International Journal of Information Management 37 (2017) 664–672

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Research Note

Liberation of public data: Exploring central themes in open government data and freedom of information research

Eric Afful-Dadzie*, Anthony Afful-Dadzie

Department of Operations and Management Information Systems, University of Ghana Business School, Accra, Ghana

ARTICLE INFO

Keywords:
Open government data (OGD)
Freedom of information (FOI)
Public data
Bibliometrics
Topic modelling
Text mining

ABSTRACT

This paper conducts a comparative literature survey of Open Government Data (OGD) and Freedom of Information (FOI), with a view to tracking the central themes in the two civil society campaigns. With seeming similarities and a growing popularity in research, the major themes framing research on the two movements have not clearly emerged. Topic modelling, text mining and document analysis methods are used to extract the themes as well as key named entities. The topics are subsequently labeled and with expert guidance, their semantic meaning are provided. The results indicate that the major theme in FOI research borders on issues relating to *disclosure, publishing, access and cost of requests*. On the other hand, themes in OGD research have largely centered on *technology* and related concepts. The approach also helped in determining key similarities and differences in the two campaigns as reported in research.

1. Introduction

Freedom of Information (FOI) and Open Government Data (OGD), are two prominent civil society campaigns championing the course of liberating government controlled data to the people. Both FOI and OGD primarily seek to make government data progressively free and easily accessible (Ubaldi, 2013). Largely, the twin but independent campaigns of FOI and OGD have been driven by (1) a global call on nations to offer a more accountable and transparent governance (2) a growing trend of sophistication in citizens' preferences and choices of government services (Holler, 2012; Lau, Patel, Fahmy, & Kaufman, 2014; Van Dooren, Bouckaert, & Halligan, 2015; Weisberg & Nawara, 2010) and (3) an opportunity to amend past policies, where data collected by government agencies tended to be the exclusive reserve of the state (Yiu, 2012). The two movements have thus become a global mouth piece of advocacy towards a more open and transparent governance. Over the years, the two campaigns have received considerable traction in the media as well as in academia (Charalabidis, Alexopoulos, & Loukis, 2016). For instance, basic statistics regarding the yearly number of Freedom of Information Act (FOIA) requests, downloads, appraisal reports, number of workshops and conferences held, together with other specific country initiatives, point to a growing interest among stakeholders (Whitmore, 2012). A similar trend is seen in yearly OGD reports, where accounts by the Independent Reporting Mechanism (IRM) of Open Government Partnership (the body responsible for the launch of OGD in 2011), indicate a steady progress by most member states (Frey, 2014). Other independent accounts in the literature also show significant progress in OGD in the US (Krishnamurthy & Awazu, 2016), UK (Tinati, Carr, Halford, & Pope, 2012), Taiwan (Wang & Lo, 2016), Spain (Carrasco & Sobrepere, 2015), and a host of many other countries including local government authorities such as cities and federal states. In addition, many global experiences have also been shared of how FOI and OGD are impacting governance particularly in the fight against corruption, economic empowerment and the quest for greater citizen engagements (Birkinshaw, 2010; Halstuk & Chamberlin, 2006; Jetzek, Avital, & Bjørn-Andersen, 2012; Shepherd, Stevenson, & Flinn, 2009; US. Senate, 2007Zeleti et al., 2016).

In the wave of the relative progress, there have also been reports of misconceptions, myths, definitional challenges and general obstacles besetting real-world practice by the two campaign groups (Camaj, 2016; Evans & Campos, 2013; Gigler, Custer, & Rahemtulla, 2011; Hubbard, 2008; Janssen, Charalabidis, & Zuiderwijk, 2012; Schartum, 1998; Zuiderwijk & Janssen, 2014). Such wide ranging experiences of "the good, bad and the ugly" of FOI and OGD, have occasioned numerous research publications covering a range of topics. However, as the fields of FOI and OGD continue to evolve, what the central themes are as far as research publications are concerned, have not clearly emerged. Furthermore, given that the two campaigns not only share similarities but differences (Ubaldi, 2013), it is imperative to understand how key concepts are unconsciously being framed in publications relating to the two notions. In view of this, this paper seeks to determine what the major themes have so far been in relation to the two

* Corresponding author.
E-mail addresses: eafful-dadzie@ug.edu.gh (E. Afful-Dadzie), aafful-dadzie@ug.edu.gh (A. Afful-Dadzie).

http://dx.doi.org/10.1016/j.ijinfomgt.2017.05.009
Received 8 August 2016; Received in revised form 20 April 2017; Accepted 19 May 2017
Available online 03 July 2017
0268-4012/

campaigns. It is our view that determining the central message shaping the two campaigns, would not only shed light on what key topics define each movement, but would also help establish whether the similarities and the seeming differences between the two concepts naturally emerge especially in research.

In the years before and after the launch of FOI and OGD, a considerable number of academic publications have been authored spanning various topics and issues. However, none so far has attempted to comparatively explore the 'running' themes in the two concepts. The few literature reviews available were conducted separately by only focusing on either of the two notions. For instance, Attard, Orlandi, Scerri, and Auer (2015) and Novais, de Albuquerque, and da Silva Craveiro (2013) conducted separate reviews of literature on open government data whiles Halstuk and Chamberlin (2006) conducted a retrospective analysis of the Freedom of Information Act from 1966 to 2006. Mendel (2008) conducted a comparative legal survey but however only centered on FOI. Furthermore, the methodological approaches adopted in review articles conducted on FOI or OGD are different from what this paper proposes. This study is therefore uniquely positioned to contribute to theory and fill research gaps in the following ways. First, key topics and associative terms 'running' in scientific discourses on the two civil movements are identified and compared. Secondly, the paper also identifies location-based named entities of interest to frame how FOI and OGD are being implemented around the world. The results provide a means to understand the central themes shaping each campaign and some potential future research directions on OGD and FOI.

The rest of the paper is organized as follows. First, a brief overview of FOI and OGD covering history and key concepts are explained. Further, differences and similarities between the two notions as captured in the literature, are presented. This is followed by the methodology which explains the approach to data collection and presents a brief introduction to text analysis concepts and their relevance to the study. Research questions guiding the study are subsequently presented. The results of the study, discussion and conclusion are further presented.

2. FOI Vs OGD

The idea of open government data (OGD) is generally viewed as an offshoot of Freedom of Information (FOI), also sometimes known as Right to Information (Ubaldi, 2013). However, the two movements come under the broader concept of *open government,* which seeks transparency and greater rights of information access for citizens (Tauberer, 2012). It must be noted that while civil resistance movements are not completely new, the quest for openness and access to government controlled information (spearheaded by FOI and OGD), is a fairly modern concept. Together, the two campaigns are contributing to opening wide the frontiers of democracy, fighting corruption and empowering citizens.

