some extent built on the endeavours of the preceding 2007-2010 Labour government and emphasised the role of government information. Alike the review by Mayo and Steinberg, the white paper had a positive tone. While the review focused more on economic and social benefits that better access to government information can bring, the white paper emphasised

---

[43] INT2_CSO. (October 2017). Personal interview. For more information, see Appendix 2.

[44] In the UK, white papers are policy documents produced by the government. Their traditional role is to provide a solid foundation for further consultation on a particular issue of interest with all affected groups. It is often a stepping-stone for drafting and adopting related legislation.

transparency as an ultimate goal that the government wants to achieve with data publication. The words "transparent", "transparency" and "openness" occurred in the document more than a hundred times. The paper suggested that the government was aware that to meet this goal, it will have to make some sacrifices and is willing to do them:

> "There is nothing easy about transparency. The formative years of open government will be tricky, difficult and uncomfortable at times. But the price is effective, personalised 21st-century democracy" (The Cabinet Office, 2012: 6).

However, at the same time, the paper made it very clear that transparency the government proposes will have certain limits. For instance, one passage stated that "we are unflinching in our belief that ***data that can be published*** [emphasis mine] should be published" (The Cabinet Office, 2012: 6). The paper did not further specify what this conditionality means and who will decide and on what grounds what can or cannot be published. Along these lines, the paper stated that "where we can make progress in opening up policy ***without damaging the policy-making process*** [emphasis mine], we will" (The Cabinet Office, 2012: 31). These and similar quotes from the paper make a case for exemptions from transparency and a room for manoeuvring when needed.

As the name of the White Paper suggests, the emphasis was on open data, i.e. proactive publication of government data in open formats. The paper was very comprehensive about the potential benefits of open data and supported the arguments with several interesting case studies from the UK and abroad. However, it was scarce about potential risks. The document highlighted the risk of compromising privacy but other than that; open data was presented largely as risk-free agenda, which was in contrast with available research at that time (Benjamin et al., 2007; Gurstein, 2011; Morozov, 2013). Researchers suggested that in some instances, publication of government data, for example, crime statistics or land ownership could reinforce already existing economic and social divides. The paper also discussed what reporting mechanisms will be in place and how government agencies will track and evaluate their progress in open data. Users' experience, their needs and preferences were emphasised in several passages. In this relation, the paper also focused on how the future uses of data could encourage innovation.

Although open data is closely linked to FOI legislation, the paper ignored it, as well as it omitted the issue of public demand for information. When the paper used the term ***demand-led*** transparency, it never referred to a right of an individual for government data, but rather to the data publication informed by expert advice. At the same time, the Parliament appointed the Justice Committee to examine FOI legislation and its workings, concerning its effectiveness, strengths and weaknesses and

accomplishments of its initial ambitions. The Committee concluded that while proactive data publication has great value, it cannot substitute the right of access to government information because it is unachievable to predict the data demand (House of Commons Justice Committee, 2012).

To facilitate demand-led transparency as outlined in the White Paper, the government established various bodies and boards. For example, the Open Data User Group was created to inform the open data delivery and increase its quality and engage with actual and potential government data users. Data Strategy Board was also set up to support open data release with a specific focus on datasets with potential economic value. In addition, the Public Sector Transparency Board was established to promote transparency agenda on all governmental levels.

In addition, another independent review of public sector information was commissioned in 2012. Stephan Shakespeare from YouGov, who was contracted to deliver the review, made a business case for open data and moved a focus from open data as a transparency policy to economic growth policy. He argued that the supply of open government data has to be more predictable for businesses to be willing to engage with it (Shakespeare, 2013). Other key recommendations of the review were adopting the National Data Strategy to ensure open data is an overarching policy, strengthening data skills and ensuring privacy is rigorously protected.

All in all, during the 2010-2015 Conservative and Liberal Democrat-coalition government, there was a political momentum for open data. Interviewees across all sectors, from the government as well as CSOs, agreed that open data was a political priority. One interviewee argued that the government wanted to be seen as a leader and massively increased the volume of government data in the public domain at that time[45]. The UK was a leader in all relevant open data rankings. In 2013 and 2014, it topped the Global Open Data Index. From 2013 to 2016 it was Open Data Barometer's number one, and it also ranked high in the OECD's OURdata index. However, as our interviewees argued, even back then, the government was not willing to take substantial measures, such as changing the legislation to improve its data release in certain areas. In particular, a change of legislation would be required for the election data to become published on the polling station level[46]. However, as interviewee argued, there was no political will to make such a change.

From 2015, the government's focus shifted to the preparation of the Brexit referendum and later on, on its outcome. Interviewees simultaneously agreed that

---

[45] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[46] INT2_CSO. (October 2017). Personal interview. For more information, see Appendix 2.

Brexit stalled progress in many policy areas. One interviewee argued that in general, policy issues do not get the previous levels of attention now, and policy on open data is "rather unfocused"[47]. The following developments related to open data agenda confirmed these views. For example, the Public Sector Transparency Board became part of the Data Steering Group in November 2015. The Open Data User Group was dissolved in 2015 despite the expectation that "close engagement with the open data user community would remain a priority for the new government" (Open Data User Group, 2015).

Perhaps the best case in point about how Brexit has affected the progress of open data policies is the National Data Strategy. Already in 2013, the Shakespeare review of public sector information recommended developing the strategy (Shakespeare, 2013). However, the first sign of the progress occurred only in June 2018 when the Department for Digital, Culture, Media and Sport (DCMS) announced the government's intention to develop the strategy. Since then it launched a call for evidence and organised over 20 roundtables. It also had planned workshops for November 2019[48], most of which had to be, however, postponed due to the upcoming elections, which were scheduled early as a result of the gridlock about the Brexit withdrawal agreement. This example demonstrates that, as one interviewee put it, open data is "a nice extra" when other things in the country are well functioning. However, if there is a crisis, "it gets to the bottom of the list"[49]. With Brexit on, civil servants have the additional workload to what they previously had with resources being constant[50]. This has also been manifested in the declined number of datasets published on the national data portal Data.gov.uk (see Figure 7-9).

---

[47] INT2_CSO. (October 2017). Personal interview. For more information, see Appendix 2.

[48] For the information about the postponed workshops concerning the National Data Strategy, see https://www.gov.uk/guidance/national-data-strategy.

[49] INT4_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[50] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

6 Journalistic use of FOIA and open government data: Evidence from expert survey and interviews

> "If you live in Britain there is no such thing as a day untouched by the ONS".
> (Cocozza, 2013)

The results of the quantitative cross-country analysis presented in chapter 4 demonstrated that the effect of government information availability on corruption is conditional upon the quality of media freedom. Where conditions for unrestricted media operation are poor, greater access to government information is unlikely to make a difference in government accountability or corruption control. These findings are consistent with the theoretical framework proposed in chapter 2. The UK has a free and diverse media sector. In this chapter, using an online expert survey and semi-structured interviews, I explore how British journalists (but also some civic activists) perceive their watchdog and information intermediary role and how they engage with government information and data. I investigate their preferences for different information sources, experience with FOI requests, but also their data analysis skills, and how they appropriated the new quantitative turn in their profession. Although this primarily is an exploratory work, I present several hypotheses based on previous research in the field. I proceed with the description of the used methods – an anonymous expert survey aimed at journalists and semi-structured qualitative interviews with CSO representatives and civil servants. I then present the findings and offer some policy recommendations derived from both survey results and interviews. Lastly, I discuss the limitations of this study.

6.1 Aims and hypotheses

The main aim of the survey and interviews was to generate data on how journalists and civic activists use FOI laws and open government data, but also how civil servants perceive these uses. While this part of my research is exploratory, I propose several hypotheses based on the previous research in the field, discussed below and partially in Chapters 1 and 2.

The debate about the relationship between journalism and technology is not new. Computer-assisted reporting (CAR) was a part of the journalistic profession already in the 1960s. In the 1990s, media scholar Garrison (1998: 19) argued that computer literacy would gain in importance and journalists would need to be able to use "spreadsheets, databases, and internet searching" to carry their jobs. He posited

that the quick pace of technological advancements widens the digital divide in the profession, in particular between the younger and older generation of journalists as those younger have an advantage of obtaining elementary digital skills in formal education. Garrison anticipated that digital skills would be essential for employability in journalism.

The present discussion about data journalism and data skills is, in its essence very similar. The main difference is that technological sophistication has increased, and so did demands on skills. Scholars acknowledge that these changes are inevitable, the question stands how they are going to transform the profession and who is going to be able to contribute (Ananny & Crawford, 2015; Howard, 2014). While journalists worked with data since the advent of CAR (Meyer, 1973), the current practice often requires working with big data, i.e. data of a very large volume. The analysis of such data prioritises certain skills, such as skills in computational methods, programming, data management and visualisation and knowledge of statistics (Appelgren & Lindén, 2020; Howard, 2014; S. C. Lewis & Westlund, 2015; Tabary, Provost, & Trottier, 2016). Some practitioners said in the study by Fink and Anderson (2015: 471) that these skills should be "part of every journalist's toolkit". Although not every open government data is necessarily big data, its effective use also requires the skills listed above.

The available research on data journalism does not suggest that these skills are prevalent in the journalist population, quite the opposite. For instance, Tabary, Provost and Trottier (2016) found that while almost a hundred of journalists contributed to selected data journalism projects realised between 2011 and 2013 in the Canadian province of Quebec, only five of them had advanced skills necessary for the data analysis and visualisation. Howard (2014) also argued that journalists are not upgrading their skills, but rather new actors with different backgrounds are entering journalism to carry data-related tasks. Moreover, he found that data journalism practices are concentrated in major news outlets, such as The New York Times, The Washington Post or The Wall Street Journal. Small regional or local outlets do not have the resources to develop data journalism practices. Other scholars have also demonstrated that data skills are a minority specialism in journalism (Schrock & Shaffer, 2017; Stoneman, 2015). Therefore, I propose a hypothesis that journalists use open data published by the government sparsely, as they continue to rely mostly on other sources of information.

h6a. Journalistic use of open government data is marginal.

Though, in the past five years, a lot has changed in terms of available training for journalism students. Higher education institutions have since established digital and data journalism courses. Moreover, most students who enter university education

now are digital natives. Most of them have always had access to digital technologies and been able to use them. They have participated in digital culture, managing their digital identities, without necessarily having a vocabulary to describe their online activities using these terms (Palfrey & Gasser, 2008). Some scholars argued that digital natives' natural adoption of new digital technologies distinguishes them from so-called digital immigrants and is a distinctive feature of the further generational divide (Prensky, 2009; Solomon & Schrum, 2007).

However, other scholars are more sceptical about whether being digital natives also means having sophisticated digital skills beyond using digital technologies for "routine forms of communication and information retrieval" (Buckingham, 2011: x). Internet scholar Hargittai (2010) demonstrated that it is not age that divides internet users but their socio-economic status. She eloquently argued that young users of digital technologies do not represent a homogeneous group. Their uses and skills vary based on their parental education, gender and race. Children of parents with higher achieved education of white or Asian American origin and males were associated with more advanced digital skills. Thus, the assumption that everyone born to the internet era is automatically better in using digital technologies is now largely recognised as a myth.

Yet, age might still matter for data journalism practices. Although the mere fact that younger journalists have been exposed to digital technologies does not implicate that they also have more sophisticated digital and data skills, it is still reasonable to expect they are better equipped to embrace the shift towards data journalism for different reasons. Kuhn's (1996) major work on how scientific paradigms change and become accepted suggests that there has always been a divide between junior and senior scientists with the former being more eager or at least less resistant to question previous practices, and thus enabling the adoption of new norms. He argued that this is because those who are "either very young or very new to the field" are "little committed by prior practise" (Kuhn, 1996: 90). His theory was applied to a number of fields, journalism included (Hellmueller, Vos, & Poepsel, 2013). I argue that newcomers to the profession, being little committed to previous norms, would be more likely to adapt to new norms. They would be more willing to engage with new data journalism practices and develop necessary data skills. I posit that fresh graduates of digital and data journalism programmes would be less resistant to changes in the profession than senior journalists and also, thanks to the recent changes of journalism courses curricula emphasising digital and data skills, they would be practically better equipped to embrace these changes in the newsrooms. Therefore, I formulate the following hypothesis:

h6b. Age matters for the level of data skills' sophistication. Junior journalists are more likely to possess more advanced data skills than their senior colleagues.

That said, it is important to note that while I generally expect junior journalists to be more digitally and data-savvy, many established reporters have always had an interest in new technologies and quickly acquired new skills. Fink and Anderson (2015) conducted interviews with journalists from the US newspapers and online-only news sites with large, medium, and small circulations to capture if the size of the outlet matters for data journalism practices. One of their interviewees mentioned that he started with Excel in the 1980s and upgraded his skills from there as the range of available tools for data analysis broadened. Still, I expect that the general trend would show more advanced digital and data skills in a younger population of journalists based on the above.

Data skills are of utmost importance for benefiting from the publication of open government data, but as Howard (2014) argued an over-emphasis on data skills might move other crucial aspects of the profession to the background. In particular, reporting skills, which are still valued the most, at least in some EU countries as Örnebring's and Mellado's (2018) research suggests. Hermida and Young (2017) documented the age gap in digital and data skills and tensions it creates concerning the discussion on what contributes to quality journalism. Their interviewees – Canadian data journalists – complained that their work does not get sufficient credit on its own and that they are often treated as assistants. This view was also repeated by the interviewees who were established journalists, i.e. in their view, the most important for journalism are excellent writing skills (ibid). They perceived data skills as a nice-to-have add-on only.

The last hypothesis I propose relates to FOI laws and open government data from a different perspective. Journalists demand greater transparency from the government. The logic behind this push for more and better government information is straightforward – if we know more about the inner workings of the government, we can hold it accountable. Some scholars argue if we can hold it accountable, we can trust it more. Although others oppose the idea that transparency implicates trust (O'Neill, 2006), generally, transparency has been seen as a remedy that could restore trust in governments. This belief has also been extended to other sectors; journalism included. According to Deuze (2005: 455), transparency in journalism means allowing "people both inside and external to journalism... to monitor, check, criticise and even intervene in the journalistic process". It represents a paradigmatic shift from objectivity, as it goes a step further and provides evidence where previously only statements were available (D. S. Allen, 2008; Diakopoulos & Koliska, 2017;

Hellmueller et al., 2013; Porlezza & Splendore, 2019). Transparency and data journalism are intertwined as the advancement of digital technologies has made publishing data about journalistic processes an easier task than it used to be in the brick-and-mortar newsrooms and printed copies of newspapers (Zamith, 2019).

However, the profession has been divided about the pros and cons of greater transparency. Those favouring the shift towards greater transparency argue that if they do not share their data with the public, they become part of the system they are challenging every day as a part of their job. Porlezza & Splendore (2019) conducted interviews with Italian data journalists and found that transparency is a value embedded in their work. In the US, news organisations, such as FiveThirtyEight specialised in data journalism, publish data and code used to produce some of their stories in GitHub repositories (Diakopoulos & Koliska, 2017). Although transparency is much praised, it still represents a rare sight in newsrooms. The examples above are exceptions, not the rule. In practice, very few newsrooms adhere to this new normative ideal of transparency. Zamith (2019) studied journalistic practices in The New York Times and The Washington Post. He found that even these big and highly professionalised outlets are modest in providing information about how their stories were produced and linking primary data sources to them. Fewer than two in ten articles linked the datasets that were used to create them or provided some methodological note. While this might be partially due to the novelty of the transparency concept and limited capacities of smaller regional and local newsrooms to publish their primary data sources, transparency competes with and contradicts other important journalistic values, mostly the demand for exclusivity.

Sharing journalistic practices and sources in the public domain might undermine competitive advantages (Diakopoulos & Koliska, 2017). Fink & Anderson (2015: 471) conducted interviews with data journalists in the US, and one of their interviewees explicitly mentioned that "he was reluctant to share his own [data journalism] experiences on the email list because he was worried that someone would steal his ideas". Other scholars (D. S. Allen, 2008; Broersma, 2013) argued that transparency might produce unintended consequences, and in terms of trust likely in the opposite direction as was expected. Broersma (2013: 33) posited that "journalism speaks with an authoritative voice that leaves no space for doubt" and allowing the public to question journalistic practices would unavoidably weaken this authority. Journalists, thus, do not have incentives to reveal the inner workings of their newsrooms and how they source stories. On the contrary, the current system encourages them to use private means of communication to minimise the risk of revealing their sources and practices. Along these lines, I hypothesise the following:

h6c. Journalists are more likely to send FOI requests through their personal emails or any other private communication tools than using publicly available online participation platforms, such as WDTK to retain exclusivity of their stories.

6.2 Survey data and interviews

6.2.1 Survey sampling strategy and data collection

The anonymous online expert survey was open for three months from 12 May 2019 to 12 August 2019. It was realised through Qualtrics and distributed using personalised email invitations and social media for further promotion. I adopted a homogenous purposive sampling strategy, as I was mostly interested in the users of FOI legislation and open government data. Thus, I decided to contact journalists who focus on politics and write investigative stories as they are likely to be FOI requesters and browse government datasets. However, non-probability sampling strategies such as this one comes with major limitations. As Battaglia (2008: 523) states, they "do not attempt to select a random sample from the population of interest", and therefore, they can introduce bias if the sample estimates substantially differ from the population's true values (Stec, 2008). As a result, reported estimates can be either underestimated or overestimated. Purposive samples do not have strong external validity, and results from the survey using such samples cannot be generalised to the population (Davern, 2008).

Another limitation is the subjectivity of this sampling strategy, i.e. another researcher studying the same topic could have easily identified some other potential respondents than I did (Battaglia, 2008). However, while random sampling where everyone in the population has a nonzero chance of being selected is a superior survey method, it is not always possible. In the case of this study, the population includes primarily investigative journalists who submit FOI requests to get information for their stories or work with publicly available government data, and as such is hard to define. Even if objective criteria for defining such population exist, compiling an up-to-date list is a Sisyphean task. Therefore, the purposive sampling strategy was selected as the best available and executable option.

To identify journalists who use information obtained through FOI requests for their reporting, I used the website Foi.directory, which was established in 2012 by a British journalist Matt Burgess. He launched the website out of his frustration with government agencies, which obfuscated the submission of FOI requests, e.g. by not publishing their FOI officers' contacts. The directory entails contact details of the

agencies with a legal obligation to disclose information under FOI legislation. In addition, the website is affiliated with a Twitter feed, which highlights noteworthy journalistic stories based on information gained through FOI requests. I used the Twitter feed to identify the journalists who authored these stories. I collected the data about the articles that were featured on the feed from 1 May 2017 to 1 August 2019. Since the website started in 2012, I could have collected the data for the last seven years instead of only two years and three months, but I deemed it unnecessary for several reasons. First, when I collected the data for 2019, I was able to notice that the names tend to repeat. A simple explanation might be that some journalists use FOI as their working instrument more than others. If they used it regularly in 2019, they likely used it in previous years as well. Second, journalism, as a profession, is characterised by constant flux. It is a very dynamic profession, with journalists changing employers relatively frequently, also due to choice but also structural changes in the sector (Deuze, 2007; Witschge & Nygren, 2009). If I collected all available data for the past seven years, it could have become complicated to identify where those journalists who changed the outlet work at the moment.

Based on this data, I created a list of potential survey participants. I added to the final list all authors of the stories featured on the FOI Directory Twitter feed from 1 May 2017 to 1 August 2019, where I was able to identify either journalist's professional email address or a Twitter handle. If they had a Twitter handle (in most cases they did), I explored their networks and identified their colleagues who do investigative reporting. This snow-balling procedure generated additional contacts.

I also used the contacts published on the website Journalism.co.uk, as the filtering allows to select freelance journalists by their interest in politics. Also, simultaneously, at the time when the survey was open, forty journalists from different Scottish media outlets had signed a letter to the Scottish Parliament's Public Audit and Post-legislative Scrutiny Committee[51] pointing out to the difficulties they still experience in accessing government information under FOIA and demanding improvements (Edwards, 2019). I used the unexpected opportunity and added the signatories to the list of potential survey participants. In addition, I used LinkedIn to search for profiles of investigative and data journalists (they had to label themselves as such). I restricted the searches exclusively to the journalists who stated the UK was their location. The searches generated over 800 profiles. However, many were

---

[51] See "Open Letter: Expand and Strengthen Freedom of Information Legislation" at The CommonSpace: https://www.commonspace.scot/articles/14313/open-letter-expand-and-strengthen-freedom-information-legislation

irrelevant, and many were private with concealed contact details. Last but not least, journalists who were awarded or nominated for recognised prizes for investigative journalism, such as The Orwell Prize, Rory Peck News Award were added to the list.

The creation of the list was an on-going process, and the contacts were added even when the survey was already open. In the list, I recorded the following: journalists' name, their affiliation, email address, twitter handle or alternative contact detail if available, and the dates when an invitation to fill the survey and reminder were sent. In total, the personalised email invitation to participate was sent out to 1093 journalists. The first page of the survey was introductory and included information about the purpose of the study, and how the collected data will be stored, managed and used (see Appendix 2 for the supplementary materials for this chapter).

I recorded *out-of-office* responses and sent journalists a reminder upon their return to work. However, the opportunities for follow-up were restricted due to the anonymous nature of the survey. While the anonymity has advantages, e.g. encouraging honest views of participants as was discussed in greater detail in Chapter 3, it does not enable to verify who filled in the survey and who did not. Apart from the respondents who got back to me, and proactively confirmed that they had participated in the research, I did not know who participated. As I was unable to distinguish between those who already participated and who did not, I decided to send only one group reminder. It was sent out to all journalists one week before closing the survey to avoid overloading them, but also to limit the possibility of being reported as a spam. The survey was also promoted on Twitter. The descriptive analysis of survey results was conducted in R (v3.5.1).

6.2.2 Interviews

The survey for journalists was complemented by semi-structured qualitative interviews with five CSO representatives as they are also likely to be users of FOI legislation and government data and two civil servants. Initially, I aimed at a higher number of interviews with civil servants. However, many of my requests for a research interview at central government departments' FOI units were met with silence or refused. The CSOs were selected based on their mission and activities. I primarily considered organisations that advocate for the right of access to information and anti-corruption organisations that use government data in various ways. The sampling strategy was again purposive. The first round of interviews was conducted in November 2017 online, and the second round of interviews was realised in December 2018 in London. In addition to these, an interview with the Scottish Information Commissioner was conducted in September 2019 in his St Andrew's office.

Prior to the interview, all interviewees were provided with participation information sheets, which discussed in detail how their information would be used (see Appendix 2 for the supplementary materials for this chapter). The interviews lasted for about forty minutes on average. They were conducted either in person or through Skype application. All were all audio-recorded and transcribed. Although some interviewees agreed to reveal their identity and professional affiliation, some of the examples and cases, they mentioned during the interviews, were shared under the condition of anonymity. Eventually, all transcripts were anonymised to minimise the risk of identifying interviewees who mentioned these examples and cases.

## 6.3 Findings: Journalists use FOIA and open government data for monitoring the government

6.3.1 Demographics

One hundred seventy-three participants started the survey, 164 granted consent to use their responses, and 152 completed the whole survey. One participant responded to the invitation to participate in the survey and stated that she was unable to complete the survey, as provided options did not fully capture her experiences. Apart from this one instance, there were no other requests for clarifications or complaints. The high completion rate (93%) shows that participants had a good understanding of what was being asked, and the questions measured what they purported to measure. Thus, the validity of survey results is strong.

As illustrated in Figure 6-1, out of 164 participants, 101 (62%) were male, and 59 (36%) were female. One (1%) survey participant identified with "other" category, and one (1%) preferred not to disclose her gender. Two (1%) omitted the question altogether, which was technically possible, as no questions in the survey were required to be answered to complete the survey. Compared to the UK population of journalists, women are underrepresented in this sample. According to the Office for National Statistics, and its results from the last Labour Force Survey, in 2018, there were 89 595 journalists in the UK with 47 852 (53%) males and 41 743 (47%) females (The Office for National Statistics, 2018).

Although the survey was intended for journalists primarily and was promoted to encourage journalists to participate, it was designed to capture views of CSO representatives as well. In sum, 124 (76%) participants claimed to be journalists, 28 (17%) data journalists, and five (3%) stated that they are civic activists. Seven (4%) ticked an "other" option and identified themselves mainly as academics and researchers (see Figure 6-2).

Figure 6-1: Survey participants by gender
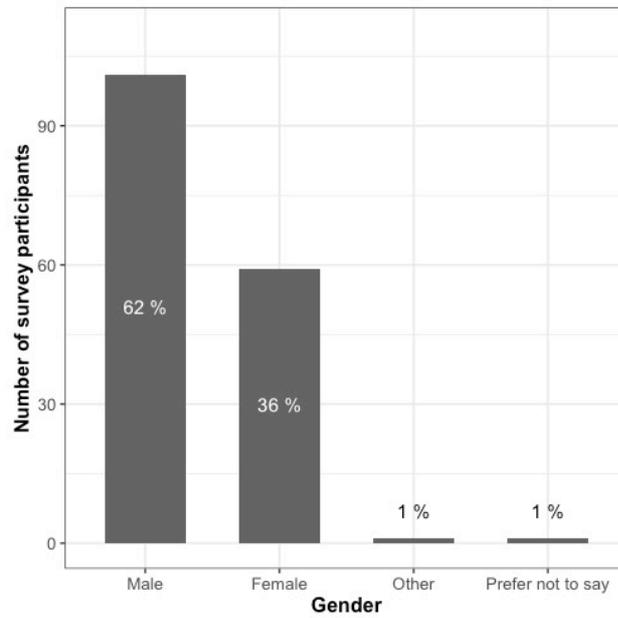
Note: Participants were asked survey Q4: What is your gender? The frequencies in the figure might not add up to 100% of survey participants (n=164), as none of the questions was required for submitting the survey.

Figure 6-2: Survey participants by profession
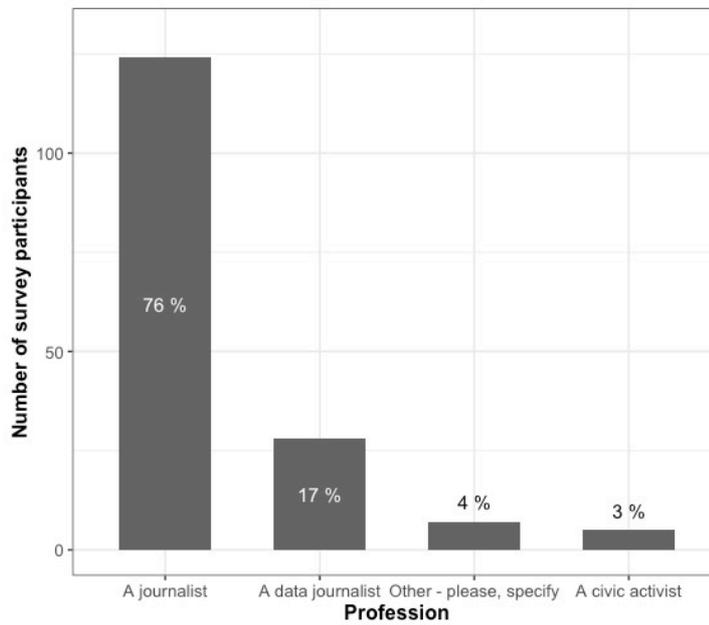
Note: Participants were asked survey Q2: What is your profession? The frequencies in the figure might not add up to 100% of survey participants (n=164), as none of the questions was required for submitting the survey.

The average age of participants was 37 years, with the youngest being 20 years old and the oldest being 72 years old. When categorised into age groups, the majority of participants was in the age bracket from 25 to 34 years, and thus junior journalists with relatively short working experience (see Figure 6-3). Due to the different framing of age groups in the surveys by the Office for National Statistics, I was unable to compare if age representation in my sample corresponds with their surveys. However, I was able to compare it with the representative survey of journalists realised by the Reuters Institute for the Study of Journalism for the age bracket from 25 to 34 years. In their representative survey from 2016, 27% of journalists are located in this age bracket (Thurman, Cornia, & Kunert, 2016). Compared to 46% in my sample, this age group is clearly overrepresented in my research, which needs to be taken into account when discussing other findings. Previous research about journalists' skills showed that similarly as in population, the digital divide between the young and the old also exists between junior digital-native journalists and senior journalists who lack digital skills (Willnat, Weaver, & Choi, 2013). The overrepresentation of younger generation in my sample might, thus, exacerbate the overall level of digital skills.

Figure 6-3: Survey participants by age
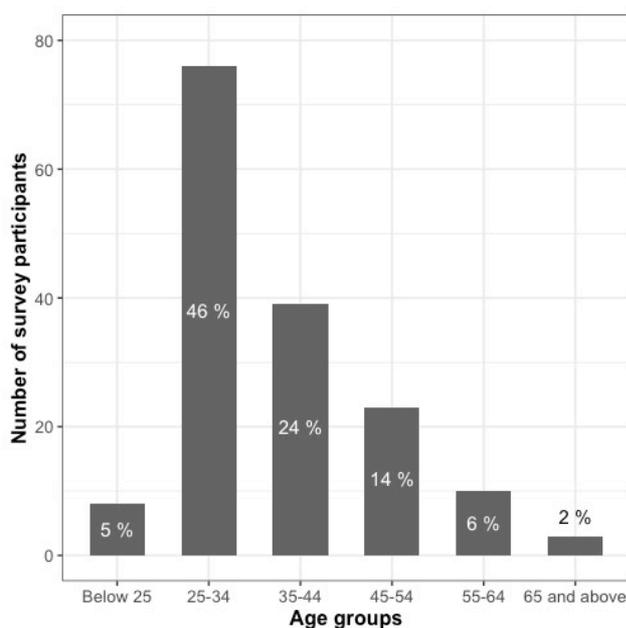
Note: Participants were asked survey Q3: What is your age? The frequencies in the figure might not add up to 100% of survey participants (n=164), as none of the questions was required for submitting the survey.

6.3.2 Journalistic sources

Survey results show, against the initial expectation, that surveyed journalists equally use FOI laws and publicly available government data as their primary source. Thus,

the hypothesis **h6a**, that journalistic use of open government data will be marginal cannot be confirmed.

As illustrated in Figure 6-4, 94 (57%) participants stated that in their reporting, they use mostly information obtained through FOI requests as well as publicly available government data. Twenty-five (15%) survey participants claimed that they use other sources than the offered survey options. In addition to sources stated in the survey, they use data from the civil society sector, companies' annual reports and information and data from courts. This is in line with the findings from interviews. A CSO representative who works on anti-corruption issues stated that he always starts his investigations with publicly available data, and only when he exhausts all possible sources, he submits an FOI request [52]. Another CSO representative who provides FOI training for journalists argued that they also recommend this practice. He further explained, "... [journalists] should use FOI to close the gaps in their stories. We discourage them from using FOI legislation in the hope of finding one"[53].

Figure 6-4: Survey participants' main information sources
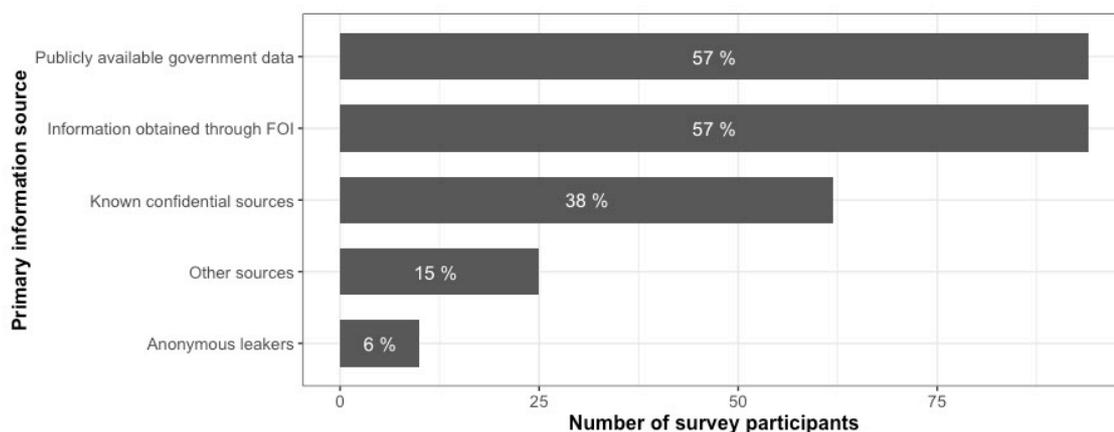
Note: Participants were asked survey Q5: What is your main source of information for your stories? (Please tick max. two that apply in most cases). The frequencies in the figure might not add up to 100% of survey participants (n=164), because this was a multiple-choice question, and none of the questions was required for submitting the survey.

Sixty-two (38%) respondents claimed to rely on their confidential sources and whistle-blowers known to them for information. Only ten (6%) participants stated anonymous leakers as their source of information. A CSO representative working on anti-corruption issues also mentioned that his organisation rarely gets information

---

[52] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[53] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

from whistle-blowers or anonymous leakers, as it does not have capacities to provide them legal support[54].

The explanation of low reliance on both known or anonymous whistle-blowers lies partially in legislation guaranteeing whistle-blowers' protection or a lack of thereof. As Smith and Cavaliere (2017: 5) argued in the latest Media Pluralism Monitor report for the UK, although the Public Interest Disclosure Act 1998 provides protection to employees from being laid off or punished for disclosing issues of public interest, it is unclear whether the same level of protection remains if they decide to disclose the information to the media. Legislative changes strengthening the protection were expected as a result of the UK OGP action plan 2013-2015, but despite several measures taken to identify weak spots in the current legislation and practice, the status quo remained (Worthy, 2017b).

There might also be other reasons behind particularly low use of information from anonymous leakers. First, anonymous sources might be difficult to verify (Lecheler & Kruikemeier, 2016). Their credibility, and thus a journalist's reputation as well could be jeopardised. Second, the latest leaks, such as Panama[55] and Paradise Papers[56], showed that the large scale and nature of the leaks require new approaches to investigations (Sambrook, 2019). In most cases, it is not attainable for a journalist to keep the exclusivity and work individually on a story. The scale of data released in the leaks requires multiple skills, including programming and data analysis skills, and thus, its use is impossible without collaboration (Léchenet, 2014).

6.3.3 Use of FOIA

In a representative survey of journalists conducted by the Reuters Institute for the Study of Journalism, 48% of its respondents stated that monitoring and scrutinising political leaders is very or extremely important in their work. Even more respondents (59%) believed in the importance of scrutiny of businesses (Thurman et al., 2016). Scrutiny is, however, not possible without access to information and the results of my survey confirm that journalists use FOI legislation to get it. Out of 164 survey

---

[54] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[55] The Panama Papers were published in 2016 and involved more than 350 reporters from 80 countries. The collaborative work was coordinated by the International Consortium of Investigative Journalists and revealed how over hundred prominent figures of political and social life concealed their wealth and dubious businesses through hard-to-trace companies, and tax havens using services of Panama-based law firm Mossack Fonseca.

[56] The Paradise Papers were published in 2017 and involved more than 380 reporters from 65 countries. The International Consortium of Investigative Journalists coordinated the efforts. The leaks again exposed offshore holdings of hundreds of politicians and celebrities all over the world.

participants who granted a consent to use their survey data, 155 (94%) claimed that they submitted an FOI request at least once in their career. One hundred and thirty-eight (84%) of them agreed that FOI legislation is instrumental for their work and stories. A handful mentioned that they do not submit FOI requests and do not see them beneficial to their work. Most of the participants perceived FOI legislation as an accountability tool that assists them in performing their watchdog role. The fact that almost every survey participant has claimed to submit an FOI request at least once in her career suggests that the legislation in the UK is accessible. An interviewee representing a CSO promoting the access to information agreed:

> "... [FOIA in the UK] is quite easy for people to make requests. Statistically, the number of requests made in the UK, which is far higher than in other countries, demonstrates that it is one of the most accessible FOI regimes"[57].

The number of FOI requests has been rising since the act has been in force (see Figure 6-5). Government agencies have previously argued that obligations under FOIA are burdensome. However, as Information Commissioners and experts oppose, it is not an easy task to establish when FOI requests are too many, as any request might be in the public interest.

Figure 6-5: FOI requests sent to the central government bodies from 2005
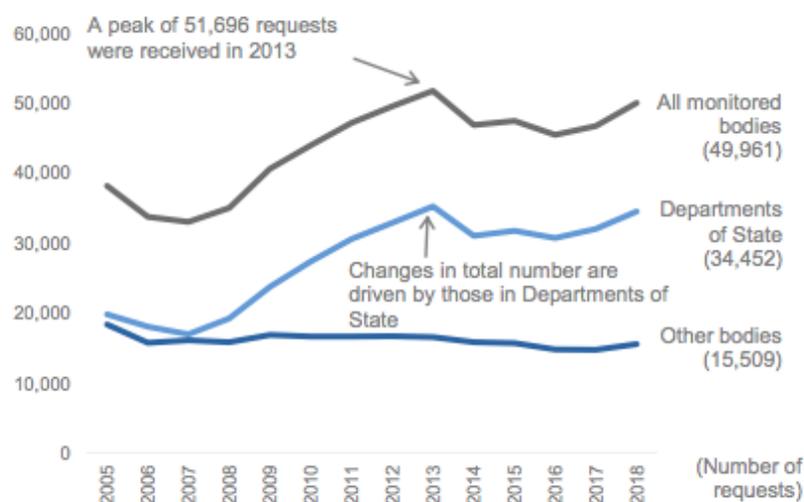
Note: The source is the Cabinet Office and its Freedom of Information statistics: annual 2018.

Figure 6-6 illustrates that only 24 (17%) survey participants submit over 100 original FOI requests per year, which accounts for two FOI requests per week.

---

[57] INT6_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

Figure 6-6: Frequency of submitting FOI requests
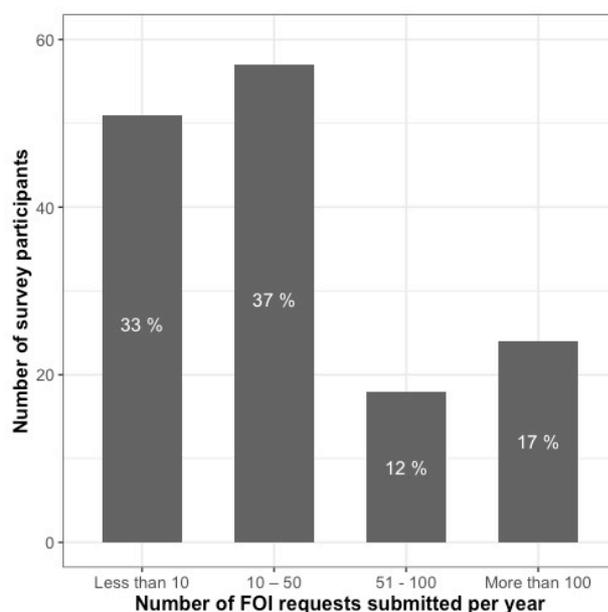
Note: Participants were asked survey Q8: On average, how many FOI requests do you file per year? This was a single-choice question. The frequencies in the figure might be lower than 100% of FOI users (n=155), as none of the questions was required for submitting the survey.

The majority of survey participants claimed that they preferably use their work or personal emails to submit FOI requests. This is consistent with the hypothesis **h6c**, which proposed that journalists are more likely to send FOI requests through their personal emails than using publicly available online participation platforms, such as WDTK. This finding is also in line with the previous research (Stoneman, 2015), which suggested that journalists are reluctant to use the public platforms for submitting FOI requests. Even though these platforms arguably simplify their work, they are concerned about the loss of exclusivity of their stories. The openness of these platforms and a legal obligation of a requester to state her name allow anyone to see in what information journalists are interested and whether they get it. Stoneman (2015) argued that it still does not necessarily reveal how a journalist planned to use the obtained information. However, it increases the risk of giving away cues and sources. He illustrated this resistance with discussions that accompanied the pilot scheme introduced by the US Federal agencies, which started to release responses to FOI requests in the public domain. Similar insights were shared in an interview with a CSO representative who worked with journalists:

> "They would not use, for example, AsktheEU.org [a platform for submitting FOI requests to the EU institutions], because they want to ensure they still have the exclusive ownership of the information. Journalists use these platforms to find

information. So, they would not want their information to be there for another journalist to steal their stories"[58].

At the same time, while journalists are not using public platforms to submit their FOI requests, they indeed use them for fishing for potentially interesting stories and exploring how to frame their requests to increase the chances of information disclosure. A CSO representative argued that WDTK is valuable precisely for this advantage it gives to its users. Prior to submitting a request, users can explore if similar requests are responded to and how[59]. Ellison, a Scottish investigative journalist, also advised fellow journalists to use these platforms as a source for their stories[60].

CSOs operating platforms for submitting FOI requests recognised journalists' requirement for confidentiality as a business opportunity and created a product, which meets this requirement. For instance, the British CSO mySociety launched in 2017 a pro version of WDTK[61] specifically targeting journalists. The pro version allows journalists to keep both requests and responses private until the story is published. In addition, it also provides tools for better management of FOI requests and collaboration in large-scale investigation requiring sending requests to multiple government agencies. As can be observed from Figure 6-7, although the participants' use of WDTK was generally low, there was a slight preference of a pro version over its public counterpart. One interviewee also mentioned that journalists might prefer the pro version of WDTK for safety reasons[62].

CSO representatives also claimed to use WDTK, both a public and a pro version. One interviewee stated that their reasons for using pro version are different from those of journalists. He stated that for his organisation, there is no harm in having their ongoing work out in the public domain. Quite the opposite, because if their FOI request is rejected in the public domain, they can then encourage their supporters to request this information again. However, they still use the pro version to file requests. In particular, when they need to send round-robin FOI requests (same requests to multiple public authorities), because a pro version includes more advanced tools for managing requests, for example, getting alerts about upcoming deadlines[63].

---

[58] INT6_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[59] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[60] https://www.bbc.co.uk/news/uk-scotland-50073052

[61] WDTK pro version is available at https://www.whatdotheyknow.com/pro.

[62] INT6_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[63] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

Six (4%) survey participants claimed to use other means to submit their FOI requests. Some of them stated what these are, i.a. email accounts specifically created for this purpose.

Figure 6-7: Primary means of submitting FOI requests
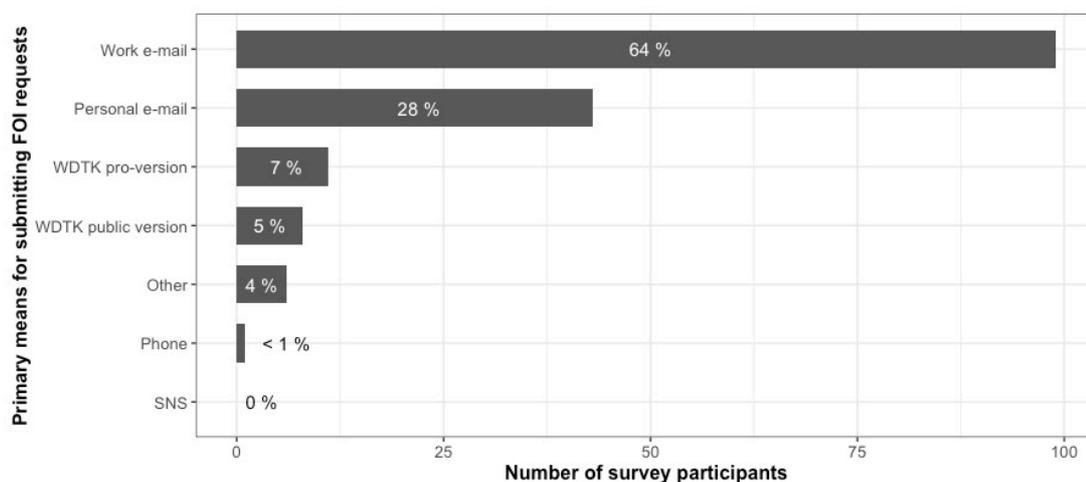
Note: Participants were asked survey Q9: How do you usually submit an FOI request in most cases? (Please tick all that apply). The frequencies in the figure might not add up to 100% of FOI users (n=155), because this was a multiple-choice question, and none of the questions was required for submitting the survey.

As was already discussed in depth in previous Chapter 5, the access to government information in the UK is governed by two pieces of legislation, which both have been in force since 1 January 2005. The Freedom of Information Act 2000 (FOIA 2000)[64] regulates access to government information in England, Northern Ireland and Wales. The Freedom of Information Scotland's Act 2002 (FOISA 2002)[65] governs access to information held by Scottish public authorities. Therefore, the survey question on the main addressees of FOI requests included both devolved Scottish government and the central British government. Since the main focus of the UK case study is not on the provision of information by local government, any differentiation was omitted at the local government level.

Figure 6-8 shows that 55 (35%) survey respondents send their FOI requests mostly to the central British government and 44 (28%) to the local government. Forty-one (26%) survey respondents stated that they address their FOI requests mostly to other government agencies than those offered as survey options, namely police forces, National Health Service's (NHS) boards and trusts, universities, Scottish Prison Service, Parole Boards for the UK and Scotland, the British public service broadcaster

[64] Legislation.gov.uk, "FOIA 2000". Available at http://www.legislation.gov.uk/ukpga/2000/36/contents

[65] Legislation.gov.uk, "FOISA 2002". Available at http://www.legislation.gov.uk/asp/2002/13/contents

BBC and the communications regulator Ofcom. Only nine (6%) survey participants claimed the Scottish government to be the main addressee of their FOI requests.

Figure 6-8: Main addressees of submitted FOI requests
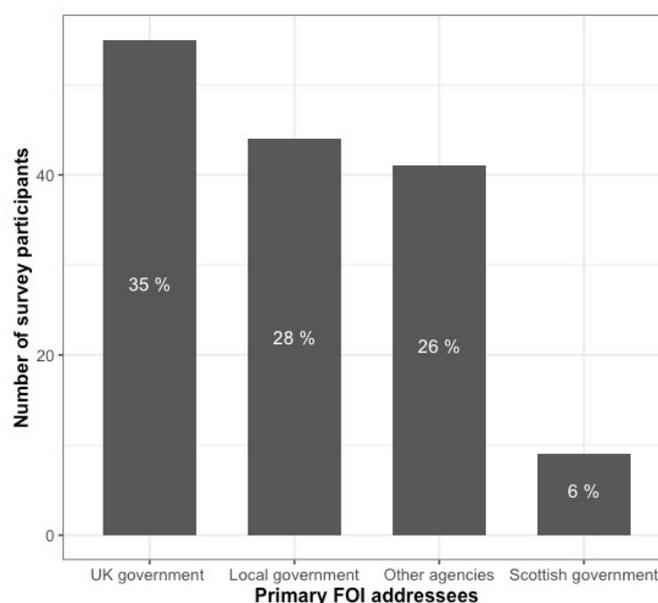
Note: Participants were asked survey Q10: Who are the addressees of your FOI requests in most cases? This was a single-choice question. The frequencies in the figure might be lower than 100% of FOI users (n=155), as none of the questions was required for submitting the survey.

The slight prevalence in focus on the central British government is understandable and in line with other available data on journalism in the UK. According to the latest report of the National Council for the Training of Journalists (NCTJ), journalism is heavily concentrated in London and the south-east of England with 33% of journalists located in London and 15% in the south-east of England (Spilsbury, 2018). The figures by the Labour Force Survey from 2018 show even greater disproportion with 52% of journalists being based in London and 13% in the south-east of England (The Office for National Statistics, 2018). However, it is also important to note that those working in London do not necessarily work for national outlets, which also explains why local governments follow the central government in the number of requests closely. In general terms, most FOI requests go to the local government. The latest study by mySociety estimated that in 2017, 467,000 FOI requests were sent to the local government (Parsons & Rumbul, 2019).

The experiences with submitting FOI are mixed. While the majority of survey participants agreed that FOI is crucial for their work, they were not satisfied with how it functions. The guidance of the UK's ICO concluded that requesters under FOIA 2000 are not required to state their motives or justify their requests in any way. At the same time, public authorities are allowed to ask what their motives are for

clarification purposes if the requests are "ambiguous or unclear, and knowing the reason for the request would help the authority identify the information the requester needs" (Information Commissioner's Office, 2015a: 8). However, it is up to the requesters to decide if they share their reasons, and their decision should not affect the likelihood of receiving the requested information. Despite authorities' inquiring about the reasons behind FOI requests should be a sporadic practice, 44 (28%) out of 155 survey participants, who submitted an FOI request, had previous experience of public authorities inquiring about the reasons behind their requests.

The main raised concerns and stated reasons for dissatisfaction with FOIA 2000 (less so with FOISA 2002) was high rejection rate. One hundred thirty-three (86%) survey participants who had previous experience with submitting an FOI request claimed that their request had been rejected at least once. While this does not indicate a problem per se, as public authorities have several legitimate reasons for rejecting to disclose the requested information, statistics of the overall UK central government's FOI performance suggest a trend towards greater opacity. The London-based think tank Institute for Government has pointed to the peak in rejected requests. They stated that the rejection rate has almost doubled since 2010 (Cheung, 2018).

These conclusions were echoed by interviewees from civil society as well. One CSO representative argued that several central government bodies "have been more secretive over time, and it does continue to be like that... If the government does not want to be transparent, it will find a way not to be"[66]. Another interviewee from a CSO advocating for FOI also shared the view that access to government information has been deteriorating, "partly due to the financial pressure on public authorities, partly due to the lack of enthusiasm for the act"[67]. An interviewee representing an anti-corruption CSO said that the main issue is inconsistency among authorities in responding to the requests with the same request being rejected by one agency and information disclosed by another one[68].

As illustrated in Figure 6-9, the first most frequent justification for rejecting FOI requests that survey participants received was a refusal on cost grounds. This experience mirrors the statistics of the central government's justification for refusals with exceeding the cost limit as the major reason (Cheung, 2018). The second most frequently cited reason for refusals was that disclosing the information will take the staff too much time, which, in legal terms, is refusal on very similar grounds as cost

---

[66] INT6_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[67] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[68] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

grounds. The guidance provided by the UK's ICO concludes that if "it would cost too much or take too much staff time to deal with the request", a public authority can reject the request (The Information Commissioner's Office, 2019). However, the public interest test must always apply, i.e. withholding information must outweigh the public interest in disclosure. The Institute for Government questions the current trend of using cost limits as justification for refusal broadly (Cheung, 2018).

As one of the interviewed CSO representatives said addressing rejections on cost ground is difficult, because as a requester "you do not know how the authority holds the information, how it searches for it"[69]. Therefore, it is difficult to narrow down the request in a way to avoid being refused on cost grounds. However, journalists and CSOs working on collaborative investigations found their ways to address this. A CSO representative mentioned that they join forces with another CSO and split the amount of requested information between them. Since the public authorities can reject requests if they represent a concerted effort, they sometimes send requests from their private email addresses to avoid being identified with the same organisation.

Figure 6-9: Most frequently cited reasons for FOI requests' refusals
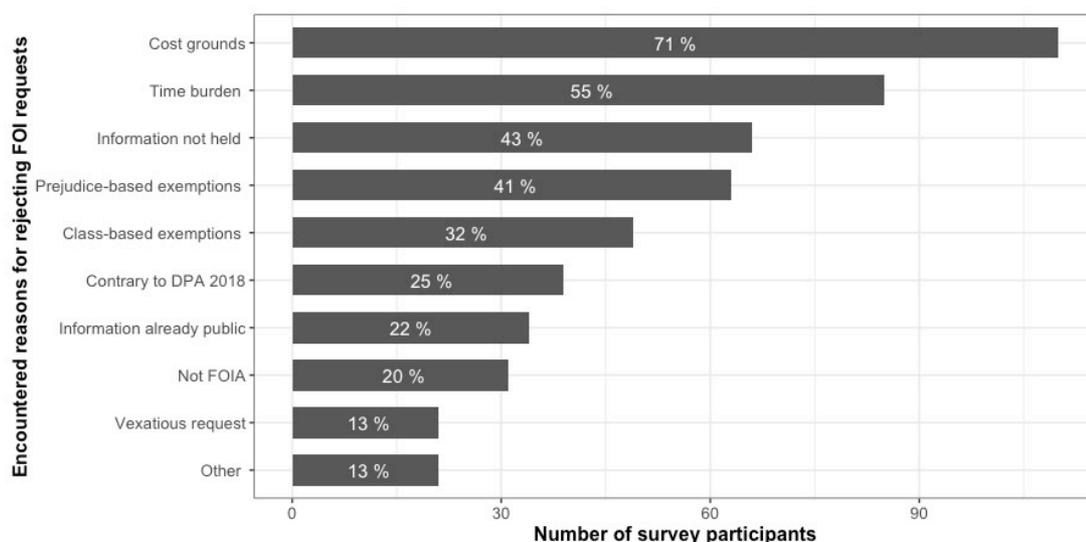
Note: Participants were asked survey Q14: On what grounds were your FOI requests rejected in most cases? (Please tick all that apply). The frequencies in the figure might not add up to 100% of FOI users (n=155), because this was a multiple-choice question, and none of the questions was required for submitting the survey.

The third most frequently cited reason for withholding information was that the information was not held. The duty to inform about the possession of information is present in both FOIA 2000 and FOISA 2002. The Scottish authorities must give

---

[69] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

requesters notice in writing that they do not hold the information. The UK FOIA 2000 distinguishes scenarios when the duty to confirm or deny does and does not arise.

Other encountered cited reasons for refusals were prejudice-based exemptions, i.e. disclosures that could create potential harm, such as information distorting someone's commercial interests and class-based exemptions, i.e. related to a particular class of information, such as information important for national security, information held for criminal investigations, information relating to correspondence with the royal family. Some participants also mentioned that their requests were refused, and information withheld due to potential infringement of the Data Protection Act 2018 and General Data Protection Regulation (GDPR) if it was disclosed. Alternatively, the information was already reasonably accessible by other means or that the authority was not a subject to FOI legislation. Twenty-one (13%) survey participants who submitted an FOI request at least once, had it rejected as vexatious.

FOIA 2000 and FOISA 2002 both state that if a request is vexatious, public authorities do not have to comply with the obligation to disclose information. However, neither FOIA 2000 nor FOISA 2002 defines what a vexatious request is. In its guidance, the UK's ICO lists several indicators of vexatious requests, e.g. abusive language, putting an undue burden on public authorities, lacking a serious purpose (The Information Commissioner's Office, 2015). The guidance provided by the Scottish Information Commissioner states similar factors, e.g. request's design to cause disruption to a public authority (Scottish Information Commissioner, 2019).

The statistics of the UK central government's reasons for refusing requests have shown that the levels of refusals on ***vexatious*** requests' grounds have been steadily low around 0 to 2% since 2010. This might come as a surprising finding, given the news coverage often suggests the opposite, and that public authorities are flooded by vexatious and frivolous requests[70]. However, Cherry and McMenemy (2013) who studied the FOI practice in 32 Scottish councils, found that the actual number of vexatious requests received was very low.

Parsons and Rumbul (2019) explored FOI practice in local councils on a much larger scale (n=418), and their findings also confirmed the low proportion of vexatious requests. They found that in 2017, 101 (33%) councils received at least one vexatious request, 61 (20%) had more than two, and 27 (9%) of councils had at least five

---

[70] For examples, see the article at the Huffington Post "FOI: Top 10 strangest requests councils have answered, from Father Christmas to zombies and aliens" available at http://bit.ly/2MNSCWN or from the Telegraph "Resident demands to know council's plans for Santa crash" available at http://bit.ly/2MKICxL or from the Sun "'Bonkers' monster requests hit the police" available at http://bit.ly/32iSHZf.

vexatious requests. In the council that provided information about vexatious requests, these represented only 0.16% of all requests. Parsons and Rumbul (ibid) estimated that the true value is even likely lower than that. The low proportion of vexatious requests have several competing explanations. First, requesters indeed file vexatious requests rarely. Second, as Cherry and McMenemy (2013) argued the subjective character of the indicators of vexatious requests make it difficult for public authorities to interpret and apply these. The Scottish Information Commissioner stated that "from [his] interviews with the number of Scottish government individuals as part of the [Scottish government] intervention there was a reluctance to use the vexatious provisions" (The Scottish Parliament, 2019: 10:21:49). He suggested that this might be due to "a high test that authorities have to meet to make the case of a request being vexatious"[71].

As illustrated in Figure 6-10, 71 (46%) survey participants agreed that previous negative experience with FOI practice, such as withheld information, has discouraged them from submitting an FOI request. Apart from the high rate of refusals, surveyed journalists complained that the time for complying with requests is long for the fast-paced work at the newsrooms. Sixty-three (41%) survey participants perceived twenty working days' time period to respond as a factor which has also deterred them from submitting an FOI request.

Figure 6-10: Deterrents from submitting FOI requests
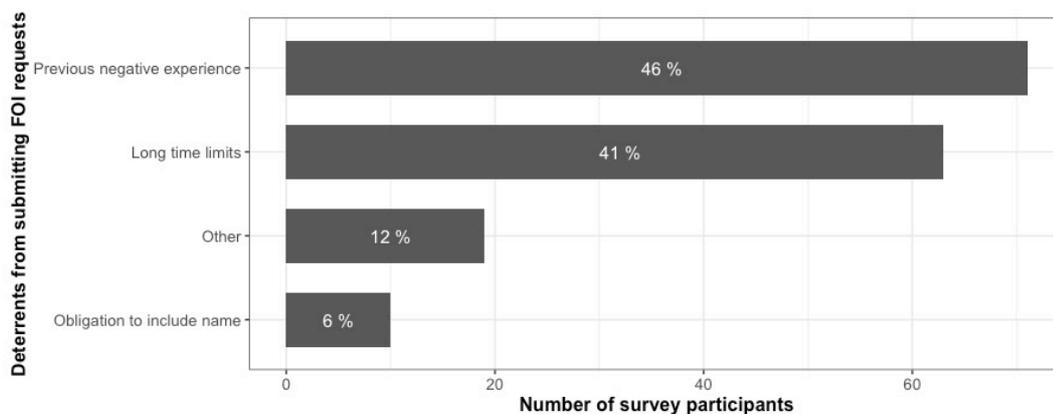
Note: Participants were asked survey Q11: Has any of the following ever discouraged you from submitting an FOI request? (Please tick all that apply). The frequencies in the figure might not add up to 100% of FOI users (n=155), because this was a multiple-choice question, and none of the questions was required for submitting the survey.