Freedom of Information was officially given a stamp of approval in a United Nations General Assembly Resolution in 1946. It was further strengthened in 1948 through *Article 19* of the Universal Declaration of Human Rights (Donnelly, 2013). The UN initiative, set the tone for the adoption of FOI Act (FOIA) in most countries. Open government data on the other hand, was first conceived in 2007 when 30 OGD campaigners met in California to deliberate on principles to guide the new movement (Chignard, 2013). This meeting and other close efforts such as the Public Sector Information (PSI) Directive of 2003, U.S. President Obama's open data initiative of 2009, helped set up the open government data partnership (Attard et al., 2015; Tauberer, 2012). More recently, the G8 Open Data Charter.[1] in 2013 have all contributed to strengthening the ideals of OGD. These current and past initiatives have helped to establish national and city-based open government data portals around the world. Since the two advocacy groups essentially lobby governments around the world into adopting the concepts, first visible signs of activity are always in the form of endorsements or declarations by sovereign governments. As at 2016, 70 countries have signed the OGD declaration and are at various actionable stages[2] On the other hand, FOI has at 2015, over 95 countries that have adopted some mode of FOI legislation.[3]

Besides similarities between FOI and OGD, differences also exist in terms of the guiding principles and their approach to liberating public data (Ubaldi, 2013). For instance, whereas FOI emphasizes on the fundamental human rights of people to information (legal arguments), OGD campaigns on the economic and social benefits that potentially accrue to people when public datasets and documents are made readily available to all for use, reuse and subsequent distribution (Janssen, 2012). It is the view of OGD advocates that, making public data free and easily accessible, not only improves democratic participation but also allows individuals to create new products and services based on informed and reliable public data (Janssen et al., 2012; Maude, 2012). The primary goals of the two campaigns, according to Janssen (2012), also influence the kinds of datasets mostly sought after in each group. Datasets such as transport data, geographic data, corporate data and general business information which have the potential to spur innovation and economic growth, are mostly of interest to OGD advocates. On the other hand, given the professional background of most FOI proponents (lawyers, media practitioners etc.), the most preferred datasets tend to be government budgets and expenditure, revenue data, legal information as well as documents covering reports and meeting minutes of key government agencies. Another dimension of OGD advocacy different from FOI, is the emphasis on deploying latest information technology tools that help in the release of public datasets. The technical dimension of OGD sets it apart from FOI particularly on how clear specifications are recommended on how to model, create, publish, store and release government data in various formats for easy accessibility. For instance, proponents of OGD tend to emphasize that datasets are stored and accessed in non-proprietary machine readable formats, such as CSV, XML, TSV, RDF (Berners-Lee, 2006). Furthermore, OGD recommends that adequate metadata and linked data technology are provided for public datasets. Thus, unlike RTI, the concept of open government data (OGD) transcends the emphasis on fundamental human rights of people to access information. Rather, focus is also on providing data management and exchange tools that make the idea of accessibility, re-use and distribution, a reality. Other recognized differences between FOI and OGD are also seen in the way they approach copyright and licensing issues. FOI implementation tends to differ from country to country especially since some countries are very restrictive on what can be published and reused. OGD on the other hand, grants an inherent license to users to freely use, reuse and distribute public data.[4] Further, because of the often legal approach adopted by FOI, its relationship with other stakeholders, especially public sector employees and politicians are sometimes antagonistic rather than collaborative as often seen in OGD approaches.[5] In the following section, the methodology adopted in the study to investigate the central themes 'running' in academic publications in the two movements is explained.

3. Methodology

The main part of the research design used topic modelling. Text mining and document analysis methods were used mainly to clean and

[1] http://opendatacharter.net/history/.

[2] http://www.opengovpartnership.org/countries.

[3] http://www.right2info.org/access-to-information-laws/access-to-information-laws#_ftnref7.

[4] http://webfoundation.org/2015/08/freedom-of-information-and-open-government-data-communities-could-benefit-from-closer-collaboration/.

[5] https://www.article19. org/data/files/pdfs/standards/righttoknow.pdf.

Table 1
Inclusion and Exclusion Search Criteria.

Search Criterion/Task	Inclusion	Exclusion
Comparative Survey	FOI; OGD	Any unrelated
Document Type	Journal articles, Conference proceedings and Book chapters.	Editorial, Doctoral dissertations, master's theses, textbooks, Letters, Erratum etc.
Language	English	Non-English
Time Period	FOI: – 2015; OGD: – 2015	After 2015
Bibliographic database	Web of Science and Scopus	All other database
Data cleaning	Main text spanning introduction to conclusion sections of each article	Title page, keywords, references, funding sources and acknowledgements sections.

transform the textual data. The methodology is conveniently segmented into three phases of *text pre-processing*, *processing* and *information extraction* as shown in Fig. 2. The three stages were however preceded by a data collection phase which primarily employed document analysis (Owen, 2014) techniques to gather the data. Particularly, the inclusion and exclusion search criteria in document analysis was used to refine the data search as shown in Table 1. The following section presents a brief introduction to the text analysis approaches used in the study and further explains what the data collection stage entailed.

3.1. Text analysis

The era of massive generation of digital textual data is impacting positively on text analysis research and practice. Robust tools and methods (Aggarwal & Zhai, 2012) that help to glean meaningful information from large amounts of documents are consistently being introduced to keep pace with a growing research field. Text analysis tools and methods are popularly used in information and document retrieval, text summarization, document dis(similarity) identification, language identification and document authorship attribution (Witten, 2005). However, irrespective of the text analysis method involved, the end goal is always about extracting high-quality information from textual data. An effective text analysis project, according to DiMaggio, Nag, and Blei (2013), must (1) be reproducible (2) automated (3) inductive to patterns and (4) provide "relationality" in different contexts. These requirements demonstrate an improvement over non-computational methods. In this study, the term text analysis is expediently used to encompass digital text computational methods such as text mining and topic modelling, even though the terms text analysis and text mining are also sometimes used synonymously. In the following section, a brief introduction to topic modelling as used in the study is presented.

3.1.1. Topic modelling

In the last decade, the concept of topic modelling has become one of the most trending research areas in text analysis (Wang & Blei, 2011; Wei & Croft, 2006). Topic modelling is basically used to summarize large corpora of text by revealing hidden themes in textual data (Blei, Ng, & Jordan, 2003; Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009; DiMaggio et al., 2013). There are three topic models namely Latent Dirichlet Allocation (LDA), Probabilistic Latent Semantic Indexing and Correlated Topic Models. The most utilized model is the LDA algorithm, which uses Bayesian statistical model to treat documents as a random bag of words over latent topics and where each topic is characterized by a distribution over words (Blei et al., 2003; Blei, 2012). Further, the technique works by identifying clusters of words that frequently co-occur in a corpus of text without regard to word order. In view of this, it is recommended that a relatively large size of textual data is used so as to help increase the likelihood of themes of text being found together. The LDA algorithm may be described formally as shown in Fig. 1 culled from Blei et al. (2003) and Yau, Porter, Newman, and Suominen (2014).