However, while journalists can prepare for twenty working days limits by planning in some cases, if these limits are not met, the stories might lose their

---

[71] INT7_SIC. (September 2019). Personal interview. For more information, see Appendix 2.

momentum. An interviewee representing an anti-corruption CSO argued that "statutory time limits are reasonable", but indefinite delays are not[72]. He mentioned his experience, in particular with local government, of not sticking to the time limits. Another CSO representative also remarked that in general, delays are "getting worse"[73].

Some of them listed a previous experience when their request was met with silence. In particular, DExEU, Ministry of Defence (MoD) and NHS were mentioned repeatedly in the survey as unresponsive, not keeping the time limits and frequently refusing their requests (SP19)[74]. Some of these examples were brought up in interviews as well, in particular, DExEU, but also Home Office and DfIT[75]. One journalist wrote that he got hold of information despite FOIA and not because of it. He mentioned an example of the HM Land Registry's overseas company dataset and UK Export Finance's internal business database, which were both released to him in error. Survey respondents also perceived the Scottish Government as opaque.

Only ten (6%) survey participants mentioned the obligation to reveal their identity as a deterrent from filing a request. One participant expressed a concern that by filing a request, he might give government agencies cues about what he is investigating (SP47), which might lead to destroying the evidence. Nineteen (12%) participants stated that factors other than those offered as survey options put them off, for example, having to appeal to the Information Commissioner and if unsuccessful to tribunals, which requires additional time. Others complained about the format in which information is provided and its quality. Mostly data journalists mentioned that often they get information in PDF format, which is not machine-readable. In cases when journalists submit identical requests to multiple public authorities, the responses are too diverse and inconsistent to be comparable.

Survey participants perceived the approach of government agencies to FOIA and FOISA as defensive. They complained that their requests "are being treated differently than those of any ordinary member of the public" (SP106). The problem of differential treatment was brought up in interviews as well. A CSO representative argued that whereas a majority of FOI requests is dealt with by regular FOI team, sensitive requests get to senior civil servants up in the departmental hierarchy[76]. Because they deal only with requests that might represent a reputational risk to the

---

[72] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[73] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[74] SP stands for survey participant number 1-164.

[75] INT6_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[76] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

department, they "see the threats, not advantages of the disclosure"[77]. The Scottish Information Commissioner confirmed that the Scottish government had a similar practice, i.e. the requests filed by journalists "went to a high-level decision-making process, so responses to journalists' requests were essentially delayed"[78]. In the Scottish government intervention, the Scottish Information Commissioner stated that this practice has to stop. He argued that "who you are should not dictate how your application is determined. The general principle [of FOIA] is that it is applicant blind"[79].

Researchers in other jurisdictions (Kwoka, 2018b) also pointed to the differential treatment of journalists' requests. Journalists among survey respondents felt more resistant approach from the Scottish government than the central British government (SP160). Survey participants argued that they are more successful in getting information from the Westminster (SP104).

Research on government's FOI practices has previously concluded that the effect of FOI legislation for governments is disruptive. However, chances of reverting FOI legislation after it has been adopted are minimal in any country. Thus, governments curb these effects by practices described and experienced above. Hayes (2009) found that agencies rely on blanket exemptions to withhold information, a practice that the UK's Information Commissioner also confirmed in its recommendation made to the Department of Health in 2008. Roberts (2005) also pointed to the practice of differential treatment of journalists' requests, unnecessarily delaying the time to respond to decrease the newsworthiness of the information. Camaj (2016b) argued that these practices have even more serious implications in countries transitioning towards democracy, where governments often use FOIA to control the access to information and set the agenda by providing exclusive access to pro-government oriented news outlets.

6.3.4 Use of publicly available government data

Although the majority of journalists still lacks the levels of digital literacy and skills needed for the current scale of investigations (Borges-Rey, 2016), it is essential to acknowledge that digital technologies have always been a part of the profession. Knight (2015), Léchenet (2014) and Parasie and Dagiral (2013), among others, documented that journalists have used CAR for exploring databases since the 1960s.

---

[77] Ibid.

[78] INT7_SIC. (September 2019). Personal interview. For more information, see Appendix 2.

[79] Ibid.

Meyer, a US journalist and one of the pioneers of applying social sciences research methods to the practice of journalism, won the Pulitzer Prize for his alternative reporting from Detroit riots in 1967, which included profiling rioters with the use of a survey (Meyer, 1973). He argued that reporting has to be built on facts and data rather than anecdotal evidence. The ideas of *precision journalism* – the term Meyer coined to emphasise the importance of rigour in journalism, have travelled to the current concept of data-driven journalism (Bounegru, 2012). Bounegru (2012) proposed that the primary role of data journalism is to assist in lowering "the barriers to understanding and interrogating data, and increasing the data literacy of their readers on a mass scale".

However, for journalists to be able to fulfil their intermediary role, they first need to have fluency in these skills. Previous research suggested that generally speaking, journalists' relationship to statistics used to be lukewarm at best (Maier, 2002), and the availability of formal training in this field was limited until very recently (Yarnall, Johnson, Rinne, & Ranney, 2008). However, in the past decade, the situation has improved with new university programmes in computational and data journalism being opened (Hewett, 2016). Although Örnebring and Mellado's (2018) research on the most valued journalistic skills in six EU countries including the UK showed that reporting skills are still valued the highest, the demand for additional skills has been on the rise. Journalists are well aware of the changes to their profession. Figure 6-11 illustrates that they perceive the strengthening importance of technical skills for performing and keeping their job (Thurman et al., 2016). Borges-Rey's (2016) research based on interviews with journalists from several UK news outlets echoes these findings. More than half of his interviewees asserted that data literacy would become a skill indispensable for journalism.

Figure 6-11: UK journalists' views on the importance of technical skills
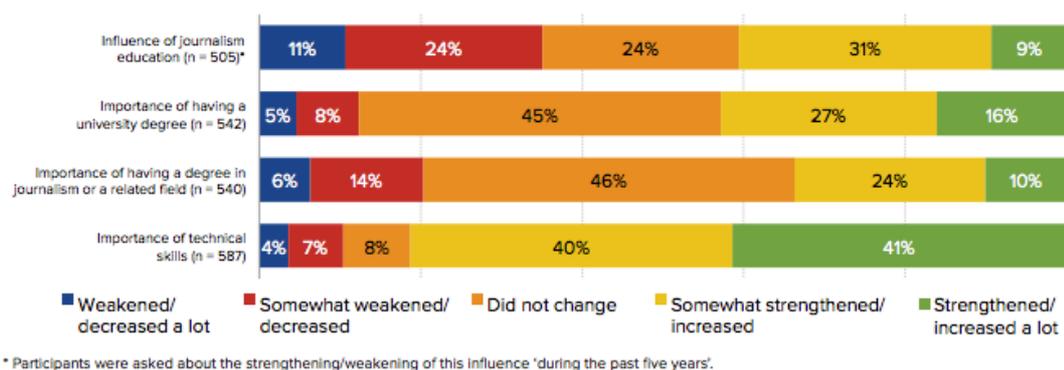
In general, my survey results demonstrate a positive approach toward data and data journalism practices. Only one survey participant expressed a strong stance against them:

> "I can't stand data journalism as a practice. I find it so boring I can scarcely find the words to describe how much I hate it. It is very worthy and important, but I am not the correct person to be doing it. I am far more interested in the active investigation than in consulting data" (SP47).

In comparison to obtaining information through FOI requests, one thing about publicly available government data came up in the answers to open-ended questions repeatedly – time efficiency and easy accessibility. Several survey participants argued that inspecting government data compared to filing an FOI request saves at least twenty working days, a statutory time limit with which public authorities must comply when responding to FOI requests. Survey participants stated this makes a big difference:

> "Waiting for FOI responses from various bodies can delay stories. Having the data readily available increases the speed at which I am able to report on things which in a stretched, local newsroom, is at a premium" (SP108).

Some argued that there is the time efficiency argument on the side of government as well, as public authorities are likely to get fewer requests as a result of proactive data publication. A CSO representative stated, "there are some local authorities who have cut the volume of FOI requests they receive substantially by identifying what people are asking for and by publishing it proactively on their website"[80]. The interviewees at the Scottish Information Commissioner Office also endorsed demand-driven proactive publication. However, they mentioned that the outcome in terms of the volume of received FOI requests is not so clear-cut. The anecdotal evidence from their discussions with public authorities was conflicting, suggesting that for some authorities, the proactive publication led to a reduction in FOI requests; for others, it triggered greater interest[81].

Other perceived benefits of open government data were its reliability and accuracy, which journalists argued is coming from the authority of its source. One journalist stated that publishing data of compromised quality would have reputation repercussions for the public authority in charge of this data. That said, many were sceptical how rigorously government data is compiled and how comprehensive it is. Some suggested that figures might be even manipulated. However, since they are coming from an official source, journalists perceived advantage in using it, because

---

[80] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[81] INT7_SIC. (September 2019). Personal interview. For more information, see Appendix 2.

they do not risk a defamation suit. The responsibility for the reliability and accuracy of the data lies with the public authority, not with the journalist.

Survey participants also appreciated the raw character of the data, which allows them "to study it without already having been ***pre-presented*** to them" (SP146). They argued that if raw data can be trusted, then they can analyse it independently and improve scrutiny. Previous research has also emphasised this aspect of raw data compared to those processed and interpreted. Baack's (2015) interviewees highlighted that raw government data in open formats allows journalists and civic activists to formulate their interpretations of this data. However, the fact that anyone can access and interpret data does not necessarily mean she will do so well and rigorously. Lugo-Ocando's and Brandão's (2016) research provided examples of journalists who misinterpreted crime data due to their lack of expertise and inability to understand this data in its complexity.

Figure 6-12: Ability to analyse data independently
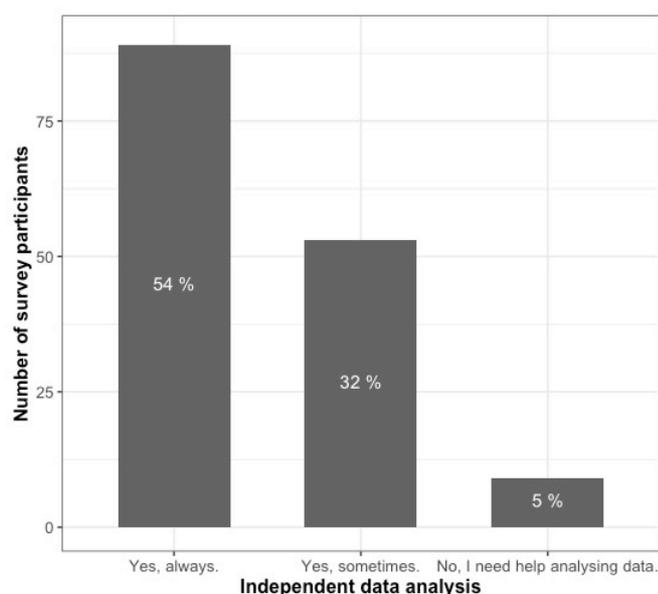
Note: Participants were asked survey Q21: Do you analyse data yourself? This question was a single-choice one. However, the frequencies in the figure might not add up to 100% (n=164), as none of the questions was required for submitting the survey.

In 2012, Rogers, the founder of the Guardian DataBlog and data enthusiast, proposed that "everyone can do [data journalism]"[82]. However, the practice has shown so far that it is not the case unless everyone has some statistical and computational

[82] Rogers, S. 2012. Anyone can do it. Data journalism is the new punk. Available at https://www.theguardian.com/news/datablog/2012/may/24/data-journalism-punk.

skills. As illustrated in Figure 6-12, over half (54%) of survey participants claimed always to analyse data themselves. Fifty-three (32%) survey participants stated that they sometimes analyse data themselves. Only nine (5%) survey participants admitted that they need help.

However, this particular question invites many interpretations, as it did not explicitly state what the term data analysis means and whether it entails analysis of both quantitative and qualitative data. One survey participant noted that he does not analyse data in a strictly statistical sense. At the same time, he chose a survey option indicating that he analyses data independently. However, another survey participant stated something along the same lines but chose an option that he always needs help in analysing data. Given that this question created ambiguities and potentially introduced measurement bias, the results need to be taken with caution.

Figure 6-13: Software used for analysing data
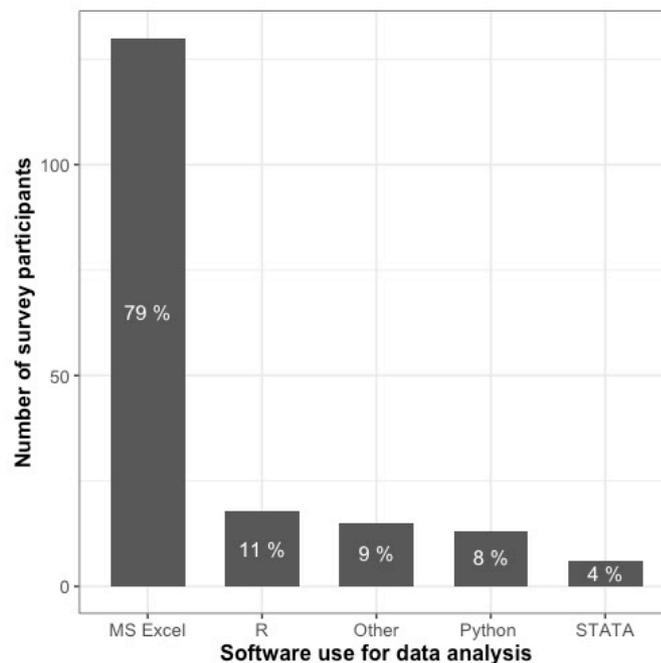
Note: Participants were asked survey Q22: What software do you mostly use for the analysis? This was a multiple-choice question. The frequencies in the figure might not add up to 100% (n=164), as none of the questions was required to be able to submit the survey.

The survey question about the use of specific software is more informative than the question about the ability to analyse data independently. As can be seen from Figure 6-13, 130 (79%) survey participants claimed to use MS Excel for data analysis. While it is not clear whether they use it to analyse or only read the data, the use of other software for data analysis indicates the level of statistical and computational skills. Out of 164 survey respondents, 18 (11%) claimed to use an open-source

programming language R mostly, 13 (8%) analyse data with Python and six (4%) work with Stata. Fifteen (9%) journalists stated that they mostly use software other than that offered as a survey option. In particular, they work with a graph platform Neo4j, OpenRefine, UltraEdit and SQL for managing data.

These numbers suggest that although the profession is changing and more journalists are acquiring new skills needed to reap the benefits from the publication of government data, data analysis in a quantitative sense is still a part of the skillset of a few rather than a more general trend. As was discussed in the introduction to this chapter, data journalism practices remain a minority specialism. These findings were also consistent with the interviews with CSO representatives who stated that they do not have an in-house data scientist's capacity in their organisations, mostly due to the lack of resources to pay for these skills[83]. Also, while they have some data analysis skills and interest to improve these, as a part of their daily work, they can allocate only a limited time to do so. That said, they argued if their skills were advanced, data analysis work on some projects could have been done more efficiently[84].

However, survey participants do not represent a monolithic group. As was already mentioned, the average age of survey participants was 37 years, with the youngest being 20 years old and the oldest being 72 years old. Perhaps contrary to the intuition and previous findings (Spilsbury, 2014), there was only a minor gap in statistical or computational skills between the younger and older generation. Therefore, my hypothesis **h6b** that junior journalists are more likely to possess advanced skills to analyse government data cannot be confirmed. In particular, Python was used by journalists of all ages. Its user was on average 39 years old (Md = 37) with the youngest being 24 years old and the oldest 65 years old. The average age of a user of R programming language and Stata was lower than of a Python user, 31 (Md=32) and 30 years old (Md=29.5) respectively. As was already mentioned, young journalists are overrepresented in this survey compared to the population. Thus, the skills within the population might look differently.

A substantial difference in the use of more sophisticated software for data analysis was neither present between genders. In contrast with available research on gender differences in journalistic digital capital (Vuyst & Raeymaeckers, 2019), in my survey, women had slightly higher rates of the use of more sophisticated software than men. While 14% of male survey participants self-identified with a data journalist

---

[83] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.
[84] Ibid.

profession, 24% of female survey participants did (see Figure 6-14 and Figure 6-15 below).

Figure 6-14: Survey participants by gender and profession (males)
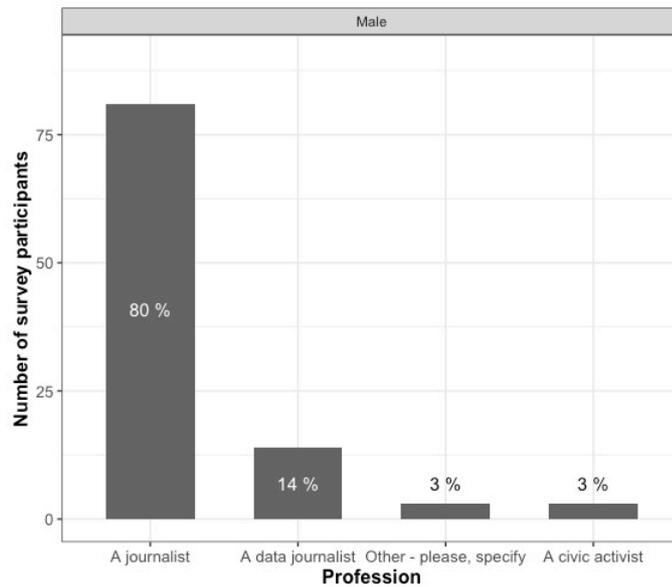
Figure 6-15: Survey participants by gender and profession (females)
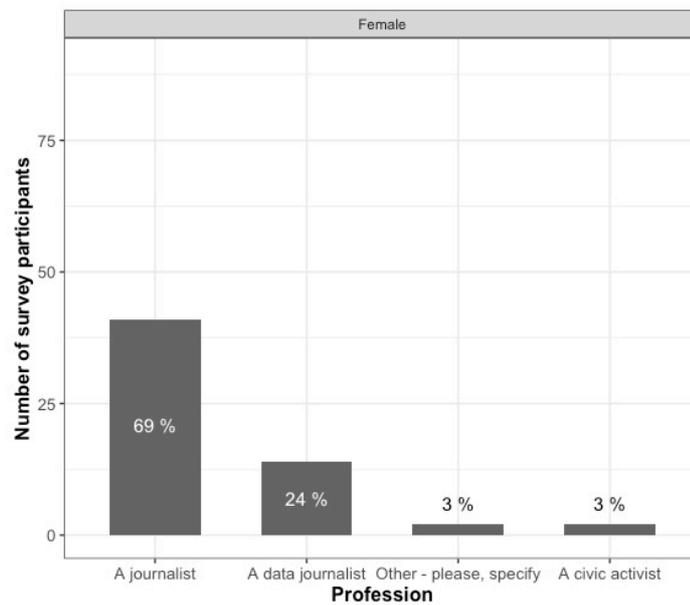
Note: The figures are a combination of survey Q2 "What is your profession?" and Q4 "What is your gender?". The frequencies in the figure might not add up to 100% of survey participants (n=164), as none of the questions was required for submitting the survey, and missing values were excluded.

In terms of the use of particular software, for instance, ten (10%) users of R programming language were men, and eight (14%) were women. Python was used almost equally by seven (7%) men and five (8%) women. Two Python users did not

declare their gender. A slight gap was in the use of Stata, with five (5%) males and only one female (2%) claiming to use the software. Given that women are underrepresented in this survey sample compared to the population of journalists, the level of statistical and computational skills in the female population of journalists might be different than in the survey sample. The absence of a gender gap is an interesting finding in itself worth further investigation.

Borges-Rey (2016) found differences in skills based on the size of the outlets with small regional news companies lacking in-house statistical and computational skills as well as the infrastructure. I was unable to explore if this holds for my survey participants, as no question about the size of respondents' outlets was included in the survey. However, other researchers concluded that the commitment to data journalism has been visible in big national news outlets, in particular, the Guardian, which has been an innovator in the field in the UK (Borges-Rey, 2016; Knight, 2015). Other big national outlets, such as the Financial Times, the Economist, BBC and the Times have also adopted data journalism practices early on. Also, outlets specialised in in-depth data investigations, such as the Bureau for Investigative Journalism, have strong in-house data scientists' capacities.

In terms of sources, the majority (78%) of survey participants stated that they had visited the Office for National Statistics[85] (ONS) website to search for the data for their stories (see Figure 6-16). The ONS is the largest producer of statistical information in the UK and provides insights into different aspects of society. For example, the statistics that ONS produces, include but are not limited to data on business and trade, economy, environment, labour market, demographic data and elections data. The fact that ONS-produced data was the most preferred data source among survey participants confirms the findings of previous research in the field. In particular, the work of Allen & Blinder (2018) who concluded that journalists prefer ONS data because they see it as independent from partisan politics. The ONS is perceived as a public institution independent of the government.

The second most visited website is the Companies House website[86], which provides data and records on companies for investigative stories on corruption and organised crime. Interviewed CSO representatives also mentioned the Companies House website as a major source of information[87]. The central e-government platform Gov.uk came up as the third most used data source. The platform provides information from all central government departments and more than 400 other public

---

[85] https://www.ons.gov.uk

[86] https://www.gov.uk/government/organisations/companies-house

[87] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

authorities. Although Gov.uk serves mostly as **_a life events directory_** for citizens, it also operates a repository of government statistical data in different areas[88].

Figure 6-16: Government websites visited to search for data
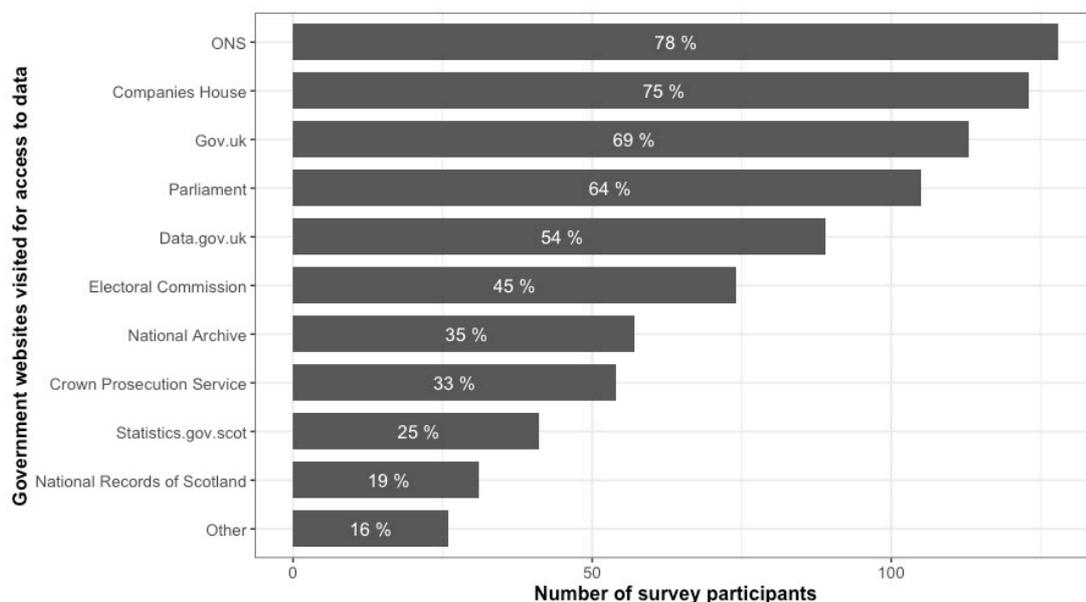
Note: The figure is a combination of survey Q16 "Have you ever visited Data.gov.uk to look for data for your stories?" and Q19 "Which other government websites have you visited to search for the data for your stories? (Please tick all that apply)". The frequencies in the figure might not add up to 100% of survey participants (n=164), as Q19 was a multiple-choice question, and none of the questions was required for submitting the survey.

The fourth most cited source of data was the Parliamentary website[89], which also holds much statistical data as a part of the Parliament's research activities and briefings. The national open data portal Data.gov.uk[90], which provides data in open formats, came up from the answers as the fifth most used source of government data. Eighty-nine (54%) survey respondents claimed to have visited Data.gov.uk to search for data for their stories. As was outlined in the previous chapter, Data.gov.uk was launched in 2010 and followed by massive proactive data publication as a part of the UK's broader government transparency and anti-corruption plans. At that time, high-level politicians suggested that the scale of open government data publication could lead to changes in FOI legislation, as additional requirements for information would become redundant. As was also described in Chapter 5, the rights under FOI legislation eventually did not change, and open government data lost its momentum. From the survey findings, it is difficult to judge if the open government data portal

---

[88] https://www.gov.uk/search/research-and-statistics

[89] https://www.parliament.uk/

[90] https://data.gov.uk/

meets the needs of journalists. However, the survey included one follow-up question, which asked if journalists found data of their interest on the portal and 80% of those who used it agreed that sometimes they find what they need.

Other government websites that survey respondents claimed to use as their data sources were the Electoral Commission[91], the National Archive[92], and the Crown Prosecution Service[93] websites. Journalists based in Scotland also claimed to look for data at Scotland's official statistics website[94] and National Records of Scotland website[95]. Survey participants also mentioned websites of other agencies they visit to obtain data for their investigations, for example, Financial Conduct Authority, Higher Education Statistics Agency, HM Land Registry, UK's ICO, NHS, and the websites of police forces.

Survey respondents and interviewees complained that they sometimes do not find data they are looking for, or its format is non-machine readable, and quality is poor. For instance, a CSO representative said that the data published on the Companies House's website is not validated. Global Witness, an international anti-corruption organisation, uncovered that some listed companies' directors had not been born yet[96]. Similarly, as for the data on the date of birth, there is no validation for the country of residence data, creating ambiguities, as some people state country, others region, which makes data analysis of this category impossible without substantial cleaning of the data[97]. Another CSO representative agreed that a lot of published data had not been validated[98]. She also argued that some data is outdated, and often data on an important topic is not collected at all. However, despite data deficiencies, survey participants and interviewees were generally positive about the idea of the proactive publication of government data.

One journalist argued that mining data is less risky compared to filing an FOI request, as "[data] enables to explore a story idea without having to alert the authorities to what is being looked at" (SP42). However, some government websites also require their users to register to be able to download the data. For instance,

---

[91] https://www.electoralcommission.org.uk/

[92] https://www.nationalarchives.gov.uk/

[93] https://www.cps.gov.uk/

[94] https://statistics.gov.scot/home

[95] https://www.nrscotland.gov.uk/

[96] For more information on the case, see Global Witness' blog at https://www.globalwitness.org/ru/campaigns/corruption-and-money-laundering/anonymous-company-owners/companies-we-keep/?accessible=true

[97] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[98] INT4_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

Companies Register has additional features for registered users. Only registered users can use **The Follow** service to subscribe to the email alerts of company transactions of interest. Downloading government datasets from Statistics.gov.uk and Data.gov.uk is possible for any user without a requirement to register and sign-in. Statistics.gov.uk offers an option to get email alerts when new statistics or a piece of research is published.

One survey question asked if an obligation to register with the government website discourages participants from using this website as a source of data. Only 34 (21%) participants answered yes and perceived such obligation as problematic. The majority did not have any major concerns. Those who did, stated several practical nuisances, such as additional logins and delays in accessing the data. However, the main source of unease and discomfort was related to a loss of anonymity. Several journalists expressed privacy concerns, i.e. that once they register, their activity on governments' platforms is being tracked. The following statement best summarises the experience that popped up in responses repeatedly:

> "Requests are flagged to press offices who then contact you attempting to obtain information and, in some cases, warning off sources" (SP65).

Some participants were convinced that tracking is a common practice, and it changes how government agencies handle data publication plans. A lack of trust in the government was pervasive. A CSO representative also argued that our knowledge of what data on the data engagement the government collects is limited. She speculated that in some political contexts, this might represent a danger for journalists[99]. As for submitting FOI requests, the Scottish government's practice was to flag up sensitive requests to high-level civil servants. Survey participants' views on this practice were strong. One survey participant argued that his "identity should not matter regarding access to information already in the public domain" (SP162). Another one concluded that "government should be transparent without needing to know who is looking at them" (SP18). The Scottish Information Commissioner confirmed this as a general FOI principle in the Scottish Government Intervention[100].

---

[99] INT4_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[100] For more information, see http://www.itspublicknowledge.info/home/AboutSIC/WhatWeDo/Intervention201702016ScottishGovernment.aspx.

6.3.5 FOI vs. open government data: benefits and limitations

Open-ended survey questions and interviews with CSO representatives provided further insights into the topic; in particular, the differences between information obtained through FOI requests and proactively published government data. Statisticians found that as for the topics discovered through open-ended questions, "recruitment of samples larger than 150 participants would not likely expand significantly the number of themes identified" (Tran, Porcher, Falissard, & Ravaud, 2016: 94). Therefore, the sample size of 164 participants of the present survey likely safeguard data saturation and covers the major themes.

One hundred and thirty-three (81%) survey respondents who answered the open-ended question about the benefits of FOIA brought up three main topics, i.e. its importance for democracy, exclusivity and accuracy of obtained information. A unison was on the FOIA's contribution to government transparency and accountability, and democracy more broadly. Over 40 journalists explicitly mentioned that they use FOIA to monitor the government's activities, expose wrongdoings (SP99) or uncover the government's ties with the private sector (SP79). They agreed that public interest stories they bring "would not have been possible to tell" without FOIA (SP39). One survey participant summarised that,

> "... [FOIA] provides an important route to uphold public offices to the values of
> transparency and ensures we can hold them to account for reasons we believe
> are in the public interest. It also gives us a route to access information that
> would not otherwise be available" (SP15).

The exclusivity of FOI-obtained information was often mentioned in the sense of gaining a piece of information that another journalist does not possess, which helps to build front-page stories. However, the exclusivity of FOI-obtained information was also discussed from another perspective, i.e. as the information that is "not routinely public" (SP23), "underreported" (SP34), "off-limits" (SP47) or otherwise "hidden from view" (SP81). Several survey respondents suggested that this information, for example, meeting minutes, could have been in the public domain, but is not on purpose. The public authorities do not publish it unless they have to, and as one survey participant argued, for this reason, FOIA is indispensable because it legally obliges the public authorities to disclose information of public interest. "FOI requests at least cannot be ignored (unlike other press inquiries). Also, sometimes the reasons for which they are denied can be illuminating", another survey participant concluded (SP87).

Several journalists also stated that the accuracy of the information obtained through FOIA is one of its significant benefits compared to other means of getting information. Since public authorities provide this information, they cannot "deny it or spin away later" (SP49). FOIA "generates documents, which can be cited as proof

supporting the article and are therefore more reliable than tip-offs or information provided off-the-record" (SP61). However, they experienced that it is not an easy task to get the information of interest. Some mentioned that if the question is not precisely formulated, the chances are the request will be rejected. Therefore, journalists often need prior knowledge that a public authority holds the particular information to be able to phrase their FOI request precisely, which means FOIA cannot be their only source. They often have to work with whistle-blowers from the authorities (SP99). Another survey participant suggested that sometimes, refusals can also indicate where to look for further information (SP163).

One hundred and twenty-five (76%) survey respondents responded to the open-ended question about the benefits of publicly available government data. The key identified topics in their responses were reliability and authority of the data, its immediate accessibility, and greater possibilities for comparative analysis. In terms of the content, survey respondents perceived data especially useful for tracking money flows between the government and private sector, monitoring government spending and detecting frauds. These topics were raised more often by survey participants who identified themselves as data journalists. Similarly, as with FOI, survey participants argued that the data provided by the government is "accurate and respectable" (SP6) and can be further used "without risk of a defamation suit" (SP92). However, although they perceived the authority of the data, they tended to question it more than the information obtained through FOI requests (SP99, SP160). One survey participant mentioned that "figures are easy for government departments to manipulate" (SP160).

Several survey participants suggested that the main advantage of open government data compared to the information obtained through FOI requests is the immediacy of getting the information (SP3, SP9, SP50, SP84, SP108, SP138, SP166). One survey participant summarised it in the following statement:

> "...[data] saves the 20-day waiting time for FOI. Most of the time, it has been
> analysed and graphed on the website, which makes it easier to dissect or
> pinpoint which information I am looking for" (SP3).

Interviewed CSO representatives also shared this view. One of them claimed always to search publicly available data first and to use FOI requests to gain information only as a last instance[101].

Another advantage of publicly available government data that survey participants often repeated was that the data could inform more in-depth stories and provide comparative analyses (SP49, SP88). Several journalists argued that the data

---

[101] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

helps them to better illustrate changes in different trends (SP6, SP14, SP26, SP51, SP78, SP88, SP149, SP153, SP156, SP161). They also mentioned that data provides a better representation of reality than single cases (SP131). In addition, if journalists reveal their sources or publish the data that informed their stories, they allow others to reproduce the stories and potentially discover new contexts. So, they also contribute to the paradigm shift towards greater transparency.

One survey participant stressed that the availability of government data is as essential for holding the government to account as FOIA:

> "analysing datasets can shed new light on trends and changes in a particular area of government policy and practice – e.g. immigration detention. Such analysis can also track government performance and whether targets are being met" (SP21).

Previous research in the field also suggested that data is useful for storytelling (Borges-Rey, 2016), and it helps to uncover and illustrate problems in society. However, an interviewee argued that it is often through FOI requests how journalists generate data:

> "I have heard from investigative journalists that they do not go to open data portals because the data is open. It is the data obtained through FOI requests that can lead to something they can investigate and draw from. Open data is often not useful for them"[102].

This is in line with the findings of Borges-Rey's research too (2016) and the way how recent big investigative stories were conducted. For example, the project from the Bureau of Investigative Journalism, which discovered the extent of publicly owned property in England sold due to austerity measures, collected the data through FOI requests submitted to more than 300 local councils. In line with my findings, the FOI requests were submitted through a pro version of WDTK[103]. In this case, journalists did not use open government data because it was not readily available, but they had to collect and compile this data. It was FOI legislation, which enabled them to do it.

This example illustrates a point that all interviewees made and emphasised, i.e. that FOI legislation and open government data are complementary. While they differed in the strength of their opinions, they all agreed that FOI and open government data are information provision mechanisms that work well together. Only

---

[102] INT4_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[103] For more details about this story, see the Bureau of Investigative Journalism's original article at https://www.thebureauinvestigates.com/stories/2019-03-04/sold-from-under-you and the background of the story at https://www.mysociety.org/2019/06/05/case-study-sold-from-under-you-by-the-bureau-of-investigative-journalism/ and https://www.thebureauinvestigates.com/blog/2019-03-04/public-spaces-are-being-soldfromunderyou-why-we-are-investigating-the-council-funding-crisis.

one of the interviewed CSO representatives was inclined to consider that if proactive publication of government data were comprehensive enough and easy to navigate, there would not be a need for FOI legislation, as Francis Maude, the Minister in 2010-2015 Conservative and Liberal Democrat coalition government, campaigned[104]. Others refused this idea ultimately as "idiotic...as firstly, open data for release is chosen by the government whereas FOI is chosen by the requester. Secondly, when it comes to open data, there is no enforcement in place. FOI is legally binding"[105].

One interviewee argued that many successes that the government attributes to open data are FOI successes, because a lot of important data that is in the public domain, got there through successful FOI requests in the first place[106]. The Bureau of Investigative Journalism's data on the public property sold by local governments, or individual heart surgery mortality rates across cardiac units in England[107] serve as examples. They also illustrate how FOI and open government data can work together to achieve common goals. The interviewees suggested that FOI could inform open government data publication practices. Public authorities could use FOI requests they receive to identify in what information the public is interested and then publish it proactively on their website so that people do not have to file more individual requests for that information[108]. The take of the Scottish Information Commissioner on the intersection of FOI legislation and proactive publication of open data is also unambiguous:

> "there is a role for proactive publication. But there is always also a role for the ability and the right to request information and to receive it unless an exemption applies. They work in tandem. If you truly want to have a system that allows people to access the information that matters to them, then they have to be able to ask for specific information over and above a hopefully very good proactive publication"[109].

---

[104] INT3_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[105] INT5_CSO. (December 2018). Personal interview. For more information, see Appendix 2.

[106] Ibid.

[107] For more details about this story, see the Guardian's original article at https://www.theguardian.com/global/2005/mar/04/sarahboseley.uk.

[108] INT5_CSO and INT6_CSO. (December 2018). INT7_SIC. (September 2019). Personal interviews. For more information, see Appendix 2.

[109] INT7_SIC. (September 2019). Personal interview. For more information, see Appendix 2.

6.4 Conclusions, discussion and limitations

The results from the survey and interviews generated interesting insights into the use of FOI legislation and proactively published government data by journalists and CSOs. As for the proposed hypotheses, two of them cannot be confirmed. In particular, a hypothesis *h6a* that the use of open government data would be marginal cannot be confirmed based on survey results only. More than half of the survey participants stated that they visited the national open data portal Data.gov.uk at least once to search for data for their stories. In addition, they claimed to use other publicly available government data from diverse government's sources and public authorities, in particular administrative data produced by the ONS.

Also, the survey data did not show any substantial difference in advanced data analysis skills between junior and senior journalists. The median values of the users' age were 32 for R users, 37 for Python users and 29.5 for Stata users, which means that advanced data analysis skills are not exclusive to digital natives only. In other words, journalists in their 20s and 30s preferred R and Stata, and those in their 30s and 40s preferred Python. Although this suggests that the level of data skills was overall somewhat higher in younger than the older generation, there was a case of 52 years old R user and 65 years old Python user. Moreover, the numbers of journalists or CSO representatives who claimed to use any advanced software were also generally low to draw any final conclusions. Therefore, the second hypothesis *h6b* that junior journalists are more likely to possess advanced skills to analyse government data cannot be confirmed based on survey results. However, the results clearly showed that data journalism or data practices in CSOs represent a minority specialism. The number of journalists using any of the aforementioned programming language (R, Python) or software (Stata) was small. The interviews too confirmed that data skills are not a common part of journalistic expertise.

In summary, on the one hand, the findings suggest that journalists use both FOI legislation and government data that is available in the public domain. On the other hand, journalists' self-reported skills do not imply that they use this data to perform data journalism practices. Very low numbers of users of more advanced data analysis software or programming languages might indicate that the majority of survey respondents use government data as stand-alone figures or statistics to better illustrate the point they are making but they do not perform any further operations with the data. A vast majority (79%) of survey participants claimed to use MS Excel for analysing data. While MS Excel performs advanced calculations, it is not suitable for analysing large volume datasets, which often are the core of data-driven investigations, such as Panama or Paradise Papers. MS Excel limits the analysis to +1,000,000 observations, which might seem as perfectly sufficient for most investigations.

However, those working with MS Excel will know that it does not handle well datasets of much smaller scale than its full capacity. Moreover, its possibilities to visualise data are limited. Therefore, it is not a tool fit for data journalism.

The conclusion that data skills are still a minority specialism in the newsrooms is important for various reasons. First, it confirms previous findings in the field (A. Schrock & Shaffer, 2017; Stoneman, 2015; Tabary et al., 2016) and thus, we can be more confident that this conclusion is reliable. Second, although few journalists have data skills, data journalism practices (at least in large national outlets) have been continuously on the rise, which means new actors are entering journalism (Ananny & Crawford, 2015; Baack, 2018b; Cheruiyot, Baack, & Ferrer-Conill, 2019; Porlezza & Splendore, 2019).

International non-governmental organisations, focusing on providing digital tools for facilitating interaction between the public and political representation on a local and national level, play a crucial role. They influence data journalism practices in different ways (Cheruiyot et al., 2019). They help to improve journalists' digital and data skills by providing training in coding, using privacy-enhancing technologies, distributing toolkits and manuals. They also support those who have the skills to collect or analyse the data more efficiently and effectively (Baack, 2018a). WDTK, an online participation platform for submitting FOI requests in the UK is just one example of the latter, as its pro-version assists journalists in sending round-robin requests, keeping track of deadlines, and managing collected information and data. At the same time, these platforms provide free access to large volume of data generated by other users. Scholars argue that civic technology organisations are intertwined with journalism and complement each other's skills (Cheruiyot et al., 2019). They pursue the same goal of following the public interest by keeping the government accountable and encouraging monitorial citizenship (Baack, 2018b, 2018a). To a great extent, data-driven investigative journalism benefits from civic technology organisations and tools they offer.

The last proposed hypothesis *h6c* was consistent with the proposed argument, i.e. journalists indeed prefer private means of communications for submitting their FOI requests over public online participation platforms, such as WDTK despite their many advantages. If they use WDTK, they choose its private pro version. As I learnt from the interviews and open-ended survey questions, there are several justifications for this choice. First and foremost, journalists want their stories to be exclusive and therefore, try to minimise possibilities for others to uncover their story leads. In some sensitive political contexts, journalists might opt for private communication means for safety reasons. Lastly, journalists prefer a private pro-WDTK version over a free

publicly available WDTK version for its extra features and information management tools.

Both survey and interviews were exploratory rather than explanatory. In addition to testing the proposed hypotheses, they generated many interesting findings. Some of them could inform further hypotheses. Others confirmed what was previously known about the topic. For instance, journalists and CSO representatives echoed in their testimonies the previous findings of an increasing rate of refusing FOI requests at the UK central government level (Cheung, 2018). Similarly, the interviews also confirmed that requests coming from journalists and CSOs are sometimes treated in a different way than other requesters, and responses to them require clearance from the high-level management. This has, obviously, negative repercussions on how promptly these requests are dealt with and what their outcome is. Another interesting finding is that although FOI is an invaluable working tool of many journalists, they use it in moderation. Some CSO representatives argued that FOI was their last resort when nothing else worked. Also, journalists' FOI requests were rarely rejected on vexatious grounds.

Survey and interviews provided with several useful insights which could improve policy practice. For instance, the problem of differential treatment can be mitigated by anonymising all requests at the first point when an FOI officer receives them before she redistributes them to other colleagues within the agency. As the Scottish Information Commissioner argued, some Scottish authorities adopted this practice to send across the message that they treat responding to FOI requests as a requester-blind practice. The wider adoption of this practice might also have an effect on the refusal rate.

The survey and interviews also generated some surprising findings, which are worthy of further exploration. For example, contrary to common woes that women are behind in their digital and data skills, data journalism practices, advanced data analysis skills and familiarity with programming languages and statistical software, as self-reported in the survey, were slightly more prevalent in women than men. Exploring whether these results owes to newly opened digital and data journalism degrees, the availability of female-only training or simply to the limits of this survey's representativeness can be an interesting new avenue in researching the changing trends in the journalistic profession. Another path worth investigating is looking at the reasons why journalists use confidential sources, both known and anonymous, minimally.

Last but not least, the majority of survey participants and all interviewees perceived FOI legislation and proactive publication of government data as crucial transparency tools that help to hold the government accountable. At the same time,

they perceived their role of information intermediaries in Benthamian spirit, i.e. that they should use the information obtained through FOI requests or scraped from government websites in the best public interest and continuously control those in power.

All in all, this study provided several interesting findings and in-depth insights into how journalists perceive and use information and data provided by the government, in what information they are interested and what digital and data skills they have. That said, it is important to acknowledge once again that both the survey results and findings from qualitative interviews are not generalisable beyond the studied sample, as non-probability sampling strategies employed for this study do not allow to extend the findings to the population reliably.

7 What the public wants to know and what it gets through FOI and open government data

The findings from the survey and interviews presented in the previous chapter have demonstrated that investigative journalists and CSO representatives use both FOIA and open government data for government oversight. The survey and interviews also provided further insights into how they use them. Compared with the Michener and Worthy's Information-Gathering Matrix (2018), introduced in Chapter 2, a significant part of self-reported journalistic use of FOIA can be identified as political and/or in the public interest. However, the journalistic use of FOIA represents only a fraction of all FOIA uses. The survey findings show that journalists prefer to keep their FOI requests from the public sight and thus, only a few of them use the public version of WDTK platform. WDTK is widely used by the public and therefore, serves as an excellent tool for researching the non-professional public demand for information. In this chapter, using unsupervised learning techniques, I analyse the content of all FOI requests submitted to the UK central government bodies through WDTK since its start in 2008 to the end of 2017, over 37,000 requests in total. First, I formulate several hypotheses. Then I describe the data and methods used for the analysis and categorisation of FOI requests. At last, I present the findings from estimated structural topic models (STMs) and discuss what implications these findings have for access to government information.

7.1 Aims and hypotheses

The main aim of this chapter is to complement the findings from the survey and qualitative interviews. While the previous chapter explored how journalists and CSOs make use of FOIA and government datasets, this chapter aspires to investigate what ordinary people want to know through their FOI requests and if this information can be equally obtained from proactively published open government data. As journalists do not tend to use online participation platforms that are publicly available, choices to research their uses of FOIA and data were limited. I used the survey as a research method and asked them directly about FOIA and open government data, keeping in mind the limitations surveys entail, e.g. social desirability bias. The possibilities for studying non-professional public demand for government information are wider. Online participation platforms, such as WDTK, offer the wealth of unbiased data, as they contain actual unredacted texts of individual FOI requests. I benefited from the

availability of this data and analyse it using quantitative text analysis, in particular, topic modelling to explore the prevalence of topics within FOI requests submitted through WDTK to the central government. I also investigate how these topics are associated with the outcome (whether the request was successful or rejected) and with governments led by different prime ministers. I also compare the topics prevalent in FOI requests with the topics covered by the proactive publication of open government data.

In the presented topic analysis, I follow the Michener's and Worthy's Information-Gathering Matrix (2018), which distinguishes between different uses of FOIA: information sought in public vs. private interest and political vs. non-political information. Legal scholar and former Scottish Information Commissioner Dunion (2011) argued that available research on the uses of FOIA suggests that the population of users is heterogeneous. FOI requesters differ in their skills as well as the interests they follow. However, politicians (Blair, 2010; The Cabinet Office, 2014) often shift the debate about FOIA as a tool for the political fight even though the available research suggests that political use represents, in fact, a minority of requests (Worthy, 2014; Worthy & Hazell, 2017). In light of this evidence, I formulate the hypothesis below:

h7a. People's use of FOIA on WDTK is more likely to be non-political and follow private rather than public interests.

As discussed in detail in chapter 1, FOIA and open government data differ in many ways, and I argue that these differences have implications for the character of the provided information as well as the access to government information more broadly. Most importantly, while FOIA is led by the public demand, in the case of open data, it is the government that ultimately decides what data will be published. As a result, I hypothesise, the information provided through FOIA will cover a greater variety of information than open government data.

h7b. Information provided through FOI requests is more wide-ranging than government data published on the national open data portal Data.gov.uk.

However, as many transparency scholars point out, the government, regardless of the leading party, does not have any interest in disclosing the information that could harm its reputation and popularity (Samaha, 2006). Therefore, I argue, the government will use the exemptions from FOIA to refuse the requests for information, which disclosure could have a detrimental impact on their reputation. I propose the following hypothesis in a similar vein:

h7c. The negative outcome i.e. rejected requests, are more likely to be associated with topics that have a high public interest value and are relevant for the government oversight, e.g. data on procurement and spending.

I also argue that FOIA is topical and reflects the ongoing political events and policy debates. People seek information in response to the need to orientate in political realities that surround them (Arendt & Fawzi, 2019). Therefore, they are more likely to inquire in their FOI requests about the current events and policies rather than the past ones. To a great extent, the events and policies are shaped by governments and prime ministers who lead them. I argue that what they see as a priority will also translate into the content of FOI requests.

In 2011, the House of Commons Political and Constitutional Reform Committee invited several experts to submit evidence on the role and powers of prime ministers. The report concluded the position of prime ministers has strengthened in past two decades as a result of the personalisation of politics, media's strong focus on party leaders, and increased authority of international decision-making bodies (The House of Commons Political and Constitutional Reform Committee, 2011). The trend towards stronger prime ministers started with the premiership of Tony Blair, who had a large majority in the House of Commons (Jones, 2016; Kavanagh, 2007). Blick and Jones (2010) stated that during Blair's premiership the number of the staff working at the Prime Minister's Office based at 10 Downing Street, but also the size of the Cabinet Office, almost tripled compared to the Thatcher and Major's premierships. Also, the Cabinet Office's role shifted from supporting the collective government to providing support to the prime minister primarily. Some scholars, in particular, Foley (2000) argued that all these measures led to an increase of the prime minister's role, which he framed as the advent of presidentialism.

Others critiqued such framing because although the influence of prime ministers in the UK politics has strengthened, they are still constrained by the internal politics of the party they represent (Bevir & Rhodes, 2006) or other externally determining factors, such as protests or pandemics (Weller, 2014). For instance, Heffernan (2013) posited that the way how Gordon Brown ousted Blair from his position is a case in point that comparing British prime ministers to presidents is misplaced. Moreover, at times, prime ministers' power is also diminished by their coalition partners, as was the case of David Cameron leading the 2010-2015 Conservative Liberal-Democrat coalition government.

That said, prime ministers still have strong agenda-setting power and are able to advocate for policies that are priorities for their parties and implement them. The dataset of FOI requests, I used for the analysis, covers three governments (Labour, Conservative-Liberal Democrat coalition and Conservative), which were distinct in

many ways. In particular, the Labour and Conservative governments differ in their ideologies and policy priorities, which, I argue, will create a substantial variation in the public demand for government information between them. One of the topics where the variation by governments and prime ministers would be most markedly visible will be foreign relations and policies. From a historical perspective, the role of prime ministers in setting foreign policies has always been crucial. They have overseen the work of the Foreign Secretary (Carter, 2015) and have also represented the government abroad (Weller, 2014).

Blair's focus on foreign policy and his interventionist approach clashed with public opinion. As Kritzinger (2003) stated, the majority of the British population was convinced about the existence of weapons of mass destruction in Iraq and initially supportive of military intervention. However, the support gradually plummeted after the Hutton Inquiry, a judicial inquiry, which investigated the circumstances of the death of former UN weapons inspector David Kelly and after the Butler Review, which concluded that the intelligence assessments to justify the military intervention were not robust (Dahlgreen, 2015). The UK's engagement in the invasion of Iraq made headlines of major news outlets long after Blair had stepped from this position of the prime minister and affected the successive years of the 2007-2010 Labour government under Brown. As a topic that dominated the media, I expect it to appear in FOI requests during Brown's premiership as well but less so under Cameron's and May's leadership when other topics will take over, in particular, the welfare reform and Brexit-related questions. Therefore, my last hypothesis is as follows:

> **h7d.** The governing party and prime minister affect the prevalence of topics in FOI requests.

7.2 Data and measurement

Large-scale studies exploring the public demand for government information (Bagozzi et al., 2019; Berliner et al., 2018) are scarce not because of a lack of interest in the topic, but data unavailability, which makes investigating information demand particularly challenging. As Michener and Worthy (2018) argued, FOIA is a requester-blind practice. Therefore, governments cannot collect data on requesters other than those necessary for responding to them, i.e. their name and contact details. While it is, undoubtedly, a reasonable measure protecting requesters' privacy and universality of the right of access to government information, at the same time, it means knowledge about requesters, their uses of FOIA and motivations for their requests is minimal.

Also, we know little about what information is of interest to requesters. Public authorities rarely publish the full text of all FOI requests they receive and responses to them. Very few governments, for example, the Mexican government, have a system in place that makes all FOI requests publicly available. In many jurisdictions, publication schemes, which oblige public authorities to publish certain classes of information, exist. Also, several FOI laws expect public authorities to publish performance statistics and disclosure logs, which include a selection of responses to previous FOI requests which asked for information of broader public interest. While these can provide some insights about trends in the FOI requests' growth or decline, rejection rates or topics of interests, they cannot give a full picture, representative of all requests and all uses.

Therefore, it is methodologically challenging to study public demand for government information in a systematic manner. The spread of online participation platforms for submitting FOI requests has offered a partial remedy to this problem, as they usually also serve as a searchable archive of submitted requests.

In this thesis, I benefit from the existence of such a platform in the UK. WDTK[110] was established in 2008 by mySociety, and to this day, it has been the only online participation platform for submitting FOI requests in the UK. With almost 600,000 submitted requests so far, it is a widely popular platform. FOI requests sent to the central government through WDTK represent only a fraction of all requests received by the central government. mySociety conducted research to explore what portion of all FOI requests, the requests sent through WDTK represent (Parsons, 2016, 2019). The study confirmed previous findings in the field, i.e. FOIA in the UK is local (Worthy & Hazell, 2014, 2017). The majority (89.8%) of WDTK requests goes to public authorities other than the central government authorities, in particular to the local government. In addition, the majority of all FOI requests to the central government[111] were sent by other communication means than WDTK. For example, in 2018, only 17.38% of all requests to the central government were sent through WDTK. Nonetheless, there is an increasing trend in the use of WDTK.

In conclusion, although FOI online participation platforms provide new valuable insights into the public demand for government information, these are not representative of all FOI requests. However, since a complete database of all FOI requests to the central government does not exist in the UK, WDTK represents the

---

[110] https://www.whatdotheyknow.com/

[111] The list of the central government bodies, included in the analysis, is available at https://research.mysociety.org/sites/foi/bodies/year/2018/

best and most comprehensive environment for studying public demand for government information that is possible in the current conditions.

In this analysis, I use the WDTK data from its start in 2008 to the end of 2017. On the whole, it is 37 397 FOI requests sent to all central government agencies except for the Home Office. The full list of all included central government agencies is available in Appendix 3. The Home Office was excluded for a variety of reasons. At first sight, it might seem like an unnecessary loss of interesting research data, as the Home Office has steadily been a central government agency with the highest number of FOI requests on WDTK since 2014.

However, as Parsons (2016, 2019) argued in his report, a large portion of these requests are under the Data Protection Act 2018 rather than under FOIA. A random selection of FOI requests on WDTK, which were sent to the Home Office, shows that people indeed use the platform to ask about their immigration status[112], immigration healthcare surcharges[113], or visa issues[114]. Despite mySociety's guidance on what can and cannot be requested from the Home Office under FOIA[115], users continue to file requests that are not covered by FOIA. The Home Office's standard procedure is that its FOI team acknowledges that the request is outside the provisions of FOIA. Then Visas and Immigration team follows it up and explains that they need to obtain further personal data from the requester to respond, and WDTK is a public platform which might expose this data. Therefore, they suggest moving the conversation to a private domain. However, requesters sometimes send their personal data through WDTK despite being advised not to do so. The Home Office data in my sample, thus, could entail several research risks. Firstly, there is a risk of exposing requesters' personal data if it was mistakenly disclosed on WDTK. Secondly, including the Home Office data could introduce biases in topic models, as the topic of immigration is likely to be overrepresented in the sample, which includes requests to the Home Office. Therefore, FOI requests sent to the Home Office were omitted from the analysis.

All in all, the data used in the analysis represents an exhaustive data on FOI requests sent to central government agencies, except for the Home Office, through online participation platform WDTK, from 22 February 2008 to 29 December 2017.

---

[112] For example, see the following request on updates about the notice of marriage at https://www.whatdotheyknow.com/request/havent_received_any_updates_afte#outgoing-962635 or at https://www.whatdotheyknow.com/request/tier_2_general_dependent_visa_2#incoming-1468643.
[113] For example, see the following request on immigration healthcare surcharge refunds at https://www.whatdotheyknow.com/request/ihs_refund_for_overlapped_years#incoming-1465133
[114]For example, see the following request on a student visa at https://www.whatdotheyknow.com/request/curtailment_letter_7#incoming-1460279
[115] https://www.whatdotheyknow.com/body/home_office

It covers FOI requests sent in the span of almost full ten years. mySociety provided the data in the following format: individual FOI requests as individual text files and a spreadsheet including all FOI requests' metadata.

The metadata contains the unique request's ID, name of public authority, unique user's ID, the time when the request was made, the time when it was updated, and the outcome of the request. The texts were paired with metadata using unique requests' ID to create a final dataset. Randomly selected requests from the dataset were checked against the live WDTK to ensure the data is reliable and was matched correctly, as the proper alignment of texts with relevant metadata is crucial (M. E. Roberts, Stewart, & Tingley, 2019). Some additional data cleaning techniques had to be performed. The final main dataset, which I used to explore the prevalence of topics within FOI requests, includes 37 397 FOI requests.

I use topic modelling for estimating the probability of finding specific topics within analysed FOI requests. Topic models are generative statistical models used in modelling large text corpora. I use STM, which is an unsupervised learning technique that produces mixed-membership models. M. E. Roberts et al. (2014: 1066-1067) specify that each document (indexed by d) is generated through the following procedure in mixed-membership models:

> "First, a distribution over topics (ϑ_d) is drawn from a global prior distribution. Then, for each word in the document (indexed by *n*), we draw a topic for that word from a multinomial distribution based on its distribution over topics ($z_{d,n} \sim \text{Mult}(\vartheta_d)$). Conditional on the topic selected, the observed word $w_{d,n}$ is drawn from a distribution over the vocabulary $w_{d,n} \sim \text{Mult}(\beta z_{d,n})$ where $\beta_{k,v}$ is the probability of drawing the *v-th* word in the vocabulary for topic *k*."