In a vocabulary indexed as $\{1,...,V\}$, words are represented as unit-basis vectors where a unit document is equal to one and the rest of the components equal to zero. Representing the components with superscripts, a *vth* word in the vocabulary is denoted by a V-vector w such

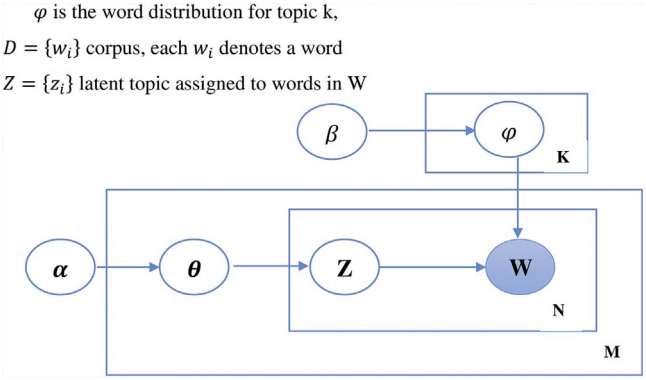

Fig. 1. The Latent Dirichlet Allocation (LDA) Model.
(Source: Blei et al., 2003)

that $w^v = 1$ and $w^u = 0$ for $u \neq v$. The following definitions are used to further explain the concept as demonstrated graphically in Fig. 1. A document is defined as a sequence of N words expressed as $W = (w_1, w_2, ..., w_N)$, where w_N is the *nth* word in the sequence.

A corpus is a collection of M documents denoted by $D = (w_1, w_2, ..., w_N)$

α is the parameter of the topic Dirichlet prior per document topic distributions

β is the parameter of the word Dirichlet prior per document topic distributions

θ is the topic distribution for document i

φ is the word distribution for topic k,

$D = \{w_i\}$ corpus, each w_i denotes a word

$Z = \{z_i\}$ latent topic assigned to words in W

Over the years, topic modelling has been used extensively in a number of areas such as fraud detection (Xing & Girolami, 2007), spam filtering (Bíró, Szabó, & Benczúr, 2008), clustering scientific documents (Yau et al., 2014), determining business topics in a source code (Maskeri, Sarkar, & Heafield, 2008), detecting phishing websites (Ramanathan & Wechsler, 2012), framing ideas in history (Hall et al., 2008), social network analysis (Ríos, Aguilera, Bustos, Omitola, & Shadbolt, 2013), understanding how the media frames certain topical issues (Afful-Dadzie, Nabareseh, Oplatková, & Klímek, 2016) among many others. Topic modelling applications have also been extended to bibliometric review, as evident in the following articles (Griffiths & Steyvers, 2004; Jiang, Qiang, & Lin, 2016; Mann, Mimno, & McCallum, 2006). The use of text analysis in bibliometric analysis is relatively new and the approach is a departure from manually analyzing documents one at a time. In this study, LDA based topic modelling, general text mining techniques and document analysis are used to extract information from research articles with a view to establishing central themes that conceptually frame research discourses on open government data (OGD) and Freedom of Information (FOI). The following section explains the data collection phase as simplified in the inclusion and exclusion search criteria in Table 1.

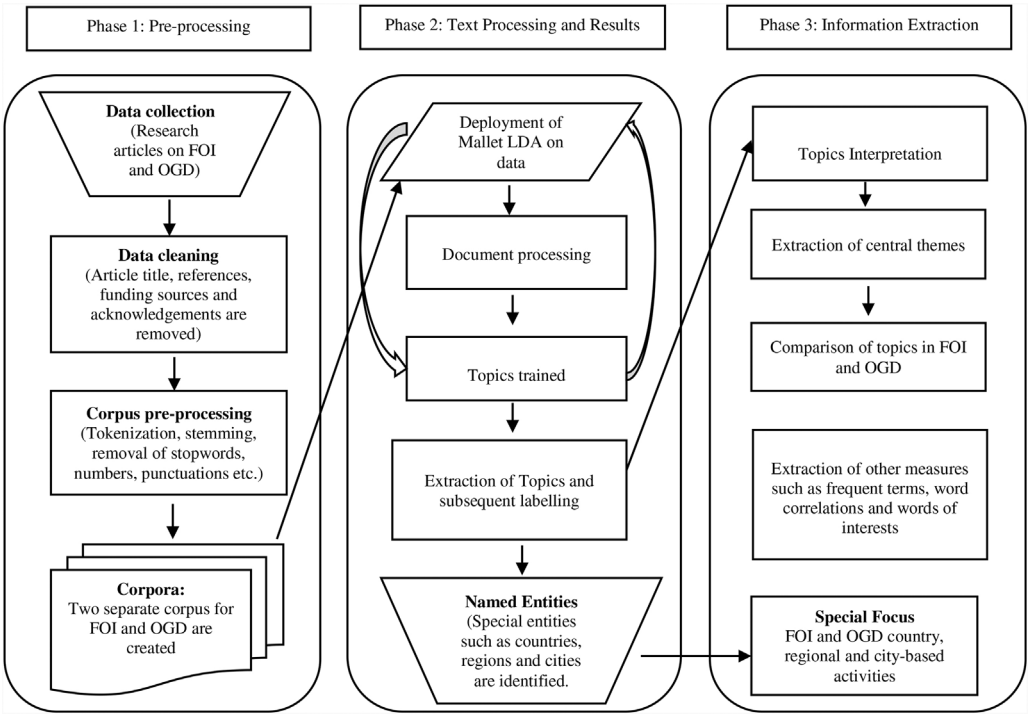

Fig. 2. Conceptual framework for the comparative survey.
(Source: authors)

4. Data collection

Research publications were selected guided by the following criteria (i) an automated search availability (ii) quality and prominence of publications and (iii) reputation of the bibliographic database. In line with the above considerations, we settled on the Web of Science and Scopus; widely recognized as the two most prominent bibliographic databases (Aghaei Chadegani et al., 2013; Wang & Waltman, 2016).

Journal articles, conference proceedings and book chapters were the only kinds of research documents deemed appropriate for the study. The document search was aided by relevant key search phrases and terms, which ensured that only articles suitable to the study were selected. For instance, search phrases such as 'freedom of information', 'freedom of information act', 'right to information', 'right to information act', 'access to information (ATI)' were used to search articles on FOI. Similarly, 'open government data', 'open data', 'government data portal', 'government public data' and 'public data portal' guided the search of articles on OGD.

In addition, a pre-condition was that, key terms or phrases must have appeared in the article title but without strict recourse to word sequencing. This meant that terms such as 'government open data' or 'public government data' were considered as guides for selection. The search strategy was partly adopted from Attard et al. (2015). After initial downloads of the articles, a document analysis was carried out to further scrutinize the articles. This approach ensured that only articles related to the concepts under study were included. For instance, there were article titles that had the phrase 'right information' but which actually had nothing to do with right to information. A thorough document analysis helped to remove all such unrelated articles. In all, 780 articles were collected on FOI and were subsequently trimmed down to 430. Similarly, the OGD article search resulted in 392 articles of which 281 were deemed appropriate for the study.

5. Research questions

To get the most out of the comparative literature survey, a set of pre-defined questions were used to guide the study. This approach was particularly useful as it helped to map the results generated to the research questions. In all, the questions were designed to aid in 'framing' trending concepts in FOI and OGD as captured in research publications. The following questions guided the research:

RQ1. What are the central themes in FOI and OGD research publications?

RQ2. How do similarities and differences identified in the topics compare with what have specifically been written in the literature?

RQ3. How do the topic labelling (classification) frame the discourses of FOI and OGD research?

RQ4. What do named entities especially occurrences of countries, regions and cities in the topics say about the campaign?