While other mixed-membership models, such as Latent Dirichlet Allocation (LDA) could be equally used for estimating the probability of topics within FOI requests, STM offers several advantages, most notably the possibility to use metadata in the estimation of topics. As M. E. Roberts, Stewart, and Tingley point out, all topic models provide a measure of topic prevalence, which "refers to how much of a document is associated with a topic" (2019: 2). However, while LDA assumes an unstructured document collection and disregards any additional metadata generated during the collection, STM takes this metadata into account and allows for hypotheses testing (M. E. Roberts et al., 2019) and investigating how selected treatments (metadata about the documents) change topical prevalence and content (M. E. Roberts et al., 2014). This metadata that "explain topical prevalence are referred to as topical prevalence covariates" (M. E. Roberts et al., 2019: 3). Given my interest in the effect of FOIs outcome and particular governments and prime ministers on topic prevalence, the use of STM is, thus, preferable. Before estimating STM, several data operations, described in M. E. Roberts, Stewart, and Tingley (2019), have to be

performed. These steps, or more precisely, how they were applied to my data (FOI requests sent to the UK central government) are described below in detail. All analyses were conducted in R (v3.5.1).

Firstly, as detailed below and in Appendix 3, I explore topics within the main dataset without using the metadata to avoid the findings to be influenced by my prior assumptions – an approach used by other researchers in the field as well (Berliner et al., 2018). I use the main dataset primarily for testing the first two hypotheses and exploring the public demand for information through FOIA and how it differs from the proactive publication of open government data.

I conducted standard pre-processing steps for STM estimation using *textProcessor* function (M. E. Roberts et al., 2019). These steps were first performed on the main dataset, which consists of all FOI requests and then on the secondary dataset with topical prevalence covariates. Words were converted to lower case and stemmed, and numbers and punctuation were removed. Words shorter than three letters were discarded. In addition, I created a character vector containing words, which I expected will occur in the text frequently, but will not add any informational value. For example, since FOI requests represent an official communication between an FOI officer and a requester, they will likely include greetings and titles (hello, dear sir, dear madam, yours sincerely, etc.) and a full name of government agencies – addressees of the requests. However, these words say nothing in particular about the content and nature of the requests. Thus, they were excluded from the analysis together with "stopwords" (most common words in a language) which are removed by default. After all pre-processing steps, the corpus consisted of 37 380 documents, 21 214 terms and 1 308 730 tokens.

The STM requires a researcher to choose the number of topics k to be estimated. Given that the quantitative text analysis is a relatively new field, there is no conclusive guidance on how to select the right number of topics k. Nonetheless, scholars agree that this choice cannot be made arbitrarily but requires a justification. Choosing an appropriate k affects the quality of topics. As Greene, O'Callaghan, and Cunningham (2014) rightly point out if the number of topics is too low, the topics are likely to be very general, and if it is too high, the topics are likely to overlap. The STM package (Roberts, Stewart, & Tingley, 2019) contains metrics, which assists in the k selection, in particular, a measure of semantic coherence and exclusivity. The topic model is semantically coherent when the topic's most probable words co-occur in the documents. However, as Roberts et al. (2014) found, justifying the k selection only by the measure of semantic coherence is not sufficient, because it can be relatively easily reached by selecting a few topics where most common words in a language are very frequent. Therefore, semantic coherence should be analysed in combination with

exclusivity, which calculates the share of top topic words which are different for a particular topic (ibid).

Following Roberts' et al. (2014) guidance, first, using the main dataset, I analysed both semantic coherence and exclusivity to select the right number of topics k. Using *searchK* function (M. E. Roberts et al., 2019), I computed diagnostic values for models with k values from two to hundred. Based on these values, I considered the k values equal to 21, 26, 33, and 51 topics. The majority of appropriate k topics, thus, lies on the interval from 10 to 40 as displayed in Figure 7-1. I estimated the model for each of these k numbers of topics and plotted semantic coherence and exclusivity for each of these models to establish the final k. As can be seen from Figure 7-2, the topic model with 33 topics has both high semantic coherence and exclusivity for the majority of topics. Exceptions are topics 15 and 25, which have a low semantic coherence and topics 2 and 11 with low exclusivity.

Figure 7-1: Diagnostic values by the number of topics (main dataset)
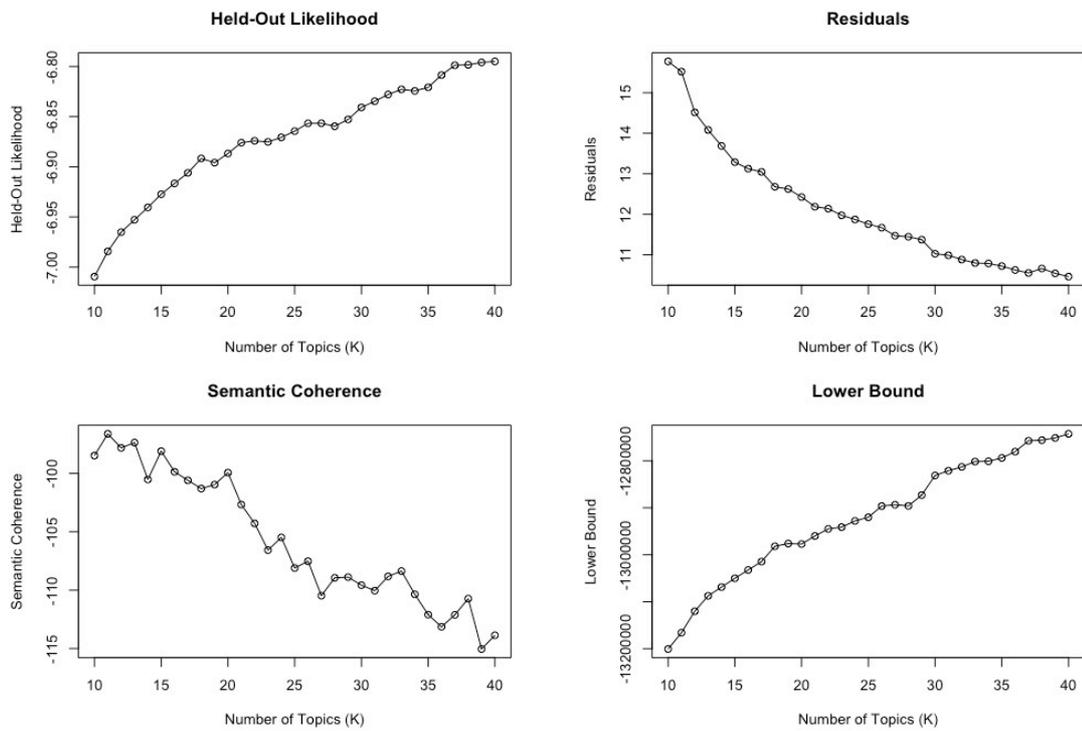
Figure 7-2: Quality of each topic in a model with k equal to 33 topics (main dataset)
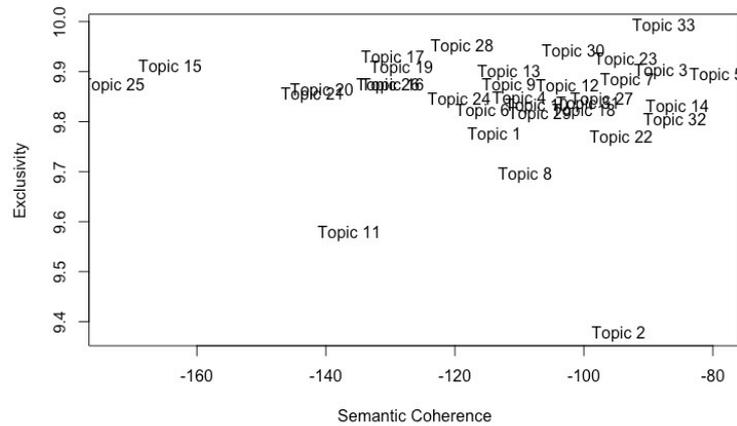
Note: This figure plots semantic coherence and exclusivity in 33-topic model.

In addition to the main dataset, I also created a derivate dataset to run STM with using some metadata as topical prevalence covariates to test the remaining hypotheses, i.e. if the negative outcome (refused requests) and the governing party and prime ministers are more associated with some topics than others.

A request can lead to different outcomes (grounded in the law), most frequently, it is successful, partially successful and rejected. These categories also represented the most frequent categories in the main WDTK dataset, which included 11 676 successful, 6 621 partially successful and 4 677 refused requests. It is important to mention that the requests' outcomes on WDTK are outcomes self-reported by requesters. They can, thus, be slightly different from how public authority would label them. However, it is highly likely that requesters and public authorities reach agreement on clear-cut outcomes, such as successful vs. refused.

There are, of course, other outcomes possible. For instance, public authorities might not hold the requested information. A requester might decide to withdraw her request for all kinds of reasons. Alternatively, she or the public authority might become unresponsive. In the first case, the outcome of the request on WDTK will be updated as "awaiting clarification", and in the second case as "awaiting response". A requester might also find the way how her request was handled as unsatisfactory and request an internal review. Most of these latter outcomes are not final, and thus not very useful for exploring their links to particular topics. However, successful requests where access to information was granted in full and refused requests, which are on the opposite end of the scale, should allow for analysing the effect of the outcomes on the expected topic prevalence.

Therefore, I filtered the dataset based on FOI requests' outcome, and only included requests that were either successful (information was fully disclosed) or

refused (information withheld from the requester). The derivate dataset includes 16 353 requests (documents) in total. I paired each document (full text of an FOI request) with the metadata provided by mySociety. These were the date when the request was sent, the date when the request was updated, the public authority (addressee) and the outcome of the request. The outcome is a binary variable in this dataset. It takes the value of 1 if the request was successful and information was provided and the value of 0 if the request was rejected.

Based on the provided "date" variable, I created new variables "govBrown", "govCameron" and "govMay" to establish under what government the request was handled (the 2007-2010 Labour under PM Gordon Brown, the 2010-2015 Conservative and Liberal Democrat Coalition, and the 2015-2016 Conservative government under PM David Cameron or the 2016-2019 Conservative government under Theresa May). This variable is again binary, and it takes the value of 1 if the particular prime minister was leading the government and 0 if he or she was not. These variables are crucial for testing the last hypothesis – the impact of the governing party and prime minister on the expected topic prevalence. Subsequently, I repeated the same standard pre-processing steps for STM estimation as I did in case of the main dataset. After all pre-processing steps, the corpus consisted of 16 349 documents, 12 089 terms and 536 965 tokens.

Using **searchK** function, I computed diagnostic values with k values from two to hundred for the following model 1:

Topic prevalence = α + βdisplay_status + βgovBrown + βgovCameron[116]+ ε

For model 1, based on these values, I considered the k values equal to 13, 18, 22, and 28 topics. The majority of appropriate k topics, thus, lies on the interval from 10 to 40 as displayed in Figure 7-3. I estimated the model for each of these numbers of topics k and plotted semantic coherence and exclusivity for each of these models to establish the final k. As can be seen from Figure 7-4, the topic model with 28 topics has both high semantic coherence and exclusivity for the majority of topics.

---

[116] display_status is a binary variable that identifies the request's outcome (1 – successful and 0 - refused). govBrown, govCameron and govMay are also binary variables and determine the government.

Figure 7-3: Diagnostic values by the number of topics (model 1)
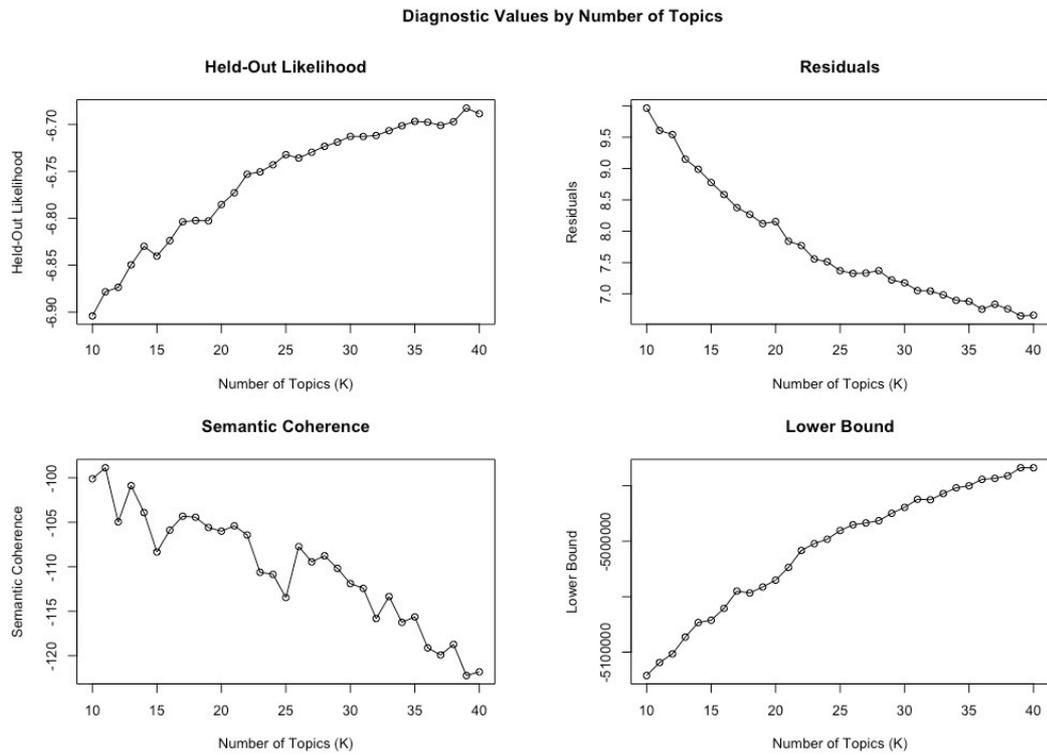
Note: Model included topical prevalence covariates: display_status (outcome), govBrown and govCameron.

Figure 7-4: Quality of each topic in a model with k equal to 28 topics
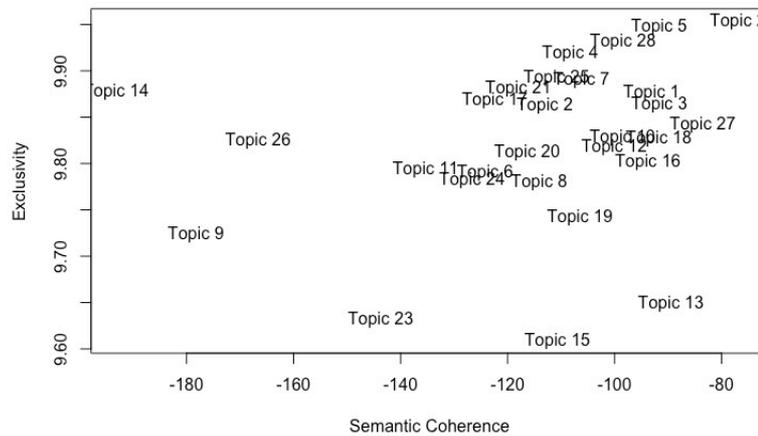
Note: This figure plots semantic coherence and exclusivity in 28-topic model with topical prevalence covariates: display_status (outcome), govBrown and govCameron.

Using **searchK** function, I then computed diagnostic values with k values from two to hundred for the following model 2:

Topic prevalence = α + βdisplay_status + βgovBrown + βgovMay + ε

187

For model 2, based on these values, I considered the k values equal to 15, 20, 22, 28, and 30 topics. The majority of appropriate k topics, thus, lies on the interval from 10 to 40 as displayed in Figure 7-5. I estimated the model for each of these numbers of topics k and plotted semantic coherence and exclusivity for each of these models to establish the final k. As can be seen from Figure 7-6, the topic model with 28 topics has both high semantic coherence and exclusivity for the majority of topics again.

Figure 7-5: Diagnostic values by the number of topics (model 2)
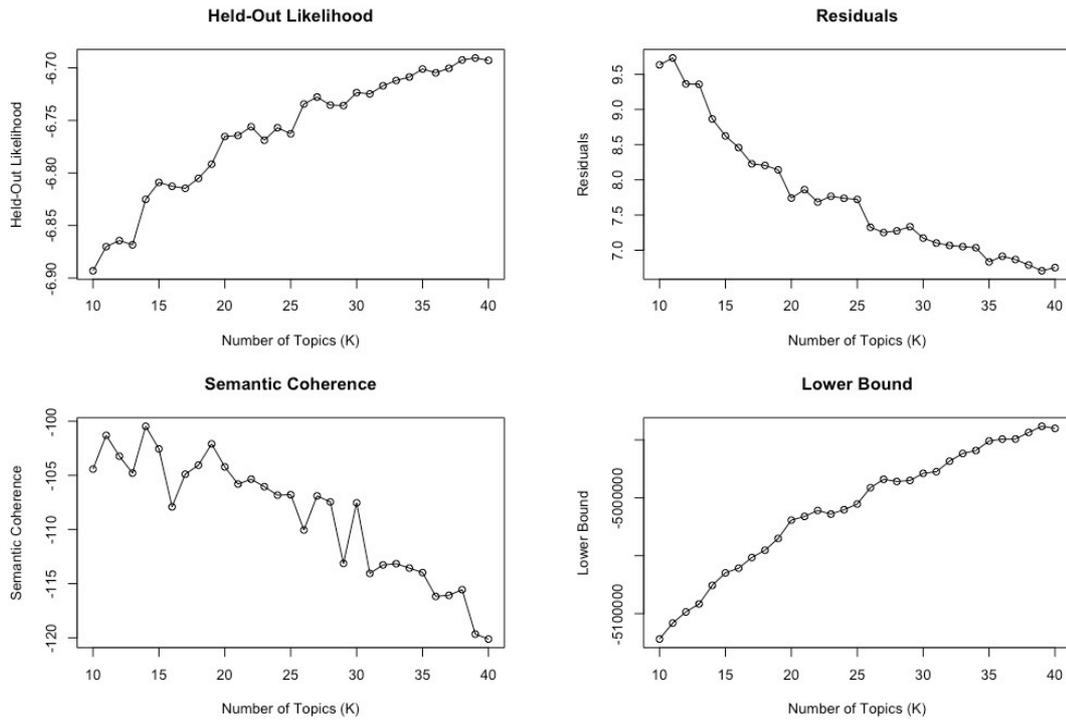
Note: This figure plots semantic coherence and exclusivity in 28-topic model with topical prevalence covariates: display_status (outcome), govBrown and govMay.

Figure 7-6: Quality of each topic in a model with k equal to 28 topics.
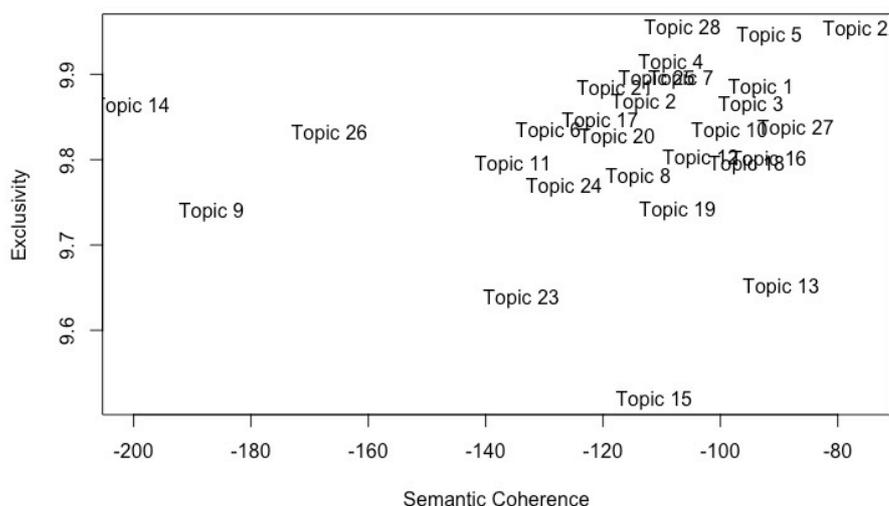
Note: This figure plots semantic coherence and exclusivity in 28-topic model with topical prevalence covariates: display_status (outcome), govBrown and govMay.

7.3 Findings I: Majority of the requested information in the public interest, unlikely to be supplied proactively as data

The usefulness of topic models in studying the public demand for information is conditional on its effectiveness to estimate meaningful topics. The model presented below identifies 33 topics that illustrate the body of all FOI requests submitted to the central government through WDTK from 2008 to 2017 most appropriately. For the proportion of these topics in the corpora of FOI requests, see Figure 7-7. I present these 33 topics of interest and how they can be classified between two dichotomies in the Information-Gathering Matrix as proposed by Michener and Worthy (2018): political vs. non-political information and information in public vs. private interest.

7.3.1 <u>Public vs. private interest and political vs. non-political topics</u>

Figure 7-7 displays the top topics from a 33-topic model and the frequency of these topics within analysed FOI requests. Before the individual topics are discussed below in great detail below, I would like to point out that first five top topics with the highest expected proportions largely represents the issues of private interest, such as individual job seekers' or benefit claimants' enquiries or Merlin reports documenting the history of ex-British Army, Navy and Royal Air Force vehicles. Only two out of

189

five topic – topic 13 and 31 are somewhat related to the public interest. We can see that high public interest topics, relevant to government accountability, such as budget spending (topic 16), land and property ownership (topic 26) represent actually a relatively minor proportion of the discourse. Procurement of government services (topic 12) and staffing policies in the government (topic 10) are somewhat more represented topics. Only spending (topic 28) and meetings minutes and official correspondence (topic 18) are topics directly relevant to government accountability, feature among more common topics.

Most of the topics out of the total of 33 topics in the STM were straightforward to interpret and label within the public vs. private interest framework. For instance, to mention a few, topics 1, 4, 5, 10, 12, 13, 18 and 30 are of clear public interest. Some of these public interest topics are represented by FOI requests which enquired about different public services, such as transport, health or education. For example, topic 1 represents FOI requests related to transport, as both highest probability words, such as "transport", "road", "rail", "route", "car, "passenger" and FREX (frequent and exclusive) words, such as "airport", "traffic", "scanner", "fare", or "bus" suggest. The majority of the requests within this topic asked for information about transport rules and regulations or very practical matters, such as the change of routes or fares. A few could also be linked to the concept of government accountability more narrowly. One example of thereof is a request stated in Appendix 3, which required statistics of pedestrians who were killed or injured on malfunctioned puffin pedestrian crossing, i.e. a clear public interest issue.

Topic 4 was also related to public service delivery, in particular in the area of health. Some of the requests within this topic asked for the Health and Safety Executive Investigation reports. Others enquired about specific medications and the availability of scientific research that investigated their impact and adverse effects. Similarly, topic 5 was related to health. FOI requests within this topic demanded documents related to the management of NHS and its performance indicators.

Figure 7-7: Expected topic proportions for the main model (33 topics)

Top Topics

Topic 14: work, pension, claimant
Topic 8: defenc, ministri, servic
Topic 13: copi, provid, report
Topic 31: school, educ, name
Topic 33: inform, follow, freedom
Topic 5: request, respons, receiv
Topic 3: person, law, legal
Topic 30: mani, time, month
Topic 23: one, take, need
Topic 28: year, total, staff
Topic 18: meet, correspond, letter
Topic 7: tax, custom, payment
Topic 22: court, justic, ministri
Topic 32: assess, work, claimant
Topic 12: contract, servic, use
Topic 10: offic, public, cabinet
Topic 27: prosecut, children, polic
Topic 24: fund, money, treasuri
Topic 4: health, nhs, trust
Topic 29: british, foreign, nation
Topic 1: transport, road, rail
Topic 19: review, level, perform
Topic 9: local, council, author
Topic 26: compani, regist, hous
Topic 16: cost, project, estim
Topic 21: scheme, manag, programm
Topic 17: form, area, site
Topic 2: famili, caus, problem
Topic 20: food, environ, affair
Topic 11: research, use, vaccin
Topic 15: train, cours, disabl
Topic 25: power, search, polic
Topic 6: autism, librari, servic

Expected Topic Proportions

Note: Words used to label the topics are the words with the highest probability to occur within that topic. The words listed in the figure are the top three words associated with the topic. For more detailed information about the STM interpretation, see also M. E. Roberts et al. (2014).

Figure 7-8: Highest word probabilities for each topic in the model with 33 topics

Other topics related to public service delivery were topics 6, 15 and 32. They all had in common a relevance to public service delivery to disadvantaged groups. As can be observed from Figure 7-8, highest probability words, such as "autism", "service", "support" suggest that FOI requests that represent topic 6 enquired about the support that the government provides to people diagnosed with Asperger's syndrome or other forms of autism. From a broader perspective, it unambiguously is a public interest topic, as the requests demand information that is in the interest of a group of citizens rather than an individual. However, it is unclear if these requests were made in a private or public capacity. Parents with children diagnosed with Asperger's syndrome or other forms of autism who were trying to find out what services they can use could have filed the requests or CSOs advocating for the rights of autistic people.

Similar to topic 6, topic 15 also entailed disability-related enquiries primarily. However, as can be seen in Figure 7-2, topic 15 has a low semantic coherence, which can also be observed from the highest probability words. The words "train", "course", "disable", "learn", "introduction", "guidance", "med" suggest that topic includes more than one main semantic meaning. A closer reading of the requests' sample reveals that another part of requests discussed the training of public servants in different areas. Topic 32 focused on work capability assessment; in particular, FOI requests enquired about the assessors' qualifications.

Requests within topic 11 focused on vaccinations. People enquired about the manufacturers of the vaccines, vaccines' effectiveness and the recommended vaccination schedule for children, for example. Several requests were specifically requesting information about the HPV vaccination. There is no doubt that vaccinations are a public interest matter, as poorly vaccinated populations represent a serious global public health threat (Alvira, 2019).

That said, it cannot be easily established from the requests whether these were made in a public or private capacity as people who filed an FOI request about vaccinations might have been very different. They might have been concerned parents looking for information to make the best-informed decision for their child or lobbyists, both in favour as well as opposed to vaccination.

Another prevalent topic related to public service delivery was devoted to education (topic 31) as the highest probability words, such as "school", "education", "name", "provide", "England", "contact", "primary" and FREX words, such "secondary", "headteacher", "nursery", "teacher" indicate. The majority of the requests within this topic were almost identical and required a list of schools in different geographical areas in the UK. This topic also demonstrates that in some cases, as in this one, proactive publication of government information might be useful, but does not guarantee that the information will reach everyone. Despite the lists of some schools being publicly

available on the Department for Education's (DfE) website[117], hundreds of requesters demanded them through FOI, which suggests they were unable to locate them through internet searches on the DfE website.

Other public interest topics were related to the environment, in particular, topic 17 and topic 20. Topic 17 had a low semantic coherence. The majority of requests discussed badger culling, e.g. the number of vaccinated badgers in different parts of England, the numbers of badgers tested for tuberculosis, cost associated with badger culling; as the highest probability words "area" and "test" and FREX words "cull", and "badger" indicate. However, a share of requests discussed health and safety inspections of different areas, e.g. drilling sites. Topic 20 with the highest probability words such as "food", "environment", "affair", "rural", "agency", "water", "waste" and FREX words such as "dog", "meat", "slaughter", "countryside", "fiduciary", "halal", "BTB" was also broadly related to the environment, in particular to animal welfare and food standards.

Topics 27 demanding statistics of convictions, mostly related to sexual offences, domestic violence or child abuse and topic 22 seeking legal advice or interpretation of legislation were also public interest topics. Topic 25 consisted of FOI requests asking for crime statistics broadly, but also seeking legal advice.

While all of these topics can be classified as topics of public interest, applying the political vs. non-political dichotomy to these topics is complicated. These topics are not political per se but can potentially be used in a political fight.

Some public interest topics also had a strong relevance for government accountability, enquiring about the government's management of resources, procurement, spending, filling government positions, and correspondence between government officials and lobbyists. For instance, requests within topic 10 enquired about staffing policies and human resources management, selection and disciplinary procedures, the expertise required for the positions filled vs. skills of selected candidates, pay scales and other. The highest probability words of Topic 21, such as "scheme", "manage", "programme", "provide", "work", "west", "placement" were similar in content, and required information about pay scales for different positions within the government departments and information about government's subcontractors.

Several topics (7, 12, 16, 24, 26, 28) were related to the government's financial accountability. Requests within topic 7 demanded information about companies on the aggressive tax evaders lists, amount of different tax credits that have been overpaid by HM Revenue and Customs (HMRC). Topic 24 with the highest

---

[117] https://www.gov.uk/government/publications/list-of-schools-by-ebacc-language-entry.

probability words such as "fund", "money", "treasury" or "bank" and FREX words, such as "borrow", "repay", "money" or "loan" enquired about the size of national debt or money paid to banks. However, as can be observed in Figure 7-2, this topic has a low semantic coherence. Apart from the requests on the state of the economy, requesters asked about departments' participation in trade unions. Topic 16 with the highest probability words, such as "cost", "project", "estimate", "include", "expense", "develop", "budget" and FREX words "cost", "incur", "expense", "estimate", "calculation", "project" and "tariff" represent requests which asked for departmental internal budgets and spending on different campaigns and projects in various fields.

Topic 26 also focused on the use of resources, but in the real estate field. The requesters asked about the cost of properties that government departments own or rent in the UK and abroad. A handful of requests within this topic was filed to follow private ends, enquiring about private sector real estate transactions.

Similarly, topic 12 and topic 28 are strongly related to the government's financial accountability. The highest probability words within topic 12 "contract", "service", "use", "system", "current", "support", "procure" and FREX words "oracle", "procure", "contract", "hardware", "sap", "payroll", "reprography" indicate that the requesters were interested in the procurement process and costs of specific products and services, in particular, in the area of information technologies. At first sight, this topic if of clear public interest, as it investigates if public resources are being spent reasonably delivering to the public the best possible value for money. However, as was already emphasised, the interpretation of the findings from unsupervised learning techniques requires human judgement. Reading a sample of requests within this topic reveals that many of them were round-robin FOI requests, i.e. identical requests sent by a requester to several government departments. That said, the requester still might be a journalist or an engaged citizen who is interested in the comparison of contracting practices between departments. Nonetheless, the level of detail in the products' and services' specification in the FOI requests rather suggests these requests were made by private companies to follow their ends. In line with Kwoka's findings (2016, 2018b), the text of these requests indicates that FOI legislation is also used by private companies to explore their business opportunities and learn about their competitors. Topic 28, which enquired about the use of public resources, for example, ministers' travel expenses, also included round-robin requests. It is, again, unclear whether they were made by journalists or private companies exploring their competition. But all in all, the highest probability words: "year", "total", "staff", "provide", "figure", "statistics" and similar FREX words "figure", "total", "statistics", "spent", "breakdown" are related to the accountability.

Other topics that can be linked to government accountability are topics 9, 13, 18, and 30. Primarily, the requests within these topics demanded government documents on different topics, for example, local government (topic 9), charities (topic 13), official correspondence, meeting minutes, government communication with representatives of private and civil society sector, and foreign governments across departments (topic 18) and statistics (topic 30), in particular, related to civil servants employed by the departments enquiring about their use of the whistle-blowers hotline or civil service hotline, and the health impacts of their work.

While the majority of topics broadly falls within the public interest, only some of them, for instance, topics related to government budgeting, spending, staffing policies and their practical implementation, can be directly linked to government accountability. These topics can also be labelled as political, as they distinctly follow the purpose to monitor the government's management of public resources.

Some topics are on the opposite end of the Information-Gathering Matrix. Requests within these topics asked for information of a non-political nature and were filed in private interest. They reveal that citizens also use FOIA for different purposes than those initially envisaged when FOIA was adopted. Often, they do not meet the legal definition of the request under FOIA. Under certain conditions, public authorities can either not respond to them at all (refuse them as vexatious) or deal with them outside of the FOIA as general enquiries or complaints. Purposely, I reduced the volume of requests of this kind by not including the requests that were sent to the Home Office. Although these are not represented in the topic model, a closer reading of their sample on WDTK shows that requesters mostly enquire about their visa and immigration status. Some of them were "the Windrush generation", i.e. Commonwealth nationals who settled in the UK after the World War II in response to labour shortages and hoped to get their immigration status confirmed through FOI requests after previous attempts through other means had failed. Given that the Home Office receives the highest number of FOI requests, and a significant part of them are this kind of requests, filed in private interest, it can be concluded that FOIA fulfils its initially intended role as a tool of democratic government oversight only partially. That said, people should have a right to obtain the government's information about themselves timely. Kwoka (2018a) argued that FOI is not fit for purpose, mostly due to its statutory time limits, in particular, in appeals against visa, asylum and immigration decisions where time is a decisive factor.

Within the estimated topic model, many topics were non-political per se, and FOI requests were filed to follow private ends. However, in a bigger picture, they raise important political issues. For instance, topic 29 included some requests, similar to those usually addressed to the Home Office, inquiring about individuals' immigration

status. An example of a request in Appendix 3 demonstrates that, in addition to the Home Office, some requesters addressed these questions to the Foreign and Commonwealth Office (FCO). The particular sought-for information by a member of the Windrush generation might be in her private interest solely, but the high number of similar requests exposes systemic problems in the government's immigration policies. These requests raise discussion on public interest issue as the immigration policies adversely affect large numbers of people. In addition, in this topic, due to its low semantic coherence, a part of requests also focused on the correspondence between the UK government and foreign governments, in particular, Nigerian and Bahraini. Requests represented in topic 3 asked for the interpretation of specific legal definitions, procedures, and internal documents of public bodies.

Topic 2 also consisted of requests made in private interest entirely. Many of them did not meet the legal definition of an FOI request, as the requesters did not ask for an actual document or information in the government's disposal. In several cases, citizens used FOI legislation to voice their complaints or concerns or to claim their other rights when all other communication platforms failed. For instance, requesters asked for contact details of public bodies they could not reach using other means of communication, complained about the quality of public services and the adverse impact of austerity measures. While this again suggests that people use FOI legislation for initially unintended ends, it also indicates that people do not have satisfactory platforms where they could articulate their issues and concerns. Similarly, as in topic 2, topic 23 also represented individual enquiries on various topics, many of which did not meet the legal definition of FOI requests. In these requests, people used FOIA to raise criticism of government policies, repeatedly of the welfare reform. For instance, in the example requests, a requester complained about how her pension age was calculated, or about nonsensical questions that are being asked during work capability assessments.

Requests within topics 14 and 19 were also made to meet private aims. Those represented in topic 14 could be summarised as jobseekers' enquiries. They enquired about specific conditions of different DWP's programmes for job seekers and procedures for interviews. Requests within topic 19 focused on education. The requesters demanded information about qualifications and examinations, as well as required performance and outcome statistics in different areas of the UK, for different genders, ethnic groups or devolved nations.

Topics 8 and 33 were perhaps the best example of requests filed in private interest. In the UK, it is possible to request the so-called "Merlin report", which provides the information to ex-British Army, Navy and Royal Air Force vehicle owners about the history of their vehicles. The majority of requests within topics 8 and 33

was related to this, i.e. requesters required Merlin reports from the Ministry of Defence (MoD) for vehicles they purchased.

On the whole, the first hypothesis (7a) that people's use of FOIA on WDTK is more likely to be non-political and follow private rather than public interest is consistent with the proposed theory only partially. While FOI requests within presented topics required the information that is not political per se, e.g. information related to a variety of public services, from transport to education; this information was primarily in the public interest. Nonetheless, topics directly linked to government oversight and accountability were only a few. Moreover, the topics with the highest expected proportions covered mostly the issues of private interest.

7.3.2 Information provided through FOI vs. open government data

In 2014, Francis Maude, the Minister for the Cabinet Office and Paymaster General in the 2010-2015 Conservative and Liberal-Democrat coalition government proposed that the proactive publication of government data could compensate for the need to request the information through FOIA (The Cabinet Office, 2014). Below, I explore some of the datasets that were released by central government departments on the UK national open data portal Data.gov.uk from 2010 to 2019. I compare the datasets with the information provided through FOI requests using some of the 33 topics from the STM above as examples to make a case in point that information covered by FOI requests is too wide-ranging to be represented by datasets.

For instance, topics 4 and 5 focused on public health and NHS management broadly. If we look at the Data.gov.uk, the first problem comes with establishing where to look for the data. Using FOIA, if public authorities do not hold the information but are aware of other bodies who might have it, the requests are likely to get redirected. While search engines on the open data portal might be helpful, a requester has to independently assess what authority might hold the data of her interest, which requires some level of knowledge of the government. For instance, the portal offers two health departments - the Department of Health and Social Care, which is the central government department and Department of Health, Social Services and Public Safety, which is a body of Northern Ireland Executive. The visitor of the open data portal must be able to recognise this and many other nuances. In contrast, FOIA removes this obligation from the requester.

The DHSC publishes data that undoubtedly contributes to greater government accountability – for example, a list of all financial transactions spending over £25,000 as well as transactions that have a single transaction value of £500 or above. However, this data is not updated regularly, although the metadata says it should be updated

every month[118]. Also, other data, important for government accountability, such as permanent secretaries and special advisers' meetings and received gifts, is outdated. The latest dataset of this nature comes from 2015, which supports arguments for the FOIA indispensableness, as FOIA guarantees access to newer data even if it is not routinely proactively published. Communication researchers argued that the outdatedness of open data is also a reason why journalists might still prefer to obtain data for their news stories via FOIA (Stoneman, 2015) or other means, such as web scraping or confidential sources (Borges-Rey, 2016). For instance, data on the prevalence of common mental health problems[119] or obesity in children[120] has the potential to generate news stories. Still, if the latest version of this data on the portal is ten years old, then its journalistic value rapidly drops to zero unless compared with up-to-date data.

Not only the already published data is not regularly updated, but the government also publishes very few new datasets. As can be observed from Figure 7-9 and Figure 7-10, the publication of government data on open data portal Data.gov.uk has been on the decline since its peak in 2013. Except for the Department for Environment, Food and Rural Affairs (Defra) and HMRC, central government departments have significantly reduced its open data publishing efforts.

Another significant difference between FOIA and open government data lies in the content of the information or data. FOIA enables citizens to ask anything of their interest, from statistics on wide-raging issues to information that is of relevance to themselves only. Data published on Data.gov.uk might provide interesting insights into different phenomena but is unlikely to answer questions pertaining to individuals. For instance, in the early years of the UK open data programme, the DHSC published many interesting datasets, e.g. on the number of available hospital beds, hospital waiting times, cancelled operations, ambulance data including numbers of urgent journeys, excess winter deaths, and prevalence of various diseases in the population.

---

[118] See https://data.gov.uk/dataset/0351cb08-2fcc-43df-914a-c1822d82b766/spend-over-25-000-in-the-department-of-health.

[119] See https://data.gov.uk/dataset/8a671aef-e9e3-4e99-a9a4-7af6c03d6190/mental-health-prevalence-of-common-mental-health-problems.

[120] See https://data.gov.uk/dataset/f05e0aa3-ac82-45e4-8f5b-55a9bcc61f63/obese-children.

Figure 7-9: The total number of central government's datasets on Data.gov.uk (2010-2019)
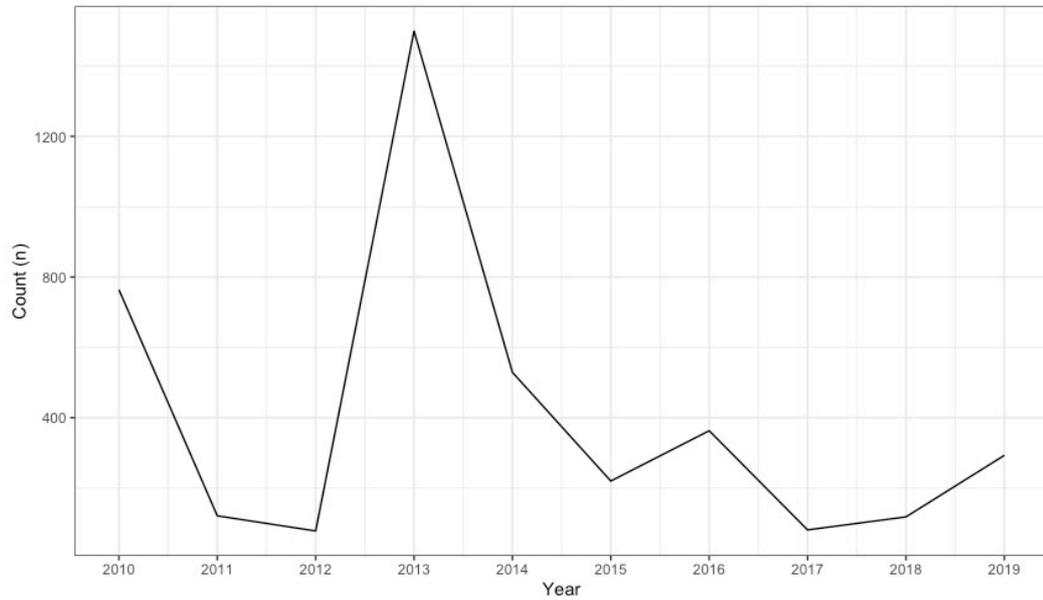
Figure 7-10: Datasets on Data.gov.uk by individual departments (2010 – 2019)
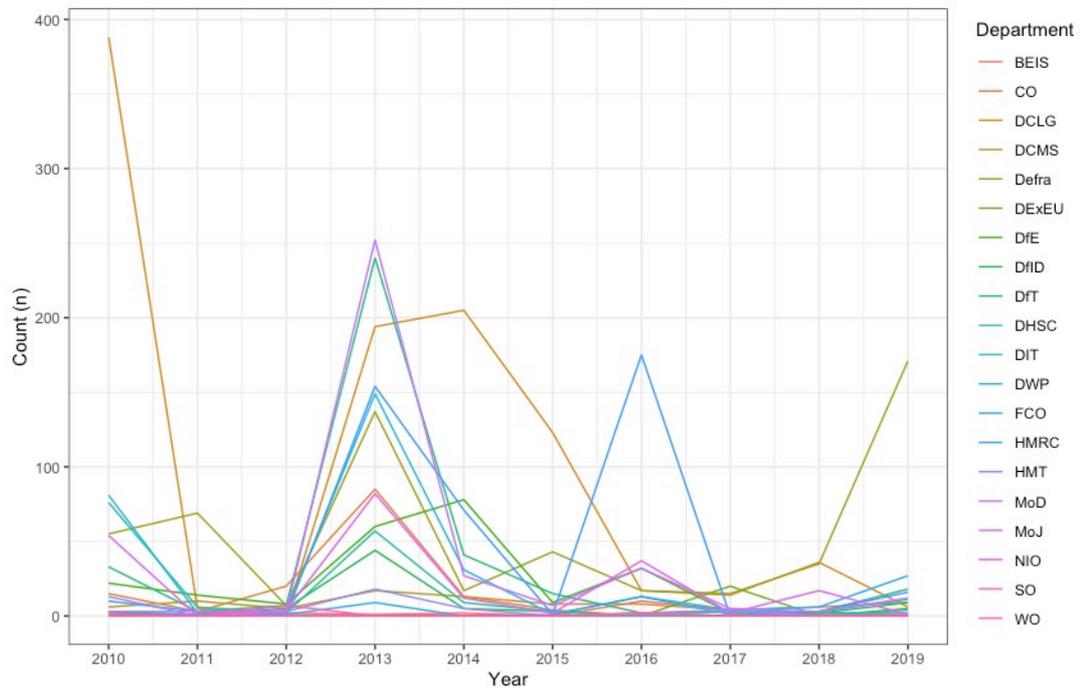
However, examining FOI requests within topic 4 and 5 reveals that the proactive publication on Data.gov.uk does not cover the majority of the questions posed through FOIA. While some of the data requested via FOIA can be released as

open data easily (e.g. death rates from various illnesses), others, such as enquiries related to the current scientific knowledge about adverse effects of specific drugs are unlikely to be answered by a dataset, as they are too specific. This difference in the content of FOI requests and published datasets is even more visible in other topics of STM presented above.

For instance, FOI requests within topic 29 demanded documents related to the UK's foreign relations, e.g. correspondence between UK ministers and foreign governments' ministers, some of them related to the UK's involvement in the invasion of Iraq in 2003. Looking at Data.gov.uk, since its launch in 2010, FCO has published only 65 datasets in total. Again, some of this data has clear relevance for government accountability, such as departmental spending over £25,000[121], staff remuneration[122], staff diversity and equality statistics[123] or development assistance allocation. However, none of the data included commonly requested information, such as official correspondence between the UK and foreign governments' representatives despite it also can be linked to government accountability. Nonetheless, it is important to note that both FOIA and FOISA include a provision (27 and 32 respectively), which exempts the information discussing international relations from disclosure if it,

> "would, or would be likely to, prejudice (a) relations between the UK and any other State, (b) relations between the UK and any international organisation or international court, (c) the interests of the UK abroad, or (d) the promotion or protection by the UK of its interests abroad... [or] if it is confidential information obtained from a State other than the UK or an international organisation or international court" (Freedom of Information (Scotland) Act, 2002; Freedom of Information Act, 2000).

Requests within topic 2, 14, 23 and topic 29 represent another example where the difference in the type of information requested through FOIA and provided as open data is stark. In requests within topic 29, requesters asked for the information on their own visa, asylum or immigration status. Since this information entails personal data, neither such requests can be responded in the public domain, nor such data can be published on the open data portal. The Home Office regularly publishes aggregated immigration statistics[124] as open data, which might be of interest to journalists and CSOs active in the area. However, this data is of no help to requesters

---

[121] See https://data.gov.uk/dataset/918e739d-5d8e-409b-b73f-bfac287dfdbf/spending-over-25-000-in-the-foreign-and-commonwealth-office.

[122] See https://data.gov.uk/dataset/78a88473-68ed-46ca-8208-26fb62567ce0/organogram-of-staff-roles-salaries.

[123] See https://data.gov.uk/dataset/79f658a1-35eb-4bfa-ae2c-72a5d7dfa5f1/fco-diversity-and-equality-report-and-gender-pay-gap.

[124] For instance, see Immigration Statistics on removals and voluntary departures, Immigration Statistics on entry clearance visas granted outside the UK or on European Economic Area (EEA).

who need to get access to their own data. Although FOIA was neither designed for this use nor governs access to one's own data, it at least enables the requester to establish contact with the authority in charge. As described above, many requests within topic 2 are not FOI requests in the strict legal sense, as instead of asking for documents, they raise various concerns or demand data that the government holds about requesters themselves. In particular, many requesters used FOIA to voice their discontent with government policies, which was also largely present in requests within topics 14 (benefit claimants' enquiries) and 23 (information of various interest). Many of these requests were addressed to the Department of Work and Pension (DWP) and complained about state pension age calculations, the severity of work capability assessments and a lack of support for people with disabilities and other disadvantaged groups.

These requests clearly demonstrate that FOIA is also used for purposes other than the control of the government in a narrow sense. This finding is in line with evidence provided in Kwoka's research (2016, 2018b, 2018a). She (2016) asks, "if FOIA is not primarily serving the newsgathering, democracy-enhancing function of informing the public debate on matters of governmental policy, what interests, then, does it primarily serve"? Other researchers also question if FOIA resources are well-spent on this kind of individual enquiries (Pozen, 2017; Pozen & Schudson, 2018). Kwoka (2018a) also argues that FOIA is not fit for the purpose of these requests and is unlikely to meet the expectations and needs of requesters – a point she eloquently illustrated on enquiries about visa, asylum and immigration statuses where a single day can make a difference in deporting the requester.

There might not be agreement about whether requests filed to follow private interests represent or not an appropriate use of FOIA. However, this use of FOIA reveals the malfunctioning of other government communication channels and the everyday impact of its policies on people. Some of these requests suggest that people used FOIA as a last resort when different ways of contacting civil servants failed. FOIA has become a platform for voicing their concerns because of the lack of alternative platforms. This is an important finding, which also confirms that open data cannot substitute FOIA, because it cannot address these requests. As such, this finding is also in support of the second proposed hypothesis (7b).

7.4 Findings II: Request outcome and governing party affect the expected topic prevalence

As outlined in the discussion about data and methods, I created a derivate dataset from the main dataset, which includes only successful and refused requests to make use of the metadata and explore the effect of the request outcome and governments and prime ministers on the expected topic prevalence. I estimated two different models with the following topical prevalence covariates: request outcome (display_status), the 2007-2010 Labour government under PM Gordon Brown (govBrown) and the 2010-2015 Conservative-Liberal Democrat coalition government and the 2015-2016 Conservative government under PM David Cameron (govCameron) and the 2016-2019 Conservative government under PM Theresa May (govMay). As I argued in the introduction of this chapter, I expect the topic prevalence will be substantially different between Labour and Conservative governments and different prime ministers:

Model 1: Topic prevalence = α + βdisplay_status + βgovBrown + βgovCameron + ε

  Model 2: Topic prevalence = α + βdisplay_status + βgovBrown + βgovMay + ε

Since the data on successful and refused requests sent to the central government is available from 22 February 2008 to 20 December 2017, it covers almost the full term of Brown's government (27 June 2007 – 11 May 2010), the entire term of the Conservative-Liberal Democrat Coalition government and a year of the Conservative government under PM David Cameron (11 May 2010 – 13 July 2016) and a half of the term of the Conservative government led by PM Theresa May (13 July 2016 – 24 July 2019).

7.4.1 The effect of the outcome on the expected topic prevalence

The first chapter, where the concepts of FOIA and open government data were introduced, also discussed that **de jure** and **de facto** FOIA could be markedly different (Berliner, 2012). I provided several examples of countries with FOI laws that have a rigorous design, but their implementation and enforcement are weak. Journalists and civic activists also raised in the survey and interviews that even democratic governments as is the UK government, try to curb the right to access government information in multiple ways. They mentioned extensive use of exemptions from FOIA as a common strategy, in particular, the exemption on cost grounds, which is difficult to appeal. The Scottish Information Commissioner also confirmed in an interview that the exemption on cost grounds "is an easier way for

organisations to manage the impact of wide-ranging requests then necessarily going down the vexatious route because [organisations] can evidence that as a firm cut off point... in the decision"[125]. I argue that governments will use legal ways to circumvent the disclosure of any information that could have damaging consequences for them. Thus, I hypothesise (7c) that the request outcome will affect the expected prevalence of topics. In particular, positive outcome (successful requests) will be associated with the prevalence of topics that are irrelevant for the control of the government and are less controversial politically. In contrast, negative outcome (refused requests) will be associated with topics that can be linked with accountability, e.g. data on procurement and spending or the minutes of meetings with lobbyists.

As the covariate of interest is binary, I used the method = "difference" option to plot the change in topic proportion shifting from one specific value (refused requests) to another (successful requests) (M. E. Roberts et al., 2019). As can be observed from Figure 7-11, I can fail to reject the hypothesis only partially. First of all, a positive outcome is associated with fewer topics than a negative outcome. In both models, the positive outcome was strongly associated with some topics that were irrelevant for the public interest. For illustration, the positive outcome is significantly associated with topics 5 and 13, which consisted mostly of requests to MoD for the history of purchased ex-military vehicles. The requests for the history of ex-military vehicles do not carry any potentially controversial meaning, and thus are easy to respond for MoD and disclose the requested information.

The positive outcome is also significantly associated with some other topics that can be assessed as politically non-controversial, e.g. topic 1 (requests for the lists of schools), topic 6 (transport-related requests) or topic 26 (requests for the bylaws of various charities). It is also significantly associated with topics 27 and 16, which consisted mostly of individual enquiries from benefit claimants and pensioners. Although many of these requests were not FOI requests in the strict legal sense, they were still responded to and requested information was provided. This is important, as first and foremost, these requests suggested that people lack platforms to raise their concerns or issues about the welfare reform.

---

[125] INT7_SIC. (September 2019). Personal interview. For more information, see Appendix 2.

Figure 7-11: Estimated effect of the request outcome on the expected topic prevalence
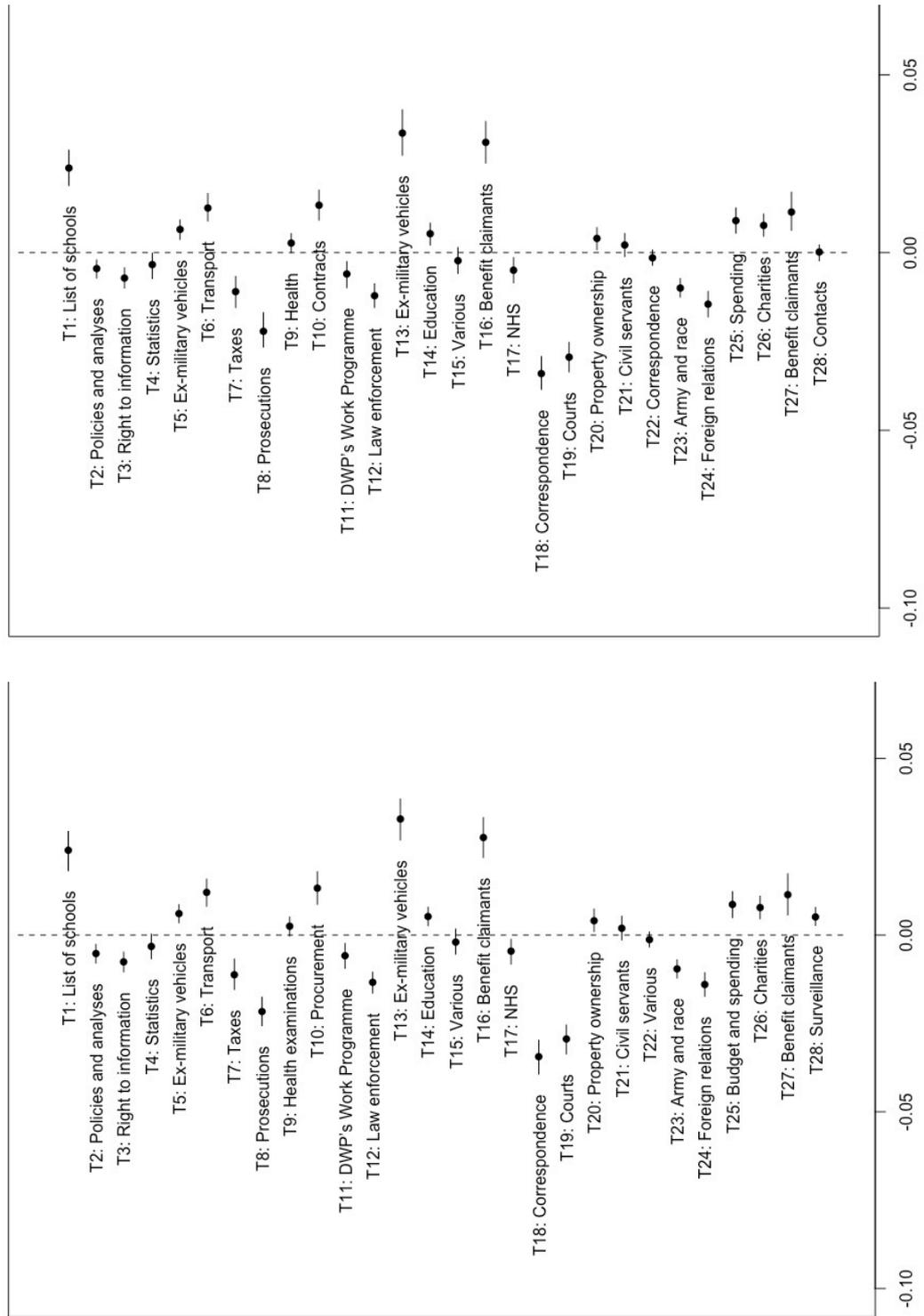
Note: The figure illustrates the statistical influence of the request outcome (display_status) on the topic prevalence. The figure on the right illustrates the effect of the outcome (display_status) on topic prevalence in the model with covariates: display_status, govBrown, govCameron. The figure on the left shows the effect of the outcome (display_status) in the model with covariates: display_status, govBrown, govMay. The estimation is based on a STM specified with 28 topics. Each dot corresponds to the estimated coefficient. The horizontal line corresponds to the 95% confidence interval.

Some of the topics significantly associated with positive outcomes were also of clear relevance to the government oversight, for example, topic 10 (requests for government contracts with private companies and procurement details), topic 20 (requests about land and property in the government's ownership or hire), and topic 25 (requests for budget and spending information). However, the effect was small, and in the case of topic 20, the confidence interval is slightly less reliable.

Moreover, negative outcome (refused requests) were also significantly associated with some government accountability topics, such as topic 18, which consisted of requests which demanded official correspondence and minutes meetings or topic 24, which also required the latter concerning the UK's foreign relations.

These results are valid for both models, i.e. models where different premierships were added as topical prevalence covariates, which suggest that governments are less likely to respond some requests than others regardless of what party is in the government and what prime minister leads it. It is also important to mention that since every FOI legislation contains exemptions from disclosure, some of the topics associated with a negative outcome (refused requests) might fall under this category. Thus, in many cases, the negative outcome does not necessarily mean that the government is arbitrarily withholding the information from the public but rather that it follows the provisions set in the legislation. From the estimated STMs, it is difficult to infer where the refusal has been legitimate and where not. However, the exemptions in FOIA 2000 and FOISA 2002 can serve as a useful orientation point. For instance, topic 24 on foreign relations might well be associated with the negative outcome for precisely this reason, i.e. matters that could harm the UK position abroad are exempt from FOI legislation.

7.4.2 The effect of the governing party on the expected topic prevalence

Several examples from Chapter 1 and 2 demonstrated that FOI requests reflect the political events. The FOIA contemporariness can also be easily spotted on the WDTK platform. In the introduction of this chapter, I discuss in great detail that prime ministers in the UK have strong agenda-setting power and how important policy discussions or changes would be reflected in the public demand for government information.

For illustration, from the more recent events, it is noteworthy to mention the public interest in the Intelligence and Security Committee's report into Russian interference in the UK concerning the EU referendum, which PM's Office had since 17 October 2019 and refused to release it before Parliament was dissolved on 6

November 2019[126]. In the following days, over 20 requests on WDTK demanded the Cabinet Office (CO) and the PM's Office to disclose the report[127]. Both the CO and the PM's Office refused the requests on the same grounds, referring to the exemption from FOIA under section 22(1), which protects information intended for future publication. They argued that as the report discusses national security issues, the process of the release has to be well planned. This decision to withhold the report is worrying as the subject matter it discussed is vital for government oversight. However, as for FOIA, the public demand for the "Russia report" is a case in point (although anecdotal only) that FOIA echoes the major political affairs. FOIA serves as an essential information source not only to journalists and civic activists, as the findings from the survey and interviews revealed in the previous chapter, but also to ordinary people. From the significant political events of an older date, MPs' expenses, the legality of the UK's invasion to Iraq, the investigation into the death of Dr David Kelly and the subsequent Hutton inquiry received attention with hundreds of requests being submitted to demand information related to these events.

I argue that FOIA because it reflects current events, can help to provide insights into people's policy preferences and the impact of policies on them during different governments. Since political parties on the opposite political spectrum put emphasis on different policies and shape political events, I argue that the governing party and prime ministers will affect the prevalence of topics within FOI requests, as people will be interested in current policies and political events (hypothesis 7d).

The topical prevalence covariates of interest are three and following: Labour government under Gordon Brown (govBrown), Conservative-Liberal Democrat coalition and Conservative government under David Cameron (govCameron) and Conservative government under Theresa May (govMay). All covariates of interest are binary, so the method = "difference" option is used to plot the change in topic proportion shifting from one premiership in contrasts with others (M. E. Roberts et al., 2019). As can be observed from Figure 7-12 and Figure 7-13, some topics are shared across parties and were prevalent in FOI requests regardless of the governing party and persona of the prime minister.

However, there were also some differences, and some topics are significantly associated with a particular party in the government and the prime minister only. For

---

[126] For the context, see https://blogs.lse.ac.uk/politicsandpolicy/intelligence-and-security-committees-report-into-russian-activities/.

[127] See the examples of requests demanding the ISC's report at https://www.whatdotheyknow.com/search/Russian%20interference/all?page=1&query=Russian+interference

instance, the effect of the Labour government under PM Brown is positive and statistically significant on the prevalence of topics, such as budget and spending (topic 25), foreign relations (topic 24), remuneration of senior civil servants (topic 21), official correspondence and minutes meetings (topic 18), contracts (topic 10). While some of the topic labels might speak for themselves to someone interested in the British politics (e.g. many requests within the topic on spending referred to the MPs' expenses scandal), other topics required further analysis of FOI requests to establish to what political events they were related. For example, reading a sample of requests sent to FCO during the 2007-2010 Labour government revealed that the public was interested in the UK's ministerial visits to Colombia and the details of the military aid that the UK provided to Colombia under the Labour government[128]. These requests occurred after the UN High Commissioner for Human Rights pointed to the widespread abuses by the Colombian armed forces. The requests for official correspondence and meeting minutes (topic 18) were heterogeneous. Requesters were interested in the correspondence related to the visit of the pope's visit to the UK in 2010 and its costs to the taxpayer, or correspondence with the devolved Scottish administration.

The Labour government under PM Brown was negatively associated with the prevalence of a few topics only – topics 5 a 13 with the requests for the history of ex-military vehicles or topic 1, which was represented mostly by the requests for lists of schools and their headteachers.

While these topics are hard to connect with a particular policy, topics 11, 16 and 27, which were negatively associated with the 2007-2010 Labour government but positively associated with Cameron's Conservative-Liberal Democrat Coalition and Conservative governments, are. Requests within these topics represent enquiries by benefits claimants and pensioners who demanded information about DWP's Work Programme, its pricing model, sanctions or raised concerns about the stricter conditionality of the eligibility for the receipt of the benefits. As can be observed from both Figure 7-12 and Figure 7-13, these requests were not prevalent during the 2007-2010 Labour government, but their proportion significantly augmented during Cameron's the 2010-2015 Conservative and Liberal Democrat Coalition and the 2015-2016 Conservative government. This yet again suggests that FOIA is sensitive to the changes in the government and the introduction of new policies. FOI requests can provide and provide valuable evidence of the government's policies and their impact on people's lives.

---

[128] For further information, see https://www.theguardian.com/world/2009/apr/29/colombia-uk-military-aid.

Figure 7-12: Estimated effect of the Labour government under PM Gordon Brown on the expected topic prevalence
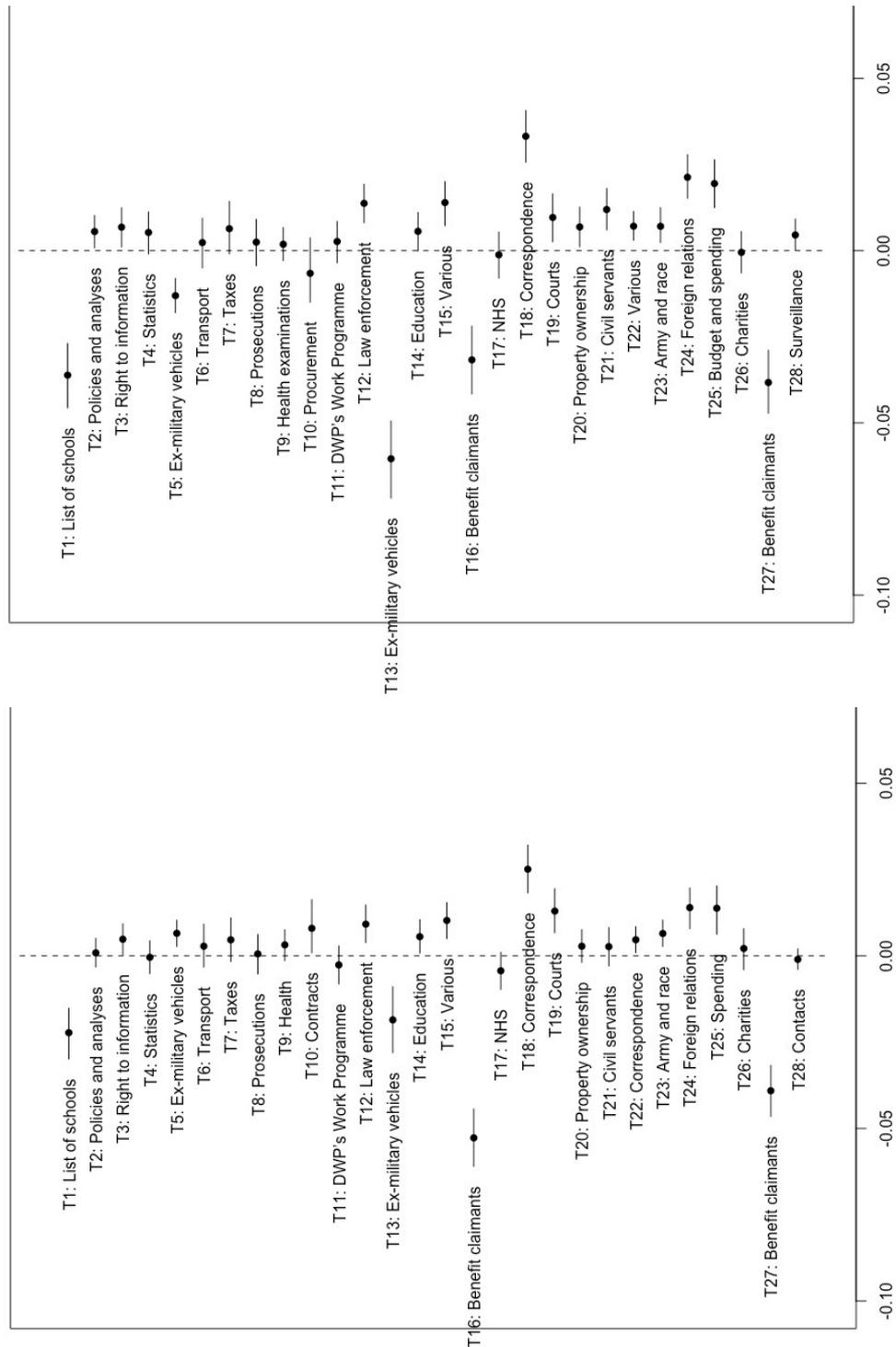
Note: The figure on the right illustrates the effect of govBrown in the model with topical prevalence covariates: display_status (outcome), govBrown, govCameron. The figure on the left shows the effect of govBrown in the model with topical prevalence covariates: display_status (outcome), govBrown, govMay. The estimation is based on a STM specified with 28 topics. Each dot corresponds to the estimated coefficient. The horizontal line corresponds to the 95% confidence interval.

Figure 7-13: Estimated effect of Cameron's governments and May's government on the expected topic prevalence
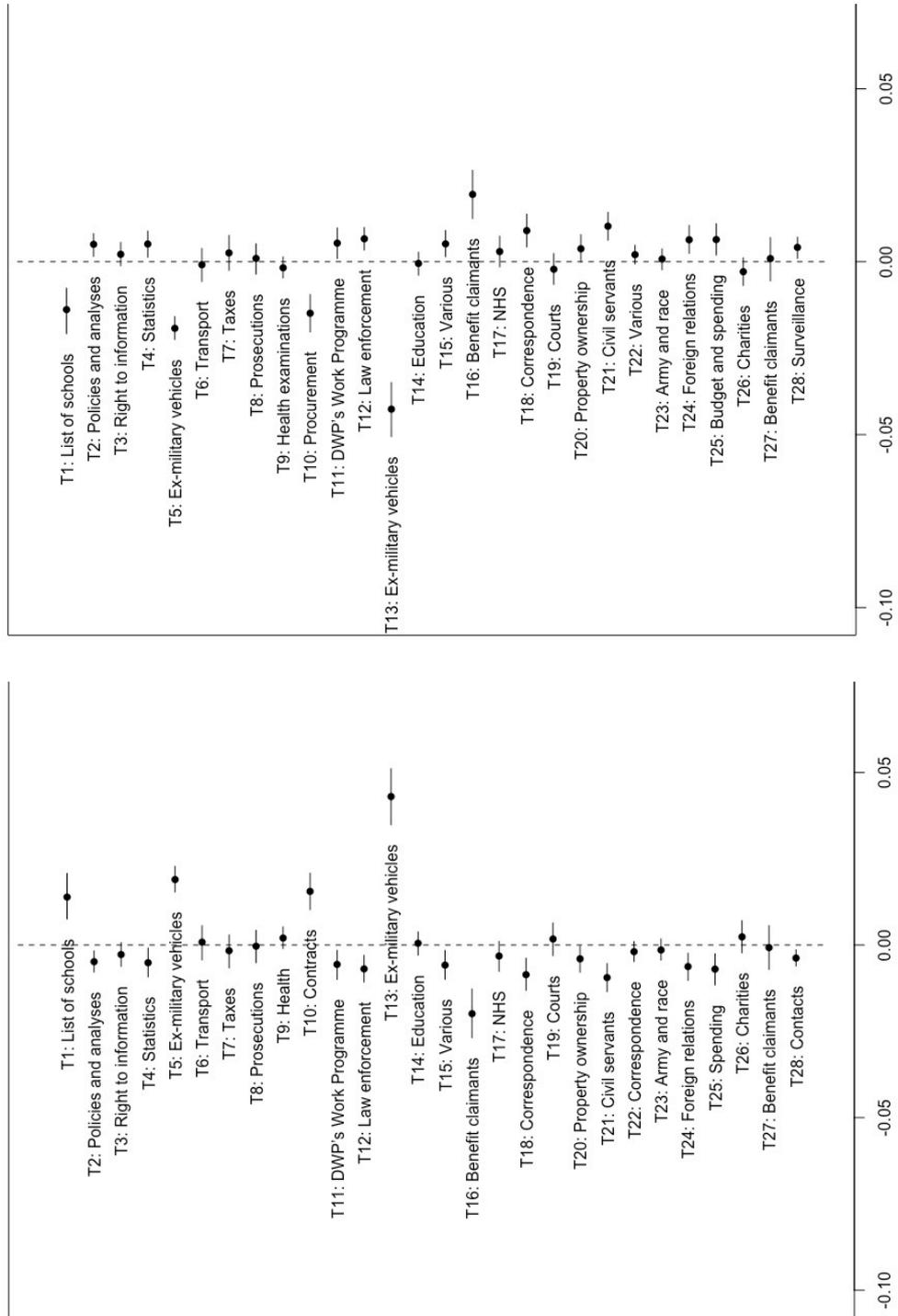
Note: The figure on the right illustrates the effect of govCameron in the model with covariates: display_status (outcome), govBrown, govCameron. The figure on the left shows the effect of govMay in the model with covariates: display_status (outcome), govBrown, govMay. The estimation is based on a STM specified with 28 topics. Each dot corresponds to the estimated coefficient. The horizontal line corresponds to the 95% confidence interval.

A closer look at the requests within the topics 11, 16 and 27, submitted under Cameron's the 2010-2015 Conservative-Liberal Democrat Coalition and the 2015-2016 Conservative government, suggests that the public reacted very sensitively to the welfare reform. The requests for DWP's guidelines and policies and clarification requests indicate that the newly introduced conditions for the receipt of the benefits were not fully understood and well received. Requests within these topics enquired about the Employment and Support Allowance (ESA), Jobseeker's Allowance (JSA), Personal Independence Payment (PIP), Work Capability Assessment (WCA) and pensions. Citizens argued that the new system, in particular, WCA is defunct, leaving vulnerable and disadvantaged groups behind, stigmatising them and deepening their marginalisation. They shared stories of being assessed as fit for work by the staff of Atos Healthcare - the company outsourced to assess people's fitness to work (now replaced by Maximus) despite having severe and lifelong conditions. FOI requests within these topics echo the findings of the research into the impacts of the welfare system reform in the UK (Fletcher, 2011; Garthwaite, 2011, 2014; Hancock & Mooney, 2013; Roulstone, 2015). Therefore, FOIA can serve as a valuable resource of people's experience of different government policies.

Cameron's governments were also positively associated with the topics that discussed budget and spending (25), foreign relations (24), civil servants, and policies that set conditions of their employment, their qualifications and remuneration (21) and official correspondence of the government department with lobbyists from private and civil society sector (18).

As for the first year and half of the Conservative government led by PM Theresa May, it was positively associated with the prevalence of three topics only: ex-military vehicles (13 and 5), contracts (10), list of schools (1).

7.5 Conclusions and discussion

One of the main aims of this chapter was to complement the previous one and provide insights into how ordinary citizens use FOIA. More precisely, I set to explore what is it that they want to know from the government and to what extent this information is readily available in the public domain in the format of open government data. The results from the main STM show that the uses of FOIA are manifold and public demand for government information is heterogeneous, as Dunion (2011) argued. However, despite this heterogeneity of the information in demand, the first hypothesis that the uses of FOIA on the online participation platform WDTK will more likely be non-political and pursue private ends cannot be confirmed.

While many of the estimated topics can be characterised as issues of public interest, they represent a relatively minor proportion of the discourse. Moreover, while the Michener and Worthy's (2018) Information-Gathering Matrix proves helpful in distinguishing between FOI requests filed in public and private interest, assessing whether the character of the requested information is political or not is challenging and reveals the limitations of empirical research of FOIA. Some FOI requests are ostensibly political. For instance, information about MPs' expenses is political, per se, as it concerns the political representation. However, information about the available support services for people with Asperger's syndrome or other forms of autism might seem as non-political, but it is political and can be used politically. In their FOI requests, requesters are rightly not obliged to reveal their motivations to demand a particular piece of information, and from the question itself, it is hard to establish the purpose. The quantitative text analysis toolkit cannot shed much light on how the information is further used once obtained. Nonetheless, it represents an excellent method for exploring and analysing the content of FOI requests systematically.

The results from the main STM also showed that some FOI requests are targeted at government oversight while others pursue individual goals. Scholars rightly argue that FOIA is used for other than initially intended purposes (Kwoka, 2016, 2018b, 2018a; Pozen & Schudson, 2018). There are situations in which FOIA cannot react to people's demands with required flexibility and promptness, and other means to address these demands would be more appropriate. However, I argue that purposes of FOIA are not set in stone and should reflect the times we live in and their challenges. During the analysis, I came across several FOI requests, which asked for information about phenomena that lacked any credible scientific evidence, such as chemtrails[129] for example. The question, if resources intended for the government oversight primarily are well spent on dealing with the requests of this kind, is legitimate. However, when misinformation is on the rise and effective ways to confront and stop it are yet to be identified (Weeks & Gil de Zúñiga, 2019), the governments could approach FOIA as another opportunity to establish contact with their constituents. FOIA represents another communication channel through which civil servants could explain the facts and refute misinformation.

The results from the main STM compared against data published on the national open data portal Data.gov.uk also demonstrated that open government data is unlikely to substitute FOIA. The datasets are unable to include and address the variety of questions people pose in their FOI requests. Moreover, the publication of

---

[129] See the request at https://www.whatdotheyknow.com/request/chemtrails_2#incoming-93300.

open government data by the central government has been on the decline in the UK (see Figure 7-9). At the same time, the numbers of FOI requests are gradually growing every year (see Figure 6-5 in the previous chapter).

However, as theorised in previous chapters, governments use the potential of FOIA sparingly. In chapter 1, I provided several examples of governments that shirk their legal obligation to disclose the information. The results of STM are partially consistent with my third hypothesis that the negative outcome of FOI requests would be associated with the prevalence of topics that have a high public interest value and are essential for control of the government. Some topics linked to accountability were indeed associated with a negative outcome, such as topic 18, which consisted of requests which demanded official correspondence and minutes meetings or topic 24, which also required the latter concerning the UK's foreign relations. At the same time, other topics that potentially advance government's oversight, as topic 10, which consisted of requests for government contracts with private companies and procurement details or topic 25 which included requests for budget and spending information were associated with the positive outcome (successful requests where the information was fully disclosed). However, all in all, positive outcome was more associated with the prevalence of politically uncontroversial topics, such as Merlin reports requested by owners of ex-military vehicles or schools' contact details.

Lastly, the results of STM are consistent with my last hypothesis that the governing party and the persona of prime minister affect the prevalence of topics. Indeed, FOI requests provide an accurate image of the government policies and their impact on people, provided in real-time. For example, under Brown's 2007-2010 Labour government, people were more interested in foreign affairs than under Cameron's governments and even more compared to the 2016- 2019 Conservative government under PM Theresa May. On the contrary, the prevalence of topic that included benefit claimants' enquiries was positively associated with Cameron's governments which introduced the welfare reform but negatively associated with the 2007-2010 Labour government. In their requests, people raised major concerns and questions about newly introduced welfare reform and repercussions it had on their lives. This is an important finding, which shows that using quantitative text analysis on the texts of FOI requests can be a new fruitful approach to explore the impact of government policies and people's attitudes toward them.

8 Conclusions and discussion

In 2013, the ECtHR published an important judgement in the case of the Youth Initiative for Human Rights v. Serbia (ECtHR, 2013). In 2005, the Youth Initiative for Human Rights, a Belgrade-based NGO requested information concerning electronic surveillance from the Serbian intelligence agency. The agency refused to disclose the information. The NGO complained to the Information Commissioner who decided in its favour and ordered the agency to provide the sought information. The agency unsuccessfully appealed to the Supreme Court and vindicated non-disclosure with the statement that "applicant does not need the information sought". Despite the dismissal of its appeal, the agency nevertheless refused to disclose the information. Eight years later, the NGO won the case at the ECtHR. If this story is unique in something, it is the persistence of the NGO. A few discontented requesters go to the Information Commissioner. Even fewer go to the courts. This story repeats itself. However, had the decision to disclose the information been up to the agency only, the NGO would have never received this information. That is just one illustration of how FOIA is different from any pro-active information provision, open data included.

Governments pledge for greater openness, they pass FOI laws and other transparency policies as cure-alls, but how do they implement them? What is the actual impact of these laws and other transparency policies? Do they strengthen corruption control as political leaders pledged? If requesters have to go through hurdles as above, do they continue to use FOI laws? Or, they turn to readily available government information instead? How they use FOI and what information they seek? I tried to provide answers to some of these questions in the preceding empirical chapters and contribute to the transparency literature with some novel insights.

Access to government information as a research topic spans across several disciplines. This thesis aimed to bring together views on access to government information from the perspective of political philosophy, comparative politics, law, public policy and journalism. The main goal of the thesis was to explain the impact of access to information as an anti-corruption policy and explore uses of FOI laws and open government data. First, I conducted a comparative cross-country analysis testing the effect of FOI laws and open government data on corruption control as a proxy for government accountability. Second, based on the results of this analysis and Lieberman's (2005) nested analysis approach, I purposefully selected a country case study (the UK), whose regression-predicted values are close to true values of corruption control. At the same time, the UK scores low on corruption and fulfils the theoretical conditions, i.e. has FOIA, scores high on the availability of open

government data, media accessibility and freedom, and judicial independence and accountability. I believe that this approach enabled me to demonstrate what impact FOI laws and open government data have when conditions are favourable. If the impact is low and FOIA and open data uses are not linked to the government oversight in these conditions; the impact will unlikely be higher and have relevance for accountability where these conditions are not met.

The thesis contributes to the field of transparency and accountability research in many ways. First, it shows that greater access to government information strengthens corruption control and accountability only when media can act upon this information freely. Therefore, while transparency policies adopted by authoritarian governments, which have a history of human rights infringements and repression of journalists might still produce positive outcomes in some areas, they are unlikely to contribute to government accountability, as authoritarian governments inhibit possibilities for media to hold it accountable.

Second, the thesis locates where the right of access to government information stands within democratic and representation theories. Many important concepts, such as "informed citizenship" or "monitorial citizenship" presuppose that citizens are able to exercise their right of access to government information. However, I find that not all information provisions are equal, they produce different outcomes, and some of them place high expectations on citizens' access and skills. In particular, I show that open government data leaves many behind due to its emphasis on technology, which many still cannot access and master. Their exclusion has also implications for their representation. If some groups are unable to seek and receive information, their interests are likely to be underrepresented in the public space. This, again, points to the role of information intermediaries, such as media and civil society. Also, inequalities embedded in technology do not only have implications for different demographics but also for different geographies. Open government data is resource-intensive policy, which requires technological and human infrastructure, and thus, it puts significantly more pressure on governments in the Global South than Global North. Transparency policies, such as FOIA and even more open government data do not require the affluent western liberal democracies to change their institutional practices and behaviours substantially. However, the scope of transformation that governments in the Global South have to undergo with limited resources is enormous.

I also demonstrate that FOIA and open government data produce different outcomes in terms of information provision. While FOIA provides "information on-demand", open data represents "information on tap". They both have advantages and disadvantages and implications for the quality vs. quantity of the provided information. While open government data makes information flow omnipresent, it is

the government who ultimately turns on and off the tap. If the government decides that some data should be withheld, citizens rarely have the means to make them disclose it. FOIA represents a narrower information provision. However, it is public demand-led, and as a result, it has higher informational value.

Below I summarise the theoretical and research design chapters and the findings of each empirical chapter in more detail. While the analysis produced answers to some of the questions, at the same time, new questions emerged which deserve further research attention. Hence, I also devote a brief section to discuss what potential avenues for further research my findings indicate. Lastly, I recognise the methodological limitations of my research.

Chapter 1 serves as an introduction to the topic. First, it traced the origins of the transparency-accountability link to the Enlightenment and discussed the work of political philosopher Bentham who argued that access to information and free press are crucial for the government oversight. Second, I introduced how the right of access to government information was legislated and addressed the disparities between the laws and their implementation and enforcement, drawing from work of transparency scholars (Berliner, 2012; Michener, 2010; Roberts, 2005, 2006; Worthy, 2017a). I argued that to minimise the gap between *de jure* and *de facto* FOIA, other safeguards, such as free press and independent oversight institutions, must be in place. If they are lacking, the incongruence between the letter of the law and implementation will be large. In fact, the above case of Serbia is a case in point. While its FOI law is assessed in the RTI Rating as the third-best in the world, its impact is questionable, as authorities are commonly reported to obstruct the rights of access to information and journalists are harassed and portrayed by political leaders as "foreign-backed propagandists" (Fredom House, 2017). At last, I discussed the differences between FOI as the demand-led provision of information and open government data as a supply-led provision of information.

Following the introduction of key concepts, I proceeded to the background for the theoretical framework of my research, presented in Chapter 2. Inspired by Bentham's (1999, 2001) political thought on publicity and Lindstedt's and Naurin's (2010) empirical research, I outline the conditions, I argue, that are necessary for transparency policies to achieve their anti-corruption goals. I agree with Lindstedt and Naurin (2010) that media access and freedom, and electoral accountability are critical for the information to have an impact. In addition, I emphasise the role of the independent and accountable judiciary and free civil society. I tested the theory's causal logic in Chapter 4. While this part of my thesis aimed to explain the causal link between transparency and corruption, other parts are exploratory in nature. Using Michener's and Worthy's Information-Gathering Matrix (2018), I also aimed to

explore how people use transparency policies and what information they seek. The matrix categorises the sought information into four categories: information in public vs. private interest, and political and non-political information. I tested this framework on the sample of FOI requests submitted to the UK central government in Chapter 7.

Chapter 3 delineated the research design and explained the choices made in examining the impact and uses of FOI laws and open government data. As mentioned above, a comparative cross-country analysis was selected to test the effect of these transparency policies on corruption control as a proxy for government accountability. Based on its results and Lieberman's nested analysis approach (2005), the UK was chosen as a single country case study. Within this case study, a combination of a survey method, interviews and quantitative text analysis was used to explore who and how uses FOI laws and government data and in what information they are interested.

8.1 Summary of findings, their implications and limitations

While several studies researching the causal link between FOI laws and corruption are available (Costa, 2013; Lindstedt & Naurin, 2010; Vadlamannati & Cooray, 2017), the effect of open government data on corruption has not yet been explored due to the novelty of this digitally-enabled transparency policy and a lack of measures. Chapter 4 provides one of the first analyses of the effect of open data on corruption levels across countries. Its results support the previous findings in the field, i.e. the utmost importance of media freedom (Camaj, 2013; Lindstedt & Naurin, 2010). The results suggest that the stand-alone effect of the availability of open data on corruption is statistically significant only in model with the interaction effect between open government data and press freedom. Moreover, the interaction effect of open data and press freedom has greater significance than the effect of open data individually.

These results were consistent across different measures of corruption and have several theoretical and policy implications worth mentioning. First, not only they support Lindstedt's and Naurin's (2010) findings and demonstrate the versatility of their proposed theoretical framework, they also show the visionary character of Bentham's (1999, 2001) political thought. He argued that giving people access to information will be toothless in securing government oversight unless the media are free to use this information without repercussions. In this regard, I believe that Bentham's work has still a lot to offer today's transparency scholars and policymakers. Second, I argue that the results invite to depart from the techno-optimist perspective on open government data as a cure that instantly solves long-standing political and

social problems, such as corruption. Just because the governments are confronted with the same problems globally does not mean that the solution is global too. The analysis broken down to multiple analyses of the effect of FOI laws and open data on corruption based on the country's income level shows that the direction of the relation between open data and corruption is not straightforward. The effect of FOI laws and open government data on corruption is positive (with an increase in access to information, corruption levels decrease) only for certain levels of economic development. Hence, the one-size-fits-all approach to transparency policies is unlikely to produce desired outcomes in all contexts. The results also showed that factors other than an increase in the access to government data are more critical for tackling corruption, in particular, free and fair elections, judicial independence and accountability, and country's economic power.

The main caveat of the analysis is that the used measures are imperfect. While measures are only proxies of researched phenomena, some reflect them better than others. The tradition of FOI laws variable, measured as the number of years since FOIA was adopted in a given country, has substantial limitations which might have also affected the analysis results. The FOI laws tradition can be conflated with democratic tradition to some extent as non-democratic countries started to pass FOI laws only recently. However, other available measures of FOIA are not perfect either. The RTI Rating, as was mentioned above, measures only *de jure* FOIA, i.e. the design of the law. I argued throughout this thesis and also using the case of Serbia just above that *de jure* FOIA can be markedly different from *de facto* FOIA. In other words, such a measure does not tell us anything about the implementation and enforcement of the law. While an excellent piece of legislation might be implemented poorly, imperfect legislation might well deliver good levels of openness. Nonetheless, other than to acknowledge this limitation, a little can be done to address it, as there are no cross-country measures of the quality of FOIA implementation available as they are notoriously difficult to develop. Another limitation of the analysis is that while it tells us that the availability of open government data is significantly associated with corruption control if media are free, it does not reveal how. To address this, I conducted a country case study looking at the uses of FOI laws and open data in the UK.

In Chapter 6, I explored the uses of FOI laws and government data through surveying journalists and interviewing CSO representatives. The survey confirmed that journalists both obtain information for their news stories through FOI requests and browsing proactively published government datasets. However, in line with the available literature on journalistic use of open data, the survey results confirmed that data journalism still is a minority specialism (Stoneman, 2015). That said, contrary to available evidence from India (Saxena & Janssen, 2017) and the US (Schrock, 2016;

Schrock & Shaffer, 2017), there were no significant differences between males and females in their use of software for advanced data analysis. Moreover, female journalists self-reported higher levels of data skills than their male colleagues. Given that women were underrepresented in the survey sample compared to the population of UK journalists, this is a finding worth further investigation. Other findings confirmed previous research in the field and proposed hypotheses. Survey participants and interviewees emphasised their use of FOI laws and government data for government oversight and considered them as tools essential for their everyday work.

Two caveats have to be considered when acquainting with the findings from this chapter. First, the survey sample was convenient, not representative and purposefully targeted news journalists who are likely to use FOI laws and government data. Second, the findings from the survey were complemented with the interviews with CSO representatives primarily. While the size of the survey sample (n=164) was satisfactory, given it was an expert survey, the future analysis would certainly benefit from a higher number of qualitative interviews.

Journalists strongly perceived their role as information intermediaries and government watchdogs. In responses to open-ended questions, they provided interesting insights into their uses of FOI laws and government data and their experiences with public authorities. In line with available research (Stoneman, 2015), their testimonies confirmed that they strongly value exclusivity of their news stories, and thus avoid using public online participation platforms, such as WDTK for submitting their FOI requests. Some mentioned that the obligation under FOIA to give out their identity prevent them sometimes from filing a request. Instead, they turn to data that the government already provided in the public domain. Nonetheless, a small number of respondents perceived obtaining data risky as well if user registration is obligatory to download the data. The findings revealed high levels of distrust towards the government.

Journalists also complained about the high refusal rate of their requests, which proves to be a justified concern when triangulated with the official statistics (Cheung, 2018; Institute for Government, 2019). Journalists also mentioned a differential treatment from the government. In line with already available research on the topic (Michener & Rodrigues, 2018; Michener et al., 2019; Roberts, 2005), they stated that they are questioned over the motivations of their requests, and FOI officers send their requests to the high-level management to decide about their outcomes. These findings suggest that FOIA is not prone to malpractices even in a country with a long tradition of democratic institutions as the UK. The law might have come into force overnight but the secretive culture, entrenched in public administration for centuries, will take some time to fade away.

One of the conclusions of Chapter 7 goes in the same direction, i.e. the UK government (notwithstanding the governing party) discloses willingly mostly information that is politically irrelevant. The survey offered insights into journalistic uses of FOIA and open data. However, they are likely to represent only a small group of all requesters. It is important to mention that it is impossible to establish exact figures and percentages, as the composition of requesters is not known because the law requires the requester to reveal her name, but not her profession or any other details. While journalists self-reported that they request information or search for datasets that are in the public interest, other FOIA and open data users might follow different ends. Using quantitative text analysis on the samples of over 37,000 and 16,000 FOI requests sent to the UK central government through an online participation platform WDTK over ten years, I explored the other uses of FOIA. As I mentioned above, journalists avoid using these public platforms. Hence, I could be confident that the requests on the public version of WDTK are primarily sent by users other than journalists.

So far, we only had findings from large-scale studies examining the content of FOI requests in Mexican (Berliner et al., 2018) and the US context (Kwoka, 2016, 2018b, 2018a). This study is the first to explore the content of FOI requests in the UK. Hence, its findings are innovative and provide new valuable insights into the public demand for government information and how the governments respond to this demand. The results from STM demonstrated that non-professional uses of FOIA are much more varied than those of journalists. While many of the estimated topics could be labelled as seeking information in the public interest, some of them clearly followed private goals, and they were proportionally more represented. Moreover, non-political topics prevailed over political ones. Overall, if FOIA use is compared against its initially intended purposes, it could be concluded that a substantial share of FOI requests is used differently, i.e. not for the control of the government. This is an important finding and has implications for FOI laws. While some scholars argue that the resources spent on FOIA uses that cannot be linked to accountability are misplaced (Pozen, 2017), I propose that FOIA should be repurposed and governments should use all opportunities, FOIA included, to establish contact with their constituents, explain new policies and counterbalance the spread of misinformation.

Another major finding is the mismatch between topics represented in FOI requests (estimated by STM) and those represented in open government data published on the national open data portal Data.gov.uk. The analysis of FOI requests received by the central government departments and datasets on Data.gov.uk clearly demonstrated that proactive publication of government datasets could not substitute FOIA precisely for the heterogeneity of the public demand for information. Nonetheless, although such a heterogeneous demand can hardly be predicted,

governments should make use of new computational methods to inform their proactive publication.

Lastly, the STM confirmed that the positive outcomes (successful FOI requests where the information was disclosed) are associated more with politically harmless information, such as the list of schools and contact details of headteachers or reports of ex-military vehicles' history requested by their new owners. Basically, this finding tells us in other words what journalists stated in the survey or CSO representatives in the interviews, i.e. the requests that matter, get often rejected. The concealed ISC's report into Russian interference in the UK concerning the EU referendum can serve as anecdotal evidence of thereof. Another finding emerged from the analysis, i.e. that the party in government and the persona of prime minister affected the prevalence of topics within FOI requests and offered a valuable methodological contribution. FOIA is a rather niche topic within political science discipline. However, the results from STM in Chapter 7 also showed that with the use of computational methods, FOIA could be fruitfully used for studying the impact of newly introduced policies or policy priorities more broadly. In the next section, I propose how the findings from the thesis could be used to advance the research on FOIA and open government data further.

8.2 Next steps

The thesis answered some of the initially set research questions. At the same time, the findings produced new ones, both in terms of theory and methods, that could serve as fruitful avenues for future research.

The findings from comparative cross-country analysis in Chapter 4 provide evidence of the effect of open government data on corruption levels. However, I found no effect of FOI laws individually. The proxy I used to measure FOIA is imperfect and reflects the FOI tradition in a country which can be conflated with democratic tradition. It does not tell us anything about how FOI requests are responded to on a daily basis. A cross-country measure of FOIA implementation would be more appropriate than the one used. Exploring how such a measure could be composed is a very challenging task given not all governments provide annual FOI statistics. Nonetheless, an effort to create such a measure could generate a valuable source of information.

The survey and interviews in Chapter 6 offer valuable insights into the use of FOI laws and open government data by journalists and civic activists. Some of them indicate paths worth investigating. For instance, it would be interesting to explore the reasons why UK journalists make limited use of confidential sources – whether it is

due to the potential ramifications from the Investigatory Powers Act 2016 or the reasonable volume of information available in the public domain. Journalists also complained about the differential treatment, as their FOI requests were flagged up to the high-level management. Some research about how gender, race and ethnicity or professional affiliation affect government responsiveness to FOI requests (Michener et al., 2019; Rodríguez & Rossel, 2018) or enquiries (Butler & Broockman, 2011) already exists. However, the existence of online participation platforms, such as WDTK, provides new opportunities to conduct experiments into how the publicity of communication and the content of FOI requests affect government responsiveness. Finally, a view from the other side – the experience of FOI officers with requests, either through interviews or an ethnographic study exploring FOI procedures in daily work, could also generate valuable findings. However, this might be a challenging task as my own experience with refusals for interviews from the public authorities proved.

Scholars who have written on content and discourse analysis argue that analysis of any text cannot be adequately executed without detailed scrutiny of conditions under which the text was produced (Hewitt, 2009; Scott, 1990). Some of the results from STM in Chapter 7 were relatively easy to connect to major political events or policy changes, for instance, the welfare reform under the 2010-2015 Conservative and Liberal Democrat Coalition government or human rights abuses by the Colombian armed forces, which were funded by the UK international aid programmes under the Labour governments in the 2000s. Future research could benefit from exploring the trends in the topic prevalence over time or comparative analyses of the FOI requests' content and content of major news media. However, it is essential to emphasise again that FOI requests from online participation platforms, such as WDTK, represent only a small share of all FOI requests sent to public authorities. If the central government published FOI data from their internal records, researchers could conduct analyses that would be representative of the whole population of FOI requests. I am confident that it would be beneficial for the governments as the results from analyses could then inform their proactive publication of both information and datasets. Hence, I would like to conclude this thesis with one policy-oriented recommendation and encourage public authorities to collect information on FOI requests systematically and make it available to the public and scientific community.

A) Appendices

Appendix 1 – supplementary materials for chapter 4

<u>The number and list of countries included in the main model</u>

Number of observations: 110

Albania, Argentina, Australia, Austria, Bangladesh, Belarus, Belgium, Benin, Bolivia, Bosnia and Hercegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Cameroon, Canada, Chile, China, Colombia, Congo DR, Costa Rica, Cote d'Ivoire, Croatia, Czech Republic, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Ethiopia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Haiti, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kosovo, Kyrgyz Republic, Latvia, Lebanon, Macedonia, Malawi, Malaysia, Mali, Mauritius, Mexico, Moldova, Montenegro, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, New Zealand, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Russian Federation, Rwanda, Saudi Arabia, Senegal, Serbia, Sierra Leone, Singapore, Slovakia, South Africa, South Korea, Spain, Sweden, Switzerland, Tajikistan, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, Ukraine, United Kingdom, United States, Uruguay, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe

Figure A-1: Mean of data availability score by the Freedom House country status
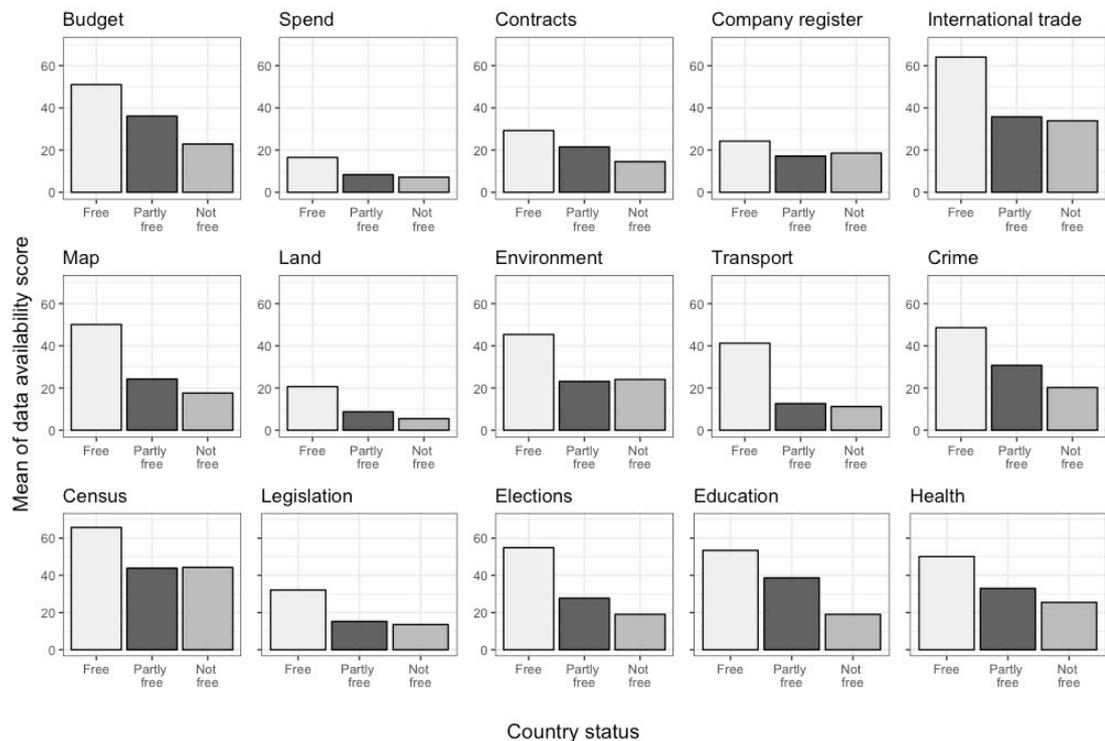
Regression diagnostic plots for all models

Figure A-2: Diagnostic plots of residuals for model 1 (no interactions)

Note: Residuals vs. Fitted plot shows no distinct pattern; the line is relatively straight, horizontal at zero; the plot indicates that there is a linear relationship between the dependent and the independent variables, i.e. that the linear relationship assumption is met. Normal Q-Q plot shows that most residuals fall along the diagonal reference line; the plot indicates that the residuals are normally distributed, i.e. that the normality assumption is met. Scale-Location plot shows no distinct pattern; the line is relatively straight, horizontal, with residuals spread approximately evenly across the range of fitted values. Residuals vs. Leverage plot shows that none of the points falls beyond Cook's distance of 1, which was suggested as one of the cut-off points to identify influential observations (e.g. outliers and/or high-leverage points).

Figure A-3: Histogram of residuals for model 1 (no interactions)
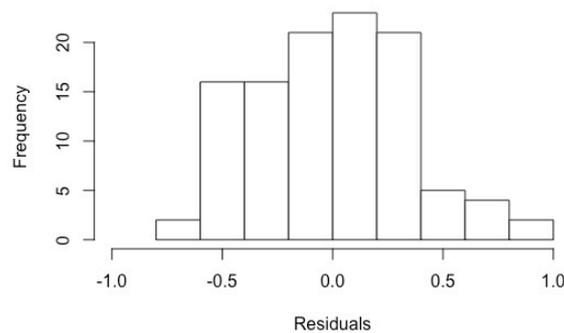
Note: The histogram shows that residuals are normally distributed, which was also confirmed by the results of a formal test. Shapiro-Wilk normality test results (p = 0.188) suggest that the null hypothesis that the data (in this case, residuals) come from a normally distributed population cannot be rejected. Breusch-Pagan test results (p = 0.485) suggest that the null hypothesis of homoskedasticity cannot be rejected (i.e., there may not be a heteroskedasticity problem).

Figure A-4: Independence of residuals plot for model 1 (no interactions)
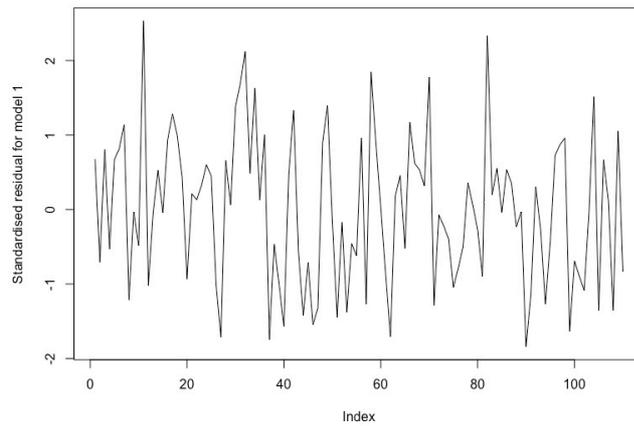
Note: The plot shows no serial pattern in the way the residuals of the model appear. It indicates that the residuals are independent of each other.

Figure A-5: Diagnostic plots of residuals for model 2 (open data-press freedom interactions)


Note: Residuals vs. Fitted plot shows no distinct pattern; the line is relatively straight, horizontal at zero; the plot indicates that there is a linear relationship between the dependent and the independent variables, i.e. that the linear relationship assumption is met. Normal Q-Q plot shows that most residuals fall along the diagonal reference line; the plot indicates that the residuals are normally distributed, i.e. that the normality assumption is met. Scale-Location plot shows no distinct pattern; the line is relatively straight, horizontal, with residuals spread approximately evenly across the range of fitted values. Residuals vs. Leverage plot shows that none of the points falls beyond Cook's distance of 1, which was suggested as one of the cut-off points to identify influential observations (e.g. outliers and/or high-leverage points).

Figure A-6: Histogram of residuals for model 2 (open data-press freedom interactions)
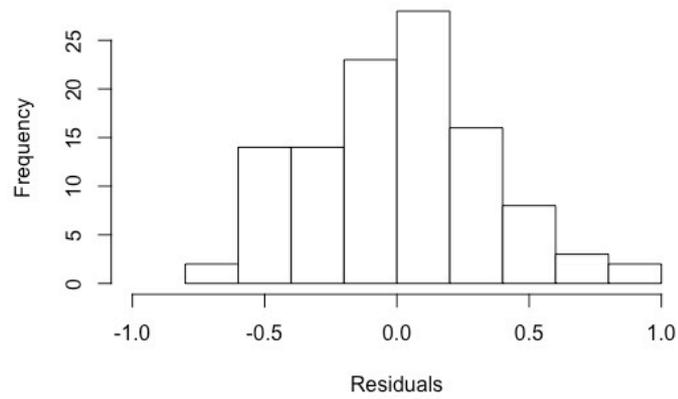

Note: The histogram shows that residuals are normally distributed, which was also confirmed by the results of a formal test. Shapiro-Wilk normality test results (p = 0.053) suggest that the null hypothesis that the data (in this case, residuals) come from a normally distributed population cannot be rejected. Breusch-Pagan test results (p = 0.761) suggest that the null hypothesis of homoskedasticity cannot be rejected (i.e., there may not be a heteroskedasticity problem).

Figure A-7: Independence of residuals plot for model 2 (open data-press freedom interactions)
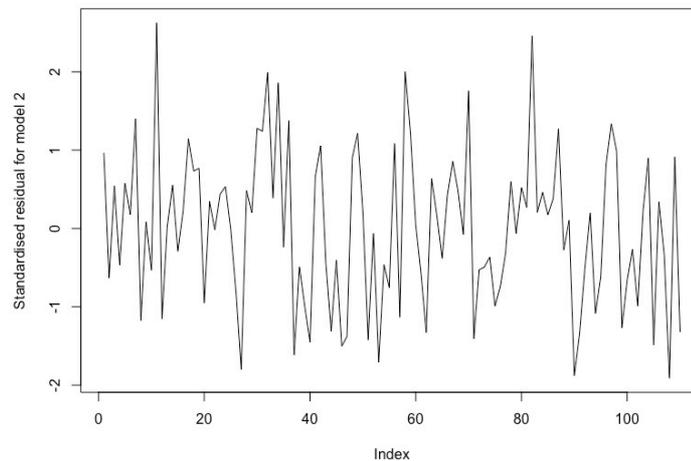

Note: The plot shows no serial pattern in the way the residuals of the model appear. It indicates that the residuals are independent of each other.

Figure A-8: Diagnostic plots of residuals for model 4a and 4b (FOI years-press freedom interactions)
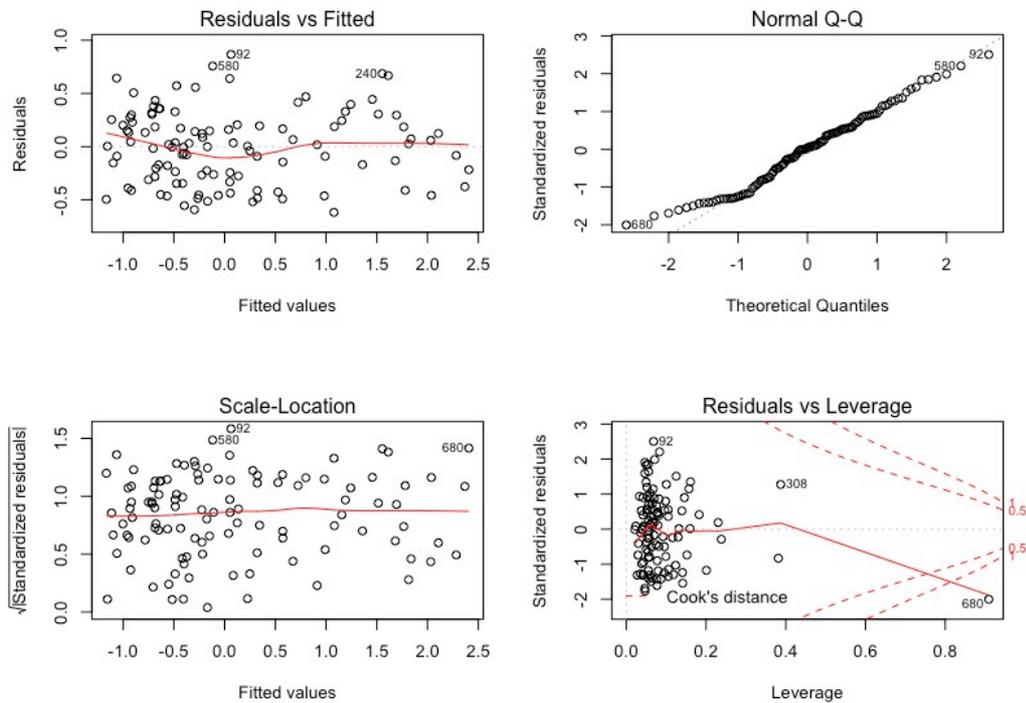

Note: Residuals vs. Fitted plot shows no distinct pattern; the line is relatively straight, horizontal at zero; the plot indicates that there is a linear relationship between the dependent and the independent variables, i.e. that the linear relationship assumption is met. Normal Q-Q plot shows that most residuals fall along the diagonal reference line; the plot indicates that the residuals are normally distributed, i.e. that the normality assumption is met. Scale-Location plot shows no distinct pattern; the line is relatively straight, horizontal, with residuals spread approximately evenly across the range of fitted values. However, as can be observed from the Residuals vs. Leverage plot, observation no. 680 (Sweden) is an influential observation because it falls beyond Cook's distance of 1, which was suggested as one of the cut-off points to identify influential observations. It might affect the slope of the regression line and have an effect on the parameter estimates. Therefore, it was removed from the data for the analysis.
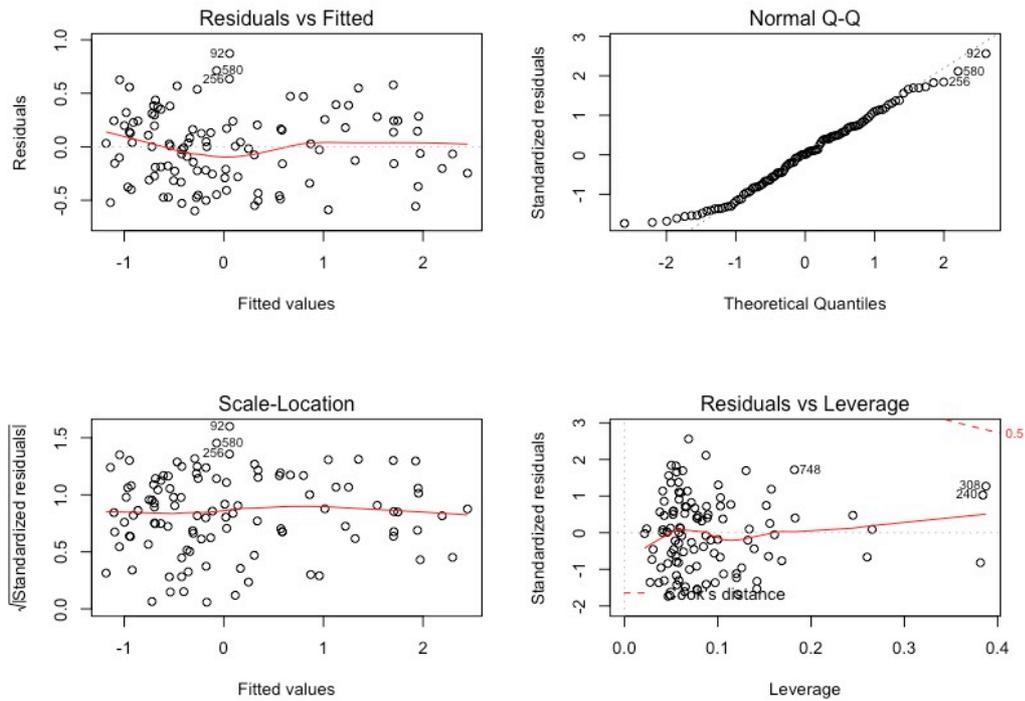
Note: Once the influential observation (no.680 - Sweden) was removed, the Residuals vs. Leverage plot shows that none of the points falls beyond Cook's distance of 1.

Figure A-9: Histogram of residuals for model 4b (FOI years-press freedom interactions)
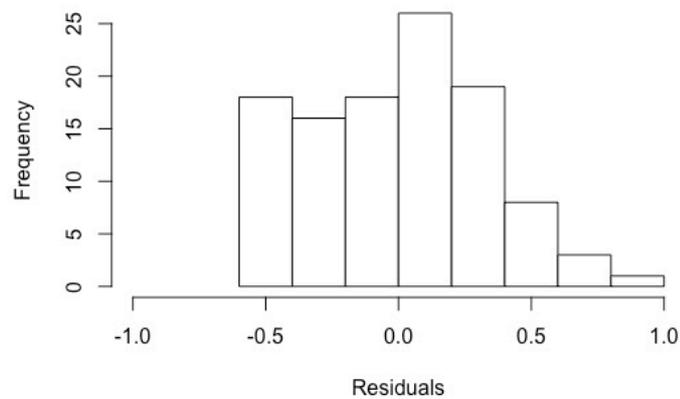
Note: The histogram shows that residuals are normally distributed, which was also confirmed by the results of a formal test. Shapiro-Wilk normality test results ($p = 0.106$) suggest that the null hypothesis that the data (in this case, residuals) come from a normally distributed population cannot be rejected. Breusch-Pagan test results ($p = 0.972$) suggest that the null hypothesis of homoskedasticity cannot be rejected (i.e., there may not be a heteroskedasticity problem).

Figure A-10: Independence of residuals plot for model 4b (FOI years-press freedom interactions)
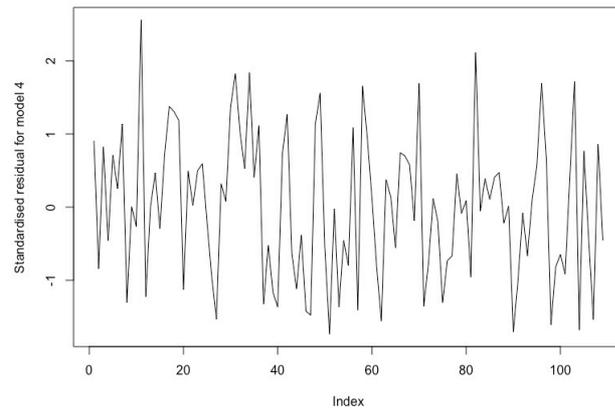

Note: The plot shows no serial pattern in the way the residuals of the model appear. It indicates that the residuals are independent of each other.

Figure A-11: Diagnostic plots of residuals for model 3 (open data-judiciary interactions)

Note: Residuals vs. Fitted plot shows no distinct pattern; the line is relatively straight, horizontal at zero; the plot indicates that there is a linear relationship between the dependent and the independent variables, i.e. that the linear relationship assumption is met. Normal Q-Q plot shows that most residuals fall along the diagonal reference line; the plot indicates that the residuals are normally distributed, i.e. that the normality assumption is met. Scale-Location plot shows no distinct pattern; the line is relatively straight, horizontal, with residuals spread approximately evenly across the range of fitted values. Residuals vs. Leverage plot shows that none of the points falls beyond Cook's distance of 1, which was suggested as one of the cut-off points to identify influential observations (e.g. outliers and/or high-leverage points).

Figure A-12: Histogram of residuals for model 3 (open data-judiciary interactions)
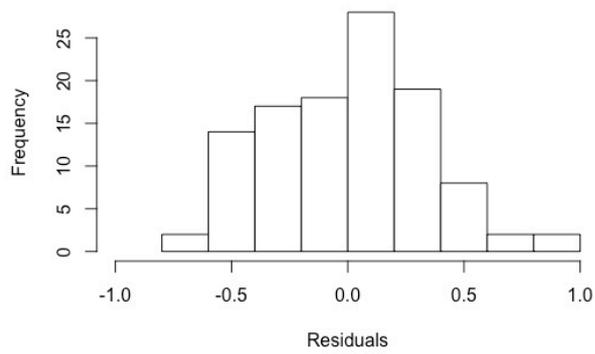
Note: The histogram shows that residuals are normally distributed, which was also confirmed by the results of a formal test. Shapiro-Wilk normality test results (p = 0.160) suggest that the null hypothesis that the data (in this case, residuals) come from a normally distributed population cannot be rejected. Breusch-Pagan test results (p = 0.566) suggest that the null hypothesis of homoskedasticity cannot be rejected (i.e., there may not be a heteroskedasticity problem).

Figure A-13: Independence of residuals plot for model 3 (open data-judiciary interactions)
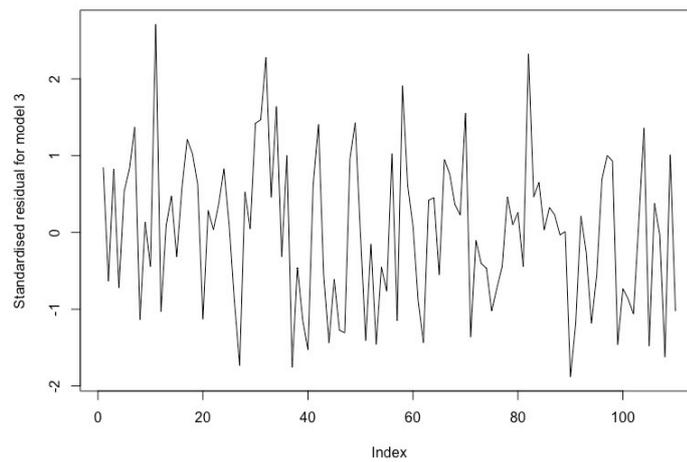
Note: The plot shows no serial pattern in the way the residuals of the model appear. It indicates that the residuals are independent of each other.