5.1. Phase 1: pre-processing

The pre-processing phase involved a number of text cleaning procedures to transform the documents into requisite formats for processing. The first procedure utilized a series of R programming codes to convert the pdf documents into text format. Subsequent codes helped to strip the articles of title pages, abstracts, keywords, references, funding sources, notes, acknowledgements and appendixes, leaving only the 'main' content beginning from "introduction to conclusion". This procedure was necessary to eliminate sections in the articles that could unduly influence the outcome of the text analysis. Following this, a number of steps were taken to transform the corpus ready for processing. These included converting the corpus into lower case characters, striping whitespaces, removing punctuations, sparse terms, numbers as well as stop-words. The study utilized relevant packages in the R Statistical Computing Software for all the text analysis including the pre-processing and the text processing phases. The main libraries or packages utilized in the study were the *tm* package (for text mining), *mallet* package (for LDA topic modelling), and *SnowballC* (for stemming).

5.2. Phase 2: text processing and results

This stage involved the deployment of the mallet LDA library together with other relevant libraries in *R* to first train, and subsequently generate the topics. The study experimented with a different number of

Table 2
The Top 15 frequent words in each topic on FOI research.

Topic	Frequent Terms (Stemmed)
Topic 1	privaci- mug public shot court interest disclosur. u.s. exempt foia feder- circuit crimin- law report
Topic 2	social societi- state polit- economy- knowledg- role process develop technolog- market mean concept model peopl-
Topic 3	freedom inform law media law press foil polit- news diffus- legisl- nation studi- country- access
Topic 4	health data clinic trial right public registr- supra note trial human privaci- drug report research
Topic 5	rti govern india public rtia state corrupt citizen peopl- local indian delhi act societi- time
Topic 6	request foi inform govern institut- feder- public access law canada law request provinci- commission atia
Topic 7	media foia news public journalist request fee waiver blogger interest request blog document interest context
Topic 8	librari research africa servic- world univers- develop local- servic- interest nation internet report communic- commiss-
Topic 9	ati/foi belief request govern belief agenc- account evid- autonomi- text work request truth agent inspect
Topic 10	right human inform internet world intern organis- communic- onlin- group activist countri- wikileak communic- work
Topic 11	foi countri- act south implement countri- africa public request law societi- studi- level develop civil
Topic 12	right court articl- adolesc- human european ecthr convent express posit public case media societi- state
Topic 13	from made time case import number part nation case issu- system includ- year found make
Topic 14	act inform public interest author commission foi request subject decis- code request exempt disclosur- held
Topic 15	foi govern cabinet disclosur- request freedom legisl- polit- account civil reform impact london flow central
Topic 16	act govern public agenc- foia record feder- agenc- congress execut- secur- law hous- u. s offic-
Topic 17	inform access freedom legal protect data make law person privat- provid- general polici- citizen privaci-
Topic 18	court agenc- foia document request disclosur- from exempt record district depart subject cir appeal u. s. c
Topic 19	corrupt foia state public offici- court strong from state year convict law local govern law
Topic 20	govern public european document council parliament bill institut- author offici- legisl- provis- open institut- administr-
Topic 21	data record manag- research request foi public request author legisl- local council nhs studi- vexati-
Topic 22	public law commiss- inform decis- disclosur- court regul- articl- applic- bodi- administr- decis- ministry- foi
Topic 23	state oil nepa pipelin- environment keyston- sand public depart climat- emiss- chang- propos- impact ghg

Table 3
The Top 15 frequent words in each topic on OGD research.

Topic	Frequent Terms (Stemmed)
Topic 1	contract qualiti- busi- portal result gdp averag- number countri- dataset tabl- reliabl- publish offici- notic-
Topic 2	govern servic- citizen app model data citizen provis- develop platform busi- crm mobil- citi- cultur-
Topic 3	govern progress web figur- social brazil websit- feder- analysi- government websit- societi- public peopl- system
Topic 4	govern transpar- matur- level open agenc- social model collabor- particip- benchmark engag- media feder- benchmark
Topic 5	dataset link ogd dataset architectur- sourc- social approach initi- portal integr- publish order metadata model
Topic 6	logd web portal dataset entiti- data. gov sourc- rdf integr- contract u. s databas- entiti- agenc- sourc-
Topic 7	visual catalog ogd catalog data portal user display system visual tool analysi- qualiti- web record
Topic 8	ogd innov- adopt benefit barrier factor organ perceiv- busi- influenc- social janssen busi- model user
Topic 9	govern open agenc- govern initi- transpar- polici- dataset portal develop countri- particip- share local privat-
Topic 10	india organis- nation polici- technolog- rti peopl- studi- act state e − govern indian societi- govern centr-
Topic 11	inform public servic- research citizen access provid- case govern make knowledg- onlin- work sweden relat-
Topic 12	busi- model municip- citi- urban user social set model group citi- dimens- dimens- strategi- privat-
Topic 13	public inform sector psi re − use direct european access polici- licens- bodi- govern licens- principl- australian
Topic 14	project releas- aid develop polici- local competit- research citi- transpar- project approach effect plan agenda
Topic 15	mechan- ogd generat- social innov- economy- energi- access open effici- resourc- particip- technic sector opow
Topic 16	ogd citi- survey implement question research polit- result depart econom- questionnair- benefit administr- respond onlin-
Topic 17	link rdf web semant- sparql dataset time ontolog- queri- result metadata forest fire servic- work
Topic 18	ogd research capabl- level dimens- user methodolog- platform evalu- measur- variabl- layer generat- model domain
Topic 19	data open base process analysi- import level user set exist framework result model nation qualiti-

topics to arrive at the ideal number appropriate for each dataset (corpus). This was done by generating posterior likelihoods on a number of models assigning different numbers of topics each time. The result indicated that a maximum of 23 and 19 topics were respectively appropriate for the FOI and the OGD corpora. The results of the topic modelling are as presented in Tables 2 and 3 respectively for FOI and OGD.

5.3. Phase 3: information extraction

The topics were interpreted based on expert knowledge and support of the literature. After generating the topics, they were subsequently labelled to help frame the central themes in the two research domains as shown in Tables 4 and 5 respectively. The following sub-sections present relevant information that were extracted to give meaning to the topics under each of FOI and OGD. Further, the results from FOI and OGD are compared for similarities, differences and general trends in the topics.

6. Topic interpretation

FOI

The topic labelling or classification was done by interpreting what a body of topics appears to convey. Guided by expert knowledge and the literature, it was realized that a number of the topics seemed to fall under some relevant issues in the two campaigns. In row 1 in Table 4 for instance, the topic label apparently frames issues relating to some FOI guiding principles and key operational terms. This is because, most authoritative texts on FOI particularly those that focus on Article 19 (Birkinshaw, 2010; Foerstel, 1999; Mendel, 2008), tend to recognize a set of guiding principles that are fundamental to the movement.

These FOI guiding principles are "*maximum disclosure, obligation to publish, promotion of open government, limited scope of exceptions (exemptions), processes to facilitate access, costs, open meetings, disclosure takes precedence and protection for whistleblowers*". The topic label *disclosure/publishing/access/costs* which has 40.0% of the topics, captures most of the above FOI principles. In addition, row 2 in Table 4 frames topic scenarios that have semblance to *law, legislation and exemptions* to

Table 4
FOI topic labelling and concept framing.