Figure A-14: Diagnostic plots of residuals for model 5 (FOI years-judiciary interactions)
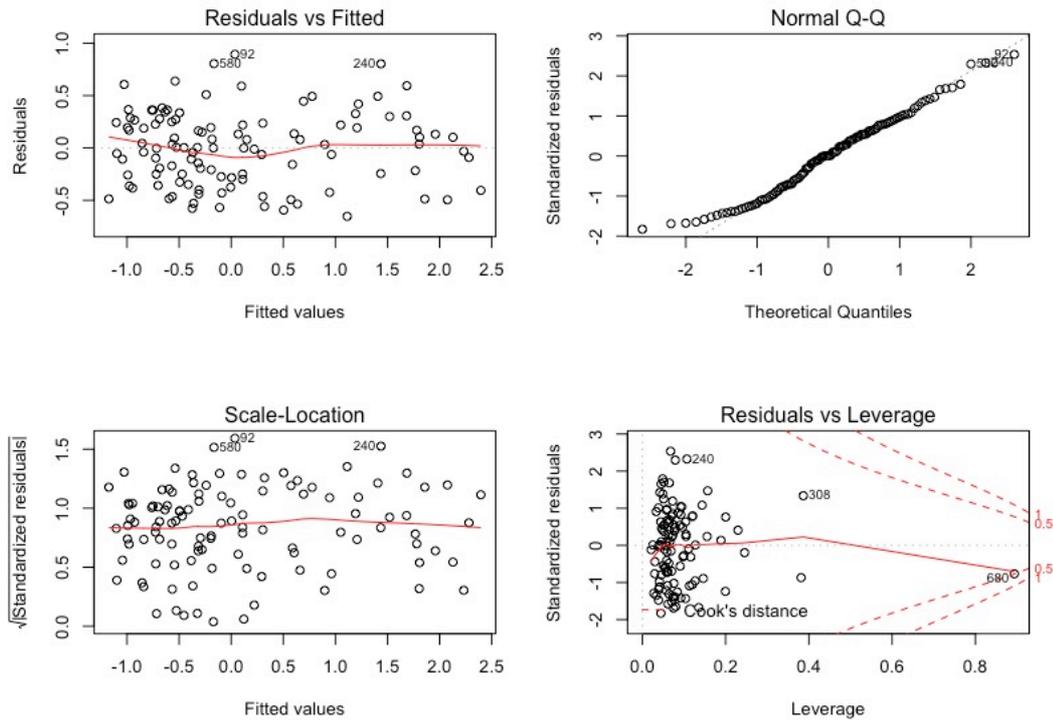
Note: Residuals vs. Fitted plot shows no distinct pattern; the line is relatively straight, horizontal at zero; the plot indicates that there is a linear relationship between the dependent and the independent variables, i.e. that the linear relationship assumption is met. Normal Q-Q plot shows that most residuals fall along the diagonal reference line; the plot indicates that the residuals are normally distributed, i.e. that the normality assumption is met. Scale-Location plot shows no distinct pattern; the line is relatively straight, horizontal, with residuals spread approximately evenly across the range of fitted values. However, as can be observed from Residuals vs. Leverage plot, observation no. 680 (Sweden) might be an influential observation. Although it does not fall beyond Cook's distance of 1, which was suggested as one of the cut-off points to identify influential observations, according to Cook (1977) any observations at values above 0.5 should be investigated as they might affect the slope of the regression line and affect the parameter estimates. Therefore, observation no. 680 (Sweden) was removed from the data for the analysis, as it was within Cook's distance of 0.5 to 1.
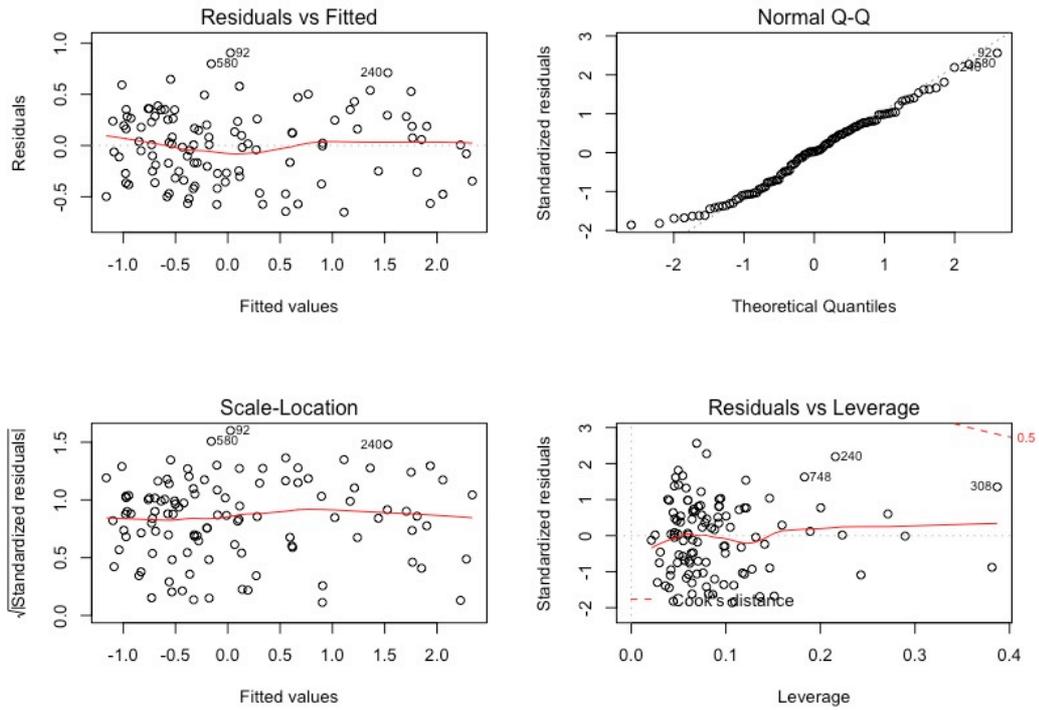
Note: Once the influential observation (no.680 - Sweden) was removed, the Residuals vs. Leverage plot shows that none of the points falls within Cook's distance of 0.5 to 1.

Figure A-15: Histogram of residuals for model 5 (FOI years-judiciary interactions)
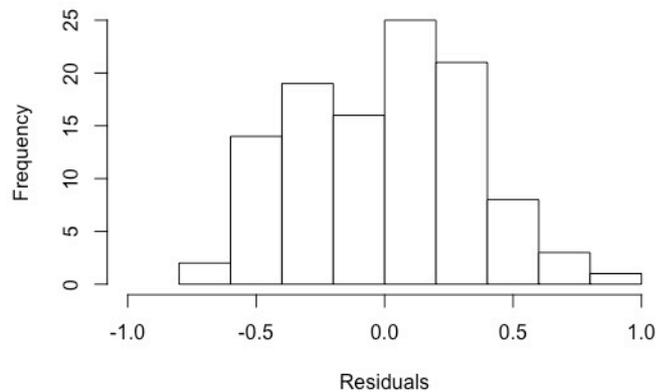

Note: The histogram shows that residuals are normally distributed, which was also confirmed by the results of a formal test. Shapiro-Wilk normality test results (p = 0.168) suggest that the null hypothesis that the data (in this case, residuals) come from a normally distributed population cannot be rejected. Breusch-Pagan test results (p = 0.398) suggest that the null hypothesis of homoskedasticity cannot be rejected (i.e., there may not be a heteroskedasticity problem).

Figure A-16: (FOI years-judiciary interactions)
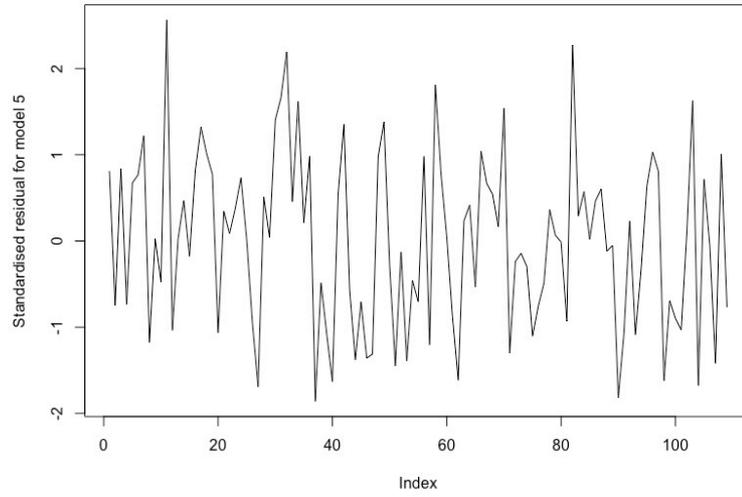
Note: The plot shows no serial pattern in the way the residuals of the model appear. It indicates that the residuals are independent of each other.

Appendix 2 – supplementary materials for chapter 6

Appendix 2.1: Information sheet for survey participants

SURVEY ON THE USE OF OPEN DATA AND FREEDOM OF INFORMATION

This survey is asking journalists and civic activists about their use of freedom of information and government datasets. The survey is online, completely anonymous and will take about 5 minutes to complete.

Who is conducting the research?

I am a PhD researcher based in the School of Government and Public Policy at the University of Strathclyde, Glasgow. My research broadly explores the impacts of government transparency. I am inviting you to take part in an anonymous online survey that asks about your use of government datasets and freedom of information legislation.

What is the purpose of this investigation?

My research looks at the impact of greater government transparency, in particular, the release of government datasets into the public domain. I explore a potential causal link between open data publication and perceived and actual corruption. The preliminary results show that the effect is conditional upon the presence of free media. In countries with relatively free media, greater availability of government data is associated with lower levels of both perceived and actual corruption. Following up on these findings, I would like to explore how journalists and civic activists in the UK use freedom of information legislation and open data in their stories.

Do you have to take part?

You do not have to take part. Participation in this survey is entirely voluntary, and you decide whether you wish to participate. In this case, simply exit the survey without submitting it. There are no potential risks to you in participating in this project.

What will you do in the project?

You will be invited to complete the online survey on your use of open data and freedom of information. A series of questions will appear on the screen, and you will be asked to check the box that best describes your answers. The survey should take no more than 5 minutes.

Why have you been invited to take part?

You have been invited to take part in this study as you are likely to use freedom of information legislation or publicly available government data.

What happens to the information in the project?

No names, identifying details or personal information will be collected. Responses will be used to inform my thesis and other publications. The survey is fully anonymised. No personal data (e.g. no IP addresses) is collected. You cannot be linked to your survey responses in any way.

What happens next?

To take part, please click on the bottom right arrows of your screen. This will direct you to a consent form, followed by the survey. If you do not wish to participate, thank you very much for your interest and for taking the time to read this information sheet.

Researcher contact details:

Maria Zuffova (PhD Candidate), School of Government and Public Policy, McCance Building, 16 Richmond Street Glasgow G1 1QX, Phone: +44 (0) 7577 877 627, Email: maria.zuffova@strath.ac.uk

Dr Heinz Brandenburg (Lead Supervisor), School of Government and Public Policy, Phone: +44 (0) 1415 484 376, Email: heinz.brandenburg@strath.ac.uk

This investigation was granted ethical approval by the School of Government and Public Policy Ethics Committee. If you have any questions/concerns, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or to whom any questions may be directed or further information may be sought from, please contact Dr Fabrizio de Francesco at Fabrizio.de-francesco@strath.ac.uk.

Appendix 2.2: The text of the survey on the use of freedom of information and open data

1. Consent: I confirm that I have read and understood the information above. I consent to take part in the survey.
o   Yes
o   No

2. What is your profession?
o   A journalist
o   A data journalist
o   A civic activist
o   Other - please, specify:

3. What is your age?

_____

4. Which gender do you identify with?
o   Female
o   Male
o   Other
o   Prefer not to say

5. What is your main source of information for your stories (Please tick max. two that apply in most cases)?
☐   Publicly available government data
☐   Information and data obtained through freedom of information
☐   Information and data obtained from confidential sources and whistle-blowers known to you
☐   Information obtained from an anonymous leaker
☐   Other - please, specify:

6. Have you ever submitted an FOI request?
o   Yes
o   No

The following questions were displayed only to those respondents who claimed to have submitted an FOI request in the past.

7. Was FOIA instrumental for any of your stories?
o   Yes
o   No

8. On average, how many FOI requests do you file per year?

o   Less than 10

o   10 – 50

o   51 - 100

o   More than 100

9. How do you usually submit an FOI request in most cases? (Please tick all that apply.)

☐   Using my personal email account

☐   Using my work email account

☐   By phone

☐   Using social networking sites

☐   Using public version of WhatDoTheyKnow.com

☐   Using a pro version of WhatDoTheyKnow.com

☐   Other - please, specify

10. Who are the addressees of your FOI requests in most cases?

o   UK government

o   Scottish government

o   Local government

o   Other - please, specify:

11. Has any of the following ever discouraged you from submitting an FOI request? (Please tick all that apply.)

☐   Twenty working days' time period to respond

☐   The obligation to include your name and contact details

☐   A previous negative experience e.g. withheld information

☐   Other - please, specify:

12. Has a public authority ever asked you for reasons of your FOI request?

o   Yes

o   No

o   Don't know / Don't remember

13. Was your FOI request ever rejected?

o   Yes

o   No

o   Don't know / Don't remember

The following question was displayed only to those respondents who claimed to have FOI request rejected in the past.

14. On what grounds were your FOI requests rejected in most cases? (Please tick all that apply)

☐ Agency does not fall under the obligation to provide information under FOIA

☐ Information was already in public domain

☐ Requested information was not held

☐ Request was treated as vexatious

☐ Disclosure would put significant time burden on staff

☐ Cost grounds

☐ Class-based exemptions (e.g. Information held for criminal investigations, information relating to correspondence with the royal family)

☐ Prejudice-based exemptions (disclosure could create potential harm)

☐ Contrary to Data Protection Act 2018

☐ Other - please, specify

15. Please, can you conclude in your own words what are the biggest benefits of FOI to your everyday work?

The following questions were again displayed to all respondents.

16. Have you ever visited Data.gov.uk to look for data for your stories?

o Yes

o No

The following questions were displayed only to those respondents who claimed to have visited Data.gov.uk in the past.

17. What data were you looking for?

18. Did you find it?

o Yes, always.

o Yes, sometimes.

o No, never.

The following questions were again displayed to all participants.

19. Which other government websites have you visited to search for the data for your stories? (Please tick all that apply)

☐ The Crown Prosecution Service website

☐ Office for National Statistics website

☐ Statistics.gov.scot

☐ National Records of Scotland website

☐ Gov.uk website

☐ The National Archive website

☐ The Electoral Commission website

☐ Companies House website

☐ Parliament website

☐ Other - please, specify:

20. Does an obligation to register/log in with the government website discourage you from looking for information there?

o Yes - please, specify your reasons:

o No

21. Do you analyse data yourself?

o Yes, always.

o Yes, sometimes.

o No, I need help analysing data.

22. What software do you mostly use for the analysis?

☐ MS Excel or equivalent

☐ R

☐ STATA

☐ Python

☐ Other - please, specify:

23. Please, can you conclude in your own words what are the biggest benefits of publicly available government datasets to your own work?

24. Please, use the space below to add any other comments:

Appendix 2.3: Information sheet for interviewees

Introduction

My name is Mária Žuffová, and I am a PhD candidate in Politics at the University of Strathclyde, Glasgow. You are being invited to take part in my PhD research project into the use of freedom of information and open government data and its effect on government accountability. Before you decide if you want to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Please, contact me at maria.zuffova@strath.ac.uk or +44 7577 877 627 if there is anything that is not clear or if you wish to receive more detailed information about the research project.

What is the purpose of this investigation?

My PhD research is divided into two parts. Using cross-national comparative data, the first part aims to explore how greater access to government data affects government accountability measured as perceived and actual corruption. It also looks at the effect of individual government datasets that may be crucial for the public to control the use of public resources, for instance, data on spending, contracts, company registration data and land data. The preliminary results show that open data availability affects levels of perceived corruption, but not of actual corruption. Another important finding is the conditionality of the effect of open data on the levels of perceived and actual corruption and their strength upon the presence of free media. The second part is an in-depth case study, which focuses on differences and similarities between freedom of information and open government data, what kind of transparency they bring, who uses it and for what purposes in the UK. The UK was systematically identified, based on large-N analysis and analysis of residuals as a crucial case. Expert interviews, in which you are asked to participate, will help me to gain important contextual information and develop a case study of a higher quality.

Do you have to take part?

No, you don't. It is important for you to know that your participation in this research is entirely voluntary. If you decide to take part and later change your mind, then you may withdraw from this research at any time without having to give any reason. If you take part, you have up until 3 days after the interview to let us know that you have changed your mind and we will remove and delete the interview.

What will you do in the project?

If you decided to take part in the project, I have arranged an in-person or online interview to hear about your experience of accessing government information and data. This interview would include discussing what kind of government information you are particularly interested in, where you search for information (whether you submit a freedom of information request, visit open data portals, or use your confidential sources). I would also be interested in how institutions have justified if they withheld information from you. The interview will take no more than 30 minutes

and will be arranged at a time to suit you. With your permission, I will audio record the interview, and a written transcript of the interview can be returned to you (if requested) to check before I use it in my analysis.

Why have you been invited to take part?

You have been invited to take part in my research because as a journalist you often request information from the government, and I am interested in your experiences. Your perspective is invaluable for my research, as it helps to balance the views of other interviewees. Apart from journalists, I aim for interviewing civil servants, CSO representatives, and researchers. To deliver high-quality, valid and reliable research, I believe that each of these groups has to be included and thus, I very much appreciate your time investment in this research project.

What are the potential risks to you in taking part?

There are no potential risks to you in participating in this project. However, if you are concerned that discussing your experience with public authorities with regard to freedom of information and open data management might have repercussions for you, please, be assured that your personal details are pseudo-anonymised automatically, and your name and affiliation will only be revealed if you choose so.

What happens to the information in the project?

All data (both audio files and transcript text files) will be stored on my laptop, and transcript text files will also be stored in Strathcloud Sharefile. My laptop is fully encrypted using FileVault, which secures the data on my disk by encrypting its contents automatically. The recovery key, which can be used to unlock the disk, is stored in a safe place. Strathcloud Sharefile is a secure file-storage and sharing application provided by the University of Strathclyde. All your details will be handled with the utmost importance and confidentially. You will not be identifiable in the project. Your personal details will be pseudo-anonymised (i.e. the raw data will be anonymised and given a code name, with the key for code names being stored in a separate location from the raw data). I will be the only one with the access to both pseudo-anonymised data and a key for code names. My supervisors, Dr Heinz Brandenburg and Dr Zachary Greene, will have access to pseudo-anonymised data only. You and other participants will be referred to by a general category (e.g. an investigative journalist from London, Glasgow etc.). Should you wish to be identified in the project using your name and affiliation, please specify this in the consent form.

The University of Strathclyde is committed to the principles and obligations set out in the General Data Protection Regulation (GDPR) and the Data Protection Act 2018 (DPA).

What happens next?

If you still are happy to take part, please, sign the attached consent form to confirm this. The findings of my research will be publicly available after passing the viva and submitting my PhD thesis. The tentative date for submitting my thesis is September

2019. If you do not want to be involved in the project, thank you for taking your time to learn about my PhD project. I very much appreciate it.

Researcher contact details:

Maria Zuffova (PhD Candidate), School of Government and Public Policy, McCance Building, 16 Richmond Street Glasgow G1 1QX, Phone: +44 (0) 7577 877 627, Email: maria.zuffova@strath.ac.uk

Dr Heinz Brandenburg (Lead Supervisor), School of Government and Public Policy, Phone: +44 (0) 1415 484 376, Email: heinz.brandenburg@strath.ac.uk

This investigation was granted ethical approval by the School of Government and Public Policy Ethics Committee. If you have any questions/concerns, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or to whom any questions may be directed or further information may be sought from, please contact Dr Fabrizio de Francesco at Fabrizio.de-francesco@strath.ac.uk.

Appendix 2.4: Topics and questions discussed during the interview

<u>Journalistic profession</u>
- experience with working on investigative stories
- use of different information sources
- use of different communication channels

<u>FOI related topics</u>
- Burn's Commission in 2015 to review FOI legislation
- RTI Rating
- FOI de jure vs. FOI de facto
- FOIA Code of Practice updated in 2018
- FOIA as a government priority
- experience with FOI legislation
- reasons for refusals of FOI requests
- adaptation strategies
- chilling effect
- use of online participation platforms

<u>Open data related topics</u>
- open data as a government priority
- open data metrics (ODB and GODI)
- use of publicly available government data, and its pros and cons
- open data users
- data quality
- data that is not being collected
- data analysis skills
- relationship between FOI and open data

Appendix 2.5: The list of interviewees

The interviews are listed in chronological order as they happened.

INT1_UKCG   Central government body representative, 2017, skype interview, 30 minutes 28 seconds.

INT2_CSO   CSO representative, 2017, skype interview, 50 minutes and 39 seconds.

INT3_CSO   CSO representative, 2018, personal interview in London, 32 minutes and 47 seconds.

INT4_CSO   CSO representative, 2018, personal interview in London, 67 minutes and six seconds.

INT5_CSO   CSO representative, 2018, personal interview in London, 50 minutes and 24 seconds.

INT6_CSO   CSO representative, 2018, personal interview in London, 58 minutes and 49 seconds.

INT7_SIC   Scottish Information Commissioner Office, 2019, personal interview in St. Andrews, 72 minutes and 29 seconds.

Appendix 3 - supplementary materials for chapter 7

<u>Appendix 3.1: Central government bodies included in the analysis</u>

AGO	Attorney General's Office
CC	Charity Commission for England and Wales
CMA	Competition and Markets Authority
CO	Cabinet Office
CPS	Crown Prosecution Service
BEIS	Department for Business, Energy and Industrial Strategy
DCLG	Ministry of Housing, Communities and Local Government
DCMS	Department for Digital, Culture, Media and Sport
DEFRA	Department for Environment, Food and Rural Affairs
DExEU	Department for Exiting the European Union
DfE	Department for Education
DfID	Department for International Development
DfT	Department for Transport
DHSC	Department of Health and Social Care
DIT	Department for International Trade
DMO	UK Debt Management Office
DWP	Department for Work and Pensions
FC	Forestry Commission
FCO	Foreign and Commonwealth Office
FSA	Food Standards Agency
GAD	Government Actuary's Department
GLD	Government Legal Department
HMRC	Her Majesty's Revenue and Customs
HMLR	Her Majesty's Land Registry
HMT	Her Majesty's Treasury
HSE	Health and Safety Executive
MOD	Ministry of Defence
MOJ	Ministry of Justice
NA	The National Archives
NCA	National Crime Agency
NIO	Northern Ireland Office
NSI	National Savings and Investments
OAG	Office of the Advocate General for Scotland
OLHC	Office of the Leader of the House of Commons
OLHL	Office of the Leader of the House of Lords
OFSTED	Office for Standards in Education, Children's Services and Skills
OFGEM	Office of Gas and Electricity Markets
OFQUAL	Office of Qualifications and Examinations Regulation
ORR	Office of Rail and Road
OS	Ordnance Survey
ONS	Office for National Statistics
OFWAT	Water Services Regulation Authority

RPA	Rural Payments Agency
SCUK	Supreme Court of the United Kingdom
SFO	Serious Fraud Office
SO	Scotland Office
WO	Wales Office

Appendix 3.2: Exploration of different topic models

The examples are structured as follows. Each example states to what public authority the request was sent, when it was made and what the final outcome of the request was. The full text of the request follows. The requests were not redacted in any way, apart from correcting grammatical errors. If the authority classified the request as vexatious, this information is also added. The requests are varied. Most notably, they differ in length. While some are just two or three sentences long, others are one and more pages long. Due to limited space, where possible I purposefully selected only short FOI requests as examples.

Detailed exploration of topics in the model with 33 topics (all requests included)

Topic 1: Transport

Highest probability words: transport, road, rail, rout, car, use, passenger
FREX words: airport, traffic, scanner, fare, bus, franchise, byelaw
Representative FOI requests:
Example 1. DfT, 19 October 2017, successful

> Further to my previous request (Ref: GT51/1/2/2/F001532), please provide full details for the following authorisations, including any associated drawings: 14/11/2005 A1114, Chelmsford Bus Lane - Motorcycle, Cycle and Taxi, 03/07/2006 Various Roads in Chelmsford Bus, Motorcycle, Taxi and Cycle, 30/06/2006 Motorcycles in bus lanes, Chelmsford, 07/07/2008 Bus lane signs for Chelmsford, Essex, and 08/11/2007 A1114 Bus Lane, Chelmsford.

Example 2. DfT, 16 November 2009, refused

> Where are the areas where passenger trains may travel at 125mph on the Great Western rail franchise? What plans are there to provide or increase the lengths of 125mph railway line on the Great Western "Berks and Hants direct line" between Reading, Westbury and Taunton?

Example 3. DfT, 20 January 2016, partially successful

> Can you please advise the number of pedestrians killed or injured on puffin pedestrian crossings, when the crossings were presumed not to be operating correctly?

Topic 2: Individual complaints

Highest probability words: family, cause, problem, see, evidence, time, life
FREX words: shrive, wreck, chum, threaten, story, away, guardian
Representative FOI requests:
Example 1. DWP, 12 May 2013, refused, classified as a vexatious request

How come when the DWP relate to impact assessments they always try to reassure that they have been done independently, when the truth is, they hide facts like this.

"BEDROOM TAX SUICIDE"

Ten days ago, Stephanie Bottrill sat in the redbrick terrace house which had been home for 18 years to write notes to her loved ones, the Sunday People reports. She ripped the pages from a spiral-bound notebook and placed them neatly in little brown envelopes. There was one for her son. Another for her daughter. Her mother. Friends. And a very special one for the year-old grandson she doted on. Then in the early hours of last Saturday Stephanie, 53, left her home for the last time, leaving her cat Joey behind as the front-door clicked shut. She crossed her road in Meriden Drive, Solihull, to drop one of her letters and her house keys through a neighbour's letterbox. Then she walked 15 minutes through the sleeping estate to Junction 4 of the M6. And at 6.15 am she walked straight into the path of a northbound lorry and was killed instantly. Stephanie Bottrill had become the first known suicide victim of the hated Bedroom Tax. In the letter to her son, Steven, 27, she had written: "Don't blame yourself for me ending my life. The only people to blame are the Government." Stephanie was tormented over having to find £20 a week to pay for the two under-occupied bedrooms she had been assessed for. Days before her death, she told neighbours: "I can't afford to live anymore." Solihull council Labour group leader David Jamieson, who knows the family well, said: "I'm absolutely appalled this poor lady has taken her own life because she was worried how she would pay the Bedroom Tax. "I hope the Government will take notice and reconsider this policy." Strange how everything that BILL GUNNYEON endorses results in another death? My question is simple, how many must die before the DWP realise that their ill-thought-out policies are taking lives?

Example 2. CPS, 30 October 2014, internal review

I have tried to find the CPS for Derby. I was told that it had now moved from Derby to Nottingham. I looked on the CPS website for Nottingham and got it on the map and went there by coach. I walked around Nottingham until I found it, but when I found it and asked the receptionist, I was told it was not there. I am on £27 a week Pension Credit, and this only leaves me £3.85 a day for food and other things, so it is costly for me to use my phone for the numbers I have been given as I am afraid the phone will not give an engaged tone but connect me to music immediately and ask me to please hold. The DWP cost me £1.50 a few days ago, but I had to put the phone down because I need to spend that money on tea and bread and not music. I hate music over the phone and prefer the engaged tone because I do not have to pay and not get a word and I don't need the music. Email is the cheapest for me as I can't afford stamps either although I am put to the expense of stamps almost every day by the DWP, and I wouldn't mind, they never answer my letters. The Nottingham CPS address is still on your web site as the address for DERBYSHIRE CPS, but it's about 4 or 5 weeks since I went to Nottingham and told I was in the wrong place.

This is the address on your website for Derbyshire and Derby CPS covering Derbyshire, Lincolnshire and Nottinghamshire: 2 King Edward Court, King

Edward Street, Nottingham, NG1 1EL, DX 729100 Nottingham 48, Tel: 01158 523 300. And this is the place I went to. If this is not the proper address for Derby/Derbyshire CPS why do you leave it on your website? I am 70 yrs. of age and have had a stroke a couple of years ago and have hyperparathyroidism, which takes calcium out of the bones and makes standing or walking painful. I need to contact CPS because I made a complaint about false statements on the DWP reports which are sent to the CPS recommending prosecution. I have complained for the best part of a year about these false allegations. My complaints have been ignored until a few days ago when the DWP FES looked at one of my complaints of false statements about the IUC and said there was no evidence in the Transcripts that the allegations were justified, that it was false, and I was given an apology.

However, the papers were sent to the CPS the "Prosecution Department", and these will still have the false allegation/s on them, and there is no evidence that, though the DWP FES has apologised to me, that they have sent a message to the CPS that they have provided documents to the CPS with false allegation/s on them. So, I feel it is left up to me to inform the CPS and ask them to retract the allegations. I have asked at the CPS Office at the Derby Crown Court, and someone was sent out to talk to me, but he said that he was a prosecutor and that he should not be talking to me nor would he and walked away.

Is there any way that I can get in touch with the CPS to inform them that there was a retraction by the DWP FES of an allegation so that it can be removed from among the other allegations which I am trying to get looked into? I suspect it will be regarded as a quibble, but there are too many false allegations, and all together, they portray me in a bad light and make it look as though I am dishonest about the main issue. If these were removed there would not be an impression that I either lied, for example, about knowing my mother's name or refusing to give the date of her death or provide a "copy" of her will, all of which portray me as being obstructive and uncooperative when in fact if these were put right then I would appear to have been honest and cooperative.

If you could give me the address of the CPS which would be in charge of papers which the DWP of Derby FIS said they had sent the file to with recommendation to prosecute or, preferably, an email address which is cheap form of communication for me but please, no phone numbers as I don't want to pay out any more money for telephone music.

Topic 3: Laws, statutes, procedures, internal documents and guidance

Highest probability words: person, law, legal, state, require, regulation, use
FREX words: legislate, procedure, legal, consent, regulation, law, breach
Representative FOI requests:
Example 1. MoJ, 16 September 2014, successful
> Can you please provide me with the definition of the word statute according to the law and the definition of the word driver according to the law and also the definition of the word person according to the law?

Example 2. SC, 25 July 2012, overdue

I have a query regarding two pieces of legislation which may be contradictory. I shall use a fictional example: One piece of legislation from 1917 states, "a man MUST wear red socks on Thursdays"; and another piece of legislation from 2002 states "a man must NOT wear red socks on Thursdays". In such an eventuality, what is the recorded procedure of the courts? Which piece of legislation is the higher authority - on the understanding that neither Act has been repealed? Does your recorded procedure give precedence to the earliest Act by date or the latter as it is most recent?

Example 3. DCLG, 10 September 2016, overdue

This is not a request for legal advice; it is a request for clarification of an administrative process. Please confirm that Schedule 2 of the Council Tax (Administration and Enforcement) Regulations 1992 remains a legislative requirement, and any deviation from the specified format, its provisions and requirements, would, upon the issue of such a document, amount to fraud in common law.

Topic 4: Public health, NHS, the safety of NHS premises

Highest probability words: health, NHS, trust, care, safety, hospital, mental
FREX words: Oxleas, captivitydetent, goddington, Dransfield, IAPT, PCT, reconfigure
Representative FOI requests:
Example 1. HSE, 19 September 2014, successful

Are the HSE investigation reports published anywhere? Please, could you supply the HSE investigation reports for the following hospital fires: Royal Marsden Hospital 2 January 2008, University College Hospital London 25 July 2008, Great Ormond Street Hospital 29 September 2008, North London Forensic District general hospital site 15 October 2008, Northwick Park Hospital District general hospital 11 February 2009, and Bath Royal United hospital 22 November 2011. Thanks in advance.

Example 2. DHSC, 29 February 2012, information not held

I would like to know how deaths have there been in the last ten years due to prescribed medication such as psychoactive and psychotropic drugs, Benzodiazepines and Tricyclic given by their GPs and Psychiatrists.

Example 3. DHSC, 14 June 2010, information not held

Is there a link between antipsychotic drugs, and pulmonary embolism and is the drug Melarill still prescribed in the UK?

Topic 5: The NHS management: finances and meeting minutes

Highest probability words: request, response, receive, within, specify, contact, forward
FREX words: request, clarify, unclear, forward, esx, exempt, prefer
Representative FOI requests:
Example 1. DHSC, 29 May 2009, successful

> Regarding Advanced HealthSpace Accounts (the ones where patients can view
> their Summary Care Record online): Please, could you confirm the total number
> of patients who have successfully registered for an Advanced HealthSpace
> Account (and who by definition have an SCR)? Of those, please could you
> confirm the total number of patients who have accessed their SCR via their
> Advanced HealthSpace Account at least once as of 29 May? (Clearly, this will
> apply only to those patients within the Bury, Bradford, Bolton, Dorset, South
> Birmingham and South West Essex PCT areas).

Example 2. DHSC, 4 July 2010, successful

> I would like to make a request under the FOI Act regarding GP Partnership
> Financial Accounts. Please, you could provide the last set of accounts filed with
> NHS Audit (and associated PCT functions) for the partnership of General
> Practitioners at Cavendish Health Centre, 53 New Cavendish Street London
> W1M 7RF, within NHS Westminster, Westminster and Chelsea PCT? This is
> for the GP Practice, Cavendish Health Centre, that has been in existence since
> 1992 as an unincorporated business NL19436845.

Example 3. HSE, 2 November 2016, partially successful

> Please, provide me with the most recent data for UK death rates from
> mesothelioma. I am interested in data that allows me to examine mesothelioma
> death by profession and by region.

Topic 6: Support for people diagnosed with Asperger's syndrome and autism

Highest probability words: autism, library, service, support, friend, organise, charter
FREX words: autism, wessex, autist, vardi, advocacy, portfield, Christchurch
Representative FOI requests:
Example 1. DfE, 15 June 2011, partially successful

> What autism/Asperger/ASC awareness training do Department for Education
> staff receive, particularly decision-makers? What form does this take? Who
> provides the training? How often is this training rolled out? What
> autism/Asperger/ASC awareness training do schools receive, from the Head
> Teacher downwards? How do you ensure schools are sufficiently autism aware
> and make the necessary and appropriate adjustments to properly and effectively
> understand and support such pupils?

Example 2. DHSC, 19 April 2012, not treated as a request under FOIA

> Where and how, currently, can adults and children in Essex obtain a proper
> informed diagnosis of autism/Asperger syndrome, if autism expertise is often

completely lacking in-county, and referrals to the outside of the county are
mainly refused? Please provide FOI contact details or web links for all Essex
PCTs (or whatever the equivalent is now – it's very confusing for the
consumer/patient.) Do you know of any other agencies within Essex or where
Essex people can go, for diagnosis?

Example 3. DWP, 7 March 2012, partially successful

What autism/Asperger syndrome awareness training do DEA's and other Job
Centre Staff currently undergo? Please describe the training and how you
ensure ALL Job Centre staff are autism aware.

Topic 7: Taxes

Highest probability words: tax, custom, payment, revenue, claim, pay, paid
FREX words: overpay, tax, caller, revenue, payment, custom, income
Representative FOI requests:
Example 1. HMRC, 7 December 2012, waiting for a response
Can you please confirm that AAC Capital Nebo Sub LP and AAC Capital Sub
Feeder LIP are on your aggressive tax evader list? They have accounts in
Guernsey.

Example 2. HMRC, 13 August 2012, internal review

In each of the four years ended 5 April 2012, how many and how much Working
Tax Credits were overpaid by HMRC? How much of the overpaid Working Tax
Credit identified at Q1 above was repaid to HMRC? How many and how much
Child Tax Credits were overpaid by HMRC? How much of the overpaid Child
Tax Credit identified at Q3 above was repaid to HMRC?

Example 3. HMRC, 4 September 2012, internal review

From the date HMRC became responsible for the administration of all Tax
Credits to the current date, how much has HMRC overpaid Tax Credits, and
written off the overpayments of Tax Credits?

Topic 8: Ex-military vehicles

Highest probability words: defence, ministry, service, military, vehicle, force, history
FREX words: military, army, rover, cadet, chassis, Land Rover, HMS
Representative FOI requests:
Example 1. MoD, 28 July 2015, successful
I have recently purchased an ex-army Land Rover 110 and an ex-army Harley
Davidson mt350 motorbike. I would like to know the history on these please if
possible. Land Rover XD tum. KJ99AA and Harley Davidson MT350. 74KL58.
Also, would I be able to have a copy of the Army mk3 command 9x9 tent
instructions and the mk3 12x12 tent instructions as well?

Example 2. MoD, 6 August 2015, successful

I own an Ex-Military Reynolds Boughton whose military registration number is 94KJ25. Could you please supply the vehicles Merlin Report which is a record of the vehicle's history when it was in military service?

Example 3. MoD, 7 July 2015, successful

I have recently purchased an ex-military Reynolds Boughton RB44. The military registration was 95KJ91. Would you please be so kind as to supply me with the vehicles Merlin report, as I would like to know its military history?

Topic 9: Local government

Highest probability words: local, council, author, community, govern, grant, county
FREX words: hamlet, bankruptcy, insolvency, Maidstone, council, region, local
Representative FOI requests:
Example 1. DCLG, 4 September 2016, information not held
Please provide, if held, the responses from the following authorities and association to your consultation published 26 March 2015 "Extending the remit of the Local Government Ombudsman to larger parish and town councils": Colney Heath Parish Council, Harpenden Rural Parish Council, Harpenden Town Council, London Colney Parish Council, Redbourn Parish Council, Sandridge Parish Council, St Albans City and District Council, St Michael Parish Council, St Stephen Parish Council, and Wheathampstead Parish Council and St Albans District Association of Local Councils.

Example 2. DCLG, 6 June 2013, partially successful

Please provide copies of the business plans that have been submitted by all local authorities that applied for the efficiency support grant in 2013. The business plans referred to are the business plans that have been submitted in support of the efficiency support grant applications.

Example 3. DCLG, 22 April 2015, request withdrawn by the requester

I would like to see all correspondence to and from Eric Pickles regarding the Khalsa Secondary Academy in Stoke Poges. From January 2015 onwards.

Topic 10: Civil servants: staffing policies

Highest probability words: office, public, cabinet, member, govern, civil, general
FREX words: Anguilla, Maude, poll, cabinet, general', attorney, inquiry
Representative FOI requests:
Example 1. DCMS, 21 July 2012, successful

Can you please provide me with a list of civil servants that have their own offices in your headquarters? For each civil servant that has their own office, I

would like their name and grade. Please, also provide the reason they have their own office. For example, all civil servants at that grade have their own office, their work is of particular sensitivity, or it was negotiated on entry etc.

Example 2. HMT, 21 June 2015, information not held

In February 2015, it was announced that Sir Howard Davies had been appointed to the Board of the Royal Bank of Scotland (to take up his position in June 2015) and to the role of Chairman of the Royal Bank of Scotland (to be taken up on 1 September 2015). Could you please inform of me the date when either the government first approached Sir Howard Davies about these two appointments to the Royal Bank of Scotland or, in the alternative, when Sir Howard Davies first expressed an interest in these positions?

Example 3. CO, 17 January 2016, partially successful

Further to my previous request around Extended Ministerial Offices, please provide further details of the appointments made to the EMO for Cabinet Office ministers, including who was appointed, their specific role and expertise in the department, and their pay grade.

Topic 11: Vaccinations

Highest probability words: research, use, vaccination, trial, effect, study, cause
FREX words: chemic, fluoride, mecf, temperature, HPV, born-digit, abnormal
Representative FOI requests:
Example 1. DHSC, 2 March 2013, partially successful

For each vaccine below, please confirm the brand of vaccine currently used in the UK recommended childhood vaccination schedule and the manufacturer of that vaccine. Please also confirm the duration of effectiveness for each of the vaccines below (i.e. the length of time from vaccination that the patient can expect to be protected from the disease against which they have been vaccinated). If there are no confirmed figures for this duration of effectiveness, please confirm the generally accepted estimates used by the Department of Health and give the sources on which these estimates are based. (Note: A list of vaccine and information requested followed, but I deleted it here for better readability)

Example 2. DHSC, 4 March 2012, successful

With regards to the HPV vaccination programme in the UK, please confirm what research is being undertaken, or planned, to ascertain if the cancer-causing strains of HPV not vaccinated against are increasing to fill the void left by the elimination of HPV16 and HPV18. Please also confirm if all cases of cervical cancer, and other cancers against which the HPV vaccine protects, and all precancerous legions are being analysed to confirm the strain of HPV present. If this is not done at present, can you confirm from what date this will happen, as it is obviously a vital indicator of the effectiveness of the HPV vaccination?

Example 3. DHSC, 29 November 2012, refused

On the NHS Choices website: HPV Vaccination page, it states that: "Research has shown that the HPV vaccine provides effective protection for at least six years after completion of the three-dose course. It is not known how long protection will last beyond this time". I understand that the HPV vaccine is ineffective if the girl/woman has already been exposed to the vaccine-specific strains of HPV, which is the argument for vaccinating at 12 years old. There would, therefore, be no benefit receiving the vaccine or boosters if infection with vaccine-specific strains of HPV has taken place.

I would like confirmation/clarification of the following:

1. When will the girls already vaccinated be told how long the vaccine will protect them for? And how will this be communicated? How will girls know if the vaccine is still offering protection against HPV?

2. Many girls vaccinated in 2008 at the age of 12, will, after six years, still be sexually naive in 2014, at the age of 18. Will these girls be offered antibody tests to confirm they are still protected against vaccine-specific strains of HPV?

3. Will girls/women be tested for infection with HPV type or HPV antibodies prior to receiving a booster, since there is no way of knowing if the vaccine was effective when first given or if the effectiveness has worn off. As all vaccinations carry some risk of serious and less serious adverse effects, if infection with HPV has occurred, vaccinating these women with booster doses will put them at unnecessary risk.

4. Will the girls who received Cervarix first time around, be offered Gardasil if a booster is required and what safety studies been conducted on girls receiving Gardasil after vaccination with Cervarix.

Please supply all copies of all correspondence/memos/reports etc. between Dept of Health and MHRA (or other agencies) regarding the above and confirm the Dept of Health policy regarding the next stage in the HPV vaccination programme to ensure continued protection against HPV. Please supply copies of safety studies/reports to support the safety of girls who may need to receive both Cervarix and Gardasil vaccines.

Topic 12: Procurement of government services

Highest probability words: contract, service, use, system, current, support, procure
FREX words: oracle, procure, contract, hardware, sap, payroll, reprography
Representative FOI requests:
Example 1. RPA, 20 April 2015, successful

I would like to ask: Do you look after your own IT or is it outsourced? How much did you pay in the last financial year for software licenses? How many computers users do you have? When do you need to renew the contract with Microsoft for software licenses? What was the value of your last contract per

year? Do you currently measure software usage versus the number of licenses purchased? If so, what is used for software usage metering? Do you use a software asset management tool? Please also provide details of IT Contracts Managers and any person(s) involved in IT Software procurement.

Example 2. DfID, 2 August 2010, partially successful

Please provide information on your corporate group buildings window-cleaning contract or contracts for in Greater London. What is the (Annual) cost of window cleaning (excl. VAT) for your contract, or contracts in Greater London? When were the corporate window cleaning contracts awarded and to whom in Greater London? When are the next review date(s) in Greater London? Where and what date/month/year will you be advertising your future contracts for window cleaning in Greater London? What will be the type of buildings and how many of them (approx.) are in the window-cleaning contract(s) in Greater London? If your main window cleaning contract is within your accommodation and building services contract, please state and then confirm if you would allow an SME supplier to tender separately for your window cleaning lot in Greater London? Is there a telephone contact number and email of the procurement office/officer that would deal with this in the future?

Example 3. ORR, 10 April 2017, partially successful

What Finance System do you use for financial management and planning? What is the monetary value of the annual support/subscription for your current Finance System? Which vendor implemented your current Finance System? Do you currently have a partner supporting your Finance System? What is the monetary value of the annual support fee for your support partner? When is the contract for the current Finance System expiring? Are you planning to renew your current Finance System or tender for a new system upon expiry?

Topic 13: Charities

Highest probability words: copy, provide, report, document, publish, commission, charity
FREX words: logo, hyperlink, copy, NOMINET, publish, commission, media
Representative FOI requests:
Example 1. CC, 4 April 2013, successful
Please provide electronic copies of governance documents you hold for Sea Shepherd UK (registered charity 1110501).

Example 2. CC, 20 September 2011, partially successful

Please provide me with an electronic copy of all correspondence between the Charity Commission and the Association of Voluntary Organisations in Wrexham (excluding any standard circulars sent to all registered charities by the Commission) since 1 November 2010.

Example 3. DCMS, 17 March 2017, successful

> Please provide the dataset collected by the Libraries Taskforce and referred to
> here: https://librariestaskforce.blog.gov.uk/2016/08/08/library-basic-data-set/.

Topic 14: Job seekers' enquiries

Highest probability words: work, pension, claimant, job, provide, benefit, sanction
FREX words: Jobcentre, JSA, Jobseeker, Jobmatch, Jobseeker', job search, UJM
Representative FOI requests:
Example 1. DWP, 18 November 2013, successful

> Can you please provide all information/guidance/policy about the support
> Jobcentre Plus provides to participants on the Work Programme? Are
> participants still interviewed by personal advisers? Can participants be
> mandated to create a profile/upload a CV on Universal Jobmatch by Jobcentre
> Plus?

Example 2. DWP, 25 January 2016, successful

> Can you please provide me with information showing that the DWP and their
> staff can make it a condition of claiming JSA that the claimant has to provide
> access to their Universal Jobmatch account and that the DWP and their staff
> can make it a condition of claiming JSA that the claimant has to log in to their
> Universal Jobmatch account at the Jobcentre so that the staff can look at the
> claimants account?

Example 3. DWP, 12 November 2017, successful

> Is it mandatory for claimants to use the Universal Credit journal to show
> evidence of job search activities, can I bring in paper copies/print-offs of jobs I
> have applied for and replies I get? Can I upload work search evidence on to my
> laptop and bring that in instead? Is it mandatory for me to upload my CV onto
> to the journal, can I not bring in a paper copy for the Jobcentre to see? Can the
> Jobcentre keep a copy of my CV without my consent?

Topic 15: Disability and training of public servants

Highest probability words: train, course, disable, learn, introduction, guidance, med
FREX words: iwm, medallion, cmep, ebm, pack, med, introduction
Representative FOI requests:
Example 1. DWP, 29 January 2016, refused

> Please provide me with a copy of the following training guidance for Health
> Care Professionals: Handout 14 Justification case scenarios WCA academy,
> Disability analysis for new entrants and introduction to ESA/revised WC,
> ESA/DLA Learning Path Pack 2, and MEDLpetM01 lima revised WCA exam
> technical manual.

Example 2. DWP, 11 January 2016, successful

> Can you please confirm if the following document: "Continuing Medical
> Educational Programme, Chronic Fatigue Syndrome/Myalgic Encephalomyelitis
> (CFS/ME) Guidelines for the Disability Analyst: Version 7 Final: Module 6:
> Reference: MED/S2/CMEP~0017" - is currently in use?

Example 3. DfE, 28 March 2012, successful

> How many people out there have a child/ren with special needs, and or the
> parents have the learning difficulties themselves? What is the standard age for
> diagnosing learning difficulties? Could you define what is classified as a learning
> difficulty, and which disorders are covered? Is there a vast difference in mild,
> moderate and severe learning difficulties?

Topic 16: Budgets and spending

Highest probability words: cost, project, estimate, include, expense, develop, budget
FREX words: cost, incur, expense, estimate, calculate, project, tariff
Representative FOI requests:
Example 1. CO, 17 November 2010, internal review
> Please disclose all your department's internal budgets set for 2010-11, providing
> the name of the budget centre (indicating whether it is an organisational unit or
> ad hoc project) and the budget limit, detailing administration costs and non-
> administration costs; with expenditure analysis for salaries, other current
> administration expenditure, capital additions, and income analysed by
> administration, other current and capital.

Example 2. DWP, 26 October 2017, successful

> I would like to know how much has cost the new campaign featuring celebrities
> and everyday people. Specifically, I'd like to know how much it had cost to hire
> Theo Paphitis and Fred Sirieix. I would also like to know why these two
> celebrities were hired.

Example 3. HMRC, 5 September 2013, refused

> Please, could you provide a summary of the following information;
>
> 1) Cost, from April 2003 to date, of failed or cancelled IT projects.
>
> 2) Cost, from April 2003 to date, of failed or cancelled building projects.
>
> 3) The estimated total cost of any incomplete IT projects to date.

Topic 17: Badger culling/health and safety inspections

Highest probability words: form, area, site, post, test, carry, operate
FREX words: survey, cull, form, fill, post, pilot, badger

Representative FOI requests:

Example 1. HSE, 29 April 2015, successful

> This request concerns the IGAS coal-bed methane pilot site Doe Green in Cheshire. Please, could you answer the following: How many times has the HSE visited this site? Was/is there an independent well inspector associated with this site? Were the HSE aware that this site is unmanned? How many wells have been drilled? What is happening at the site now – are wells being dewatered? are they producing gas? Is this being fed into the national supply?

Example 2. Defra, 5 October 2014, information not held

> With regard to stated Policy of creating a "buffer zone" where badger vaccination will help minimise the effects of perturbation from culling can you please confirm how many badgers were vaccinated as part of Badger Cull Pilots in Year 1, in Somerset and Gloucestershire? A breakdown for each area, please.

Example 3. Defra, 18 December 2016, partially successful

> Please, can you tell me the total number of badgers tested for TB in the 2016 culls and the results of the post-mortems? The total number of badgers tested in 2013, 2014 and 2015 culls and the results? All information from 2013,14,15 and 16 of how badger population was estimated in each cull zone? The total cost so far spent on the badger cull from year one until now (end of cull 2016)? Is this cost paid by the taxpayer or paid by NFU, breakdown, please?

Topic 18: Meetings minutes and official correspondence

Highest probability words: meet, correspond, letter, sent, date, minute, communicate
FREX words: attendees, email, correspond, minutes, meet, discuss, letter
Representative FOI requests:
Example 1. HMT, 14 October 2008, information not held

> I request a copy of all emails, text messages sent from Catherine Macleod, an adviser to the chancellor of the exchequer, to Robert Peston, Business Editor for BBC News, for the time period 1 October 2008 to 12 October 2008.

Example 2. DHSC, 22 December 2015, refused

> I would like to request any documents relating to consultations with Mars UK prior to the establishment of the Food Network HLSG. This may include meeting minutes, agendas, discussion papers and action notes.

Example 3. FCO, 24 March 2009, refused

> I would like to request a list of all meetings between FCO officials and representatives of the Colombian Government in 2006.

Topic 19: Education: qualifications and examinations

Highest probability words: review, level, perform, grade, award, rate, group
FREX words: candid, GCSE, GCSES, perform, IGCS, select, qualify
Representative FOI requests:
Example 1. WO, 31 May 2008, information not held

I would like a breakdown of the total qualifiers 2006/07 for Welsh domiciled students from all HEI's throughout the UK as such: postgraduate research, postgraduate taught, PGCE, first degree, foundation degree, another undergraduate HE (all broken down by gender: male/female).

Example 2. OFQUAL, 11 September 2012, partially successful

Please, supply documents showing the percentage deviation in GCSE English results from the expected outcome, for each of the English and Welsh Exam Boards for June 2011, February 2012 and June 2012. For clarification, we know that EDEXCEL's results deviated 6-7% above OFQUAL's expected outcome for this exam. We do not know about the other five Boards.

Example 3. DfE, 9 April 2016, successful

Can you provide the percentage of 5+ GCSEs A*-C grades and 5+ GCSEs A*-C grades including English and Maths achieved in 2015 for each of the London boroughs for the following groups: White British, Indian, Pakistani and Bangladeshi? In addition to this, can you provide the same percentages achieved for those eligible for Free School Meals for the ethnic groups mentioned?

Topic 20: Environment: animal welfare and food standards

Highest probability words: food, environ, affair, rural, agency, water, waste
FREX words: dog, meat, slaughter, countryside, fiduciary, halal, BTB
Representative FOI requests:
Example 1. Defra, 6 March 2017, successful

I am writing to request a list of Fox Hunts in the UK, who have had hunt dogs and horses with tuberculosis. Or any other hunt.

Example 2. Defra, 17 March 2013, successful

What protection in law do swans have from being attacked by uncontrolled/dangerous dogs?

Example 3. FSA, 16 March 2013, information not held

I note that Frank Mallon of the Irish meat production company Liffey Meats was found guilty and fined for 25 charges relating to illegal growth promoters in cattle in back in 1996. (Liffey Meats is the company that has been supplying beef products that contain horse meat to Tesco and Ireland.) Does the FSA know how many kilos or tons of meat produced using illegal growth promoters by this company, or other companies involved in the use of illegal growth

promoters are estimated to have been sold for consumption in the UK before the problem was discovered? Could the FSA explain why these growth promoters as used by Liffey Meats are illegal, what specifically they were, and what the potential effects on human health are? Could the FSA tell us how many other meat producers that have been found using illegal growth promoters in their cattle are still operating and supplying meat to the UK?

Topic 21: Civil servants and independent service providers: selection procedures and remuneration

Highest probability words: scheme, manage, programme, provide, work, west, placement
FREX words: CPA, ingeus, seetec, subcontractor, scheme, MWA, Yorkshire
Representative FOI requests:
Example 1. DfID, 22 December 2016, refused

> Could you please provide the minimum and maximum pay scales for the four main analytical professions broken down by profession, grade, and location, working within your organisation: Economists (GES), Social Researchers (GSR), Statisticians (GSS), Operational Researchers (GORS). In addition, could you give the median salary broken down by analytical profession, grade, and location?

Example 2. DCLG, 20 December 2016, successful

> Could you please provide the minimum and maximum pay scales for the four main analytical professions broken down by profession, grade, and location, working within your organisation: Economists (GES), Social Researchers (GSR), Statisticians (GSS), Operational Researchers (GORS). In addition, could you give the median salary broken down by analytical profession, grade, and location?

Example 3. DWP, 16 November 2015, successful

> What are the names of all of the DWP's Providers Subcontractors (Subs) for Community Work Placements across the UK? Please provide this Subs information divided by contract package area and name of the DWP prime provider for each of the Subs.

Topic 22: Courts

Highest probability words: court, justice, ministry, case, judge, prison, order
FREX words: HMP, judiciary, CPR, case man, IAO, IPP, court'
Representative FOI requests:
Example 1. MoJ, 11 February 2013, information not held

> District Judge Simmonds is operating as a family division district judge in the PRFD, High Holborn, London and does not appear on the

http://www.judiciary.gov.uk list of judges. Could you please either rectify the judiciary website or explain why an unregistered judge is practising in court?

Example 2. MoJ, 26 May 2017, successful

Can you please confirm how many individuals have been convicted of perjury in the following Courts: The County Court, The Family Court, The Magistrates' Court, The Crown Court, The High Court, The Court of Appeal, The UK Supreme Court, The Court of Justice of the European Union?

Example 3. MoJ, 19 September 2012, overdue

What are the procedural rules for granting Interim Charging Orders? When Interim Charging Orders are issued by a County Court, is a court hearing required and are both parties informed by the court or does an ex parte hearing take place? Does the Court issue and serve the order or is it sent to the complainant to serve on the other party? What are the criteria for being classed as served with a court order?

Topic 23: Information of various interest (austerity measures)

Highest probability words: one, take, need, ask, now, find, say
FREX words: say, yet, take, one, don't, now, put
Representative FOI requests:
Example 1. DWP, 29 July 2017, waiting for a response

Why does my friend born in March 1953 get her state pension now, so much earlier than me (13 October 1953) who on earth calculated this! When I ask this question, all I get is the standard reply. Firstly, I understand why this is happening, so I don't need that standard reply, it does make sense, but what I don't understand is the total unfairness in the way it is happening and why so little notice. Luckily, I have a job, it isn't exactly the job I would have done over the years, but I resent someone six months older than me receiving their pension years before me when I've probably put the same in the pot. Nobody is answering the unfairness of how it has been worked out!

Example 2. DWP, 30 May 2013, successful

What action is the DWP now going to take toward Atos for asking silly or insensitive questions in their assessments? These questions indicate insufficient training has been given and no qualified doctor would ask such things (unless there is an agenda at Atos?) Questions such as: How long have you had Asperger's Syndrome? Do you expect your arm to grow back soon? Why aren't you dead yet? How long have you had spina bifida?

Example 3. DWP, 12 August 2015, information not held

My husband was diagnosed with Parkinson's when he was 64 but was not told about the mobility car scheme until he was 65 by which time, he was told it was too late! Surely this is unfair as patients get older, their need for suitable transport increases. My husband is now 76, and I am 68 and struggle with his

mobility. Therefore, he stays most of the time indoors as there is no other option other than taxis, public transport is out of the question as he needs a wheelchair.

Topic 24: National debt/trade unions

Highest probability words: fund, money, treasury, govern, bank, majesty', account
FREX words: borrow, GMP, treasury, repay, money, majesty', loan
Representative FOI requests:
Example 1. HMT, 14 December 2014, partially successful

How much money has been given from the Westminster Government to the Scottish Government between the years 2010 (April) to 2014 (April), can I have it broken down from each year, please? Since the introduction of austerity measures under the Conservative/Liberal Democrats Government, how much has been saved from allocation of Westminster Government Funding to the Scottish Government? How much of the money that has been cut goes to pay for the National Debt?

Example 2. HMT, 4 September 2010, successful

Under the terms of the Freedom of Information Act, I wish to receive the following information from your department: the amount of money paid back to the government, during 2009-10, by each individual bank that received government funding during the current economic downturn, and the repayment plans for each individual bank that received government funding during the current economic downturn for future years.

Example 3. HMT, 10 February 2011, successful

Under the terms of the Freedom of Information Act, I wish to make a request for the following information from your department: Is your department currently paying any money to any trade union? Has your department paid any money to trade unions since 1997? If so, in what form has the department funded, or is funding, trade unions? Which trade unions have received or are receiving funding from the department? How much money has been paid to each of these trade unions and when was the money paid? How much money do you expect to give each trade union this year? Do you plan to continue funding trade unions at the same level?