Concept framing (Topic labelling)	Topic	Topic proportion (%)
Disclosure/Publishing/Access/Costs	Topic 1, Topic 5, Topic 6, Topic 7, Topic 9, Topic 10, Topic 15, Topic 16, *Topic 18*, *Topic 21*	40.0
Law/Legislation/Exemptions	Topic 3, Topic 12, Topic 14, Topic 17, *Topic 18*, Topic 19, Topic 20, *Topic 21*, Topic 22	35.0
Service provision	Topic 2, Topic 8, Topic 11, *Topic 21*,	15.0
Health/Environmental	Topic 4, *Topic 21*, Topic 23,	10.0

Table 5
OGD topic labelling and concept framing.

Concept framing (Topic labelling)	Topic	Topic proportion (%)
Transparency/Collaboration/ Participation	Topic 4, *Topic 8*, Topic 9, *Topic 14*,	12.5
Technology	Topic 2, Topic 3, Topic 5, Topic 6, Topic 7, Topic 17, Topic 18, Topic 19	50.0
Economic/Social/Innovation	Topic 1, *Topic* 8, Topic 10, *Topic 14*, Topic 15, *Topic 16*	18.75
Citizen engagement	Topic 11, Topic 12, Topic 13, *Topic 14*, *Topic 16*	18.75

FOI requests and disclosures. The law, as already explained in the introduction, is a major driving force behind the FOI campaign. Finally, rows 3 and 4 label topics that cover a number of other key issues such as FOI service provision, and health and environmental issues respectively.

It can be recognized that Topics 1, 5, 6, 7, 9, 10, 15, 16, 18 and 21 embody themes that relate to most of the key principles in FOI. For instance, Topic 1 addresses the issue of access to mug shots information which so often raises conflicting issues about privacy on one hand, and access, disclosures and exemptions on the other hand. Topic 1 specifically points to the many instances of legal tussles fought in US. courts regarding whether or not there should be exemptions to requests for mug shot booking information. Terms such as "*privaci-*", "*mug*", "*interest*", "*disclosure*", "*exempt*", "*crimin-*" and "*law*" capture the sentiments surrounding the issue particularly the controversy about some companies profiting from the sale of mugshot photos (Rostron, 2013) – an attack on individual privacy. While Topic 6 focuses on general FOI requests and the law at federal and provincial levels in Canada, topic 7 with terms such as "*media*", "*journalist*", "*fee*", "*waiver*" and "*blog*", appears to be addressing the issue of costs to requesters of freedom of information, as evidenced in many countries including the US[7]. Topics 9, 10, 15, 16 and 18 also carry similar themes about FOI requests, disclosures, access, politics and the reach of the law regarding FOI implementations. Topic 21 with words such as "*data*", "*research*", "*request*", "*nhs*", "*public*" and "*vexati-*" seems to be addressing general frustrations (vexations) regarding access to National Health Service (NHS) data even for the purposes of research. Though no specific country is mentioned in Topic 21, there appears to be a general apprehension among stakeholders regarding the use and re-use of sensitive clinical data under the FOI Geissbuhler et al. (2013), posits that while concerns about access to clinical data is genuine and must be regulated by governments, sharing health data advances public health research and improves patient care.

In the category of *law/legislation/exemptions*, Topics 3, 12, 14, 17, 18, 19, 20, 21 and 22 frame a general theme around the law, legislation, the courts, and privacy protection. Topic 12 in particular mentions "*ecthr*" which is the European Court of Human Rights and therefore appears to be case settlements regarding FOI. Topics 19 and 22 under the same category have terms like "*corrupt*", "*state*", "*public*", "*offici-*", "*convict*", "*commisi-*", "*court*" and "*legisl-*" appear to be raising issues of corruption and the need for the law to prosecute state officials implicated in corrupt deals.

In the health/environment category, two topics, 4 and 23, capture the discourses around health and environmental information disclosures. Topic 4 with words like "*health*", "*data*", "*clinic*", "*drug*", "*privaci-*", "*report*" and "*research*" is similar to Topic 21 and appears to focus on health information disclosures, access and attendant privacy issues. Topic 23 is loaded with lots of environmental issues and concerns relating to freedom of information. The terms "*pipelin-*", "*keyston-*", "*oil*" and "*nepa*" readily bring to mind the oil pipeline system in Canada and the United States, the National Environmental Policy Act (NEPA) and related concerns. The other terms in the category like "*environment*", "*climat-*", "*emiss-*", "*chang-*", "*impact*" and "*ghg*" apparently address issues about how the keystone pipeline project would impact climate change through emissions and related concerns about greenhouse gas (GHG) emissions. The study further found that most of the key terms in the various topics happen to have a strong correlation among themselves. Terms such as '*access*', '*media*', '*exempt*', '*legisl-*', '*privaci-*', '*request*' and '*law*' had at least a correlation of 0.70 among themselves. This gives credence to their regular occurrences in the topics. Topic 13 seemed to be an outlier as no relevant interpretation could be drawn from it. In determining the proportion of topics under each label, Topic 13 is not counted as well as topics that appeared in more than one category, such as Topics 18 and 21.

Overall, we recognize from the proportion of topics under each classification that the central themes in FOI research have largely been on issues relating to the core principles of FOI; be it issues surrounding disclosures, publishing, access, costs, exemptions among others.

OGD

Since the launch of open government data in 2009, the campaign to liberate public data has largely been shaped by the terms *Transparency, Openness, Participation and Collaboration* (Krishnamurthy & Awazu, 2016; Lathrop & Ruma, 2010; McDermott, 2010; Veljković, Bogdanović-Dinić, & Stoimenov, 2014). Surprisingly however, many of the topics apparently do not capture these widely accepted pillars of OGD. As shown in Table 5, only two topics representing 12.5% of the total topics implicitly talk about such. Other labels identified in the topics were *Technology; Economic/Social/Innovation and Citizen engagement.* As many as 50.0% of the topics centre on technology related matters. This seems to affirm the position by Janssen (2012) that OGD unlike FOI/RTI is technology driven rather than a rights seeking campaign. For instance, Topic 2 contains the terms "*app*", "*crm*", "*mobil*", "*citizen*" and "*platform*" suggesting the deployment of technologies such as mobile apps, customer relationship management (CRM) solutions and other platforms to engage citizens in a democracy. Topic 5 is rich with OGD technology related terms such as, "*link*", "*ogd*", "*dataset*", "*architectur-*", "*portal*" and "*metadata*" which appear to rehash the technology expectations of a functional open government data platform.

Topic 6 contains other unique technological terms like "*logd* (linked open government data)", "*web*", "*rdf* (Resource Description Framework)", "*databas-*" and "entiti-" which readily point to calls by OGD proponents for structured non-proprietary machine-readable technologies that support linked data and also aid in data access use and redistribution. Topic 7 under the same category also mentions several technology related terms. It can further be seen that Topic 17 in particular focuses on semantic ontology technologies as evidenced in terms such as "*rdf*", "*sparql*", "*semant-*", "*ontolog-*", "*queri-*" and

Table 6
Key named entities identified in the various topics.