Topic 25: Crime statistics, warrants, prisoners

Highest probability words: power, search, police, warrant, SAI, RSPB, complaint
FREX words: SAI, RSPB, baba, investigatory, warrant, intercept, power
Representative FOI requests:
Example 1. DHSC, 3 December 2012, partially successful

Please, can you clarify the law on UK Prisoners wishing to become living "direct" (to the known donor) or "non-direct" (to unknown donor) kidney

donors? Please, can you also explain what effect being in Prison may have on a kidney patient's ranking on the organ transplant waiting list? I am looking for any proposals or notable documentation that may shed light on Prisoner's rights to donate and receive organs whilst incarcerated.

Example 2. MoJ, 2 April 2012, overdue

I have some questions regarding Warrants, both civil and criminal. Do Warrants (civil or criminal) need to be signed by a Judge or Magistrate according to Legislation or Case Law? Do Warrants (civil or criminal) need to be dated according to Legislation or Case Law? Do Warrants (civil or criminal) have to have the name of the court that issued them displayed on the document according to Legislation or Case Law? By Warrants I mean Arrest Warrants, Distress Warrants, Warrants of Possession, Warrant of Restitution, Warrants of Execution, Search Warrants, Bench Warrants and any other type of warrants that I may have missed from my list.

Example 3. MoJ, 1 April 2010, information not held

Will you please advise me of the number of knives that have been used/seized in recorded crimes or recorded incidents in the latest five years for each Police authority in the UK and NI, where you have statistics? Will you please advise me of the % of the total for each authority for each of the following categories: legally held knives – non-locking folding knives less than a 3-inch blade, kitchen knives, home-made weapons/knives, tools or other bladed instruments, and lockable folding knives and sheath knives?

Topic 26: Land and property ownership

Highest probability words: company, register, house, land, private, property, limit
FREX words: registry, duchi, freehold, landlord, tenant, own, abv
Representative FOI requests:
Example 1. MoD, 19 February 2009, successful
Please, list properties owned by the MoD in Cornwall. Also, please, list properties where the MoD has an interest in Cornwall. Also, can you list all the properties that the MoD has disposed of in Cornwall for each of the last ten years? Also, can you list all of the properties that MoD is about to dispose of in Cornwall over the two years?

Example 2. FCO, 9 December 2013, successful

Could you please provide information on the following? The number of residential properties in Belgium that are owned and rented by the FCO. The total annual spend on rent, rental services and/ or maintenance for residential properties in Belgium that are owned and rented by the FCO. The total number of FCO staff currently occupying residential properties in Belgium that are owned and rented by the FCO. The total number of other residents in residential properties in Belgium that are owned and rented by the FCO. The average number of rooms in residential properties in Belgium that are owned and rented by the FCO. The average number of bedrooms in residential

properties in Belgium that are owned and rented by the FCO. The number of residential properties in Belgium that are owned and rented by the FCO that are currently unoccupied.

Example 3. HMLR, 6 July 2016, overdue

Please, could you inform me why the sale price information for the following property is not available on your website: Stables and Land adjoining Forest Mead and Butterwood, Linchmere, Haslemere, GU27 3NE? I believe this property was sold in October or November 2015?

Topic 27: Prosecution and convictions

Highest probability words: prosecute, children, police, service, investigate, child, case
FREX words: convict, prosecutor, guilty, CRB, arrest, prosecute, assault
Representative FOI requests:
Example 1. CPO, 27 January 2011, partially successful

For the years 2008, 2009 and 2010 in relation to cases going to trial, please could you provide year by year breakdown of all convictions there have been under the following Acts and what the sentence was: Children and Young Persons Act 1933, sect 1; Infanticide Act 1938; Indecency with Children Act 1960; Sexual Offences Act 2003; Protection of Children Act 1978, sect 1; Child Abduction Act 1984; Domestic Violence Crime and Victims Act 2004, sect 5; Mental capacity Act 2005, sect 44? Please, could you also indicate where there was a "not guilty" verdict (no conviction)? For the same period, if possible, please, could you provide the number of cases which did not go to trial?

Example 2. CPO, 11 December 2012, information not held

I would like to know how many people have been charged with offences related to female genital mutilation.

Example 3. CPO, 29 October 2011, successful

Could you please provide your current guidelines either relied upon or given to local police forces when dealing with offenders suspected of theft offences concerning thefts from their employers? I am interested to know who decides if an offender should be offered a caution for a theft offence and on what basis and guidelines would a decision to caution be made?

Topic 28: Spending

Highest probability words: year, total, staff, provide, figure, statist, per
FREX words: figure, total, statist, spent, breakdown, year, staff
Representative FOI requests:
Example 1. Defra, 3 August 2009, successful

I am interested in obtaining information concerning your procurement of printing consumables, specifically inkjet cartridges and laser toner cartridges.

Can you please provide me with the following: your total spent on inkjet cartridges during the previous fiscal year; your total spent on laser toner cartridges during the previous fiscal year; your top 3 suppliers of inkjet cartridges and their % share of your total inkjet cartridge spend for the previous fiscal year; your top 3 suppliers of laser toner cartridges and their % share of your total laser toner cartridge spend for the previous fiscal year; the percentage of inkjet cartridges (in value and/or quantity) procured during the previous fiscal year that was "re-filled" or "re-conditioned" as opposed to brand new; the percentage of laser toner cartridges (in value and/or quantity) procured during the previous fiscal year that was "re-filled" or "re-conditioned" as opposed to brand new; any procurement guidelines and/or contracts pertaining to the purchase of printing consumables; and any documentation, addressing the procurement of "re-conditioned" or "re-filled" printing consumables.

Example 2. HMT, 22 October 2013, partially successful

Please, could I have details of the annual amount spent on taxis for HMT Ministers for each year 2013- 2010. If possible, please could you break the figures down as follows: a total spent outside of London, and a total spent in London, within Westminster.

Example 3. HMRC, 21 January 2016, partially successful

Please, provide the following information for the last five years, together with projected expenditure to 2019-20 for which figures are available: total Travel and Subsistence expenditure for HMRC and as a percentage of total expenditure: UK and by region, total Travel and Subsistence expenditure for HMRC by employee grade: UK and by region, and highest and lowest single claim for Travel and Subsistence: UK, region and grade of employee.

Topic 29: Citizenship enquiries/Foreign relations

Highest probability words: British, foreign, nation, country, unit, commonwealth, right

FREX words: foreign, commonwealth, embassy, EEA, republic, Nigeria, Bahrain

Representative FOI requests:

Example 1. FCO, 12 November 2010, successful

Please provide the text of all bilateral treaties between the UK and the Republic of France agreed during the year 2010.

Example 2. FCO, 20 July 2016, refused

I would like a copy of Telegram 1256, titled "Iraq: Dividing the Spoils", which was sent to the FCO on 1 October 2002 by the UK Ambassador to the United States, Sir Christopher Meyer.

Example 3. FCO, 20 April 2014, information not held

I was born in Jamaica in 1979. I came here in 1991 at the age of 11, am I a Commonwealth citizen?

Topic 30: Statistics

Highest probability words: many, time, month, tell, since, period, made
FREX words: many, month, tier, age, tell, period, compensate
Representative FOI requests:
Example 1. DWP, 29 September 2016, refused

> How many people die before reaching retirement age? How many people die
> before claiming their state pension? How many of these are men and how many
> are women? What percentage of people died before claiming the state pension in
> 2015?

Example 2. DWP, 6 October 2015, information not held

> How many of the department's staff contact either the civil service whistle-
> blowers hotline or the civil service hotline in 2013, 2014, year to date? For the
> same periods by hotline how many of the complaints were found to be fully
> upheld, partially upheld or rejected?

Example 3. CO, 18 July 2016, successful

> Please, can you tell me how many staff you have, at the time of writing, on long
> term sick due to stress?

Topic 31: List of schools

Highest probability words: school, education, name, provide, England, contact,
primary
FREX words: secondary, headteacher, nursery, lea, mis, teacher, contacts head
Representative FOI requests:
Example 1. DfE, 12 September 2017, successful

> Please, can you send me an excel list of all UK schools, infants, juniors,
> secondary, nursery, and academies with their name, email address, name of the
> headteacher and deputy's head, headteacher/deputy's head and office email
> address, and phase of education, for the 2017/2018 academic year?

Example 2. DfE, 10 December 2012, successful

> Can you, please, provide a list of Special Educational Needs Schools within
> Nottingham, Nottinghamshire, Derby and Derbyshire, including mainstream
> schools with SEN provisions or PRU provision?

Example 3. DfE, 22 June 2017, successful

> I am looking to acquire an up-to-date datasheet (Excel) list of all secondary
> schools in Greater London and Kent with headteacher's names, email address
> and telephone contact details.

Topic 32: Work capability assessments

Highest probability words: assess, work, claimant, decision, ESA, ATOS, medic
FREX words: ATOS, PIP, WCAS, reconsider, ESA, DMS, assess
Representative FOI requests:
Example 1. DWP, 2 July 2014, successful

> I would like to make an FOI request for the following information: What is the
> number of claimants that have directly migrated from incapacity benefit and or
> income support or any other benefit to ESA support group WITHOUT a face to
> face assessment conducted by Atos being deemed necessary? How many
> claimants have directly migrated to ESA support group without a face to face
> assessment where MENTAL ILLNESS is the main factor for the decision that a
> face to face assessment is unnecessary? What are the diagnosed conditions of
> the claimants who were granted direct migration without a face to face
> assessment to ESA support group on mental health grounds? Please, provide a
> break-down percentages of the mental health conditions per claimant accepted
> by the DWP as a legitimate reason (good reason) for granting benefit without a
> face to face assessment?

Example 2. DWP, 14 April 2013, partially successful

> When a person appeals the decision of the DWP, for DLA and ESA which is
> based on an Atos work capability assessment, who is responsible for obtaining
> medical evidence from their GP; the DWP or the person who is making the
> appeal? Exactly how much experience does the Atos HCP need to have and
> exactly what training do they undertake before s/he is qualified to assess a
> person with Fibromyalgia or any other specific condition?

Example 3. DWP, 4 January 2015, partially successful

> What specific training do ATOS and other ESA/PIP assessors undertake in
> relation to a ruptured Aortic Abdominal Aneurysm?

Topic 33: Ex-military vehicles

Highest probability words: inform, follow, freedom, provide, relate, record, supply
FREX words: freedom, inform, VRN registration, held, relate, follow, hold
Representative FOI requests:
Example 1. MoD, 9 September 2016, successful

> I would like to make a freedom of information request for any information held
> in relation to an ex-MoD vehicle with the following VRN/registration:
> HH94AA.

Example 2. MoD, 27 March 2017, successful

> I would like to make a freedom of information request for any information held
> in relation to ex-MoD vehicles with the following VRN/registrations: 29KK04,
> 29KK07, 29KK12, 29KK14, 29KK17, 29KK20, 29KK21, 29KK23, 29KK24,
> 29KK25.

Example 3. MoD, 20 July 2017, successful

I would like to make a freedom of information request for any information held in relation to ex-MoD vehicles with the following VRN/registrations: 34KK23.

Detailed exploration of topics in the model with 28 topics (successful and rejected requests included)

Model 1: Topic prevalence = α + βdisplay_status + βgovBrown + βgovCameron+ ε

Topic 1: List of schools

Highest probability words: school, education, name, provide, format, primary, pupil
FREX words: headteacher, pupil, school, mis, education, spreadsheet, secondary
Representative FOI requests:
Example 1. DfE, 9 October 2017, successful

> Please, ca you send me a excel list of all UK schools, with their name, email address, address, other contact details, website, name of the head teacher, head teacher email address, and phase of education, preferably in a spread sheet format using Microsoft Excel to my email address.

Example 2. DfE, 7 September 2017, successful

> Please, can you send me an excel list of all UK schools, with their name, email address, name of the head teacher, headteacher email address, and phase of education, for the 2017/2018 academic year?

Example 3. DfE, 5 December 2012, successful

> Under the freedom of information act could you please send me a list of all primary schools in England with the addresses (including full post codes), DFES ID's, contact telephone numbers, name of head teacher and an email address for each school. If this can be provided in excel format that would be very much appreciated.

Topic 2: Policies, analyses and reports

Highest probability words: report, review, publish, see, recent, page, research
FREX words: impact, target, publish, audit, page, report, review
Representative FOI requests:
Example 1. DfE, 20 April 2014, refused

> I am requesting the "leaked" DfE report mentioned in this article: http://www.telegraph.co.uk/education/educationnews/10775755/State-schools-isolate-non-Muslims.html. The report focuses on Park View School and its sister schools, Golden Hillock and Nansen. The report was compiled in 2014. If multiple versions of the same document exist, please include them.

Example 2. CO, 25 September 2017, refused

> Has any research or investigation been commissioned by the Cabinet Office or other major departments of government into the impact on living standards (as measured by the Retail or Consumer Price Indices or comparable indices) as a

result of the UK leaving the European Union? If such a commission has been undertaken what was the conclusion of that assessment including any estimate into the likely increase in the Retail or Consumer Price Index or comparable indices in the run-up to the UK leaving the European Union and the first 5 years after the UK's departure.

Example 3. ONS, 18 June 2015, successful

There have been two main changes affecting the way the economic activity of disabled people is calculated: one in 2010, with the rewording of the introduction to the section of the survey covering disabilities which lead to an increase in the number of economically active people with a disability of around 300,000 and another one in 2013. The questions pre and post-2013 are indicated in the ONS statistics release. I would like to know what the difference was between the two introductions pre and post-2010. Was there also a change either in the questions asked, or in criteria? Was there at any other times, a rewording of the introduction to the section of the survey covering disabilities, of the questions asked or of the criteria? If yes, which one and when? What was the net effect of the changes in 2013? Is there a document which explains in detail these changes?

Topic 3: Right to information

Highest probability words: request, limit, date, section, reason, record, time
FREX: request, receipt, ICO, FOIA, limit, amend
Representative FOI requests:
Example 1. HSE, 21 March 2014, successful
I should like to request a list of all registered firework storage sites throughout England, Wales and Scotland, that have a current licence with yourselves.

Example 2. DCMS, 6 September 2016, refused

Please, could you disclose to me the contents of the folder "POL/034/002/006/4 Phorm Briefings/Submissions"? I understand from an earlier request (handled by Brigitte Charles) that the folder and files currently sit within the Ministry of Justice filing system, TRIM (*).

Example 3. DHSC, 21 September 2009, successful

With reference to the case that resulted in the Information Commissioner's recent decision notice FS50195059 please could you provide me with copies of all the information you supplied to the applicant prior to decision notice being issued. Please supply me with a copy of the letter you have now been told to release as part of the Information Commissioner's decision notice FS50195059.

Topic 4: Statistics

Highest probability words: many, year, month, figure, tell, since, time
FREX words: many, figure, IGCS, past, statistics, month, year
Representative FOI requests:
Example 1. MoJ, 30 July 2010, refused

> Most public bodies list the freedom of information requests on their website
> with the status of what was released. You do not supply this data on your
> website under FOI. Please give me an excel (NOT as a pdf) list of all FOI
> requests you have had since it became law. For each FOI request, please list the
> date of request, date of reply, reference number/ID, subject/title of the request.
> Each request on the list should include the subject/title of the request, status
> [whether the information was fully released, partially released or withheld].

Example 2. MoD, 29 June 2009, refused

> What is the number of visitors (persons not on active duty in the theatre) to
> MoD stations in Iraq for every month since March 2003? How many man-hours,
> broken down by month for every month since March 2003, were spent hosting
> visitors?

Example 3. ONS, 22 June 2012, successful

> Could you tell me the current size of the 12-17-year-old population in the UK
> and projections for the next five years? And similarly, for the 18-25 and 26-30
> age groups for the same period. For the 18-25-year-old group, could you also
> give me a breakdown of how many are in education, how many are in work, and
> how many are unemployed?

Topic 5: Ex-military vehicles I

Highest probability words: inform, provide, follow, freedom, relate, held, make
FREX: VRN registration, inform, freedom, held, trade, union, hold
Representative FOI requests:
Example 1. MoD, 10 April 2017, successful

> I would like to make a freedom of information request for any information held
> relating to ex-MOD vehicles with the following VRN/registrations: 35KK04,
> 35KK06, 35KK09, 35KK12, 35KK13, 35KK14, 35KK20, 35KK21, 35KK24,
> 35KK26, 35KK27,35KK30, 35KK35, 35KK39, 35KK50, 31KK73, 26KK21,
> 34KK15, 63KC23.

Example 2. MoD, 27 March 2017, successful

> I would like to make a freedom of information request for any information held
> relating to ex-MOD vehicles with the following VRN/registrations: 29KK04,
> 29KK07, 29KK12, 29KK14, 29KK17, 29KK20, 29KK21, 29KK23,29KK24,
> 29KK25.

Example 3. MoD, 15 February 2017, successful

> I would like to make a freedom of information request for any information held relating to ex-MOD vehicle with the following VRN/registration: 29KK57.

Topic 6: Transport

Highest probability words: transport, road, approve, authorise, rail, park, train
FREX: rail, traffic, railway, fare, franchise, highway, pedestrian
Representative FOI requests:
Example 1. DfT, 14 April 2011, successful

> When are warning signs illuminated and/or displayed on the motorways to warn drivers of adverse weather conditions, in particular, torrential rain?

Example 2. DfT, 16 November 2009, refused

> Where are the areas where passenger trains may travel at 125mph on the Great Western rail franchise? What plans are there to provide or increase the lengths of 125mph railway line on the Great Western "Berks and Hants direct line" between Reading, Westbury and Taunton?

Example 3. DfT, 22 December 2015, successful

> Has the Department for Transport ever been approached by the Office of Rail and Road to produce an official document specifically giving Cambrian Heritage Railways Limited (as created in 2009) permission to operate a passenger railway service in Oswestry in the light of the charities' current application for a transfer Order?

Topic 7: Taxes

Highest probability words: tax, custom, revenue, payment, rate, pay, income
FREX words: loan, repay, borrow, tax, earn, rate, revenue
Representative FOI requests:
Example 1. HMT, 13 May 2011, successful

> Can you please provide me with the historical figures for the top rate of income tax, the basic rate of income tax, rate of corporation tax and rate of capital gains tax for each year since 1945?

Example 2. HMRC, 3 June 2014, successful

> HMRC has hired at least 12 private debt collectors to pursue tax credits overpayments. 1. Have you sold any tax credit debt to any debt collector? 2. How much do you pay debt collectors to arrange collection of tax credits overpayments?

Example 3. HMRC, 8 October 2013, refused

In October 2006 Farepak Food and Gifts Limited went into administration, and European Home Retail Plc also went into administrative receivership. HMRC was a creditor for both of these companies mentioned above. Therefore, I'd like to request the following: 1) How much was FFG Ltd owe you as a creditor for outstanding taxes? 2) How much was EHR Plc owe you as a creditor for outstanding taxes? 3) How much was FFG Ltd owe you as a creditor for employee national insurance contributions? 4) How much was EHR Plc owe you as a creditor for employee national insurance contributions?

Topic 8: Prosecutions and convictions

Highest probability words: justice, prosecute, ministry, case, service, police, child
FREX words: HMP, custody, prosecute, prison, child, shrive, crime
Representative FOI requests:
Example 1. MoD, 19 January 2017, successful
Please, could you supply the number of current serving Soldiers within the Army Medical Services with an Unspent/ Spent Criminal Conviction for any offences within the Offences against the Person Act 1861 including offences for Section 18 Assault Grievous Bodily Harm (GBH) Wounding (with intent), Section 20 Assault Grievous Bodily Harm (GBH) Wounding (without intent), Section 47 Assault Actual Bodily Harm (ABH), Section 39 Assault Common Assault?

Example 2. MoJ, 2 February 2015, successful

Can you please provide a list of all prisoners with the sentence "life sentence without parole" and details for them like name, age, what crime they have committed, date of conviction, and in what prison are located these prisoners?

Example 3. CPS, 26 October 2015, successful

In cases of nudity, and naked rambling, can you send the guidelines that the CPS uses to decide whether it is in the public interest to prosecute. In the cases of Stephen Gough since 2000, how many times has CPS chosen to decide to prosecute, and what in each case where the factors taken into account? Is the cost of imprisonment taken into account and if so, how much is the approximate cost of years imprisonment? Does the CPS have to take a decision to prosecute when a person is under an ASBO?

Topic 9: Health examinations/vaccinations

Highest probability words: test, record, use, build, transfer, facility, vaccine
FREX words: vaccine, born-digit, syndrome, HPV, brain, fatigue, abnormality,
Representative FOI requests:

Example 1. DWP, 9 November 2014, successful

Can you please confirm if the following document: "Continuing Medical Educational Programme, Chronic Fatigue Syndrome/Myalgia Encephalomyelitis (CFS/ME) Guidelines for the Disability Analyst: Version 7 Final: Module 6: Reference: MED/S2/CMEP~0017" - is currently in use?

Example 2. MoD, 29 January 2010, successful

Following reports of animal experimentation at the British government's Porton Down research facility, I would like to request that you confirm whether or not any primates (of any description) have been subject to animal testing/experimentation over the last ten years. If so, how many were subject to this treatment, and what exactly was the nature of this testing?

Example 3. FSA, 7 December 2008, successful

It has recently been reported that Irish pork exceeds the "safe" level of dioxin contamination. I would like to know how the "safe" levels for this, and other substances are set - in particular, whether "safe" levels are based on the expectation of the impact of a single dose, lifetime dose or some other basis.

Topic 10: Procurement and contracts (IT systems)

Highest probability words: contract, use, service, current, manage, system, framework
FREX words: procure, contract, oracle, framework, renew, photocopiers MFD, reprography
Representative FOI requests:
Example 1. DWP, 6 August 2015, refused

Do you currently make use of any Oracle products? If so, which? What is the quantity of licensed users split by each Oracle product used by your organisation? What is the value of your current Oracle contracts, split by a product? Do you outsource your Oracle projects, services or maintenance to an external supplier? If so, whom? When does that contract expire? In 2013, the Cabinet Office appointed an Open Standards Board to drive the use of Open Source across Government. Which Open Source databases have you deployed since 2013 and have firm plans to deploy within the next 18 months as alternatives to Oracle products?

Example 2. DCMS, 25 April 2017, successful

What Finance system do you use for Finance management? What system do you use for Planning and Budgeting? Which vendor/System Integrator (SI) implemented your current Finance System? Who is your current Finance system support partner? When is the contract for the current Finance System expiring? Are you planning to renew your current finance System or tender for a new system upon expiry? If tender, when are you planning to publish for a tender? Which framework do you use for procurement? What is your budget for IT spend for 2017/2018? How much do you currently spend on IT systems? Are you planning to join a Shared Service in the next 1 - 2 years?

Example 3. BEIS, 1 April 2015, successful

Do you look after your own IT or is it outsourced? How much did you pay in last financial year for software licenses? How many computers users do you have? When do you need to renew the contract with Microsoft for software licenses? What was the value of your last contract per year? Do you currently measure software usage versus the number of licenses purchased? If so, what is used for software usage metering? Do you use a software asset management tool? Please also provide details of IT Contracts Managers and any person(s) involved in IT Software procurement.

Topic 11: DWP's Work Programme

Highest probability words: work, scheme, programme, ltd, west, pension, placement
FREX words: Yorkshire, CPA, Avanta, Ingeus, placement, north, Seetec
Representative FOI requests:
Example 1. DWP, 29 August 2013, refused

Since the start of the work programme I would like a record of all complaints made about the work programme and also what the complaints were about, but not who made the complaint about Avanta, Fifth Floor, Beckwith House, Wellington Road, Stockport, Cheshire, SK4 1AF.

Example 2. DWP, 5 September 2013, refused

Please, can you list the organisations that have provided placements on the Day One Support for Young People trailblazer through Maximus Employment and Training Ltd and Careers Development Group since the contract began?

Example 3. DWP, 25 October 2012, refused

How many Mandatory Work activity placements have started since 1 April 2012 in CPA 7 North West, and what are the names of the organisations that delivered the 4-week full-time unpaid community benefit work placements?

Topic 12: Law enforcement

Highest probability words: public, right, interest, state, legal, legislate, protect
FREX: legislate, ensure, public, breach, right, protect, party
Representative FOI requests:
Example 1. DWP, 18 April 2012, successful

Today, Jack Straw MP acknowledged that (some) senior Government officials and civil servants take steps to avoid potential repercussions from the FOI legislation deliberately. Specifically, he said that minutes were not taken of some high-level meetings in order not to leave a paper trail while other important decisions would be made by text message, reducing accountability rather than increasing it. He said that senior civil servants feared horrific detail from their notebooks about their ministers' streams of consciousness would be made public under the transparency law, which he passed as Home Secretary in 2000. What

information do you have that clarifies DWP's attitude to such practices; shows that it is clear to its staff that such practices are wholly unacceptable and how this message is regularly reinforced and confirms the disciplinary steps that would be taken for infringement? How many cases have arisen over the past five years where employees have been found to be circumventing FOI legislation to avoid the risk of disclosure?

Example 2. MoJ, 1 September 2009, refused

I understand that the Lord Chancellor is responsible for the Data Protection Act and the FOI act, and also the ICO. How does the Lord Chancellor ensure that these acts are enacted impartially and fairly within other departments who are within his control? Should there be allegations that some of his departments are neither meeting the requirements or spirit of the acts and such departments are not being properly disciplined, by the ICO, how would such allegations be dealt with impartially and fairly?

Example 3. HMT, 25 September 2012, successful

I am aware of a public body that has awarded over £6m contract to a number of suppliers without conducting any proper tendering. My questions are: What consequences a public body will face if, without any justification, it ignores the HMT's guidelines regarding public procurements? If the above public body is not willing to take any action against those responsible, which government body will investigate the matter? Should the police be informed? Are the rules the same for Scotland and England?

Topic 13: Ex-military vehicles II

Highest probability words: ministry, defence, vehicle, service, military, history, registry
FREX words: rover, cadet, veteran, medal, chassis, trailer, Land Rover
Representative FOI requests:
Example 1. MoD, 17 October 2016, successful
I recently purchased a former Army Land Rover Defender "Snatch", military registration 30 KK 45 and would like to know its military history. Please, could you email me its Merlin Card details?

Example 2. MoD, 25 October 2014, successful

Do you have any history of the deployment of my Ex Military Land Rover Defender (Truck Uty Med FFR Plain Hardtop 12/24 4x4 Land Rover Defender 110 (DSL)) Military Registration 59KK65 Chassis Number SALLDHAC5JA917505?

Example 3. MoD, 5 October 2013, successful

Hi, I would be most grateful if you could send me a copy of the Merlin report for my Ex Military Land Rover its Military registration number is 37KL01. This

will be very helpful so that when I take my Land Rover to events during the year, I can display it in its correct markings.

Topic 14: Education

Highest probability words: training, name, course, academy, study, qualify, advertise
FREX words: Gove, NOMINET, campaign, sponsor, advertise, exam, course
Representative FOI requests:
Example 1. DfE, 12 November 2014, successful

Please advise on what happens with regards to a teacher reference number allocated to a trainee teacher on a course of initial teacher training leading to QTS when that trainee withdraws from or fails the course of ITT and subsequently enrols on a new course of ITT leading to the QTS: do they (a) retain the teacher reference number allocated to them through their original course; or (b) get allocated a new teacher reference number through their new course?

Example 2. BEIS, 1 April 2009, successful

Can you tell me the number of job vacancies currently advertised within your department on this day broken down into those advertised internally and those advertised externally?

Example 3. DfE, 28 July 2011, successful

Your website states that on 30 November 2010, Michael Gove accompanied Rebekah Brooks (News International), Will Lewis (News International), James Harding (The Times) on a visit to an Academy School. Under the Freedom of Information Act, I would like to request the following information: The name (or names) and location (or locations) of the Academy school (or schools) visited on 30 November 2010 by Michael Gove, Rebekah Brooks, Will Lewis, and James Harding.

Topic 15: Various areas of interest

Highest probability words: one, time, now, need, say, take, want
FREX words: scanner, seem, think, water, don't, something, vet
Representative FOI requests:
Example 1. Defra, 10 December 2008, successful

Could you please protect the freshwater eel under the countryside act as soon as possible because they are endangered, and in need of some help even their sad position on the endangered list they are still being fished for in their thousands, and it must stop.

Example 2. DfT, 20 April 2012, successful

> I will be going on holiday from Manchester Airport in June. I have had both my
> hips replaced and am always selected to go through the body scanner because of
> this. I feel very uncomfortable when It's a male reading the scan. Can I request
> a female to read my scan? Or better still, can I request not to go through the
> body scanner? I do understand the reasons for going through, but it spoils my
> holiday thinking about it.

Example 3. DfT, 26 September 2016, successful

> I am writing on behalf of Stop Killing Cyclists to find out what amount of time
> and money was spent on the THINK! Cycle safety advert which shows an HGV
> lorry coming up from behind and overtaking a cyclist in a 20mph zone and does
> what is known as a left hook and crushes the cyclist. This advert has been put
> up on twitter with "Cyclists: Don't get caught between a lorry and a left-hand
> turn. Watch and share our new #THINK! cycle safety ad." As well as finding
> out what time and money was spent on this advert, I would like to know who
> signed off this advert.

Topic 16: Benefit claimants' and pensioners' enquiries I

Highest probability words: work, claimant, pension, job, provide, universal, sanction
FREX: Jobcentre, Jobseeker, Jobmatch, toolkit, UJM, JSA, job search
Representative FOI requests:
Example 1. DWP, 16 November 2012

> If a person claiming job seekers Allowance refuses to give DWP permission to
> access their universal job match account, will they be sanctioned? Even though
> they have set up and are using the Universal Jobmatch account what law gives
> DWP the right to force a person to sign a form (permission to access their
> Universal Jobmatch account) under threat of a sanction?

Example 2. DWP, 6 December 2013, successful

> When signing up to Universal Jobmatch do you have to give access to your
> account to the Jobcentre advisors? Or, can you provide a written job search? As
> I see it from a legal perspective, you do not have to tick the box to allow DWP
> to have access to your account.

Example 3. DWP, 5 March 2017, successful

> I ask for details of the regulations by which a Work Coach/Advisor can refuse
> to provide a Jobseekers Allowance claimant with a Jobseekers Direction. In
> particular, with respect to a situation where the JSA claimant has been required
> to apply for a specific vacancy which they feel is unsuitable and have then asked
> for a Jobseekers Direction with regard to the vacancy in question. The Work
> Coach/Advisor is refusing this request. Furthermore, on what basis can the
> Work Coach/Advisor claim that adding the vacancy information to the
> Claimant Commitment is just the same as a Jobseekers Direction?

Topic 17: NHS, health and safety

Highest probability words: health, NHS, safety, care, execute, trust, hospital
FREX words: Ellison, safety, health, NHS, NLP, smoke, IAPT
Representative FOI requests:
Example 1. HSE, 25 May 2016, refused

> Since the investigation into the Alton Towers incident has now concluded, and liability has been admitted I am writing to request your investigation report and associated document held by the HSE pertaining to this incident and subsequent investigation work carried out by the HSE.

Example 2. DHSC, 26 April 2016, refused

> Please provide all correspondence between Jane Ellison (Minister for Public Health) and Martin John Jarvis (Professor of Health Psychology at UCL) since the appointment of Ms Ellison in October 2013.

Example 3. HSE, 7 April 2014, successful

> Please, could you provide the number of investigations regarding deaths in the workplace that are still under investigation/review for the periods 2012 and 2013? Please, could you provide the incident names in these statistics (companies/individual names under investigation not required)?

Topic 18: Meetings minutes and correspondence

Highest probability words: copy, meet, govern, correspond, minister, document, secretary
FREX words: meet, email, minutes, correspond, assent, committee, secretary
Representative FOI requests:
Example 1. MoJ, 8 October 2014, refused

> Please, could you release the official portraits of all the Ministers who have served in this department under the coalition government since 2010? Could these images be released under the Open Government Licence, so that the images may be added to their respective articles on Wikipedia?

Example 2. HMT, 26 June 2015, refused

> Can you provide me with all correspondence including letters, emails and details about meetings between George Osborne and Boris Johnson regarding the Garden Bridge over the past four years?

Example 3. DCMS, 7 October 2011, refused

> Please provide copies of internal correspondence, minutes of meetings and conversations within DCMS, between 22 June 2011 and 10 September 2011 in respect of the DCMS proposal to exempt regulated entertainment from the provisions of the Licensing Act 2003.

Topic 19: Courts and legislation

Highest probability words: court, law, justice, order, legal, judge, appeal
FREX words: supreme, litigation, court, judge, kingdom, magistrate, judgment
Representative FOI requests:
Example 1. MoJ, 24 March 2010, refused
> Can you confirm that Article 61 of The Magna Carta Of 1297 is still in force
> and that as such it is still legal to enter into lawful rebellion against parliament?

Example 2. MoJ, 20 February 2017, successful

> Please provide a list of the names of every Deputy District Judge who sits in
> the Magistrates' Court. Please also include which circuit they sit on.

Example 3. MoJ, 8 September 2017, successful

> Please, would you confirm whether Mr Justice Cobb has jurisdiction to sit as a
> Court of Appeal judge and if so when was he appointed?

Topic 20: Land and property ownership

Highest probability words: local, council, author, community, house, govern, property
FREX words: DCLG, Iain, freehold, property, PFI, CLG, council
Representative FOI requests:
Example 1. HMLR, 22 July 2015, successful
> The owner of Ingleside, Vicarage Road, Yalding, Kent, ME18 6DT extended the
> area of land at the bottom of its garden by about 12ft - 15ft sometime after
> May 2006. Neighbouring properties who have not extended their garden, in the
> same way, are currently in dispute with the Diocese of Rochester who is
> claiming that it owns the land. Can you, please, let me know the date that the
> land at the bottom of Ingleside's garden was extended together with details of
> the transaction including the name of the original owner of said land?

Example 2. MoD, 22 January 2009, refused

> Please, list properties owned by the MoD in Nottinghamshire. Also please list
> properties where the MoD has an interest in Nottinghamshire. Also, can you list
> all the properties that the MoD has disposed of in Nottinghamshire for each of
> the last ten years? Also, can you list all of the properties that MoD is about to
> dispose of in Nottingham over the next two years?

Example 3. HMLR, 16 March 2015, successful

> Under the Freedom of Information Act 2000, please provide me with the
> following; purchase prices of 3-bed apartments on Grove Avenue Sutton, SM1
> between August 2014 and March 2015.

Topic 21: Civil servants I: policies, qualifications and remuneration

Highest probability words: staff, employ, member, complaint, civil, employee, service
FREX words: employee, staff, servant, role, SCS, salary, team
Representative FOI requests:
Example 1. DfT, 9 July 2011, successful

> I request confirmation of the median salary of FTE staff of the following grades
> as it applied on 31 March 2005, 31 March 2006, 31 March 2007, 31 March 2008,
> 31 March 2009, 31 March 2010, 31 March 2011. The relevant grades are AA,
> AO, EO, HEO, SEO, G7, G6, SCS (exclusive of permanent Secretaries). All
> figures may include location, Shift disturbance and RRA allowances but should
> not include overtime. I am seeking only the above grades, not equivalent
> specialist grades.

Example 2. MoD, 30 May 2009, successful

> Could you please provide details of any reward or remuneration awarded to
> members of qualified staff performing the role of a First Aider within your
> department.

Example 3. DCLG, 17 July 2011, successful

> I request the median full-time equivalent salary for staff broken down by grades
> for all DCLG staff within the AA-Grade 6 pay ranges as it applied on
> 31/03/2006, 31/03/2007, 31/03/2008, 31/03/2009, 31/03/2010 and 31/03/2011.
> The grades I request the information for are AAs, AOs, EOs, HEOs, SEOs,
> Grade 7s and Grade 6s.

Topic 22: Various areas of interest

Highest probability words: receive, response, look, within, write, contact, forward
FREX words: write, forward, assist, look, receive, clarify, contact
Representative FOI requests:
Example 1. DfE, 28 July 2009, successful

> Under the Freedom of Information Act 2000, I seek the following information
> about the Department's staffing and funding for publicity and communications.
> I would like this information broken down annually for the past three years up
> to the time this request is answered. The years may be financial or calendar,
> whichever is convenient, but please keep them consistent. Specifically, I need
> the number of staff (full-time equivalent) working in marketing, external
> communications and press offices; the number of dedicated Freedom of
> Information officers (full-time equivalent); total staffing cost of marketing,
> external communications and press officers; total staffing cost of dedicated
> Freedom of Information staff; total spending on marketing, external
> communications and press offices; total spending on Freedom of Information;
> and if the department produces a newsletter then the name of this publication
> and all associated costs.

Example 2. DfT, 5 November 2011, refused

> Under the Freedom of Information Act 2000, I seek a copy of all consultation documents on the High-Speed Rail proposal submitted by people who live in the Camden borough ward of Regent's Park.

Example 3. DCMS, 28 October 2010, successful

> I would like to make a request under the terms of the Freedom of Information Act 2000. Please, provide me with copies of any correspondence the department has received from Chelgate or records of meetings which have been held where representatives from Chelgate have attended for the period from 1 January 2009 to the present. My preferred format to receive this information is by electronic means.

Topic 23: Army/race and ethnicity

Highest probability words: force, arm, army, service, group, personnel, request
FREX words: cartridge, rape, ethnic, martial, block, religion, arm
Representative FOI requests:
Example 1. MoD, 4 July 2015, refused
> What percentage of ethnic minorities within the Armed Forces are non-combatants and/or officers?

Example 2. MoD, 6 November 2013, refused

> I would like to know how many members of the Armed Forces have been arrested by civilian (i.e., non-military) police over the last five years for the following offences - rape of a female, sexual assault of a female, attempted rape of a female, attempted sexual assault of a female. Please, can you separate the figures in your response by year?

Example 3. MoD, 23 April 2012, successful

> I would like to request information regarding how the decisions are made to which health conditions are a bar to military service, more specifically pertaining to depression, why they are chosen and by whom. I feel some cases of depression are not a bar to service within the armed forces.

Topic 24: Foreign relations

Highest probability words: office, cabinet, nation, foreign, British, commonwealth, country
FREX words: foreign, commonwealth, embassy, ambassador, oversea, matrimonial, diplomat
Representative FOI requests:

Example 1. FCO, 14 July 2010, refused

> On 8 November 1978, there was a raid on Yieldingtree Farm, Lusaka, Zambia, by forces loyal to Joshua Nkomo. During this raid, three foreign nationals were kidnapped and brutally beaten, including my mother, Patricia Taylor, a British citizen, since deceased. I would like to know what records the Foreign Office has of this event and if I could have any copies of said records.

Example 2. FCO, 22 April 2012, successful

> Do you know how I can get the information/files on the visit by Nicolae Ceausescu to the UK (I believe in 1978)?

Example 3. FCO, 18 August 2010, successful

> I would like to request information on the number of Malaysian nationals issued with British Overseas Citizen (BOC) passports since 4 July 2002. Also, I would like to request details of the location where those passports were issued.

Topic 25: Budget and spending

Highest probability words: cost, provide, fund, total, money, finance, amount
FREX words: food, spent, rural, finance, fund, cost, environment
Representative FOI requests:
Example 1. DfID, 28 October 2014, refused

> How much money (£ total) has DFID spent specifically funding sustainable agriculture projects (including organic farming and other agroecological farming techniques such as IPM, push-pull and agroforestry) in Africa in the last ten years? How much money (£ total) has DFID spent funding all agriculture-related projects (crop growing, harvesting, processing) in Africa in the last ten years?

Example 2. DHSC, 29 October 2012, successful

> I'd like to know how much you spend in a year (either the last 12 months to date, the last financial year or the last calendar year) on media monitoring. This would include either payments to third parties that provide monitoring services or subscriptions to online tools. This request covers traditional media such as newspapers, online media and social media such as twitter. Please provide a breakdown, if possible, on the amount spent on those three areas (traditional/online/social). Could you also please briefly tell me why you use media monitoring tools/services?

Example 3. DCMS, 31 May 2014, successful

> I was wondering if you could tell me how much the Department spent on purchasing newspapers and magazines in the calendar years of 2012, 2013 and to-date in 2014, and the number of newspapers and magazines by name that are delivered to the Department on a daily basis.

Topic 26: Charities

Highest probability words: commission, charity, England, site, Wales, copy, design
FREX words: logo, poll, trustee, mail, ordnance, Ashmount, site
Representative FOI requests:
Example 1. CC, 15 March 2017, successful

> Please, could I request a copy of the governing document for the Incorporated
> Trustees of the No 1 Trust Fund Charity No 275353?

Example 2. CC, 10 February 2015, successful

> Please provide me with an electronic copy of the Constitution of 1048511 -
> CAPRICORN ANIMAL RESCUE AND SANCTUARY (INC ASTON,
> HAWARDEN ANIMAL AID)

Example 3. CC, 2 January 2017, successful

> Please can I request a copy of the Governing Document for Ryedale Dog
> Rescue. Registration number 1117972.

Topic 27: Benefits claimants' and pensioners' enquiries II

Highest probability words: work, pension, claim, benefit, assess, claimant, ESA
FREX: ESA, ATOS, PIP, WCA, HCP, incapacity, reassess
Representative FOI requests:
Example 1. DWP, 28 February 2015, successful

> As the SDP is dependent on PIP/DLA awards, can you confirm that as long as
> you have the qualifying disabled benefit (Daily living PIP or middle/high rate
> care DLA) and that nobody is claiming carers allowance, you also live alone or
> are in a couple whom both receive the qualifying disabled benefit then the SDP
> is added to your ESA regardless of whether it's (IR) or (C)?

Example 2. DWP, 3 October 2013, successful

> I am in receipt of my state pension which includes SERPS pension and would
> like to know if my basic and SERPS pension will continue to receive increases
> after the new Single Tier Pension starts on 6 April 2016.

Example 3. DWP, 3 March 2016, successful

> Please treat this as a request for information under the Freedom of Information
> Act. I wish to know the number of ESA claimants in the Work-Related Activity
> Group who also receive PIP Standard Mobility Component only, PIP Enhanced
> Mobility Component only, PIP Standard Daily Living Component only, PIP
> Enhanced Daily Living Component only, both Standard components, both
> Enhanced components, one enhanced and one standard component. As above,
> but for claimants in the ESA Support Group.

Topic 28: Government's records about citizens/IT systems

Highest probability words: provide, policy, whether, avail, use, guidance, detail
FREX: avail, policy, call, guidance, whether, website, access
Representative FOI requests:
Example 1. DWP, 3 February 2012, successful

> I would like to inquire as to whether the Department for Work and Pensions'
> telephone system has an automatic call logging feature. In particular, I wish to
> know whether the system will automatically log the telephone numbers of
> callers to the DWP, regardless of whether there is any note of what may have
> been discussed in the call. On the assumption that this system is in place, please
> confirm when the system was put in place, and in particular, whether it was in
> operation in 2006.

Example 2. DWP, 18 April 2013, successful

> Could you please answer the following questions under this freedom of
> information request? Does Atos healthcare record telephone calls, when they
> contact Service users and vice versa? If the answer to question 1 is yes, can the
> Service user, get copies of all telephone audios under section 7 of the data
> protection act 1998, via submitting a data access request. Do the DWP, record
> telephone calls, when they speak to service users and vice versa? If the answer is
> yes to question 3, can the service user get copies of all telephone audios, under
> section 7 of the data protection act 1998, via submitting a data access request?
> What happens to all of these audios, are they kept for a certain time and then
> disposed of.

Example 3. DWP, 25 March 2009, refused

> I understand that Jobcentre Plus's website has an interface (API) by which
> JCP job listings are displayed on Directgov, as well as on the websites of
> partners selected by DWP. I would like to request: the documentation that
> describes the fields that are provided by this API, a list of the bodies which
> have access to this API, whether they be government departments, agencies or
> private companies.

Figure A-17: Expected topic proportions for the model with 28 topics (covariates: display_status (outcome), govBrown, govCameron)
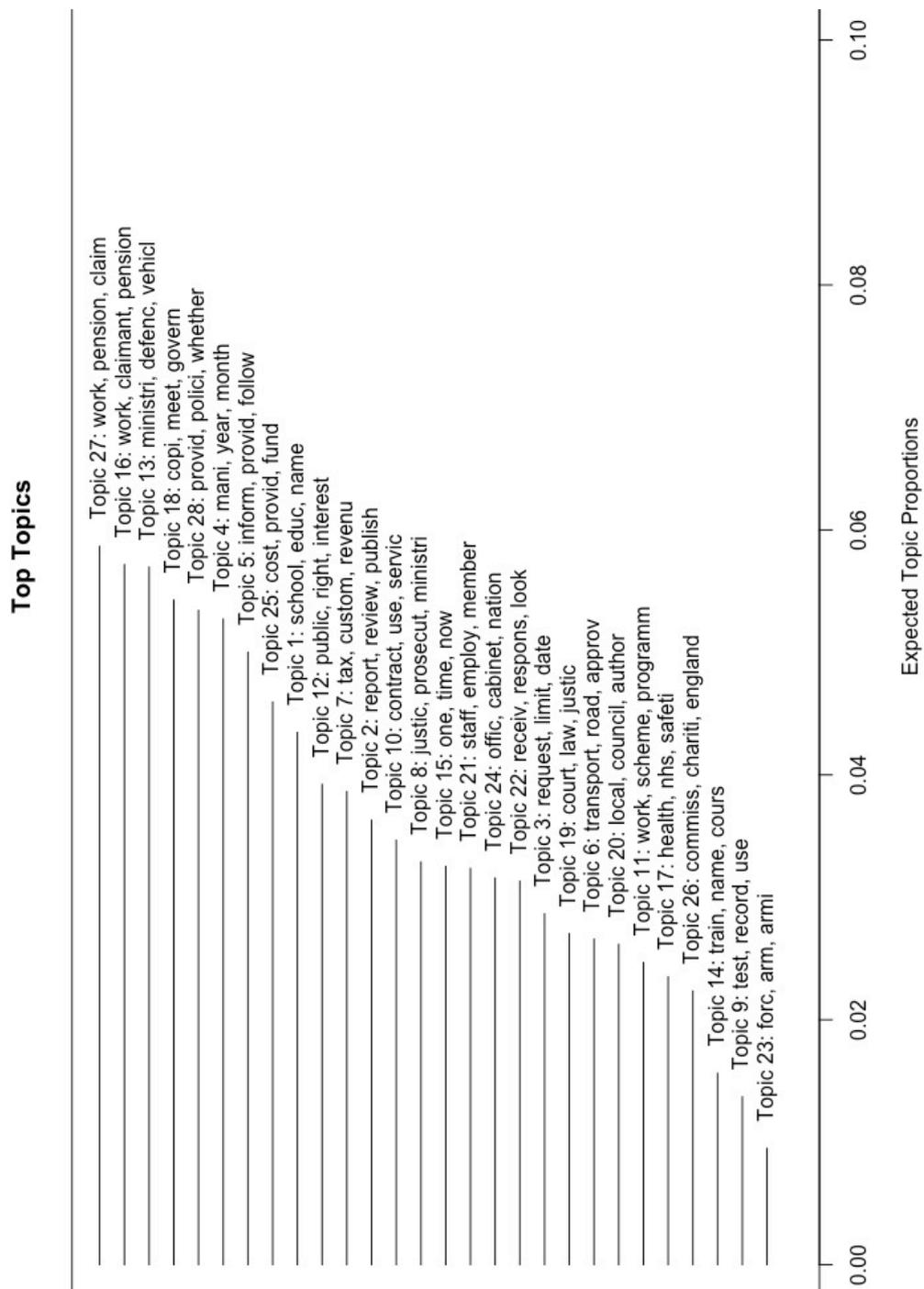
Note: Words used to label the topics are the words with the highest probability to occur within that topic. The words listed in the figure are the top three words associated with the topic. For more detailed information about the model see M. E. Roberts et al. (2014).

Figure A-18: Highest word probabilities for each topic in the model with 28 topics, covariates: display_status (outcome), govBrown, govCameron
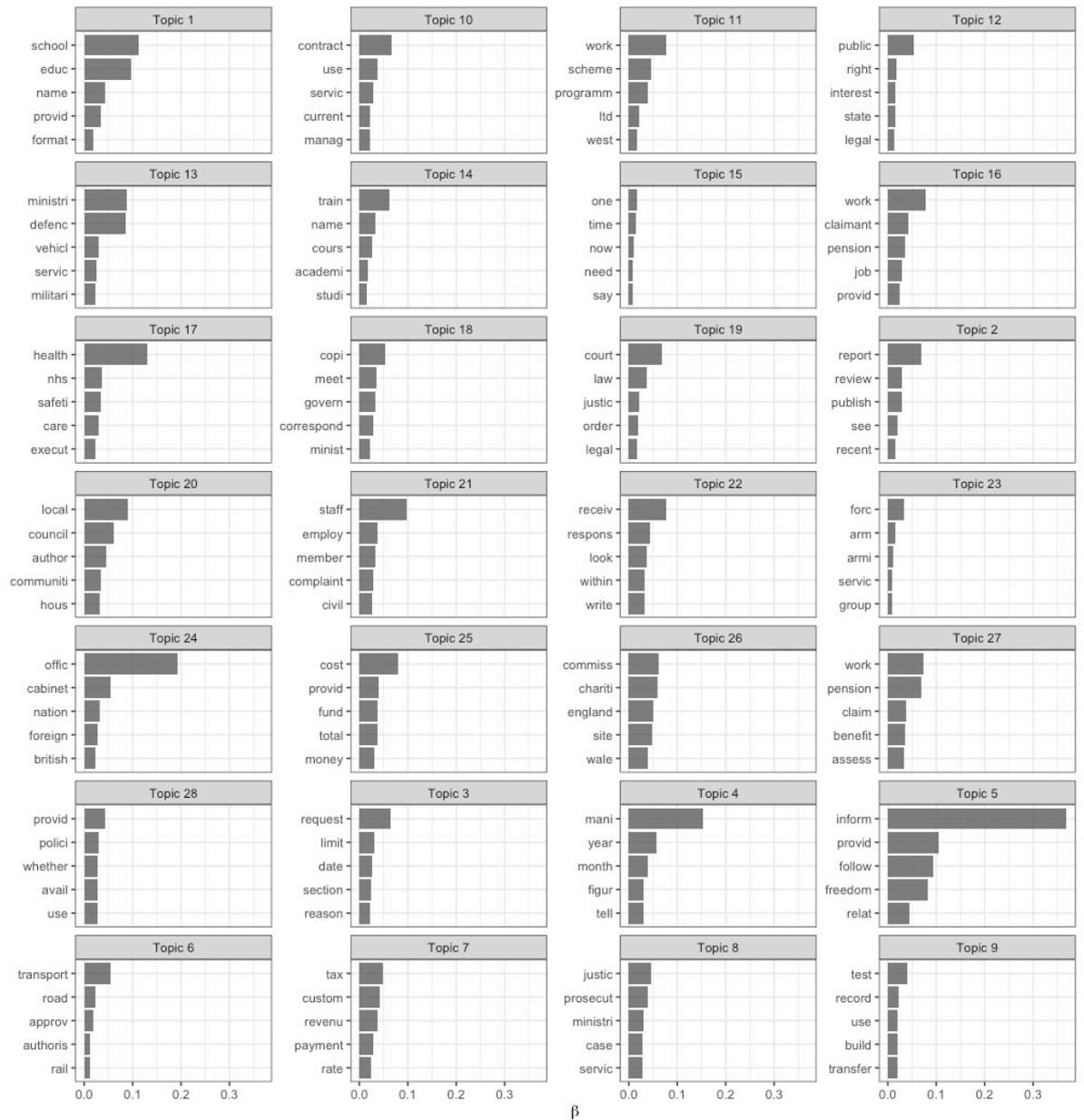
Note: The words listed in the figure are the top five words associated with the topic. For more detailed information about the model see M. E. Roberts et al. (2014).

Detailed exploration of topics in the model with 28 topics (successful and rejected requests included)

Model 2: Topic prevalence = α + βdisplay_status + βgovBrown + βgovMay + ε

Topic 1: List of schools

Highest probability words: school, education, name, provide, format, primary, pupil
FREX words: pupil, mis-, headteacher, school, education, spreadsheet, hyperlink
Representative FOI requests:
Example 1. DfE, 7 September 2017, successful

> Please, can you send me an excel list of all UK schools, with their name, email
> address, address, other contact details, website, name of the headteacher,
> headteacher email address, and phase of education?

Example 2. DfE, 31 January 2017, successful

> Could you please send me a list of all schools, their address and the
> headteachers name, headteachers email address, Establishment Status, Type of
> Establishment Number, LA Number, UKPRN, Telephone number, Postcode?

Example 3. DfE, 9 March 2015, successful

> Can you please send me a list of all primary and secondary schools in England
> and Wales with the addresses, contact telephone numbers, name of the
> headteacher and an email address for each headteacher. If this can be provided
> in excel format that would be appreciated.

Topic 2: Policies, analyses and reports

Highest probability words: report, publish, review, see, recent, page, link
FREX words: publish, target, impact, page, audit, report, review
Representative FOI requests:
Example 1. CO, 25 September 2017, refused

> Has any research or investigation been commissioned by the Cabinet Office or
> other major departments of government into the impact on living standards (as
> measured by the Retail or Consumer Price Indices or comparable indices) as a
> result of the UK leaving the European Union? If such a commission has been
> undertaken what was the conclusion of that assessment including any estimate
> into the likely increase in the Retail or Consumer Price Index or comparable
> indices in the run-up to the UK leaving the European Union and the first five
> years after the UK's departure.

Example 2. DfE, 11 November 2014, successful

> In the recent Consultation paper: "Performance descriptors for use in key stage 1 and 2 statutory teacher assessment for 2015/2016", published in October 2014, there is a reference on Page 5, paragraph 8 to a number of experts who worked on the draft – "The performance descriptors..." have been drafted with experts, including teachers, representatives from Local Authorities, curriculum and subject experts. I request the names of the experts, mentioned in this paragraph, who worked on drafting the Performance Descriptors".

Example 3. CO, 7 April 2015, successful

> Please reveal the findings of the examination of the EU powers - 32 reports as described in the Guardian's article:
> http://www.theguardian.com/world/2015/mar/28/lords-accuse-tories-burying-eu-powers-review.

Topic 3: Right to information

Highest probability words: request, record, limit, date, section, reason, time
FREX words: request, receipt, ICO, FOIA, limit, amend
Representative FOI requests:
Example 1. CO, 6 April 2016, refused

> The 1st Tier Tribunal has now ordered disclosure of the 1992 edition of the Precedent Book subject to some redactions. Please provide this. If you are relying on the section 21 exemption, please provide precise instructions of how I can access it. If you are relying on the section 22 exemption, please provide details of the proposed timescale.

Example 2. FCO, 5 May 2009, successful

> I would like to request that you send me a copy of all the documents that have been released pursuant to the ICO Decision Notice case reference FS50083726, and all documents that have been released before the escalation to the ICO in this case. I understand that a selected number of documents has been released at various times before the ICO issued a decision notice and I ask for a copy of these too. I have searched on the FCO's Access to Information (Disclosure Log) website but have been unable to locate a copy of the files. If you have uploaded *all of the files* (ref. above) now, I'd be grateful if you could just deal with this request informally by sending me a link to the documents and I will then confirm that I withdraw my request. This is to avoid the necessity of you sending me a formal Section 21 refusal notice.

Example 3. DHSC, 21 September 2009, successful

> With reference to the case that resulted in the Information Commissioner's recent decision notice FS50195059 please could you provide me with copies of all the information you supplied to the applicant prior to decision notice being issued. Please supply me with a copy of the letter you have now been told to release as part of the Information Commissioner's decision notice FS50195059.

Topic 4: Statistics

Highest probability words: many, year, month, figure, since, tell, total
FREX words: figure, many, IGSC, past, statistics, month, percentage
Representative FOI requests:
Example 1. MoD, 29 June 2009, refused

> What is the number of visitors (persons not on active duty in the theatre) to MoD stations in Iraq for every month since March 2003? How many man-hours, broken down by month for every month since March 2003, were spent hosting visitors?

Example 2. SCUK, 4 July 2016, successful

> Please, can you tell me the following: How many writs of Habeas Corpus have been submitted from 01/01/15 to the 01/01/16? How many of these writs were successful? How many of the unsuccessful writs were appealed?

Example 3. CO, 8 February 2016, refused

> Please, can you tell me the following information: How many files your department have - which were created before 1945 - are still classified to the time of writing (07/02/16)? If you still have these files, then what is the longest period of time that they are classified?

Topic 5: Ex-military vehicles

Highest probability words: inform, provide, follow, freedom, relate, held, make
FREX words: VRN registration, freedom, inform, union, trade, held, hold
Representative FOI requests:
Example 1. MoD, 18 April 2017, successful

> I would like to make a freedom of information request for any information held relation to ex-MoD vehicles with the following VRN/registrations: SH30AA, SH31AA, SH32AA, SH33AA, SH34AA, SH35AA, SH36AA, SH37AA, SH38AA, SH39AA, SH40AA.

Example 2. MoD, 30 January 2017, successful

> I would like to make a freedom of information request for any information held relating to the ex-MoD vehicle with the following VRN/registration 77KF99.

Example 3. MoD, 14 February 2017, successful

> I would like to make a freedom of information request for any information held relating to the ex-MoD vehicle with the following VRN/registration 26KK81.

Topic 6: Transport

Highest probability words: transport, road, approve, train, travel, rail, park
FREX words: railway, fare, franchise, highway, pedestrian, rail, south-eastern
Representative FOI requests:
Example 1. DfT, 14 April 2011, successful

> When are warning signs illuminated and/or displayed on the motorways to
> warn drivers of adverse weather conditions, in particular, torrential rain?

Example 2. ORR, 22 July 2015, successful

> Is the ORR aware that Shropshire Council (owner of the disused railway line
> between Gobowen and Blodwel) will be allowing Cambrian Heritage Railways
> Limited to construct and operate a new level crossing on the emergency access
> route to the Cambrian Medical Centre on Thomas Savin Road in Oswestry?

Example 3. DfT, 9 April 2013, successful

> Can you please provide a copy of the Highway Code rules or Regulation/s for
> alighting passengers from cars and taxis? Can you please provide a copy of the
> Highway Code rules or Regulation/s for Loading and unloading from cars and
> taxis?

Topic 7: Taxes

Highest probability words: tax, custom, revenue, payment, rate, pay, treasury
FREX words: borrow, loan, tax, repay, earn, rate, RPI
Representative FOI requests:
Example 1. HMT, 13 May 2011, successful

> Can you please provide me with the historical figures for the top rate of income
> tax, the basic rate of income tax, rate of corporation tax and rate of capital
> gains tax for each year since 1945?

Example 2. HMT, 20 March 2012, successful

> I would be extremely interested to see what evidence, studies and projections
> you have based your prediction upon, that lowering the top rate of UK income
> tax will increase overall yield from those paying the highest rate.

Example 3. HMRC, 21 January 2013, refused

> In relation to all transfer pricing investigations, settled throughout 2012, on
> companies operating in the UK: how much money did HMRC collect and how
> much money did it originally demand?

Topic 8: Prosecutions and convictions

Highest probability words: justice, prosecute, ministry, case, service, police, prison
FREX words: HMP, custody, prosecute, prison, shrive, convict, child
Representative FOI requests:
Example 1. MoJ, 2 March 2009, successful

> Please, can you tell me the number of escapes from HM Prison, Norwich in 2008
> and the nature of the escape? Please include escapes from the prison, escapes
> from prison escort, escapes from court escort and those who absconded.

Example 2. MoJ, 21 October 2016, refused

> Now that ex-policeman Gordon Anglesea has been found guilty, please send a
> copy of the Macur review with the redactions made about Gordon Anglesea
> removed.

Example 3. CPS, 20 September 2017, refused

> How many cases of people between the ages of 13 and 17 did you refuse to
> prosecute because it wasn't in the public interest to do so in relation to section
> 13 of the sexual offences act 2003 in the year 2016?

Topic 9: Health (focus on vaccinations)

Highest probability words: test, record, use, transfer, build, facility, evidence
FREX words: born-digit, syndrome, HPV, abnormal, EBM, vaccine, MECF
Representative FOI requests:
Example 1. DHSC, 22 September 2013, refused
> Please, could you provide me with a full list of the ingredients in the following
> two vaccination shots: the children's seasonal flu vaccination for winter
> 2013/2014 Fluenza® and the seasonal flu vaccination being administered to the
> over 65 during winter 2013/2014.

Example 2. DWP, 29 August 2016, successful

> Please, can you send me the training material, known as Self-Directed Learning
> Packs, provided by the medical training and development team at the DWP,
> and used by HCPs working for Atos Healthcare, for these specific conditions:
> Asperger's (or Autism Spectrum Disorder), Depression, Chronic Cirrhosis, and
> Osteoporosis.

Example 3. DHSC, 22 November 2008, successful

> The Department of Health is currently implementing a nationwide
> immunisation program for girls to protect them from HPV (human
> papillomavirus). This immunisation program will cost the British government
> £9 million per year. Could you send me the plan submitted to the British
> government for the introduction of the HPV vaccination program in the UK?

Topic 10: Contracts (focus on IT systems)

Highest probability words: contract, use, service, current, system, manage, framework
FREX words: contract, procure, oracle, framework, photocopiersmfd, reprography, renew
Representative FOI requests:
Example 1. OFGEM, 9 September 2015, successful

> Do you currently make use of any Oracle products? If so, which? What is the quantity of licensed users split by each Oracle product used by your organisation? What is the value of your current Oracle contracts, split by a product? Do you outsource your Oracle projects, services or maintenance to an external supplier? If so, whom? When does that contract expire? In 2013, the Cabinet Office appointed an Open Standards Board to drive the use of Open Source across Government. Which Open Source databases have you deployed since 2013 and have firm plans to deploy within the next 18 months as alternatives to Oracle products? Please note, by Oracle products I mean commercially licensed Oracle software products, for example, any of the Oracle Database products (https://www.oracle.com/database/products.html) or any of their Middleware products such as Oracle WebLogic.

Example 2. MoD, 5 May 2015, refused

> How many computers are still running Windows XP? When do you anticipate you will transition all of these XP machines to a new operating system? Which operating system are you switching these machines to? What parts of the department are these machines mainly used in? How are you securing the XP machines in the interim period before their operating system is replaced? Have you taken out an Extended Support deal with Microsoft to update these XP machines? What is the cost of this Extended Support deal? When does this Extended Support deal expire?

Example 3. HMRC, 16 April 2015, successful

> Do you look after your own IT or is it outsourced? How much did you pay in the last financial year for software licenses? How many computers users do you have? When do you need to renew the contract with Microsoft for software licenses? What was the value of your last contract per year? Do you currently measure software usage versus the number of licenses purchased? If so, what is used for software usage metering? Do you use a software asset management tool? Please also provide details of IT Contracts Managers and any person(s) involved in IT Software procurement.

Topic 11: DWP's Work Programme

Highest probability words: work, scheme, programme, ltd, west, pension, placement
FREX words: CPA, Avanta, Yorkshire, placement, north, Seetec, west
Representative FOI requests:

Example 1. DWP, 30 August 2013, refused

> Since the start of the work programme I would like a record of all complaints made about the work programme and also what the complaints were about, but not who made the complaint about Avanta, Suite G1, Ground Floor, Blackfriars House, Blackfriars Street, Manchester, M3 2JA.

Example 2. DWP, 29 August 2013, refused

> Since the start of the work programme I would like a record of all complaints made about the work programme and also what the complaints were about, but not who made the complaint about Avanta, Elizabeth House, 38 Bold Street, Leigh, Lancashire, WN7 1AG.

Example 3. DWP, 4 August 2013, refused

> I would be grateful if you would provide the following information under the FOI Act: the names and locations of organisations participating in the Work Programme in the Redcar and Middlesbrough Package Area by providing mandatory work placements through the DWP's prime providers G4S and Avanta Enterprise Limited and their subcontractors.

Topic 12: Law enforcement

Highest probability words: public, interest, right, state, policy, legal, legislation
FREX words: ensure, legislation, public, breach, procedure, protect, party
Representative FOI requests:
Example 1. DWP, 7 January 2013, successful

> Audio-recording WCAs: RPF 27 to RPF 33. The DWP's full knowledge of its continuing illegality in obstructing the recording of WCAs was exposed when it disclosed the DWP\Atos memo of Sept.\Oct. 2010, entitled RPF 27: http://www.whatdotheyknow.com/request/131400/response/339826/attach/4/ RFP%2027.PDF.pdf. You have also published RPF 33. Please now disclose RPF 28, RPF 29, RPF 30, RPF 31 and RPF 32.

Example 2. DWP, 17 May 2008, successful

> When DWP employees call members of the public by telephone, highly confidential information is often discussed. Could you please tell me what policies, procedures and training DWP has put in place to allow members of the public to be confident that they are talking to a DWP employee, rather than a potential identity thief?

Example 3. MoJ, 13 April 2015, successful

> Re: Public Sector Data Sharing - Guidance on the Law, April 2011

> In section 6, Common Law and statutory restrictions on disclosure of data, you state that a breach of statutory duty may be an offence. Under the provisions of the Freedom of Information Act 2000, I should be grateful if you could clarify

the following. Should a violation of the NHS Statutory Restrictions on data disclosure, namely the NHS (Venereal Regulations) 1974 occur, disclosing individual's HIV+ status without his consent or knowledge, for purposes different from those permitted by this Act, will it be a civil or a criminal offence? To whom should such a breach be reported?

Topic 13: Ex-military vehicles II

Highest probability words: ministry, defence, vehicle, service, military, history, register
FREX words: rover, cadet, veteran, chassis, Land Rover, FFR, ex-army
Representative FOI requests:
Example 1. MoD, 25 October 2014, successful

Do you have any history of the deployment of my Ex Military Land Rover Defender (Truck Uty Med FFR Plain Hardtop 12/24 4x4 Land Rover Defender 110 (DSL)) Military Registration 59KK65 Chassis Number SALLDHAC5JA917505?

Example 2. MoD, 17 October 2016, successful

I recently purchased a former Army Land Rover Defender "Snatch", military registration 30 KK 45 and would like to know its military history. Please, could you email me its Merlin Card details?

Example 3. MoD, 20 April 2016, successful

Please, can I request a Merlin report on vehicle 20HG63 this is my ex-military Land Rover?

Topic 14: Education

Highest probability words: train, name, course, academy, advertise, qualify, learn
FREX words: Nominet, Gove, advertise, campaign, sponsor, trainee, exam
Representative FOI requests:
Example 1. DfE, 1 February 2012, successful

Please, can you tell me a list of the vocational courses that will no longer be equivalent to GCSEs and the ones that will after the announcement by Michael Gove on 31 January 2012: http://www.bbc.co.uk/news/education-16789215?

Example 2. DfE, 19 September 2011, successful

Could you please direct me to information on KS4 pupils who were not submitted for GSCE exams (i.e. those that sat "equivalent" qualifications. I require the information to include: the total number of students attending at KS4, the total number of students who passed GCSE subjects (any number), the total number of students who failed to achieve a pass grade at GCSE level, and the total number of students who were not entered into GCSE exams (i.e. students who sat other qualifications). I need this information to be broken down by LEA areas.

Example 3. DfE, 16 September 2014, successful

> In July, I wrote and asked for the number of students enrolled at the Trinity Academy, Brixton. You replied and said that, at the time, 25 were enrolled, but that applications were still possible. I am now asking for an update on the number of students currently enrolled at the school.

Topic 15: Various areas of interest

Highest probability words: time, one, need, want, now, take, say
FREX words: scanner, water, think, seem, something, cull, don't
Representative FOI requests:
Example 1. 25 June 2012, successful

> It has come to my attention that on 21 December 2012 a large-scale event is supposed to happen. As a very cautious person, I wish to know what the procedure would be for civilians in such an event, would the London underground be used as a shelter, or are there designated civilian shelters?

Example 2. Defra, 18 June 2013, successful

> Owen Paterson and Nigel Gibbens chief vet have said that there is pressure from the EU/Brussels to get BTB under control and that a badger cull must be part of this. Please advise which EU directive states the explicit requirement for a badger cull (or documentary evidence of this pressure).

Example 3. DfT, 6 June 2010, successful

> Last week we had beautiful clear blue skies over London, from dawn until dusk. On Saturday there were heavy chemtrails over the London skies, which continued for the whole day and resulting in a thick hazy cloud. I would like to request information, under the freedom of information act, as to what chemtrails are and their purpose. I would appreciate it if you do not reply with a whitewash answer on contrails as this is something completely different than the subject of chemtrails.

Topic 16: Benefit claimants' and pensioners' enquiries I

Highest probability words: work, claimant, pension, provide, job, universal, sanction
FREX words: jobcentre, jobseeker, Jobmatch, toolkit, UJM, JCP, job search
Representative FOI requests:
Example 1. DWP, 26 June 2016, successful

> To put my request in context, I have been told by a JCP work coach that the Jobcentre is now legally empowered to retain or photocopy claimant's work search activities without the claimant's consent. Is there any legal requirement now in place for claimants in receipt of Jobseeker's Allowance or the Jobseeker's Allowance element of Universal Credit that allows the Jobcentre to retain or photocopy a claimant's work search activities without the claimant's consent?

Example 2. DWP, 30 October 2017, successful

> Is it mandatory for Universal Credit claimants to use the Universal Credit journal to provide evidence of their work search and evidence of a CV, or can they use their own Tablet, iPad, Laptop, Netbook or any other digital medium of their choosing? Or paper copies as evidence as well?

Example 3. DWP, 4 April 2015, successful

> Could you please provide me with the most up to date (as of 04/04/2015) Universal Jobmatch Toolkit, either as an attachment, or a direct link!

Topic 17: NHS, health and safety

Highest probability words: health, NHS, safety, care, execute, trust, hospital
FREX words: Ellison, safety, health, NLP, NHS, smoke, dental
Representative FOI requests:
Example 1. DHSC, 26 April 2016, refused

> Please provide all correspondence between Jane Ellison (Minister for Public Health) and Martin John Jarvis (Professor of Health Psychology at UCL) since the appointment of Ms Ellison in October 2013.

Example 2. HSA, 19 January 2017, successful

> I wish to request the number of RIDDOR incidents reported by Multimesh UK from 01/01/2015 to 19/01/2017.

Example 3. DHSC, 23 April 2016, refused

> Please provide correspondence between Jane Ellison (Minister for Public Health) and the Royal College of Physicians since the appointment of Ms Ellison in October 2013.

Topic 18: Meeting minutes and correspondence

Highest probability words: copy, document, govern, meet, correspond, include, minister
FREX words: meet, email, minute, correspond, committee, document, secretary
Representative FOI requests:
Example 1. DfT, 30 October 2014, successful

> Please, could you release the official portraits of all the Ministers who have served in this department under the coalition government since 2010? Could these images be released under the Open Government Licence, so that the images may be added to their respective articles on Wikipedia?

Example 2. DHSC, 8 March 2017, refused

> On 6 March 2017 Department of Health released the above consultation. in section 2.5 the Infected Blood Reference Group is mentioned. This advisory group made up of experts brought together to help inform and shape the policy on reforms. The group's members include beneficiary representatives. Who are these beneficiary representatives? Can I request all minutes from any meetings this group have also held correspondence between group members? Also, any correspondence between any members of this group and the Department of Health from the date this group formed until present date.

Example 3. DHSC, 22 December 2015, successful

> I would like to request any documents relating to consultations with Tesco, prior to the establishment of the Food Network HLSG. This may include meeting minutes, agendas, discussion papers, submissions to the DH, and action notes.

Topic 19: Courts

Highest probability words: court, law, justice, order, legal, appeal, judge
FREX words: supreme, litigation, court, kingdom, judge, magistrate, judgment
Representative FOI requests:
Example 1. SCUK, 17 December 2010, refused

> I would like to know how many people were extradited from the UK to Sweden, by the Supreme Court, between the dates 1 January 2009 and 31 December 2009?

Example 2. SCUK, 8 August 2012, successful

> Can the supreme court confirm if a member of the public can motion the court to hear an amicus curiae and if so, what is the formal procedure to file such a motion?

Example 3. MoJ, 9 May 2011, refused

> Please, could you let me know how many claims were brought in the English courts using the European Small Claims procedure in 2010?

Topic 20: Land and property ownership

Highest probability words: local, council, author, community, house, govern, property
FREX words: DCLG, Iain, freehold, PFI, property, CLG, council
Representative FOI requests:
Example 1. HMLR, 8 May 2012, refused

> Do you have the details of who owns 27 Foxwood Lane, Acomb, York?

Example 2. HMLR, 6 February 2014, refused

> I would like to see the boundary lines for the Hastings Borough Council owned
> land running parallel to White Rock Road and continuing parallel into White
> Rock Gardens. I would also like to see the boundary lines for the property 26
> White Rock TN34 1JY.

Example 3. HMLR, 23 August 2010, successful

> Please, can you supply written details of the process for registering unregistered
> a parcel of unregistered land?

Topic 21: Civil servants: policies, qualifications and remuneration

Highest probability words: staff, employ, member, complaint, civil, employee, service
FREX words: employee, staff, servant, role, SCS, salary, team
Representative FOI requests:
Example 1. MoD, 30 May 2009, successful
> Could you please provide details of any reward or remuneration awarded to
> members of qualified staff performing the role of a First Aider within your
> department.

Example 2. DHSC, 8 March 2012, successful

> Can you please tell me how many workers you have re-employed in the last year
> that were previously made redundant by your department? Specifically, is there
> anyone who was recruited and employed this year, who was previously made
> redundant from another post in the department?

Example 3. HMRC, 23 July 2010, successful

> Please, could you provide me with a structure chart for your Legal Team, and
> advise me which posts are held by temporary members of staff?

Topic 22: Correspondence II (focus on lobbying)

Highest probability words: receive, response, look, within, write, contact, forward
FREX words: write, forward, assist, look, receive, clarify, outline
Representative FOI requests:
Example 1. MoD, 4 February 2011, refused
> All correspondence between your organisation and Migration Watch since May
> 2010. Please indicate whether minutes or official accounts of meetings were
> made and if so, please release these. Please also give details of any Home Office
> involvement in Migration Watch and any details of funding given to Migration
> Watch.

Example 2. HMT, 30 May 2011, refused

> Please, can you provide the private bodies, organisations, and individuals who have lobbied the Financial Secretary to the Treasury (Mark Hoban) and/or the Treasury over legislation regarding commodities derivatives reforms from May 2010? Please also provide the situation where this lobbying occurred (e.g. over dinner, at multi-stakeholder meetings, etc.). Please also state when these lobbying events occurred.

Example 3. OS, 28 June 2008, successful

> My request is for copies of all correspondence in 2007 and 2008 between Ordnance Survey and Mandate Communications Ltd, part of the Engine Group of companies.

Topic 23: Army/race and ethnicity

Highest probability words: force, arm, army, group, service, personnel, request
FREX words: cartridge, rape, ethnic, religion, LSM FOI, race, command
Representative FOI requests:
Example 1. MoD, 29 March 2013, refused
> Could you please confirm if the service number Q8514862 (RAF) was issued to a male or female? This should not cause any problems regarding ID and privacy as until recently the RN and Army service numbers were clearly distinguished between male and female service members, i.e. 246***** or WO811***.

Example 2. MoD, 4 July 2015, refused

> What percentage of ethnic minorities within the Armed Forces are non-combatants and/or officers?

Example 3. MoJ, 15 July 2012, successful

> Please provide me with a breakdown by race of all the people convicted of murder in the UK in 2011. Please provide me with a breakdown by race of all the people convicted of rape in the UK in 2011. Please provide me with a breakdown by race of all the people convicted of gun crime in the UK in 2011. Please provide me with a breakdown by race of all the people convicted of knife crime in the UK in 2011. Please provide me with a breakdown by race of all the people convicted of street crimes in the UK in 2011. Please provide me with a breakdown by race of all the people convicted of robbery in the UK in 2011.

Topic 24: Foreign relations

Highest probability words: office, cabinet, nation, foreign, British, commonwealth, country
FREX words: commonwealth, embassy, ambassador, foreign, matrimony, diplomat, republic

Representative FOI requests:

Example 1. FCO, 6 April 2013, successful

> I was wondering if you could inform me the number of staff in the British Embassies in Dublin, Washington DC, Mexico City, Tokyo, Abu Dhabi, Madrid and Paris as of 31 December 2009, 2010, 2011, 2012 and 1 April 2013.

Example 2. FCO, 14 July 2010, refused

> On 8 November 1978, there was a raid on Yielding tree Farm, Lusaka, Zambia, by forces loyal to Joshua Nkomo. During this raid, three foreign nationals were kidnapped and brutally beaten, including my mother, Patricia Taylor, a British citizen, since deceased. I would like to know what records the Foreign Office has of this event and if I could have any copies of said records.

Example 3. FCO, 27 March 2016, refused

> Please send me minutes of the meetings where Palestinian issues were discussed with Prime Minister Netanyahu in September 2015.

Topic 25: Spending

Highest probability words: cost, fund, total, provide, money, finance, amount
FREX words: food, spent, rural, environment, fund, cost, finance
Representative FOI requests:

Example 1. DfID, 28 October 2014, refused

> How much money (£ total) has DFID spent specifically funding sustainable agriculture projects (including organic farming and other agroecological farming techniques such as IPM, push-pull and agroforestry) in Africa in the last ten years? How much money (£ total) has DFID spent funding all agriculture-related projects (crop growing, harvesting, processing) in Africa in the last ten years?

Example 2. DMCS, 31 May 2014, successful

> I was wondering if you could tell me how much the department spent on purchasing newspapers and magazines in the calendar years of 2012, 2013 and to-date in 2014. The number of newspapers and magazines by name that is delivered to the department on a daily basis.

Example 3. HMT, 21 January 2012, successful

> Please, could you tell me who the "procurement lead" in charge of getting the best value for money when purchasing services and equipment is within your department?

Topic 26: Charities

Highest probability words: commission, charity, England, site, Wales, copy, provide
FREX words: logo, poll, trustee, site, mail, ordnance, commission
Representative FOI requests:
Example 1. CC, 1 November 2010, successful

> Re Charity 1098004 - Narconon London, the 2008 accounts are more than a
> year overdue for this charity. Both the trustees of this charity are also trustees
> of charity 267386 (Narconon), which has a history of late filing. What action
> has the Charity Commission taken against this charity or its trustees?

Example 2. CC, 1 May 2016, successful

> Can you please confirm whether or not Friends of Prostate Sufferers is a
> registered charity with the Charity Commission for England and Wales?
> According to information on the FOPS website (www.thefops.org.uk), they were
> granted recognition as a charity by HMRC in March 2012 with reference
> XT33456. However, when I search for this organisation on the charity
> commission website (www.charitycommission.gov.uk) either by name (FOPS or
> Friends of Prostate Sufferers) or by reference (XT33456) no match is found.

Example 3. CC, 13 October 2017, successful

> Please provide me with an electronic copy of the constitution and all Governing
> Documents of RUGBY MOSQUE SOCIETY (no. 503021).

Topic 27: Benefit claimants' and pensioners' enquiries II

Highest probability words: work, pension, claim, assess, benefit, claimant, ESA
FREX words: ESA, Atos, pip, WCA, hcp, incapacity, reassess
Representative FOI requests:
Example 1. DWP, 3 October 2013, successful

> I am in receipt of my state pension which includes SERPS pension and would
> like to know if my basic and SERPS pension will continue to receive increases
> after the new Single Tier Pension starts on 6 April 2016.

Example 2. DWP, 13 May 2013, successful

> In relation to claims for Employment and Support Allowance, could you please
> provide the relevant "Guidelines for the Disability Analyst" for the following
> condition – Myelofibrosis?

Example 3. DWP, 26 September 2016, successful

> I have new disabilities: dyslexia, dyspraxia, dyscalculia, cognitive memory loss,
> along with physical and other disability. Due to memory and cognitive
> deficiency, I require extra time to complete the new PIP review form. Please,
> can you inform me due to reasonable adjustments under the Equality Act 2010
> for disabled people, how much extra time do you allow for completed review
> paperwork?

Topic 28: Contacting authorities

Highest probability words: company, avail, answer, individual, detail, whether, website

FREX words: company, call, individual, answer, avail, private, website

Representative FOI requests:

Example 1. DWP, 2 December 2011, successful

> The only phone number I have is 0845 6094904, which uses an automated call handling and queuing system. What call handling performance information is available, including calls abandoned?

Example 2. DWP, 19 September 2015, refused

> Please will you let me know the following: the name of your telephone system, how many months or years telephone recordings are saved, and if someone wanted a copy of all their calls (Subject Access Request), then what information do you need to enable this? Do you need time and date of calls? Or, do you just need the inbound telephone number in order to trace the call.

Example 3. HMRC, 10 August 2010, successful

> Has the HMRC considered replacing or clearly advertising 0300 rate phone numbers in addition to 0845 lo-call numbers? If not, when will this be considered and if yes, what was the outcome and please supply any documents or feasibility study into the use of 0300 numbers at HMRC including the Benefits Office/Job Centre Plus and National Insurance Contributions Centre?

Figure A-19: Expected topic proportions for the model with 28 topics (covariates: display_status (outcome), govBrown, govMay)
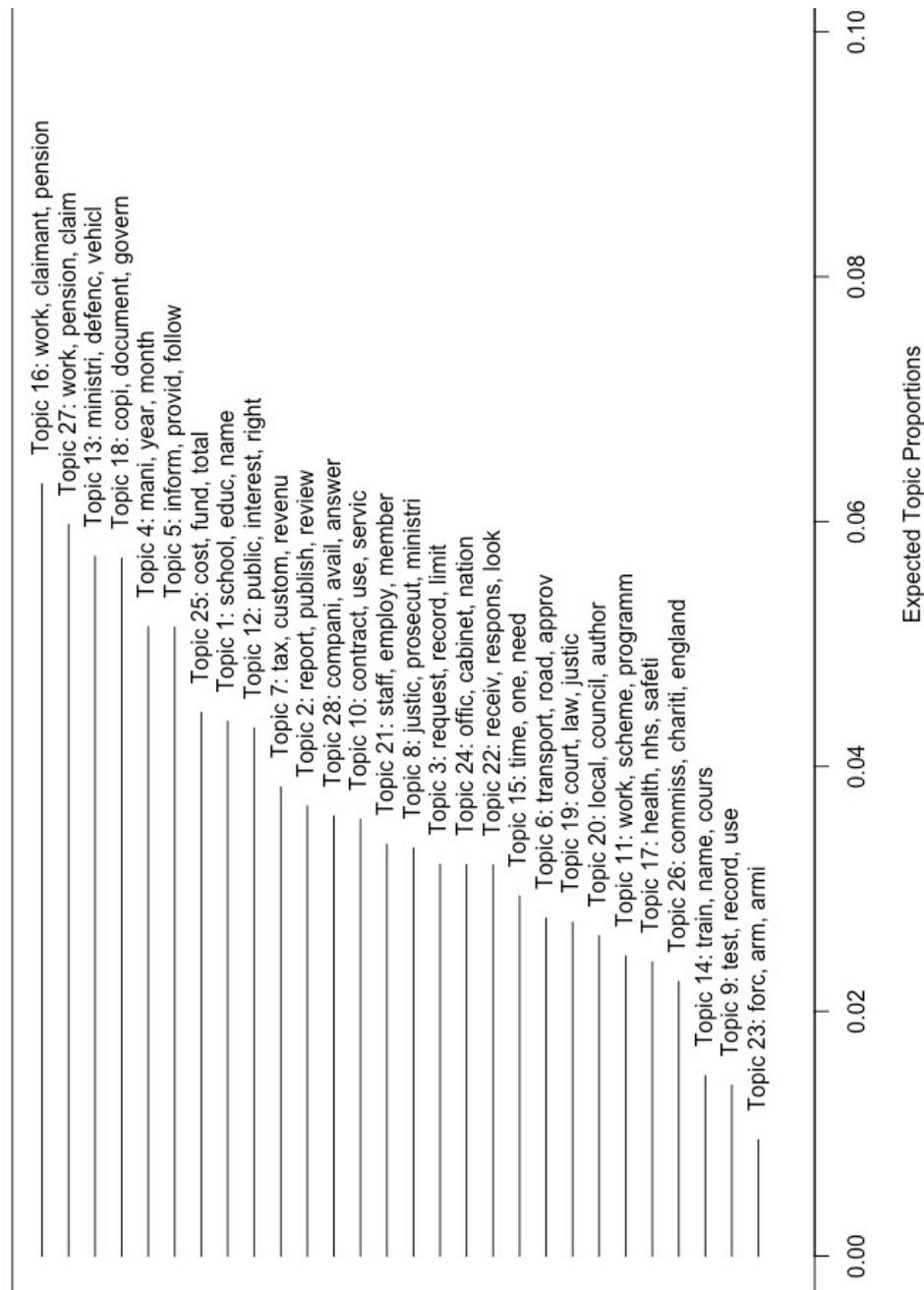
Note: Words used to label the topics are the words with the highest probability to occur within that topic. The words listed in the figure are the top three words associated with the topic. For more detailed information about the model see M. E. Roberts et al. (2014).

Figure A-20: Highest word probabilities for each topic in the model with 28 topics covariates: display_status (outcome), govBrown, govMay
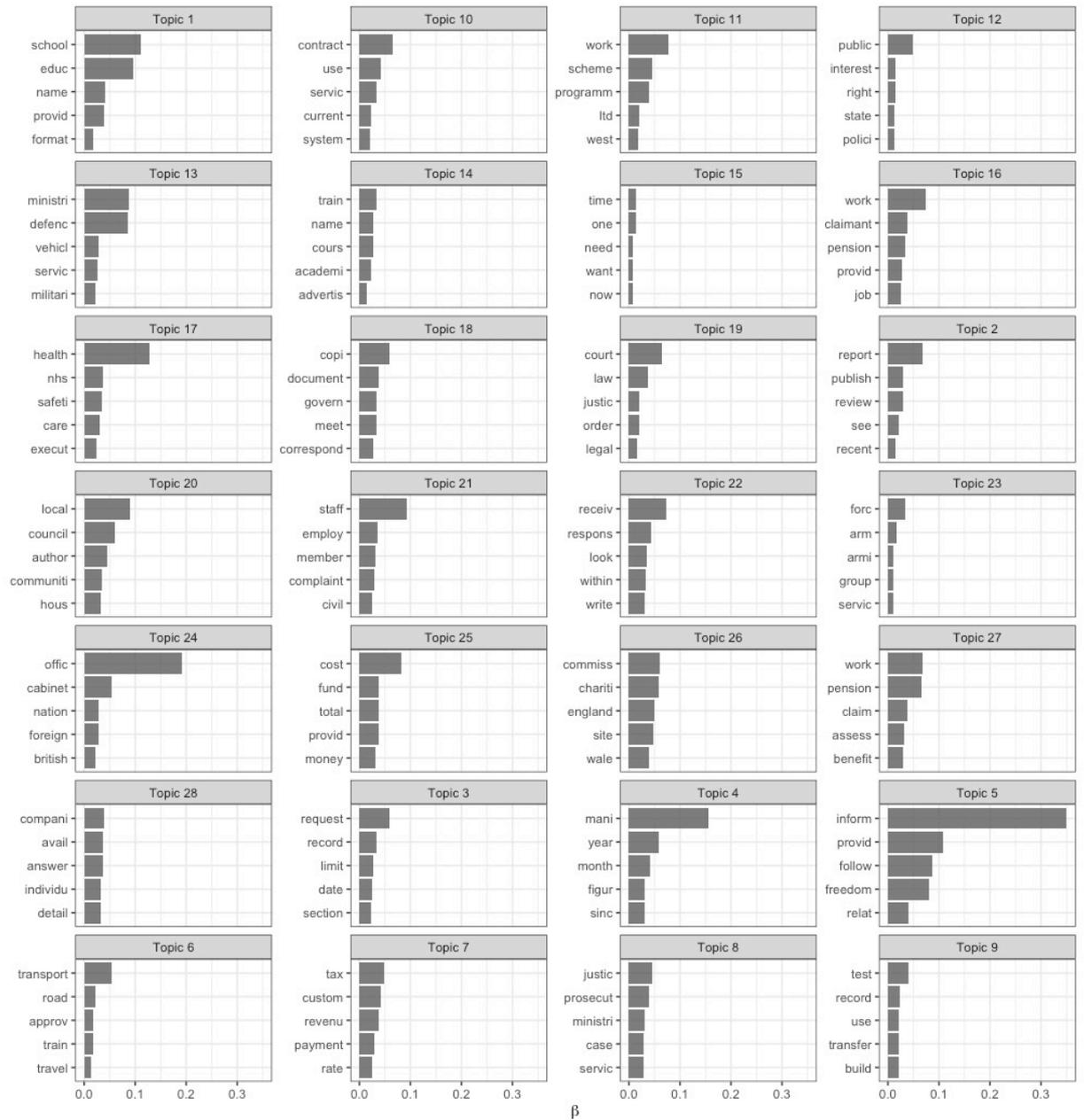
Note: The words listed in the figure are the top five words associated with the topic. For more detailed information about the model see M. E. Roberts et al. (2014).

Bibliography

Access Info Europe, & Centre for Law and Democracy. (2017a). Methodology. Retrieved February 14, 2018, from http://www.rti-rating.org/methodology/

Access Info Europe, & Centre for Law and Democracy. (2017b). Rating. Retrieved February 24, 2018, from http://www.rti-rating.org/country-rating/

Access to Medical Reports Act (1988). Retrieved from http://www.legislation.gov.uk/ukpga/1988/28/contents

Access to Personal Files Act (1987). Retrieved from http://www.legislation.gov.uk/ukpga/1987/37/enacted

Ackerman, J. (2004). Co-governance for accountability: Beyond "exit" and "voice." *World Development*, *32*(3), 447–463. https://doi.org/10.1016/j.worlddev.2003.06.015

Afful-Dadzie, E., & Afful-Dadzie, A. (2017). Liberation of public data: Exploring central themes in open government data and freedom of information research. *International Journal of Information Management*, *37*(6), 664–672. https://doi.org/10.1016/j.ijinfomgt.2017.05.009

Aghekyan, E., Bhatia, R., Dunham, J., O'Toole, S., Repucci, S., & Roylance, T. (2018). Freedom in the world 2018: democracy in crisis. Retrieved February 14, 2018, from https://freedomhouse.org/report/freedom-world/freedom-world-2018

Allen, D. S. (2008). The Trouble with Transparency. *Journalism Studies*, *9*(3), 323–340. https://doi.org/10.1080/14616700801997224

Allen, W. L., & Blinder, S. (2018). Media Independence through Routine Press-State Relations: Immigration and Government Statistics in the British Press. *International Journal of Press/Politics*, *23*(2), 202–226. https://doi.org/10.1177/1940161218771897

Almond, G. A. (1966). Political Theory and Political Science. *The American Political Science Review*, *60*(4), 869–879. https://doi.org/10.2307/1953762

Alvira, X. (2019). Vaccine hesitancy is a global public health threat. Are we doing enough about it? Retrieved January 2, 2020, from https://www.elsevier.com/connect/vaccine-hesitancy-is-a-global-public-health-threat-are-we-doing-enough-about-it

Ananny, M., & Crawford, K. (2015). A Liminal Press: Situating news app designers within a field of networked news production. *Digital Journalism*, *3*(2), 192–208. https://doi.org/10.1080/21670811.2014.922322

Andersson, S., & Heywood, P. M. (2009). The politics of perception: Use and abuse of transparency international's approach to measuring corruption. *Political Studies*, *57*(4), 746 – 767. https://doi.org/https://doi.org/10.1177/0010414013489081

Anduiza, E., Gallego, A., & Muñoz, J. (2013). Turning a Blind Eye: Experimental Evidence of Partisan Bias in Attitudes Toward Corruption. *Comparative Political Studies*, *46*(12), 1664–1692. https://doi.org/10.1177/0010414013489081

Angrist, J. D., & Pischke, J. S. (2015). *Mastering 'metrics: The path from cause to effect. Mastering 'Metrics: The Path from Cause to Effect*. Princeton, Oxford: Princeton University Press. https://doi.org/10.1093/erae/jbv011

Appelgren, E., & Lindén, C.-G. (2020). Data Journalism as a Service: Digital Native Data Journalism Expertise and Product Development. *Media and Communication*, *8*(2), 62–72. https://doi.org/10.17645/mac.v8i2.2757

Arendt, F., & Fawzi, N. (2019). Googling for Trump: investigating online information seeking during the 2016 US presidential election. *Information Communication and Society*, *22*(13), 1945–1955. https://doi.org/10.1080/1369118X.2018.1473459

Arnold, H. J., & Feldman, D. C. (1981). Social Desirability Response Bias in Self-Report Choice Situations. *Academy of Management Journal*, *24*(2), 377–385.

https://doi.org/10.2307/255848

Article 19, Freedom Now, Index on Censorship, Institute for Reporters' Freedom and Safety, International Media Support, PEN International, & Reporters without Borders. (2017). Joint submission to the universal periodic review of Azerbaijan. Retrieved February 14, 2018, from https://www.article19.org/wp-content/uploads/2017/10/Azerbaijan-UPR-Submission-Oct-2017.pdf

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 1–11. https://doi.org/10.1177/2053951715594634

Baack, S. (2018a). *Knowing what counts: How journalists and civic technologists use and imagine data*. University of Groningen. Retrieved from https://www.rug.nl/research/portal/en/publications/knowing-what-counts(4c94668a-c25c-43cb-9b36-5d54e3ff3c2e).html

Baack, S. (2018b). Practically Engaged: The entanglements between data journalism and civic tech. *Digital Journalism*, *6*(6), 673–692. https://doi.org/10.1080/21670811.2017.1375382

Bac, M. (2001). Corruption, connections and transparency: does a better screen imply a better scene? *Public Choice*, *107*, 87 – 96. https://doi.org/10.1023/A:1010349907813

Bågenholm, A. (2013). Throwing the rascals out? The electoral effects of corruption allegations and corruption scandals in Europe 1981-2011. *Crime, Law and Social Change*, *60*, 595–609. https://doi.org/10.1007/s10611-013-9482-6

Bagozzi, B. E., Berliner, D., & Almquist, Z. W. (2019). When does open government shut? Predicting government responses to citizen information requests. *Regulation and Governance*. https://doi.org/10.1111/rego.12282

Banisar, D. (2004). *The Freedominfo.org Global Survey: Freedom of Information and Access to Government Record Laws around the World*.

Banisar, D. (2006). *Freedom of Information around the World 2006: A Global Survey of Access to Government Information Laws*. London. Retrieved from http://www.freedominfo.org/documents/global_survey2006.pdf

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, *19*(1–2), 129–152. https://doi.org/10.3233/IP-140327

Bartlett, G., & Everett, M. (2017). *The Official Secrets Acts and Official Secrecy* (No. Number CBP07422). London, United Kingdom.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, *6*(3), 285–327. https://doi.org/10.1177/1476127008094339

Battaglia, M. P. (2008). Nonprobability sampling. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 523–526). Thousand Oaks, California: SAGE Publications Ltd.

Baume, S., & Papadopoulos, Y. (2018). Transparency: from Bentham's inventory of virtuous effects to contemporary evidence-based scepticism. *Critical Review of International Social and Political Philosophy*, *21*(2), 169–192. https://doi.org/10.1080/13698230.2015.1092656

Bayramov, G., Ibadoghlu, G., & Imanov, R. (2016). *Shadow report*. Baku.

Bello, O., Akinwande, V., Jolayemi, O., & Ibrahim, A. (2016). Open data portals in Africa: An analysis of open government data initiatives. *African Journal of Library Archives and Information Science*, *26*(2), 97–106.

Benjamin, S., Bhuvaneswari, Rajan, P., & Manjunatha. (2007). Bhoomi: 'E-Governance', or, an Anti-Politics Machine Necessary to Globalize Bangalore?

Bentham, J. (1999). Political Tactics. In M. James, C. Blamires, & C. Pease-Watkin (Eds.)

(1st ed., p. 267). Oxford: Oxford University Press.

Bentham, J. (2001). *Writings on the poor laws*. (M. Quinn, Ed.) (1st ed.). Oxford: Claredon Press.

Berliner, D. (2012). *Institutionalizing Transparency: The Global Spread of Freedom of Information in Law and Practice*. University of Washington. Retrieved from https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/21770/Berliner_washington_0250E_10985.pdf?sequence=1&isAllowed=y

Berliner, D. (2014). The political origins of transparency. *Journal of Politics*, *76*(2), 479–491. https://doi.org/10.1017/S0022381613001412

Berliner, D. (2016). Transnational advocacy and domestic law: International NGOs and the design of freedom of information laws. *Review of International Organizations*, *11*(1), 121–144. https://doi.org/10.1007/s11558-015-9228-6

Berliner, D., Bagozzi, B. E., & Palmer-Rubin, B. (2018). What information do citizens want? Evidence from one million information requests in Mexico. *World Development*, *109*, 222–235. https://doi.org/10.1016/j.worlddev.2018.04.016

Berliner, D., & Erlich, A. (2015). Competing for transparency: Political competition and institutional reform in Mexican states. *American Political Science Review*, *109*(1), 110–128. https://doi.org/10.1017/S0003055414000616

Besley, T., & Burgess, R. (2002). The political economy of government responsiveness: theory and evidence from India. *The Quarterly Journal of Economics*, *117*(4), 1415–1451. https://doi.org/10.1162/003355302320935061

Besley, T., Burgess, R., & Prat, A. (2002). Mass media and political accountability. In R. Islam (Ed.), *The right to tell: the role of mass media in economic development* (1st ed., pp. 45–60). Washington, DC: World Bank Institute. Retrieved from http://documents.worldbank.org/curated/en/957661468780322581/pdf/multi0page.pdf

Besley, T., & Prat, A. (2006). Handcuffs for the grabbing hand? Media capture and government accountability. *American Economic Review*, *96*(3), 720–736. https://doi.org/10.1257/aer.96.3.720

Bevir, M., & Rhodes, R. A. W. (2006). Prime ministers, presidentialism and westminster smokescreens. *Political Studies*, *54*(4), 671–690. https://doi.org/10.1111/j.1467-9248.2006.00632.x

Bichard, J.-A., & Knight, G. (2012). Improving public services through open data: public toilets. In *Proceedings of the Institution of Civil Engineers-Municipal Engineer* (Vol. 165, pp. 157–165). https://doi.org/10.1680/muen.12.00017

Birchall, C. (2014). Radical transparency? *Cultural Studies - Critical Methodologies*, *14*(1), 77–78. https://doi.org/10.1177/1532708613517442

Birchall, C. (2015). "Data.gov-in-a-box": Delimiting transparency. *European Journal of Social Theory*, *18*(2), 185–202. https://doi.org/10.1177/1368431014555259

Birkinshaw, P. (1997). Freedom of information and open government: The European community/union dimension. *Government Information Quarterly*, *14*(1), 23–27. https://doi.org/10.1016/S0740-624X(97)90050-2

Birkinshaw, P. (2006a). Freedom of information and openness: Fundamental human rights? *Administrative Law Review*, *58*(1), 177–218.

Birkinshaw, P. (2006b). Transparency as a human right. In *Transparency: The Key to Better Governance*. Retrieved from http://oxfordindex.oup.com/view/10.5871/bacad/9780197263839.003.0003

Birkinshaw, P. (2010). Freedom of information and its impact in the United Kingdom. *Government Information Quarterly*, *27*(4), 312–321. https://doi.org/10.1016/j.giq.2010.06.006

Blair, H. (2000). Participation and accountability at the periphery: Democratic local

governance in six countries. *World Development*, *28*(1), 21–39.

Blair, T. (2010). *A journey* (1st ed.). London: Hutchinson.

Blick, A., & Jones, G. (2010). *The Centre of Central Government* (LSE Public Policy Group No. 3). London, United Kingdom.

Borges-Rey, E. (2016). Unravelling Data Journalism: A study of data journalism practice in British newsrooms. *Journalism Practice*, *10*(7), 833–843. https://doi.org/10.1080/17512786.2016.1159921

Bounegru, L. (2012). Data Journalism in Perspective. In J. Gray, L. Bounegru, & L. Chambers (Eds.), *The Data Journalism Handbook*. European Journalism Centre. Retrieved from https://datajournalism.com/read/handbook/one/introduction/data-journalism-in-perspective

Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework. *European Law Journal*, *13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Bovens, M., Schillemans, T., & Hart, P. T. (2008). Does public accountability work? An assessment tool. *Public Administration*, *86*(1), 225–242. https://doi.org/10.1111/j.1467-9299.2008.00716.x

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370. https://doi.org/10.1007/BF02294361

Brace, I. (2008). *Questionnaire design: how to plan, structure and write survey material for effective market research* (2nd ed.). London: Kogan Page.

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design for market research, political polls, and social and health questionnaires*. San Francisco, CA: John Wiley & Sons, Incorporated.

Broad, E. (2015). Appeal to the FOI Commission: don't confuse open data with FOI. Retrieved February 2, 2018, from https://theodi.org/blog/appeal-to-the-foi-commission-dont-confuse-open-data-with-foi

Broersma, M. (2013). A Refractured Paradigm: Journalism, hoaxes and the challenge of trust. In C. Peters & M. Broersma (Eds.), *Rethinking Journalism: Trust and Participation in a Transformed News Landscape* (1st ed., pp. 28–44). Oxford, UK: Routledge.

Brooke, H. (2009, May 15). Unsung hero. *The Guardian*. Retrieved from https://www.theguardian.com/politics/2009/may/15/mps-expenses-heather-brooke-foi

Brugger, J., Fraefel, M., Riedl, R., Fehr, H., Schöneck, D., & Weissbrod, C. S. (2016). Current barriers to open government data use and visualization by political intermediaries. In *Proceedings of the 6th International Conference for E-Democracy and Open Government, CeDEM 2016* (pp. 219–229). https://doi.org/10.1109/CeDEM.2016.18

Brunetti, A., & Weder, B. (2003). A free press is bad news for corruption. *Journal of Public Economics*, *87*(7–8), 1801–1824. https://doi.org/10.1016/S0047-2727(01)00186-4

Bruno, J. R. (2017). Vigilance and Confidence: Jeremy Bentham, Publicity, and the Dialectic of Political Trust and Distrust. *American Political Science Review*, *111*(2), 295–307. https://doi.org/10.1017/S0003055416000708

Buckingham, D. (2011). Foreword. In M. Thomas (Ed.), *Deconstructing digital natives: Young people, technologies and the new literacies* (pp. ix–xi). Routledge.

Butler, D. M., & Broockman, D. E. (2011). Do politicians racially discriminate against constituents? A field experiment on state legislators. *American Journal of Political Science*, *55*(3), 463–477. https://doi.org/10.1111/j.1540-5907.2011.00515.x

Camaj, L. (2013). The Media's Role in Fighting Corruption: Media Effects on Governmental Accountability. *International Journal of Press/Politics*.

https://doi.org/10.1177/1940161212462741

Camaj, L. (2016). Governments' Uses and Misuses of Freedom of Information Laws in Emerging European Democracies: FOI Laws' Impact on News Agenda-Building in Albania, Kosovo, and Montenegro. *Journalism & Mass Communication Quarterly*, *93*(4), 923–945. https://doi.org/10.1177/1077699015610073

Camaj, L. (2018). Blurring the boundaries between journalism and activism: A transparency agenda-building case study from Bulgaria. *Journalism*, *19*(7), 994–1010. https://doi.org/10.1177/1464884916677757

Cameron, D. (2009, May 25). A new politics: Electoral reform. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2009/may/25/david-cameron-a-new-politics1

Carter, B. E. (2015). The Prime Minister and the Conduct of Foreign Affairs and War. In *Office of the Prime Minister* (pp. 295–330). Princeton, NJ: Princeton University Press.

Cherry, M., & McMenemy, D. (2013). Freedom of information and "vexatious" requests - The case of Scottish local government. *Government Information Quarterly*, *30*(3), 257–266. https://doi.org/10.1016/j.giq.2013.02.004

Cheruiyot, D., Baack, S., & Ferrer-Conill, R. (2019). Data Journalism Beyond Legacy Media: The case of African and European Civic Technology Organizations. *Digital Journalism*, *7*(9), 1215–1229. https://doi.org/10.1080/21670811.2019.1591166

Cheung, A. (2018). We need to know why so many FoI requests are being refused. Retrieved October 16, 2019, from https://www.instituteforgovernment.org.uk/blog/we-need-know-why-so-many-foi-requests-are-being-refused

Chowdhury, S. K. (2004). The effect of democracy and press freedom on corruption: An empirical test. *Economics Letters*, *85*(1), 93–101. https://doi.org/10.1016/j.econlet.2004.03.024

Chung, J., & Monroe, G. S. (2003). Exploring Social Desirability Bias. *Journal of Business Ethics*, *4*(4), 291–302. https://doi.org/10.1023/A:1023648703356

Clark, M., & Grech, A. (2017). *Journalists under pressure - Unwarranted interference, fear and self-censorship in Europe*. Retrieved from https://book.coe.int/en/human-rights-and-democracy/7284-journalists-under-pressure-unwarranted-interference-fear-and-self-censorship-in-europe.html

Coase, R. H. (1990). *The Firm, the Market, and the Law*. Chicago: University of Chicago Press.

Cocozza, P. (2013). Stats of the nation: how the state measures our lives. Retrieved October 18, 2019, from https://www.theguardian.com/uk-news/2013/jul/31/stats-nation-office-national-statistics

Committee to Protect Journalists. (2018). CPJ's database of attacks on the press: journalists killed. Retrieved February 14, 2018, from https://cpj.org/data/killed/2018/

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, *19*(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493

Cooley, A., & Snyder, J. (2015). *Ranking the world: Grading states as a tool of global governance*. (A. Cooley & J. Snyder, Eds.), *Ranking the World: Grading States as a Tool of Global Governance*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781316161555

Coppedge, M. (1999). Thickening Thin Concepts and Theories: Combining Large N and Small in Comparative Politics. *Comparative Politics*, *31*(4), 465–476. https://doi.org/10.2307/422240

Cornford, J., Wilson, R., Baines, S., & Richardson, R. (2013). Local governance in the new information ecology: The challenge of building interpretative communities. *Public*

Money and Management, *33*(3), 201–208. https://doi.org/10.1080/09540962.2013.785705

Costa, S. (2013). Do freedom of information laws decrease corruption? *Journal of Law, Economics, and Organization*, *29*(6), 1317–1343. https://doi.org/10.1093/jleo/ews016

Council of Europe. (1976). Motion for a recommendation: Introduction of a freedom of information act by member countries of the Council of Europe. Parliamentary Assembly. Retrieved from http://www.assembly.coe.int/nw/xml/XRef/Xref-DocDetails-EN.asp?FileID=4105&Lang=EN

da Silva Craveiro, G., & Albano, C. (2017). Open data intermediaries: Coproduction in budget transparency. *Transforming Government: People, Process and Policy*, *11*(1), 119–131. https://doi.org/10.1108/TG-12-2015-0057

Dahl, R. A. (1966). *Political Oppositions in Western Democracies*. (R. A. Dahl, Ed.) (1st ed.). New Haven, Connecticut: Yale University Press.

Dahlgreen, W. (2015). Memories of Iraq: did we ever support the war? Retrieved June 5, 2020, from https://yougov.co.uk/topics/politics/articles-reports/2015/06/03/remembering-iraq

Dai, J., & Li, Q. (2016). Designing audit apps for armchair auditors to analyze government procurement contracts. *Journal of Emerging Technologies in Accounting*, *13*(2), 71–88. https://doi.org/10.2308/jeta-51598

Davern, M. E. (2008). Representative sample. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 720–721). Thousand Oaks, California: SAGE Publications Ltd.

Davies, C., & Harding, L. (2018, February 28). Slovakian journalist "was investigating political links to Italian mafia." *The Guardian*. Retrieved from https://www.theguardian.com/world/2018/feb/28/slovakian-journalist-was-investigating-political-links-to-italian-mafia

Davies, T. (2010). *Open data, democracy and public sector reform: A look at open government data use from data.gov.uk*. Retrieved from http://www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf

Davies, T. (2014). Open data policies and practice: An international comparison. In *Paper for the European Consortium for Political Research Conference*.

de Fine Licht, J., & Naurin, D. (2016). Transparency. In C. Ansell & J. Torfing (Eds.), *Handbook on Theories of Governance* (pp. 217–224). Cheltenham, UK; Northampton, MA: Edward Elgar Publishing.

De Saulles, M. (2013). Open data and open government in the UK: how closely are they related? In *Proceedings of the 13th European Conference on E-Government* (pp. 160–165).

de Smaele, H. (2004). Limited access to information as a means of censorship in post-communist Russia. *Javnost*, *11*(2), 65–82. https://doi.org/10.1080/13183222.2004.11008854

de Tocqueville, A. (2002). *Democracy in America*. (trans. H. C. Mansfield and D. Winthrop, Ed.) (2nd ed.). Chicago, IL: University of Chicago Press. https://doi.org/10.2307/2679211

de Vaus, D. A. (2001). *Research design in social research*. SAGE Publications Ltd.

Dempsey, J. (2006). Hungary's Leader, Under Siege Over Lies, Refuses to Resign. Retrieved May 1, 2019, from https://www.nytimes.com/2006/09/20/world/europe/20hungary.html

Denscombe, M. (2009). *Ground rules for social research: guidelines for good practice*. Maidenhead Berkshire: McGraw Hill/Open University Press.

Deuze, M. (2005). What is journalism? Professional identity and ideology of journalists reconsidered. *Journalism*, *6*(4), 442–464. https://doi.org/10.1177/1464884905056815

Deuze, M. (2007). *Media Work*. Cambridge, UK: Polity Press.

Devarajan, S. (2013). Africa's statistical tragedy. *Review of Income and Wealth*, *59*, 9–15. https://doi.org/10.1111/roiw.12013

Diakopoulos, N., & Koliska, M. (2017). Algorithmic Transparency in the News Media. *Digital Journalism*, *5*(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053

Dilulio, J. D. (1994). Principled agents: the cultural bases of behavior in a federal government bureaucracy. *Journal of Public Administration Research and Theory*, *4*(3), 277–318. https://doi.org/10.1093/oxfordjournals.jpart.a037210

Djankov, S., McLiesh, C., Nenova, T., & Shleifer, A. (2003). Who owns the media? *The Journal of Law and Economics*, *46*(2), 341–382. https://doi.org/10.1086/377116

Dohoo, I. R., Ducrot, C., Fourichon, C., Donald, A., & Hurnik, D. (1997). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive Veterinary Medicine*, *29*(3), 221–239. https://doi.org/10.1016/S0167-5877(96)01074-4

Donnelly, L. (2015, February 15). Revealed: more than 500,000 home care visits last less than five minutes. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/news/health/news/11302534/Revealed-more-than-500000-home-care-visits-last-less-than-five-minutes.html

Downie Jr., L., & Rafsky, S. (2013). The Obama administration and the press. Retrieved February 16, 2018, from https://cpj.org/reports/2013/10/obama-and-the-press-us-leaks-surveillance-post-911.php

Dugan, E. (2014). Thousands with degenerative conditions classified as "fit to work in future" – despite no possibility of improvement. *The Independent*. Retrieved from http://www.independent.co.uk/news/uk/politics/thousands-with-degenerative-conditions-classified-as-fit-to-work-in-future-despite-no-possibility-of-9811910.html

Dunion, K. (2011). *Freedom of Information in Scotland in Practice.* Dundee: Dundee University Press.

Dyer, J. (2016). Fifty Years of FOIA. Retrieved July 18, 2019, from https://niemanreports.org/articles/fifty-years-of-foia/

Eccleston, R., & Woodward, R. (2014). Pathologies in International Policy Transfer: The Case of the OECD Tax Transparency Initiative. *Journal of Comparative Policy Analysis: Research and Practice*, *16*(3), 216–229. https://doi.org/10.1080/13876988.2013.854446

ECtHR. Case no. 68416/01, Steel and Morris v. the United Kingdom (2005). Retrieved from http://hudoc.echr.coe.int/eng?i=001-68224

ECtHR. Case no. 64752/01, Voskuil v. The Netherlands (2008). Retrieved from http://hudoc.echr.coe.int/eng?i=001-83413

ECtHR. Case no. 31475/05, Kenedi v. Hungary (2009). Retrieved from http://hudoc.echr.coe.int/eng?i=001-92663

ECtHR. Case No. 48135/06, Youth Initiative for Human Rights v. Serbia. (2013). Retrieved from http://hudoc.echr.coe.int/eng?i=001-120955

Edwards, R. (2019). Journalists call for reform to stop public agencies blocking freedom of information. Retrieved June 12, 2019, from https://theferret.scot/journalists-public-agencies-freedom-of-information/

Elliott, J. (2016, March 11). The federal government no longer cares about disclosing public information. *The Washington Post*. Retrieved from https://www.washingtonpost.com/opinions/the-federal-government-no-longer-cares-about-disclosing-public-information/2016/03/10/7e0bf1bc-e631-11e5-a6f3-

21ccdbc5f74e_story.html?utm_term=.1fe2fe442b9c

Environment and Safety Information Act (1988). Retrieved from http://www.legislation.gov.uk/ukpga/1988/30/contents

Escaleras, M., Lin, S., & Register, C. (2010). Freedom of information acts and public sector corruption. *Public Choice*, *145*(3–4), 435–460. https://doi.org/10.1007/s11127-009-9574-0

Feld, L. P., & Voigt, S. (2003). Economic growth and judicial independence: Cross-country evidence using a new set of indicators. *European Journal of Political Economy*, *19*(3), 497–527. https://doi.org/10.1016/S0176-2680(03)00017-X

Fenster, M. (2006). The opacity of transparency. *Iowa Law Review*, *91*(3), 885–949.

Ferraz, C., & Finan, F. (2008). Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, *123*(2), 703–745. https://doi.org/10.1162/qjec.2008.123.2.703

Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States. *Journalism Studies*, *16*(4), 467–481. https://doi.org/10.1080/1461670x.2014.939852

Fletcher, D. R. (2011). Welfare Reform, Jobcentre Plus and the Street-Level Bureaucracy: Towards Inconsistent and Discriminatory Welfare for Severely Disadvantaged Groups? *Social Policy and Society*, *10*(4), 445–458. https://doi.org/10.1017/S1474746411000200

Foley, M. (2000). *The British Presidency: Tony Blair and the Politics of Public Leadership*. Manchester, UK: Manchester University Press.

Fowler Jr., F. J., & Cosenza, C. (2008). Design and Evaluation of Survey Questions. In L. Bickman & D. J. Rog (Eds.), *The SAGE Handbook of Applied Social Research Methods* (2nd ed.). London, United Kingdom: SAGE Publications Ltd. https://doi.org/10.4135/9781483348858

Fox, J. (2000). Civil Society and Political Accountability: Propositions for Discussion. In *Institutions, Accountability and Democratic Governance in Latin America*. Notre Dame: University of Notre Dame.

Fox, J. (2007). The uncertain relationship between transparency and accountability. *Development in Practice*, *17*(4–5), 663–671. https://doi.org/10.1080/09614520701469955

Fray, W. C., & Spar, L. A. (1996a). Constitution of the United States: Bill of Rights. *The Avalon Project at the Yale Law School: Documents in Law, History and Diplomacy*. New Haven, Connecticut: University of Yale.