	Topics	Name entity: country/region/city	Other Entities
FOI/RTI	Topic 1, Topic 5, Topic 6, Topic 8, Topic 11, Topic 15, Topic 16, Topic 20; Topic 2, Topic 7, Topic 10, Topic 18, Topic 23	U.S, India, Canada, Africa, London, Europe	Politicians, citizen(s), media, journalist, wikileak, court, parliament, nhs (National Health Service), key stone pipeline
OGD	Topic 3, Topic 4, Topic 6, Topic 10, Topic 12, Topic 13; Topic 2, Topic 8,	Brazil, U.S, India, Sweden, Europe, Australia	Citizen(s), media, user, janssen

"*metadata*". Semantic ontologies add value to an otherwise unconnected data by offering endless possibilities with linked data such as utilizing it across platforms and organizations. Linked data also help in generating data analysis, reports, publications and maps (Geiger & von Lucke, 2012; Shadbolt et al., 2012). Topics 18 and 19 also appear to mention technological terms all related to open government data.

Geiger and von Lucke (2012) argue that, beyond the often narrowest definition of OGD which revolves around the terms *Transparency, Openness, Participation and Collaboration,* other terms that matter are *social innovations and economic development.* In this regard, Topics 1 and 15 implicitly capture such themes in the discourses of open government data. With terms such as "*gdp*", "*reliabl-*", "*publish*", "busi-" and "*countri-*", Topic 1 mildly seems to address the socio economic dimension of OGD and its inherent potential to trigger business innovations among citizens (Janssen et al., 2012; Maude, 2012). The themes in Topic 15 also appears to reinforce the social, economic and innovation dimensions of OGD. Another concept often talked about in relation to OGD, is citizen engagement. Topics 11, 12, 13 and 16 contain words that seem to describe some form of government-citizen engagement.

The study further investigated whether some of the key terms as identified in the topics correlate strongly among themselves. Terms that are strongly associated with at least a 0.70 correlation measure were '*access*', '*media*', '*exempt*', '*innov-*', '*transpar-*', '*link*', '*ogd*', '*technolog-*', '*dataset*' and '*format*'. Overall, the proportion of topics indicate that the major central theme in OGD research have so far centered on technology and related issues. So much has been written on data publishing standards and technologies especially as relating to data formats, open linked data, the architecture and general data management standards as used in open government data portals. The following section compares the key themes identified in each campaign for similarities and differences. Particularly, we find out whether there are any similarities or differences not already mentioned in popular literature.

6.1. Topic comparison

The topics extracted and their apparent meanings bring out several familiar themes in FOI and OGD research. They also further reveal some similarities and differences between the two concepts as discussed earlier. For instance, the topics implicitly affirm that the FOI campaign has largely thrived on law whiles much of OGD's focus has been about open data technologies. The identified major theme in OGD research confirms the fear by Geiger and von Lucke (2012) that, care must be taken to not narrow or reduce the focus of OGD to data access technologies. Some named entities identified in the sets of topics under FOI and OGD, also appear to reinforce the belief that OGD has for the most part concentrated on the use of technology whiles FOI focuses on the law. For instance, terms (names) that refer to an individual involved in an FOI activity, tended to use words such as *citizen*, *public*, *person* and *author* as seen in Table 4. However in the OGD topics, the term "user" which is predominantly used to describe a kind of computer or information systems operator such as "end-user" frequently appeared.

The topics and their subsequent labelling further reveal that, unlike OGD that stresses on a free at no cost public data, data under FOI cannot be said to be entirely free since requesters in most countries including the U.S, U.K, Ireland, Australia, Canada and Scotland continue to charge fees to administer FOI requests (Goodspeed, 2011). Again, in spite of the fact that FOI charges fees, the process of obtaining an information is also comparatively laborious than with OGD. For instance a requester may have to first identify a relevant agency that has the information, compose a formal letter, pay a request fee and follow up on the application.[6] For the most part however, OGD requesters only have to access the needed data directly from national or city-based dedicated web portals in various data formats at no cost.

The comparison also brings to the fore similarities especially with names of recognized stakeholders under the two campaigns. The FOI topics make mention of names such as *journalist*, *politicians*, *public officials*, *citizens* and by inference *lawyers* (from the many occurrences of the terms, "laws" and "courts"). The topics under OGD also mentions entity names like *government*, *media* and citizens which come close to those identified in the FOI topics. This may seem to suggest that the two campaigns have a similar core group of stakeholders.

6.2. Named entities: country/regional/city-based initiatives

Named Entity Recognition (NER) is an information extraction task that seeks to identify and subsequently classify noun phrases (entities) found in textual data (Downey, Broadhead, & Etzioni, 2007). Typically, the noun phrases may refer to persons, locations, organizations, time, money or any other entity of interest. In this study, the topic modelling identified some key entities particularly location-based entities that may be worthy of mention. These entities under FOI and OGD are classified as *named entity: country/region/city* as shown in Table 6. It must be noted that while the named entities alone would not reveal a complete information, they often represent an activity, progress, impediment or an initiative relating to either FOI or OGD.

In FOI research, some location-based (countries, regions and cities) entities can be recognized in the topics as observed in Table 2 and subsequently classified in Table 6. Some of these entities are the *U.S*, *India*, and *Canada* (in Topics 1, 5 and 6 respectively); *Africa* and *Europe* (respectively in Topics 8 and 12); *Delhi* and *London* (in Topics 5 and 15). In the case of the *US* as a named entity, we recognize that while numerous publications address a range of topics, most have centered on the core issue of 'how much' information can be disclosed. This issue is often contested in various levels of the US judicial system and in congressional hearings (Relly & Schwalbe, 2016). On India, much of the FOI articles have focused on the law, access, and the impact of FOI on fights against corruption in public office (Calland & Bentley, 2013; Roberts, 2010). Other entities refer to key FOI stakeholders such as the *media*, *journalists*, *politicians*, *citizen*, *parliament*, *courts*, *library*, etc. The *National Environmental Policy Act* (NEPA), *WikiLeaks* are also some of the named entities mentioned in FOI topics as shown in Table 2. Similarly in OGD research, various entities mostly location-based can be identified in the topics generated. These are *Brazil, U.S, Sweden, Australia, India and Europe*. The mention of Sweden in Topic 11 in Table 3 is not particularly surprising since most texts that trace the history of open government often make reference to Sweden as one of the first countries to pass a legislation over 200 years ago, to make public

[6] https://www.icij.org/resources/2012/04/freedom-of-information.

information accessible to citizens (Janssen, 2012; Mendel, 2008). One other unique named entity appears in Topic 8 by way of "Janssen". This may be so because Katleen Janssen and Marijn Janssen have contributed so much to research in open government with lots of citations between them. Two of their most cited articles are "Benefits, adoption barriers and myths of open data and open government" by Marijn and "The influence of the PSI directive on open government data: an overview of recent developments" by Katleen.