Fray, W. C., & Spar, L. A. (1996b). Declaration of Independence - July 4, 1776. *The Avalon Project at the Yale Law School: Documents in Law, History and Diplomacy*. New Haven, Connecticut: University of Yale. Retrieved from http://avalon.law.yale.edu/18th_century/declare.asp

Fray, W. C., & Spar, L. A. (1996c). Declaration of the Rights of Man - 1789. *The Avalon Project at the Yale Law School: Documents in Law, History and Diplomacy*. New Haven, Connecticut: University of Yale. Retrieved from http://avalon.law.yale.edu/18th_century/rightsof.asp

Fredom House. (2017). Freedom of the Press 2017: Serbia profile. Retrieved December 22, 2019, from https://freedomhouse.org/report/freedom-press/2017/serbia

Freedom House. (2017a). Azerbaijan: Freedom on the Net 2017. Retrieved February 21, 2018, from https://freedomhouse.org/report/freedom-net/2017/azerbaijan

Freedom House. (2017b). Freedom of the Press 2017 Methodology. Retrieved May 14, 2018, from https://freedomhouse.org/report/freedom-press-2017-methodology

Freedom of Information (Scotland) Act (2002). Scottish Parliament. Retrieved from http://www.legislation.gov.uk/asp/2002/13/section/32

Freedom of Information Act (2000). Retrieved from http://www.legislation.gov.uk/ukpga/2000/36/section/27

Freeguard, G. (2015). Freedom of information and government effectiveness. Retrieved February 2, 2018, from https://www.instituteforgovernment.org.uk/blog/freedom-information-and-government-effectiveness

French, D. (1978). Spy Fever in Britain, 1900-1915. *The Historical Journal*, *21*(2), 355–370. https://doi.org/10.1017/S0018246X00000571

Fukuyama, F. (2001). Social capital, civil society and development. *Third World Quarterly*, *22*(1), 7–20. https://doi.org/10.1080/713701144

Fung, A. (2013). Infotopia: unleashing the democratic power of transparency. *Politics & Society*, *41*(2), 183–212. https://doi.org/10.1177/0032329213483107

Galka, M. (2017). Who Uses FOIA? – An Analysis of 229,000 Requests to 85 Government Agencies. Retrieved July 22, 2019, from https://foiamapper.com/who-uses-foia/

Galnoor, I. (1977). *Government secrecy in democracies*. New York: Harper and Row.

Garrison, B. (1998). Coping with New Technologies in the Newsroom. *Convergence: The International Journal of Research into New Media Technologies*, *4*(4), 18–25. https://doi.org/10.1177/135485659800400403

Garthwaite, K. (2011). "The language of shirkers and scroungers?" Talking about illness, disability and coalition welfare reform. *Disability and Society*, *26*(3), 369–372. https://doi.org/10.1080/09687599.2011.560420

Garthwaite, K. (2014). Fear of the brown envelope: Exploring welfare reform with long-term sickness benefits recipients. *Social Policy and Administration*, *48*(7), 782–798. https://doi.org/10.1111/spol.12049

Gatti, R. (2004). Explaining corruption: Are open countries less corrupt? *Journal of International Development*, *16*, 851–861. https://doi.org/10.1002/jid.1115

Gay, O. (1997). *The Code of Practice on Access to Official Information* (Research Paper No. 97/69). London, United Kingdom.

Gellner, E. (1994). *Conditions of Liberty: Civil Society and Its Rivals*. London: Hamish Hamilton.

Geroge, Alexander y Bennet, A. (2005). Case Studies and Theory Development in the Social Sciences. *MIT Press*, 432.

Gerring, J. (2007). Is There a (Viable) Crucial-Case Method? *Comparative Political Studies*, *40*(3), 231–253. https://doi.org/10.1177/0010414006290784

Gerring, J. (2012). *Social science methodology: A unified framework* (2nd ed.). Cambridge, UK: Cambridge University Press.

Gerstein, J. (2012, March 5). Obama's muddy transparency record. *Politico*. Retrieved from https://www.politico.com/story/2012/03/obamas-transparency-stumbles-073606?paginate=false

Gillham, B. (2000). *The research interview*. London: : Continuum.

Goertz, G. (2008). Concepts, Theories, and Numbers: A Checklist for Constructing, Evaluating, and Using Concepts or Quantitative Measures. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford Handbook of Political Methodology* (p. 880). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199286546.001.0001

Goodall, J., & Gay, O. (2010). *Freedom of Information: the first five years*. London, United Kingdom.

Government of Albania. (2018). *The Open Government Partnership Fourth Open Government National Action Plan for Albania 2018-2020*. Retrieved from https://www.opengovpartnership.org/wp-content/uploads/2019/01/Albania_Action-Plan_2018-2020_EN.pdf

Government of Canada. (2014). *Canada's action plan on open government 2014 - 2016*. Ottawa. Retrieved from https://open.canada.ca/en/content/canadas-action-plan-open-government-2014-16

Granickas, K. (2014). *Open Data as a Tool to Fight Corruption*. Retrieved from https://ofti.org/wp-content/uploads/2014/05/221171136-Open-Data-as-a-Tool-to-Fight-Corruption.pdf

Greene, D., O'Callaghan, D., & Cunningham, P. (2014). How many topics? Stability analysis for topic models. *Lecture Notes in Computer Science*, *8724*(1), 498–513. https://doi.org/10.1007/978-3-662-44848-9_32

Greene, J. C. (2007). *Mixed methods in social inquiry* (1st ed.). San Francisco, CA: Jossey-Bass.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Grimes, M. (2013). The contingencies of societal accountability: examining the link between civil society and good government. *Studies in Comparative International Development*, *48*(4), 380–402. https://doi.org/10.1007/s12116-012-9126-3

Grimmelikhuijsen, S. G., & Welch, E. W. (2012). Developing and Testing a Theoretical Framework for Computer-Mediated Transparency of Local Governments. *Public Administration Review*. https://doi.org/10.1111/j.1540-6210.2011.02532.x

Grimmer, J. (2010). A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis*, *18*(1), 1–35. https://doi.org/10.1093/pan/mpp034

Grimmer, J., & King, G. (2011). General Purpose Computer-Assisted Clustering and Conceptualization. In *Proceedings of the National Academy of Sciences* (pp. 1–8). https://doi.org/10.1073/pnas.1018067108

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Grønbech-Jensen, C. (1998). The scandinavian tradition of open government and the European union: Problems of compatibility? *Journal of European Public Policy*, *5*(1), 185–199. https://doi.org/10.1080/13501768880000091

Gurstein, M. B. (2011). (2011). Open data: Empowering the empowered or effective data use for everyone? | Gurstein | First Monday. *First Monday*, *16*(2), 1–8. https://doi.org/10.5210/fm.v16i2.3316

Habermas, J. (1984). *The theory of communicative action* (1st ed.). Boston, MA: Beacon Press.

Habermas, J. (1998). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. (trans. W. Rehg, Ed.) (1st ed.). Boston, MA: MIT Press. https://doi.org/10.2307/2941077

Haider Malik, W. (2007). *Judiciary-Led Reforms in Singapore: Framework, Strategies, and Lessons*. Washington, DC. Retrieved from http://documents.worldbank.org/curated/en/641511468300664743/pdf/387790SG0Judic101OFFICIAL0USE0ONLY1.pdf

Hallin, D. C., & Papathanassopoulos, S. (2002). Political clientelism and the media: southern Europe and Latin America in comparative perspective. *Media, Culture & Society*, *24*(2), 175–195. https://doi.org/10.1177/016344370202400202

Hallin, D., & Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics*. (Cambridge University Press, Ed.). Cambridge, UK.

Hamilton, J. T. (2016). *Democracy's Detectives: The Economics of Investigative*

Journalism. Cambridge, MA: Harvard University Press.

Hamilton, J. T. (2018). FOIA and Investigative Reporting: Who's Asking What, Where, and When and Why It Matters. In D. E. Pozen & M. Schudson (Eds.), ***Troubling Transparency: The History and Future of Freedom of Information*** (pp. 116–134). New York, NY: Columbia University Press. https://doi.org/10.1017/jli.2019.7

Hancock, L., & Mooney, G. (2013). "Welfare Ghettos" and the "Broken Society": Territorial Stigmatization in the Contemporary UK. ***Housing, Theory and Society***, *30*(1), 46–64. https://doi.org/10.1080/14036096.2012.683294

Hansen, G. (1996). ***Constituencies for Reform: Strategic Approaches for Donor-Supported Civic Advocacy Programs***. Washington, DC.

Hargittai, E. (2010). Digital Na(t)ives? Variation in internet skills and uses among members of the "net Generation." ***Sociological Inquiry***, *80*(1), 92–113. https://doi.org/10.1111/j.1475-682X.2009.00317.x

Hayes, A. F. (2018). ***Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*** (2nd ed.). New York, NY: Guilford Press.

Hayes, J. (2009). ***A Shock to the System: Journalism, Government and the Freedom of Information Act 2000***. Oxford, UK. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-11/A Shock to the System.pdf

Hayo, B., & Voigt, S. (2014). Mapping Constitutionally safeguarded judicial independence-a global survey. ***Journal of Empirical Legal Studies***, *11*(1), 159–195. https://doi.org/10.1111/jels.12038

Heald, D. (2006). Varieties of transparency. In C. Hood & D. Heald (Eds.), ***Transparency The Key to Better Governance?*** (pp. 25–43). Oxford: Oxford University Press. https://doi.org/10.5871/bacad/9780197263839.001.0001

Heffernan, R. (2013). There's no need for the "-isation": The prime minister is merely prime ministerial. ***Parliamentary Affairs***, *66*(3), 636–645. https://doi.org/10.1093/pa/gss058

Heimstädt, M. (2017). Openwashing: A decoupling perspective on organizational transparency. ***Technological Forecasting and Social Change***, *125*, 77–86. https://doi.org/10.1016/j.techfore.2017.03.037

Hellberg, A.-S., & Hedström, K. (2015). The story of the sixth myth of open data and open government. ***Transforming Government: People, Process and Policy***, *9*(1), 35–51. https://doi.org/10.1108/TG-04-2014-0013

Hellmueller, L., Vos, T. P., & Poepsel, M. A. (2013). Shifting journalistic capital? Transparency and objectivity in the twenty-first century. ***Journalism Studies***, *14*(3), 287–304. https://doi.org/10.1080/1461670X.2012.697686

Hermida, A., & Young, M. L. (2017). Finding the Data Unicorn. ***Digital Journalism***, *5*(2), 159–176. https://doi.org/10.1080/21670811.2016.1162663

Hewett, J. (2016). Learning to teach data journalism: Innovation, influence and constraints. ***Journalism***, *17*(1), 119–137. https://doi.org/10.1177/1464884915612681

Hewitt, S. (2009). ***Discourse Analysis and Public Policy Research*** (Centre for Rural Economy Discussion Paper Series No. 24). ***Centre for Rural Economy Discussion Paper Series***. Newcastle, UK. Retrieved from https://www.ncl.ac.uk/media/wwwnclacuk/centreforruraleconomy/files/discussion-paper-24.pdf

Heywood, P. M., & Rose, J. (2013). Close but no cigar: the measurement of corruption. ***Journal of Public Policy***, *34*(3), 507–529. https://doi.org/10.1017/S0143814X14000099

Hohfeld, W. N. (1913). Some Fundamental Legal Conceptions as Applied in Judicial Reasoning. *The Yale Law Journal*, *23*(1), 16–59.

Holzinger, K., & Knill, C. (2005). Causes and conditions of cross-national policy convergence. In *Journal of European Public Policy* (Vol. 12, pp. 775–796). https://doi.org/10.1080/13501760500161357

House of Commons Justice Committee. (2012). *Post-legislative scrutiny of the Freedom of Information Act 2000*. London, United Kingdom. Retrieved from https://publications.parliament.uk/pa/cm201213/cmselect/cmjust/96/96.pdf

Howard, A. B. (2014). The Art and Science of Data-Driven Journalism. *Tow Centre for Digital Journalism*. Columbia Journalism School. https://doi.org/10.1017/CBO9781107415324.004

Howe, J. (2006). Crowdsourcing: A definition. Retrieved January 29, 2018, from http://crowdsourcing.typepad.com/

Information Commissioner's Office. (2015). *Consideration of the identity or motives of the applicant - Freedom of Information Act and Environmental Information Regulations*. Wilmslow, United Kingdom. Retrieved from https://ico.org.uk/media/for-organisations/documents/1043418/consideration-of-the-identity-or-motives-of-the-applicant.pdf

Information Commissioner's Office. (2016). Intellectual property rights and disclosures under the Freedom of Information Act. Retrieved January 30, 2018, from https://ico.org.uk/media/for-organisations/documents/1150/intellectual_property_rights_and_disclosures_under_the_foia.pdf

Institute for Government. (2019). Freedom of Information. Retrieved December 22, 2019, from https://www.instituteforgovernment.org.uk/explainers/freedom-of-information

Inter-American Court of Human Rights. Claude-Reyes et al. v. Chile (2006). Retrieved from http://www.corteidh.or.cr/docs/casos/articulos/seriec_151_ing.pdf

IRM Staff. (2016). *Azerbaijan: 2012-2015 End of term report*. Washington, DC. Retrieved from https://www.opengovpartnership.org/sites/default/files/Azerbaijan_EOTR_2012-2015_for-public-comment.pdf

Jackman, R. W. (1985). Cross-National Statistical Research and the Study of Comparative Politics. *American Journal of Political Science*, *29*(1), 161–182. https://doi.org/10.2307/2111217

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption Barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Jones, G. (2016). *The Power of the Prime Minister: 50 Years On*. London, United Kingdom.

Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. *Strategic Management Journal*, *39*(8), 2362–2385. https://doi.org/10.1002/smj.2783

Kant, I. (2015). *Perpetual Peace*. North Charleston, South Carolina: CreateSpace Independent Publishing Platform 2015.

Kassen, M. (2017). Open data in Kazakhstan: Incentives, implementation and challenges. *Information Technology & People*, *30*(2), 301–323. https://doi.org/10.1108/ITP-10-2015-0243

Katz, J. M. (1969). Games Bureaucrats Play: Hide and Seek under the Freedom of Information Act. *Texas Law Review*, *48*, 1261.

Kavanagh, D. (2007). The Blair premiership. In A. Seldon (Ed.), *Blair's Britain 1997-2007*

(pp. 3–15). Cambridge, UK: Cambridge University Press.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing Social Inquiry. Designing Social Inquiry*. Princeton, NJ: Princeton University Press.

Kitchin, R. (2014a). Open and linked data. In *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences* (pp. 48–66). London: SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

Kitchin, R. (2014b). *The data revolution: big data, open data, data infrastructures & their consequences*. London: SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

Knight, M. (2015). Data journalism in the UK: A preliminary analysis of form and content. *Journal of Media Practice*, *16*(1), 55–72. https://doi.org/10.1080/14682753.2015.1015801

Kornberger, M., Meyer, R. E., Brandtner, C., & Höllerer, M. A. (2017). When bureaucracy meets the crowd: Studying "open government" in the Vienna city administration. *Organization Studies*, *38*(2), 179–200. https://doi.org/10.1177/0170840616655496

Kritzinger, S. (2003). Public opinion in the Iraq crisis: Explaining developments in Italy, the UK, France and Germany. *European Political Science*, *3*(1), 30–35.

Kuhn, K. (2011). Open government data and public transportation. *Journal of Public Transportation*, *14*(1), 83–97. https://doi.org/10.5038/2375-0901.14.1.5

Kuhn, T. (1996). *The Structure of Scientific Revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.

Kwoka, M. (2016). FOIA, Inc. *Duke Law Journal*, *65*, 1361–1437.

Kwoka, M. (2018a). First-person FOIA. *Yale Law Journal*, *127*(8), 2204–2585.

Kwoka, M. (2018b). The Other FOIA Requesters. In D. E. Pozen & M. Schudson (Eds.), *Troubling Transparency: The History and Future of Freedom of Information* (pp. 73–90). New York, NY: Columbia University Press. https://doi.org/10.1017/jli.2019.7

Laperrière, A. (2019, June). Open data and food security. *The Ecologist*. Retrieved from https://theecologist.org/2019/jun/07/open-data-and-food-security

Larkins, C. M. (1996). Judicial Independence and Democratization: A Theoretical and Conceptual Analysis. *American Journal of Comparative Law*, *44*(4), 605–626. https://doi.org/10.2307/840623

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*. https://doi.org/10.1017/S0003055403000698

Lecheler, S., & Kruikemeier, S. (2016). Re-evaluating journalistic routines in a digital age: A review of research on the use of online sources. *New Media and Society*, *18*(1), 156–171. https://doi.org/10.1177/1461444815600412

Léchenet, A. (2014). *Global Database Investigations: The role of the computer-assisted reporter*. Oxford, UK. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Global%2520 Database%2520Investigations%2520-%2520The%2520role%2520of%2520the%2520computer-assisted%2520reporter.pdf

Lee, R. M. (2001). Research Uses of the U.S. Freedom of Information Act. *Field Methods*, *13*(4), 370–391. https://doi.org/10.1177/1525822X0101300404

Levitsky, S., & Way, L. A. (2002). Elections Without Democracy: The rise of competitive authoritarianism. *Journal of Democracy*, *13*(2), 51–65. https://doi.org/10.1353/jod.2002.0026

Levy, J. S. (2008). Case studies: Types, designs, and logics of inference. *Conflict Management and Peace Science*, *25*(1), 1–18.

https://doi.org/10.1080/07388940701860318

Lewis, S. C., & Westlund, O. (2015). Big Data and Journalism. *Digital Journalism*, *3*(3), 447–466.

Lewis, W. A. (1965). *Politics in West Africa*. London, United Kingdom: George Allen & Unwin.

Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review*, *99*(3), 435–452. https://doi.org/10.1017/S0003055405051762

Lindstedt, C., & Naurin, D. (2010). Transparency is not enough: Making transparency effective in reducing corruption. *International Political Science Review*, *31*(3), 301–322. https://doi.org/10.1177/0192512110377602

Lourenço, P. R., Piotrowski, S., & Ingrams, A. (2017). Open data driven public accountability. *Transforming Government: People, Process and Policy*, *11*(1), 42–57. https://doi.org/10.1108/TG-12-2015-0050

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, *32*(3), 323–332. https://doi.org/10.1016/j.giq.2015.05.006

Lugo-Ocando, J., & Brandão, R. F. (2016). Stabbing News. *Journalism Practice*, *10*(6), 715–729. https://doi.org/10.1080/17512786.2015.1058179

Maier, S. R. (2002). Numbers in the News: A mathematics audit of a daily newspaper. *Journalism Studies*, *3*(4), 507–519. https://doi.org/10.1080/1461670022000019191

Margetts, H. (2006). Transparency and digital government. In C. Hood & D. Heald (Eds.), *Transparency The Key to Better Governance?* (pp. 197–207). Oxford: Oxford University Press. https://doi.org/10.5871/bacad/9780197263839.001.0001

Martin, C. (2014). Barriers to the open government data agenda: taking a multi-level perspective. *Policy & Internet*, *6*(3), 217–240. Retrieved from http://doi.wiley.com/10.1002/1944-2866.POI367

Mason, J. (2002). *Qualitative researching*. London: SAGE Publications Ltd.

Maude, F. (2009). Ready to deliver change. Retrieved July 25, 2019, from https://conservative-speeches.sayit.mysociety.org/speech/601301

Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics*, *110*(3), 681–712. https://doi.org/10.2307/2946696

Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London: John Murray Publishers.

Mayernik, M. S. (2017). Open data: Accountability and transparency. *Big Data & Society*, *4*(2), 1–5. https://doi.org/10.1177/2053951717718853

McCarthy, D. R., & Fluck, M. (2016). Leaky politics: the false promise of transparency. Retrieved January 30, 2018, from https://www.opendemocracy.net/wfd/daniel-mccarthy-matthew-fluck/leaky-politics

McClean, T. (2011). *Shackling Leviathan: a comparative historical study of institutions and the adoption of Freedom of Information.* London School of Economics and Political Science.

McGee, R., & Edwards, D. (2016). Introduction: Opening governance – change, continuity and conceptual ambiguity. *IDS Bulletin: Transforming Development Knowledge*, *47*(1), 1–22. https://doi.org/10.19088/1968-2016.103

McKenna, C. (2017, November 28). The British Columbians who die waiting for an ambulance. *Vice*. Retrieved from https://www.vice.com/en_ca/article/ne38z7/the-british-columbians-who-die-waiting-for-an-ambulance

McMillan, F. (2018). *The Meaning of Mandates: Party Competition and the Programme-to-Policy Linkage*. University of Strathclyde.

Meijer, A., 't Hart, P., & Worthy, B. (2018). Assessing Government Transparency: An Interpretive Framework. *Administration and Society*, *50*(4), 501–526. https://doi.org/10.1177/0095399715598341

Meyer, P. (1973). *Precision Journalism: A Reporter's Introduction to Social Science Methods*. Bloomington, Indiana: Indiana University Press.

Michener, G. (2010). *The surrender of secrecy: explaining the emergence of strong access to information laws in Latin America*. University of Texas, Austin. Retrieved from https://repositories.lib.utexas.edu/handle/2152/ETD-UT-2010-05-1112

Michener, G. (2011). FOI laws around the world. *Journal of Democracy*, *22*(2), 145–159. https://doi.org/10.1353/jod.2011.0021

Michener, G., & Rodrigues, K. F. (2018). A necessidade de identificação como barreira ao acesso à informação: evidências e práticas no Brasil e no mundo. *Administração Pública e Gestão Social*, *10*(4), 226–327.

Michener, G., Velasco, R. B., Contreras, E., & Rodrigues, K. F. (2019). Googling the Requester: Identity - Questing and Discrimination in Public Service Provision. *Governance*, 1–19. https://doi.org/10.1111/gove.12416

Michener, G., & Worthy, B. (2018). The Information-Gathering Matrix: A Framework for Conceptualizing the Use of Freedom of Information Laws. *Administration and Society*, *50*(4), 476–500. https://doi.org/10.1177/0095399715590825

Miller, C. (2015, June 10). Shocking figures reveal police have no officers available for thousands of 999 calls. *The Mirror*. Retrieved from https://www.mirror.co.uk/news/uk-news/shocking-figures-reveal-police-no-5859698

Millner, J. (2014, April 14). How open data is being used to hack climate change. *The Guardian*. Retrieved from https://www.theguardian.com/media-network/media-network-blog/2014/apr/14/climate-change-flooding-open-data

Mills, L. C. (2017). Catching the 'big fish': The (ab)use of corruption-related prosecutions across sub-Saharan Africa. *Development Policy Review*, *35*, 160–177. https://doi.org/10.1111/dpr.12326

Moore, M. (2011). The limits of transparency. *Political Quarterly*, *82*(4), 506–508. https://doi.org/10.1111/j.1467-923X.2011.02248.x

Morozov, E. (2013). *To save everything, click here: technology, solutionism, and the urge to fix problems that don't exist*. London: Penguin Books Ltd.

Mullin, B. (2016). More than 250 staffers at the Guardian have accepted buyouts. Retrieved January 30, 2018, from https://www.poynter.org/news/more-250-staffers-guardian-have-accepted-buyouts

Mutua, M. W. (1996). The Ideology of Human Rights. *Virginia Journal of International Law*, *63*(3), 589–657.

Nation. (2015, March 18). Obama administration sets new record for withholding FOIA requests. *PBS NewsHour*.

Naurin, D. (2006). Transparency, Publicity, Accountability - The Missing Links. *Swiss Political Science Review*, *12*(3), 90–98.

Nelson, D. (1999). Scotland to take lead on freedom of information. *The Guardian*. Retrieved from https://www.theguardian.com/politics/1999/nov/21/freedomofinformation.uk

Neuman, L. (2016). The right of access to information: exploring gender inequities. *IDS Bulletin: Transforming Development Knowledge*, *47*(1), 83–98. https://doi.org/10.19088/1968-2016.108

No author. (2007). The Annotated 8 Principles of Open Government Data. Retrieved January 29, 2018, from https://opengovdata.org/

Norris, P. (2008). The fourth estate. In 1 (Ed.), *Driving democracy: do power-sharing institutions work?* (p. 320). New York/Cambridge: Cambridge University Press.

Noveck, B. S. (2009). *Wiki government: how technology can make government better, democracy stronger, and citizens more powerful.* Washington, DC: Brookings Institution Press. https://doi.org/10.5860/choice.47-2862

Noveck, B. S. (2017). Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency. *Yale Human Rights and Development Law Journal*, *19*(1), 1–45.

Noveck, B. S. (2018). Open Data: The Future of Transparency in the Age of Big Data. In D. E. Pozen & M. Schudson (Eds.), *Troubling Transparency: The History and Future of Freedom of Information* (1st ed., pp. 206–225). New York, NY: Columbia University Press. https://doi.org/10.1017/jli.2019.7

O'Neill, O. (2006). Transparency and the ethics of communication. In C. Hood & D. Heald (Eds.), *Transparency The Key to Better Governance?* (pp. 75–90). Oxford: Oxford University Press. https://doi.org/10.5871/bacad/9780197263839.001.0001

OECD. (2003). *Fighting Corruption: What Role for Civil Society? The Experience of the OECD.* Paris. Retrieved from https://www.oecd.org/daf/anti-bribery/anti-briberyconvention/19567549.pdf

OECD. (2016). *Compendium of good practices on the publication and reuse of open data for Anti-corruption across G20 countries: Towards data-driven public sector integrity and civic auditing.* Paris. Retrieved from https://www.oecd.org/gov/digital-government/g20-oecd-compendium.pdf

Ofcom. (2019). The Ofcom Broadcasting Code. London, United Kingdom: Ofcom. Retrieved from https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code

Open Data User Group. (2015). April 2015: Minutes of the Open Data User Group. London, United Kingdom. Retrieved from https://www.gov.uk/government/groups/open-data-user-group

Open Government Partnership. (2018). Eligibility Criteria. Retrieved February 25, 2019, from https://www.opengovpartnership.org/resources/eligibility-criteria

Origlia, C., Cersosimo, D., Bianchi, T., & Fortunato, A. (2016). Assessing an open government data initiative. the case of OpenCoesione. In *Proceedings of the 6th International Conference for E-Democracy and Open Government, CeDEM 2016* (pp. 164–171). https://doi.org/10.1109/CeDEM.2016.26

Örnebring, H., & Mellado, C. (2018). Valued skills among journalists: An exploratory comparison of six European nations. *Journalism*, *19*(4), 445–463. https://doi.org/10.1177/1464884916657514

Ostrom, E. (1996). Crossing the Great Divide : Synergy , and Development. *World Development*, *24*(6), 1073–1087. https://doi.org/10.1016/0305-750X(96)00023-X

Palfrey, J., & Gasser, U. (2008). *Born Digital: Understanding the First Generation of Digital Natives.* New York, NY: Basic books.

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media and Society*, *15*(6), 853–871. https://doi.org/10.1177/1461444812463345

Parsons, A. (2016). *WhatDoTheyKnow and Central Government FOI: 2016 Report.* Retrieved from https://research.mysociety.org/html/foi-report-2016/#home-office-and-immigration

Parsons, A. (2019). *Public FOI: Freedom of Information in central government and WhatDoTheyKnow.*

Parsons, A., & Rumbul, R. (2019). *Freedom of Information in Local Government.* Retrieved from https://research.mysociety.org/media/outputs/mysociety-local-government-foi_lFa9d6X.pdf

Parycek, P., Schöllhammer, R., & Schossböck, J. (2016). "Each in their own garden": Obstacles for the implementation of open government in the public sector of the German-speaking region. In *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance (ICEGOV '15-16)* (pp. 291–300). https://doi.org/10.1145/2910019.2910105

Paterson, M. (2015). The Media and Access to Government-Held Information in a Democracy. *Oxford University Commonwealth Law Journal*, *8*(1), 3–23. https://doi.org/10.1080/14729342.2008.11421487

Peisakhin, L., & Pinto, P. (2010). Is transparency an effective anti-corruption strategy? Evidence from a field experiment in India. *Regulation and Governance*, (4), 261 – 280. https://doi.org/10.1111/j.1748-5991.2010.01081.x

Peixoto, T. (2013). The uncertain relationship between open data and accountability: A response to Yu and Robinson's The New Ambiguity of "Open Government." *UCLA Law Review Discourse*, *60*, 200–213.

Pemstein, D. et al. (2015). V-Dem Codebook.

Pitkin, H. F. (1967). *The Concept of Representation*. Berkeley: University of California Press.

Ponting, C. (1990). *Secrecy in Britain* (1st ed.). Oxford, UK; Cambridge, MA, USA: B. Blackwell.

Popper, K. R. (2001). *All life is problem-solving* (1st ed.). New York: Routledge.

Porlezza, C., & Splendore, S. (2019). From Open Journalism to Closed Data: Data Journalism in Italy. *Digital Journalism*, *7*(9), 1230–1252. https://doi.org/10.1080/21670811.2019.1657778

Posner, E. A. (2014). *The Twilight of Human Rights Law.* (1st ed.). Oxford: Oxford University Press.

Pozen, D. E. (2017). Freedom of Information Beyond the Freedom of Information Act. *University of Pennsylvania Law Review*, *165*, 1097–1158.

Pozen, D. E., & Schudson, M. (2018). *Troubling Transparency: The History and Future of Freedom of Information*. (D. E. Pozen & M. Schudson, Eds.) (1st ed.). New York, NY: Columbia University Press.

Prat, A. (2005). The Wrong Kind of Transparency. *American Economic Review*, *95*(3), 862–877. https://doi.org/10.1257/0002828054201297

Prensky, M. (2009). H. sapiens digital: From digital immigrants and digital natives to digital wisdom. *Innovate: Journal of Online Education*, *5*(3). Retrieved from https://nsuworks.nova.edu/innovate/vol5/iss3/1

Prime Minister's Office. (2010). PM's podcast on transparency. Retrieved November 22, 2019, from https://www.gov.uk/government/news/pms-podcast-on-transparency

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, *54*(1), 209–228. https://doi.org/10.1111/j.1540-5907.2009.00427.x

Raab, C. D. (1994). Open government: policy information and information policy. *The Political Quarterly*, *65*(3), 340–347.

Ragin, C. C. (2014). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. University of California Press*.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of Data Collection, Perceptions of Risks and Losses, and Motivation to Give Truthful Answers to Sensitive Survey Questions. *Applied Cognitive Psychology*, *13*(5), 465–484. https://doi.org/10.1002/(SICI)1099-0720(199910)13:5<465::AID-ACP609>3.0.CO;2-Y

Reinikka, R., & Svensson, J. (2005). Fighting corruption to improve schooling: evidence from a newspaper campaign in Uganda. *Journal of the European Economic*

Association, *3*(2–3), 259–267. https://doi.org/10.1162/jeea.2005.3.2-3.259

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. Close-ended Questions in Web Questionnaires. In *metodološki zvezki - Advances in Methodology and Statistics* (pp. 159–175).

Relly, J. E., & Schwalbe, C. B. (2013). Watchdog journalism: India's three largest English-language newspapers and the Right to Information Act. *Asian Journal of Communication*, *23*(3), 284–301. https://doi.org/10.1080/01292986.2012.729149

Relly, J. E., & Schwalbe, C. B. (2016). How business lobby networks shaped the U.S. Freedom of Information Act: An examination of 60 years of congressional testimony. *Government Information Quarterly*, *33*(3), 404–416. https://doi.org/10.1016/j.giq.2016.05.002

Remuzat Rennó, L. (2008). Rewarding the corrupt? Reelection and Scandal Involvement in the Brazilian 2006 Legislative Elections. *Colombia Internacional*, *68*(July/December), 98–106. https://doi.org/10.7440/colombiaint68.2008.04

Reporters without borders. (2019). United Kingdom. Reporters without borders. Retrieved from https://rsf.org/en/united-kingdom

Rios-Fígueroa, J. (2012). Justice system institutions and corruption control: evidence from Latin America. *Justice System Journal*, *33*(2), 195–214.

Roberts, A. S. (2005). Spin control and freedom of information: Lessons for the United Kingdom from Canada. *Public Administration*, *83*(1), 1–23. https://doi.org/10.1111/j.0033-3298.2005.00435.x

Roberts, A. S. (2006). *Blacked out: Government secrecy in the information age*. Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9781139165518

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Rodríguez, R. P., & Rossel, C. (2018). A field experiment on bureaucratic discretionary bias under FOI laws. *Government Information Quarterly*, *35*(3), 418–427. https://doi.org/10.1016/j.giq.2018.06.001

Rog, D. J., & Bickman, L. (2008). *The SAGE Handbook of Applied Social Research Methods* (2nd ed.). SAGE Publications Ltd. https://doi.org/10.4135/9781483348858

Rose, R., & Mishler, W. (2010). Experience versus perception of corruption: Russia as a test case. *Global Crime*, *11*(2), 145 – 163. https://doi.org/10.1080/17440571003669175

Rose, R., & Peiffer, C. (2019). *Bad governance and corruption*. (D. Hough & P. M. Heywood, Eds.). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-92846-3

Rosenn, K. S. (1987). Protection of Judicial Independence in Latin America. *The University of Miami Inter-American Law Review*, *1*(19), 1–35.

Roulstone, A. (2015). Personal Independence Payments, welfare reform and the shrinking disability category. *Disability and Society*, *30*(5), 673–688. https://doi.org/10.1080/09687599.2015.1021759

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, *34*(1), 45–52. https://doi.org/10.1016/j.giq.2017.01.001

Samaha, A. (2006). Government Secrecy, Constitutional Law, and Platforms for Judicial Elaboration. *UCLA Law Review*, *53*(4), 909–997.

Sambrook, R. (2019). *Global Teamwork: The Rise of Collaboration in Investigative Journalism*. Oxford, UK: The Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-03/sambrook_e-ISBN_1802.pdf

Sandberg, J., & Alvesson, M. (2011). Ways of constructing research questions: gap-spotting or problematization? *Organization*, *18*(1), 23–44.

Santiso, C., & Roseth, B. (2017). Data disrupts corruption. *Stanford Social Innovation Review*, *15*(2), 50–55.

Savage, A., & Hyde, R. (2014). Using freedom of information requests to facilitate research. *International Journal of Social Research Methodology*, *17*(3), 303–317. https://doi.org/10.1080/13645579.2012.742280

Saxena, S., & Janssen, M. (2017). Examining open government data (OGD) usage in India through UTAUT framework. *Foresight*, *19*(4), 421–436. https://doi.org/10.1108/FS-02-2017-0003

Sberna, S., & Vannucci, A. (2013). "It's the politics, stupid!". The politicization of anti-corruption in Italy. *Crime, Law and Social Change*, *60*(5), 565–593. https://doi.org/10.1007/s10611-013-9480-8

Schillemans, T., Van Twist, M., & Vanhommerig, I. (2013). Innovations in accountability. *Public Performance & Management Review*, *36*(3), 407–435. https://doi.org/10.2753/pmr1530-9576360302

Schmidt, S. (2017). Why hundreds of New York Times employees staged a walkout. Retrieved January 30, 2018, from https://www.washingtonpost.com/news/morning-mix/wp/2017/06/30/the-new-york-times-is-eliminating-its-copy-editing-desk-so-hundreds-of-employees-walked-out/?utm_term=.2b037780011c

Schmitter, P. C. (2009). The nature and future of comparative politics. *European Political Science Review*, *1*(1), 33–61. https://doi.org/10.1017/S1755773909000010

Schnell, S. (2014). *International Policy Diffusion and Domestic Policy-Making- the Case of Two Transparency and Anticorruption Policies in Romania*. The Faculty of The Columbian College of Arts and Sciences of The George Washington University.

Schnell, S. (2018). Cheap talk or incredible commitment? (Mis)calculating transparency and anti-corruption. *Governance*, *31*(3), 415–430. https://doi.org/10.1111/gove.12298

Schofield, P. (2006). Publicity, responsibility, and the architecture of government. In Oxford University Press (Ed.), *Utility and Democracy: The political thought of Jeremy Bentham* (pp. 250–271). Oxford. https://doi.org/10.1093/acprof:oso/9780198208563.001.0001

Schofield, P. (2009). *Bentham: A Guide for the Perplexed*. London: Continuum International Publishing Group.

Schrock, A. R. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media & Society*, *18*(4), 581–599. https://doi.org/10.1177/1461444816629469

Schrock, A., & Shaffer, G. (2017). Data ideologies of an interested public: A study of grassroots open government data intermediaries. *Big Data & Society*, *4*(1), 1–10. https://doi.org/10.1177/2053951717690750

Schumpeter, J. A. (2010). *Capitalism, socialism and democracy*. *Capitalism, Socialism and Democracy*. London: Routledge. Retrieved from https://www.dawsonera.com/abstract/9780203857090

Scott, J. (1990). *A Matter of Record: Documentary Sources in Social Research*. Cambridge, UK: Polity Press.

Scottish Information Commissioner. (2019). *FOISA Guidance: Vexatious or repeated requests*. St Andrews, United Kingdom. Retrieved from http://www.itspublicknowledge.info/Law/FOISA-EIRsGuidance/Section14/Vexatious_or_repeated_requests.aspx

Scrollini, F. M. (2015). *Right-to-information arenas: Exploring the right to*

information in Chile, New Zealand and Uruguay. London School of Economics. Retrieved from http://etheses.lse.ac.uk/3361/

Semadeni, M., Withers, M. C., & Trevis Certo, S. (2014). The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal*, *35*, 1070–1079. https://doi.org/10.1002/smj.2136

Shakespeare, S. (2013). *Shakespeare review: an independent review of public sector information*. London, United Kingdom. Retrieved from https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information

Sharafutdinova, G. (2010). What explains corruption perceptions? The dark side of political competition in Russia's regions. *Comparative Politics*, *42*(2), 147–166. https://doi.org/10.5129/001041510X12911363509431

Shkabatur, J. (2012). Transparency with(out) accountability: open government in the United States. *Yale Law and Policy Review*, *31*(1), 79–140.

Shkabatur, J., & Peled, A. (2016). Sustaining the open government data movement worldwide: Insights from developing countries. In *Proceedings of the 6th International Conference for E-Democracy and Open Government, CeDEM 2016* (pp. 172–179). https://doi.org/10.1109/CeDEM.2016.10

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*(3), 308–315. https://doi.org/10.1016/j.giq.2015.05.003

Smith, R. C., & Cavaliere, P. (2017). *Country Report: United Kingdom*. Florence, Italy. Retrieved from https://cadmus.eui.eu/bitstream/handle/1814/61160/2018_United_Kingdom_EN.pdf?sequence=1&isAllowed=y

Smulovitz, C., & Peruzzotti, E. (2000). Societal Accountability in Latin America. *Journal of Democracy*, *11*(4), 147–158. https://doi.org/10.1353/jod.2000.0087

Solaz, H., De Vries, C. E., & de Geus, R. A. (2019). In-Group Loyalty and the Punishment of Corruption. *Comparative Political Studies*, *52*(6), 896–926. https://doi.org/10.1177/0010414018797951

Solis, J. A., & Antenangeli, L. (2017). Corruption is bad news for a free press: reassessing the relationship between media freedom and corruption. *Social Science Quarterly*, *98*(3), 1112–1137. https://doi.org/10.1111/ssqu.12438

Solomon, G., & Schrum, L. (2007). *Web 2.0: new tools, new schools*. Washington, DC: International Society for Technology in Education.

Sovey, A. J., & Green, D. P. (2011). Instrumental variables estimation in political science: a readers' guide. *American Journal of Political Science*, *55*(1), 188–200. https://doi.org/10.1111/j.1540-5907.2010.00477.x

Spilsbury, M. (2014). *Emerging skills for journalists*. Essex, United Kingdom. Retrieved from https://www.nctj.com/downloadlibrary/NCTJ Emerging Skills FINAL.pdf

Spilsbury, M. (2018). *Journalists at work*. Essex, United Kingdom. Retrieved from https://www.nctj.com/downloadlibrary/JaW Report 2018.pdf

Splichal, S. (2002). The principle of publicity, public use of reason and social control. *Media, Culture & Society*, *24*(1), 5–26. https://doi.org/10.1177/016344370202400101

Stec, J. A. (2008). Bias. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 56–60). Thousand Oaks, California: SAGE Publications Ltd.

Steffek, J. (2010). Public accountability and the public sphere of international governance. *Ethics & International Affairs*, *24*(1), 45–67. https://doi.org/10.1111/j.1747-7093.2010.00243.x

Steinberg, T., & Mayo, E. (2007). *The Power of Information: An independent review*.

Stephenson, M. C. (2004). Court of public opinion: government accountability and judicial independence. *Journal of Law, Economics, and Organization*, *20*(2), 379–399. https://doi.org/10.1093/jleo/ewh038

Stiglitz, J. E. (1999a). Knowledge as a global public good. In I. Kaul, I. Grunberg, & M. Stern (Eds.), *Global public goods: international cooperation in the 21st century* (1st ed., pp. 308–325). Oxford: Oxford University Press.

Stiglitz, J. E. (1999b). On Liberty, the right to know, and public discourse: the role of transparency in public life. Retrieved July 25, 2019, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.594.93&rep=rep1&type=pdf

Stoneman, J. (2015). *Does open data need journalism?* Oxford. Retrieved from https://reutersinstitute.politics.ox.ac.uk/our-research/does-open-data-need-journalism

Styrin, E., Luna-Reyes, L. F., & Harrison, T. M. (2017). Open data ecosystems: an international comparison. *Transforming Government: People, Process and Policy*, *11*(1), 132–156. https://doi.org/10.1108/TG-01-2017-0006

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys* (1st ed.). Chicago: Aldine Publishing Company.

Sunstein, C. R. (2018). Output Transparency vs. Input Transparency. In D. E. Pozen & M. Schudson (Eds.), *Troubling Transparency: The History and Future of Freedom of Information* (1st ed., pp. 187–205). New York, NY: Columbia University Press. https://doi.org/10.1017/jli.2019.7

Surowiecki, J. (2005). *The Wisdom of Crowds* (1st ed.). New York: Anchor Books.

Susha, I., Zuiderwijk, A., Janssen, M., & Grönlund, Å. (2015). Benchmarks for Evaluating the Progress of Open Data Adoption: Usage, Limitations, and Lessons Learned. *Social Science Computer Review*, *33*(5), 613–630. https://doi.org/10.1177/0894439314560852

Susman, T. (2016). Independent assessment finds federal agency release rate for FOIA much lower than claimed. Retrieved February 16, 2018, from https://sunlightfoundation.com/2016/08/26/independent-assessment-finds-federal-agency-release-rate-for-foia-much-lower-than-claimed/

Svensson, J. (2005). Eight Questions about Corruption. *Journal of Economic Perspectives*, *19*(3), 19–42. https://doi.org/10.1257/089533005774357860

Tabary, C., Provost, A. M., & Trottier, A. (2016). Data journalism's actors, practices and skills: A case study from Quebec. *Journalism*, *17*(1), 66–84. https://doi.org/10.1177/1464884915593245

Telegraph Reporters. (2009, May 8). MPs' expenses: Full list of MPs investigated by The Telegraph. Retrieved from https://www.telegraph.co.uk/politics/0/mps-expenses-full-list-mps-investigated-telegraph/

Thacker, P. D. (2013, March 12). Where the sun don't shine. *Slate*. Retrieved from https://www.politico.com/story/2012/03/obamas-transparency-stumbles-073606?paginate=false

The Cabinet Office. (1993). *Open Government*. London, United Kingdom. Retrieved from https://discovery.nationalarchives.gov.uk/details/r/C13340010

The Cabinet Office. (1997). *Your right to know: the Government's proposals for a Freedom of Information Act*. London, United Kingdom. Retrieved from https://www.gov.uk/government/publications/your-right-to-know-the-governments-proposals-for-a-freedom-of-information-act

The Cabinet Office. (2007). *The Government's Response to The Power of Information: An independent review by Ed Mayo and Tom Steinberg*. London, United Kingdom. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachme

nt_data/file/243238/7157.pdf

The Cabinet Office. (2012). ***Open Data White Paper: Unleashing the Potential*** (No. White papers). London, United Kingdom. Retrieved from https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential

The Cabinet Office. (2014). Francis Maude speech on open data and transparency. Retrieved January 29, 2018, from https://www.gov.uk/government/speeches/francis-maude-speech-on-open-data-and-transparency

The Cabinet Office. (2019). National Statistics: Freedom of Information statistics - July to September 2019. Retrieved December 22, 2019, from https://www.gov.uk/government/collections/government-foi-statistics#2019

The Cabinet Office, & Independent Commission on Freedom of Information. (2016). Independent Commission on Freedom of Information report.

The Conservative Party. (1992). 1992 Conservative Party General Election Manifesto: The Best Future for Britain. Retrieved from http://www.conservativemanifesto.com/1992/1992-conservative-manifesto.shtml

The Conservative Party. (2010). 2010 Conservative Party Manifesto. Retrieved from https://www.conservatives.com/~/media/Files/Manifesto2010

The Economist. (2010). Data, data everywhere. Retrieved January 26, 2018, from http://www.economist.com/node/15557443%5Cnhttp://www.economist.com/node/15557443/print

The Guardian. (1999, September 13). Tidal wave of support for the Guardian's campaign. Retrieved from https://www.theguardian.com/politics/1999/sep/13/freedomofinformation.uk

The Guardian. (2015, October 11). The Guardian view on the freedom of information commission: a very British farce. Retrieved from https://www.theguardian.com/commentisfree/2015/oct/11/the-guardian-view-on-the-freedom-of-information-commission-a-very-british-farce

The House of Commons. (1979). HC Debate vol. 960 cc2131-213. London, United Kingdom. Retrieved from https://api.parliament.uk/historic-hansard/commons/1979/jan/19/official-information-bill

The House of Commons. (1981). HC Debate vol 998 cc514-74. London, United Kingdom. Retrieved from https://api.parliament.uk/historic-hansard/commons/1981/feb/06/freedom-of-information-bill#S5CV0998P0_19810206_HOC_22

The House of Commons. (1984). HC Debates vol. 55 cc738-41. London, United Kingdom. Retrieved from https://api.parliament.uk/historic-hansard/commons/1984/mar/06/freedom-of-information#S6CV0055P0_19840306_HOC_205

The House of Commons Political and Constitutional Reform Committee. (2011). ***Role and powers of the Prime Minister***. London, United Kingdom.

The Information Commissioner's Office. (2015). ***Dealing with vexatious requests (section 14)***. Wilmslow, United Kingdom. Retrieved from https://ico.org.uk/media/for-organisations/documents/1198/dealing-with-vexatious-requests.pdf

The Information Commissioner's Office. (2019). When can we refuse a request for information? Retrieved October 16, 2019, from https://ico.org.uk/for-organisations/guide-to-freedom-of-information/refusing-a-request/

The Labour Party. (1974). ***October 1974 Labour Party Manifesto: Britain Will Win With Labour***. Retrieved from http://www.labour-party.org.uk/manifestos/1974/Oct/1974-oct-labour-manifesto.shtml

The Labour Party. (1997). 1997 Labour Party Manifesto: New Labour because Britain deserves better. Retrieved from http://www.labour-party.org.uk/manifestos/1997/1997-labour-manifesto.shtml

The Labour Party. (2005). The Labour Party manifesto 2005: Britain forward not back. Retrieved December 22, 2019, from http://news.bbc.co.uk/1/shared/bsp/hi/pdfs/13_04_05_labour_manifesto.pdf

The Labour Party. (2010). The Labour Party manifesto 2010: A future fair for all. Retrieved December 22, 2019, from http://www.cpa.org.uk/cpa_documents/TheLabourPartyManifesto-2010.pdf

The National Archives. (2016). *Freedom of Information, Copyright and Copying Of records held in archives*. Retrieved from https://www.nationalarchives.gov.uk/documents/foi-copyright-and-copying-of-records.pdf

The Office for National Statistics. (2018). Dataset EMP04: Employment by occupation. Retrieved October 15, 2019, from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/employmentbyoccupationemp04

The Scottish Parliament. (2019). Public Audit and Post-legislative Scrutiny Committee. Edinburgh, United Kingdom: The Scottish Parliament. Retrieved from https://www.scottishparliament.tv/meeting/public-audit-and-post-legislative-scrutiny-committee-january-10-2019

Thurman, N., Cornia, A., & Kunert, J. (2016). *Journalists in the UK*. Oxford, UK. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Journalists%2520in%2520the%2520UK.pdf

Timm, T. (2012). Under Obama, the Freedom of Information Act is still in shackles. Retrieved February 16, 2018, from https://www.eff.org/deeplinks/2012/01/under-obama-administration-freedom-information-act-still-shackles

Tran, V. T., Porcher, R., Falissard, B., & Ravaud, P. (2016). Point of data saturation was assessed using resampling methods in a survey with open-ended questions. *Journal of Clinical Epidemiology*, *80*, 88–96. https://doi.org/10.1016/j.jclinepi.2016.07.014

Transparency Azerbaijan. (2016). *Monitoring report on the implementation of the Open government initiative national action plan for 2012-2015*. Baku. Retrieved from http://transparency.az/alac/files/OGP_report_2016_FINAL_EN.pdf

Truex, R. (2011). Corruption, Attitudes, and Education: Survey Evidence from Nepal. *World Development*, *39*(7), 1133–1142. https://doi.org/10.1016/j.worlddev.2010.11.003

Underwood, P. G., & Darch, C. (2010). *Freedom of Information and the Developing World: The Citizen, the State and Models of Openness*. Oxford: Chandos Publishing.

United Nations. (1948). Universal Declaration of Human Rights. Retrieved February 22, 2019, from http://www.un.org/en/universal-declaration-human-rights/

Vadlamannati, K. C., & Cooray, A. (2017). Transparency pays? Evaluating the effects of the freedom of information laws on perceived government corruption. *Journal of Development Studies*, *53*(1), 116–137. https://doi.org/10.1080/00220388.2016.1178385

van Zyl, A. (2014). How civil society organizations close the gap between transparency and accountability. *Governance*, *27*(2), 347–356. https://doi.org/10.1111/gove.12073

Vark, C. van. (2013, November 25). How might open data in agriculture help achieve food security? *The Guardian*.

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open

data quality measurement framework: Definition and application to open government data. *Government Information Quarterly*, *33*(2), 325–337. https://doi.org/10.1016/j.giq.2016.02.001

Vleugels, R. (2008). Overview of all 86 FOIA Countries. Retrieved July 25, 2019, from https://www.statewatch.org/news/2008/sep/foi-overview-86-countries-sep-2008.pdf

Vrushi, J., & Hodess, R. (2017). *Connecting the dots: Building the case for open data to fight corruption*. Berlin. Retrieved from https://www.transparency.org/whatwedo/publication/connecting_the_dots_building_the_case_for_open_data_to_fight_corruption

Vuyst, S. De, & Raeymaeckers, K. (2019). Is Journalism Gender E-Qual? A study of the gendered accumulation and evaluation of digital capital in journalism. *Digital Journalism*, *7*(5), 554–570. https://doi.org/10.1080/21670811.2017.1369357

Walby, K., & Luscombe, A. (2017). Criteria for quality in qualitative research and use of freedom of information requests in the social sciences. *Qualitative Research*, *17*(5), 537–553. https://doi.org/10.1177/1468794116679726

Wang, H.-J., & Lo, J. (2016). Adoption of open government data among government agencies. *Government Information Quarterly*, *33*(1), 80–88. https://doi.org/10.1016/j.giq.2015.11.004

Wang, H., & Rosenau, J. N. (2001). Transparency international and corruption as an issue of global governance. *Global Governance*, *7*(1), 25–49. https://doi.org/10.1163/19426720-00701005

Weeks, B. E., & Gil de Zúñiga, H. (2019). What's Next? Six Observations for the Future of Political Misinformation Research. *American Behavioral Scientist*, 1–13. https://doi.org/10.1177/0002764219878236

Weinstein, J., & Goldstein, J. (2012). The benefits of a big tent: Opening up government in developing countries a response to Yu & Robinson's the new ambiguity of "Open Government." *UCLA Law Review Discourse*, *60*, 38–48.

Welch, S., & Hibbing, J. R. (1997). The effects of charges of corruption on voting behavior in congressional elections, 1982-1990. *Journal of Politics*, *59*(1), 226–239. https://doi.org/10.2307/2998224

Weller, P. (2014). The Variability of Prime Ministers. In R. A. W. Rhodes & P. 't Hart (Eds.), *The Oxford Handbook of Political Leadership* (1st ed., p. online). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199653881.013.024

White, Patrick. (2017). *Developing Research Questions. Developing Research Questions* (2nd ed.). New York: Palgrave Macmillan.

White, Peter. (1980). Official Secrets and Government Openness in Britain. *The Australian Library Journal*, *29*(1), 20–28. https://doi.org/10.1080/00049670.1980.10755408

Whitten-Woodring, J., & James, P. (2012). Fourth estate or mouthpiece? a formal model of media, protest, and government repression. *Political Communication*, *29*(2), 113–136. https://doi.org/10.1080/10584609.2012.671232

Willnat, L., Weaver, D. H., & Choi, J. (2013). The global journalist in the twenty-first century: A cross-national study of journalistic competencies. *Journalism Practice*. https://doi.org/10.1080/17512786.2012.753210

Wirtz, B. W., Piehler, R., Thomas, M. J., & Daiser, P. (2016). Resistance of public personnel to open government: A cognitive theory view of implementation barriers towards open government data. *Public Management Review*, *18*(9), 1335–1364. https://doi.org/10.1080/14719037.2015.1103889

Witschge, T., & Nygren, G. (2009). Journalism: a profession under pressure? *Journal of Media Business Studies*, *6*(1), 37–59. https://doi.org/10.1080/16522354.2009.11073478

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

World Bank. (2018). World Bank Statement on Amendments to Tanzania's 2015 Statistics Act. Retrieved February 10, 2018, from https://www.worldbank.org/en/news/statement/2018/10/02/world-bank-statement-on-amendments-to-tanzanias-2015-statistics-act

World Wide Web Foundation. (2016). *Open Data Barometer - 4th edition - ODB Methodology. Open Data Barometer*. Retrieved from https://opendatabarometer.org/4thedition/methodology/

Worthy, B. (2014). A powerful weapon in the right hands? How members of parliament have used freedom of information in the UK. *Parliamentary Affairs*, *67*(4), 783–803. https://doi.org/10.1093/pa/gss091

Worthy, B. (2015a). The impact of open data in the UK: Complex, unpredictable and political. *Public Administration*, *93*(3), 788–805. https://doi.org/10.1111/padm.12166

Worthy, B. (2015b, June 10). Open data on council spending is largely unread by voters. *The Guardian*. Retrieved from https://www.theguardian.com/public-leaders-network/2015/jun/10/council-finance-open-data-unread-by-voters

Worthy, B. (2017a). *The politics of freedom of information: how and why governments pass laws that threaten their power. The politics of freedom of information*. Manchester: Manchester University Press.

Worthy, B. (2017b). *United Kingdom End-of-term report 2013–2015*. Retrieved from https://www.opengovpartnership.org/wp-content/uploads/2017/05/UK_EOTR_2013-2015_0.pdf

Worthy, B., & Hazell, R. (2014). The impact of the Freedom of Information Act in the UK. In N. Bowles, J. T. Hamilton, & D. Levy (Eds.), *Transparency in Politics and the Media: Accountability and Open Government* (pp. 31–45). London: L.B. Tauris.

Worthy, B., & Hazell, R. (2017). Disruptive, dynamic and democratic? Ten years of FOI in the UK. *Parliamentary Affairs*, *70*(1), 22–42. https://doi.org/10.1093/pa/gsv069

Xezonakis, G., Kosmidis, S., & Dahlberg, S. (2016). Can electors combat corruption? Institutional arrangements and citizen behaviour. *European Journal of Political Research*, *55*, 160–176. https://doi.org/10.1111/1475-6765.12114

Yang, T.-M., Lo, J., & Shiang, J. (2015). To open or not to open? Determinants of open government data. *Journal of Information Science*, *41*(5), 596–612. https://doi.org/10.1177/0165551515586715

Yang, T. M., & Wu, Y. J. (2016). Examining the socio-technical determinants influencing government agencies' open data publication: A study in Taiwan. *Government Information Quarterly*, *33*(3), 378–392. https://doi.org/10.1016/j.giq.2016.05.003

Yarnall, L., Johnson, J. T., Rinne, L., & Ranney, M. A. (2008). How Post-secondary Journalism Educators Teach Advanced CAR Data Analysis Skills in the Digital Age. *Journalism & Mass Communication Educator*, *63*(2), 146–164. https://doi.org/10.1177/107769580806300204

Yazaki, Y. (2017). Newspapers and political accountability: evidence from Japan. *Public Choice*, *172*(3–4), 311–331. https://doi.org/10.1007/s11127-017-0444-x

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Los Angeles: SAGE Publications Ltd.

Yu, H., & Robinson, D. G. (2012). The new ambiguity of "open government." *UCLA Law Review*, *178*, 180–208.

Zamith, R. (2019). Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. *Digital Journalism*, *7*(4), 470–489. https://doi.org/10.1080/21670811.2018.1554409

Zhu, J., Lu, J., & Shi, T. (2012). When Grapevine News Meets Mass Media: Different Information Sources and Popular Perceptions of Government Corruption in Mainland China. *Comparative Political Studies*, *46*(8), 920–946. https://doi.org/10.1177/0010414012463886

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003

Software

Hlavac, M. (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. http://CRAN.R-project.org/package=stargazer

Long, J. A. (2019). interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions. R package version 1.1.0, https://cran.r-project.org/package=interactions.

Roberts, M.E., Stewart, B.M., Tingley, D. (2019). stm: An R Package for Structural Topic Models. Journal of Statistical Software, 91(2), 1–40.

Wickham, H., François, R., Henry, L., Müller, K. (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.6. https://CRAN.R-project.org/package=dplyr

Wickham, H., Chang, W., Henry, L., Lin Pedersen, T., Takahashi, K., Wilke, C., Woo, K., Yutani, H., RStudio. (2019) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. R package version 3.2.1. https://cran.r-project.org/web/packages/ggplot2/index.html

Yang, J., Small, D. (2014). ivpack: Instrumental Variable Estimation. R package version 1.2. https://cran.r-project.org/web/packages/ivpack/index.html