7. Discussion and conclusion

Several decades have passed since the freedom of information act (FOIA) was conceived as a means to providing access to public data and with a view to entrenching the values of democracy. After many years of global successes and challenges in implementation, a similar movement in the form of open government data (OGD) was launched to help support the idea of greater openness and accountability in governance. Though run independently, the two campaigns continue to draw the world's attention to the importance of establishing a free accessible public data regime to augment the values of democracy. While the two concepts continue to receive considerable attention, the major themes that run in scientific written discourses have not clearly emerged. In this paper, topic modelling, text mining and document analysis methods were harnessed to determine the major topics running through FOI and OGD research publications and to establish whether these central themes help frame the ideologies in the two concepts.

The text analysis approach used in this comparative bibliometric analysis is a departure from traditional approaches where manual document analysis methods are used in extracting the metrics. In addition, traditional literature review approaches habitually tends to focus on information such as yearly number of publications, citation index, leading authors and affiliations, top journals and domain areas. Text analysis methods are now providing a computational alternative to bibliometric analysis through document authorship attribution, language identification, document retrieval and clustering.

The topic modelling not only helped in establishing the central themes in FOI and OGD research but also helped to clearly define similarities and differences in the two campaigns. The topic classification was carried out to determine the proportion of topics that fall under each of the labels. Since the topic labels mostly reflected key concepts in each campaign, a topic label with a significant proportion of the overall topics, give an indication of the central theme in that subject. In this respect, the results indicated that the central theme in FOI research have largely centered around issues of *disclosure*, *publishing*, *access* and *cost* of requests. The next major theme identified in FOI research hovers around issues of the *law*, *legislation* and *exemptions* relating to the FOI act. These two themes closely reflect most of the guiding principles in FOI as particularly enshrined in Article 19 of the Universal Declaration of Human Rights. On the other hand, the major theme in OGD research as shown in the topics seems to center on *technology* and related subjects. The next major running theme on OGD research is about issues relating to citizen engagements. Unlike FOI, the major themes identified in OGD research do not closely reflect much of its touted principles and key operational terms. This is because the topic label *Transparency/Collaboration/Participation* which are terms which have mostly occasioned the OGD campaign, conspicuously had very few proportion of the topics. This reinforces the point that much of the focus on OGD research have been on technological issues relating to data access and management of public data.

In respect of research question 2, the results by the topic modelling particularly revealed some other differences between freedom of information and open government data that have not yet been mentioned in the literature like those addressed by Geiger and von Lucke (2012), Janssen (2012) and Ubaldi (2013). One of such difference identified in this study, is the issue of cost to an individual accessing a public data. Though the focus in the two campaigns is to progressively make public data free at no cost, FOI requesters are still charged requests fees even in many advanced countries. On the other hand, OGD data sets appear to be completely free of costs to citizens who have an added luxury of accessing the materials in many formats. It is not clear whether the kinds of data requested under FOI warrants the charges. The study also shows that, comparatively, accessing an OGD data is easier than the often strenuous processes that requesters go through to obtain FOI data. In terms of the topic labelling, research question 3 is answered since the use of expert knowledge and the literature helped to frame the discourses surrounding the two subjects and subsequently extracted valuable information that give a general trend of the themes in each campaign.

Research question 4 sought to understand how named entities identified in the topics comment about each campaign. The key named entities give an indication for further research. For instance, a thorough review of publications revealed that, whiles numerous country-level FOI activities have been reported in the literature, only a few have focused on country-to-country comparisons. It would be interesting to see future research focus on comparing countries over some FOI performance measures such as the level of involvement by civil societies and the media, the impact of FOI on corruption fights among others. Though this approach should not be intended as a score card on countries, it would give a general trend of FOI performances worldwide. Similarly, we also realized that whereas much has been written about individual country performances, initiatives and general happenings on OGD, research seems to be silent on city or state based activities. At present, only a few OGD city-based scientific publications have been written such as on Chicago, Vienna, Rotterdam, Bologna and Trentino. This appears to give OGD a narrow scope since open government data is not only meant to be practiced at the level of central governments. For instance, while over 70 international countries are involved in OGD programmes, there are also 164 international cities and regions practicing OGD with independently run data web portals.[2,7] In view of this, future research on OGD should focus on various city-based activities to help broaden the scope and the understanding of what OGD truly entails.

References

Afful-Dadzie, E., Nabareseh, S., Oplatková, Z. K., & Klímek, P. (2016). *Framing media coverage of the 2014 sony pictures entertainment hack: A topic modelling approach. Proceedings of the 11th international conference on cyber warfare and security: ICCWS2016*. Academic Conferences and publishing limited (p. 1).

Aggarwal, C. C., & Zhai, C. (2012). *Mining text data*. Springer Science & Business Media.

Aghaei Chadegani, A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., et al. (2013). A comparison between two main academic literature collections: Web of science and Scopus databases. *Asian Social Science*, *9*(5), 18–26.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. http://dx.doi.org/10.1016/j.giq.2015.07.006.

Bíró, I., Szabó, J., & Benczúr, A. A. (2008). Latent dirichlet allocation in web spam filtering. *Proceedings of the 4th ACM international workshop on adversarial information retrieval on the web*, 29–32.

Berners-Lee, T. (2006). *Linked data-design issues*. http://www.w3.org/DesignIssues/LinkedData.html.

Birkinshaw, P. (2010). Freedom of information and its impact in the United Kingdom. *Government Information Quarterly*, *27*(4), 312–321. http://dx.doi.org/10.1016/j.giq.2010.06.006.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(January), 993–1022.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. http://dx.doi.org/10.1145/2133806.2133826.

Camaj, L. (2016). From 'window dressing'to 'door openers'? Freedom of Information legislation, public demand, and state compliance in South East Europe. *Government Information Quarterly*, *33*(2), 346–357. http://dx.doi.org/10.1016/j.giq.2016.03.001.

Carrasco, C., & Sobrepere, X. (2015). Open government data an assessment of the spanish municipal situation. *Social Science Computer Review*, *33*(5), 631–644. http://dx.doi.org/10.1177/0894439314560678.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Proceedings in Advances in Neural*

[7] https://www.data.gov/open-gov/.

Information Processing Systems, 288–296.
Charalabidis, Y., Alexopoulos, C., & Loukis, E. (2016). A taxonomy of open government data research areas and topics. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 41–63. http://dx.doi.org/10.1080/10919392.2015.1124720.
Chignard, S. (2013). *A brief history of open data*. Paris Tech Review29.
DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, *41*(6), 570–606. http://dx.doi.org/10.1016/j.poetic.2013.08.004.
Donnelly, J. (2013). *Universal human rights in theory and practice*. Cornell University Press.
Downey, D., Broadhead, M., & Etzioni, O. (2007). Locating complex named entities in web text. *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, 2733–2739.
Evans, A. M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management*, *32*(1), 172–185. http://dx.doi.org/10.1002/pam.21651.
Foerstel, H. N. (1999). *Freedom of information and the right to know: The origins and applications of the Freedom of Information Act*. Greenwood Publishing Group.
Frey, L. (2014). *Open government partnership four-year strategy 2015–2018*. http://www.opengovpartnership.org/sites/default/files/attachments/4YearAP-Online.pdf.
Geiger, C. P., & von Lucke, J. (2012). Open government and (linked)(open)(government)(data). *JeDEM-eJournal of eDemocracy and Open Government*, *4*(2), 265–278.
Geissbuhler, A., Safran, C., Buchan, I., Bellazzi, R., Labkoff, S., Eilenberg, K., et al. (2013). Trustworthy reuse of health data: A transnational perspective. *International Journal of Medical Informatics*, *82*(1), 1–9. http://dx.doi.org/10.1016/j.ijmedinf.2012.11.003.
Gigler, B. S., Custer, S., & Rahemtulla, H. (2011). *Realizing the vision of open government data: Opportunities, challenges, and pitfalls*. World Bank.
Halstuk, M. E., & Chamberlin, B. F. (2006). The Freedom of Information Act 1966–2006: A retrospective on the rise of privacy protection over the public interest in knowing what the government's up to. *Communication Law and Policy*, *11*(4), 511–564. http://dx.doi.org/10.1207/s15326926clp1104_3.
Power, voting, and voting power. In M. J. Holler (Ed.), Springer Science & Business Media.
Hubbard, P. (2008). China's regulations on open government information: Challenges of nationwide policy implementation. *Open Government: a Journal on Freedom of Information*, *4*(1), 1–34.
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.
Janssen, K. (2012). Open government data and the right to information: Opportunities and obstacles. *The Journal of Community Informatics*, *8*(2).
Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2012). The value of open government data: A strategic analysis framework. *Proceedings of the 2012 pre-ICIS workshop*.
Jiang, H., Qiang, M., & Lin, P. (2016). A topic modeling based bibliometric exploration of hydropower research. *Renewable and Sustainable Energy Reviews*, *57*, 226–237.
Krishnamurthy, R., & Awazu, Y. (2016). Liberating data for public value: The case of data.gov. *International Journal of Information Management*, *36*(4), 668–672.
Lathrop, D., & Ruma, L. (2010). *Open government: Collaboration, transparency, and participation in practice*. O'Reilly Media, Inc.
Lau, R. R., Patel, P., Fahmy, D. F., & Kaufman, R. R. (2014). Correct voting across thirty-three democracies: A preliminary analysis. *British Journal of Political Science*, *44*(02), 239–259.
Mann, G. S., Mimno, D., & McCallum, A. (2006). Bibliometric impact measures leveraging topic analysis. *Proceedings of the 6th ACM/IEEE-CS joint conference on digital libraries*, 65–74.
Maskeri, G., Sarkar, S., & Heafield, K. (2008). Mining business topics in source code using latent dirichlet allocation. *Proceedings of the 1st India software engineering conference* (pp. 113–120).
Maude, F. (2012). *Open data white paper-unleashing the potential. the stationary office limited on behalf of HM government*. London, United Kingdom: Cabinet Office.
Mendel, T. (2008). *Freedom of information: A comparative legal survey*. Paris: Unesco.
Novais, T., de Albuquerque, J. P., & da Silva Craveiro, G. (2013). *An account of research on open government data (2007–2012): A systematic literature review. EGOV/ePart ongoing research* (pp. 76–83).
Owen, G. T. (2014). Qualitative methods in higher education policy analysis: Using interviews and document analysis. *The Qualitative Report*, *19*(26), 1.
Ríos, S. A., Aguilera, F., Bustos, F., Omitola, T., & Shadbolt, N. (2013). Leveraging social network analysis with topic models and the Semantic Web (extended). *Web Intelligence and Agent Systems: An International Journal*, *11*(4), 303–314.
Ramanathan, V., & Wechsler, H. (2012). Phishing Website detection using latent Dirichlet allocation and AdaBoost. *Proceedings of 2012 IEEE conference on intelligence and security informatics*, 102–107.
Relly, J. E., & Schwalbe, C. B. (2016). How business lobby networks shaped the US Freedom of Information Act: An examination of 60years of congressional testimony. *Government Information Quarterly*, *33*(3), 404–416. http://dx.doi.org/10.1016/j.giq.2016.05.002.
Roberts, A. (2010). A great and revolutionary law? The first four years of India's Right to Information Act. *Public Administration Review*, *70*(6), 925–933. http://dx.doi.org/10.1111/j.1540-6210.2010.02224.x.
Rostron, A. (2013). The mugshot industry: Freedom of speech, rights of publicity, and the controversy sparked by an unusual new type of business. *Washington University Law Review*, *90*(4).
Schartum, D. W. (1998). Access to government-held information: Challenges and possibilities. *The Journal of Information, Law and Technology*. http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/1998_1/schartum/.
Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data.gov.uk. *IEEE Intelligent Systems*, *27*(3), 16–24.
Shepherd, E., Stevenson, A., & Flinn, A. (2009). The impact of freedom of information on records management and record use in local government: A literature review. *Journal of the Society of Archivists*, *30*(2), 227–248. http://dx.doi.org/10.1080/00379810903445000.
Tauberer, J. (2012). *History of the movement. Open government data: The book*https://opengovdata.io/2014/history-the-movement/.
Tinati, R., Carr, L., Halford, S., & Pope, C. (2012). *Exploring the impact of adopting open data in the UK government. Digital futures*. Aberdeen, GB: Web & Internet Science 3pp.
US. Senate (2007). *Open government: Reinvigorating the freedom of information act: Hearing before the committee on the judiciary, United States senate, one hundred tenth congress, first session, march 14, 2007*. Washington: U.S. G.P.O. DIANE Publishing. https://www.gpo.gov/fdsys/pkg/CHRG-110shrg35801/pdf/CHRG-110shrg35801.pdf.
Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives*. Paris, France: OECD.
Van Dooren, W., Bouckaert, G., & Halligan, J. (2015). *Performance management in the public sector*. Routledge.
Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, *31*(2), 278–290. http://dx.doi.org/10.1016/j.giq.2013.10.011.
Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 448–456).
Wang, H. J., & Lo, J. (2016). Adoption of open government data among government agencies. *Government Information Quarterly*, *33*(1), 80–88. http://dx.doi.org/10.1016/j.giq.2015.11.004.
Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, *10*(2), 347–364. http://dx.doi.org/10.1016/j.joi.2016.02.003.
Wei, X., & Croft, W. B. (2006). LDA-based document models for ad-hoc retrieval. *Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval*, 178–185.
Weisberg, H. F., & Nawara, S. P. (2010). How sophistication affected the 2000 presidential vote: Traditional sophistication measures versus conceptualization. *Political Behavior*, *32*(4), 547–565. http://www.jstor.org/stable/40960954.
Whitmore, A. (2012). Extracting knowledge from US department of defense freedom of information act requests with social media. *Government Information Quarterly*, *29*(2), 151–157.
Witten, I. H. (2005). *Text mining. Practical handbook of internet computing, 14-1. CRC-computer-and-information-science-series/book-series/CHCOMINFSCI*. Chapman-Hall.
Xing, D., & Girolami, M. (2007). Employing latent Dirichlet allocation for fraud detection in telecommunications. *Pattern Recognition Letters*, *28*(13), 1727–1734. http://dx.doi.org/10.1016/j.patrec.2007.04.015.
Yau, C. K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*, *100*(3), 767–786. http://dx.doi.org/10.1007/s11192-014-1321-8.
Zeleti, F. A., Ojo, A., & Curry, E. (2016). Exploring the economic value of open government data. *Government Information Quarterly*, *33*(3), 535–551. http://dx.doi.org/10.1016/j.giq.2016.01.008.
Zuiderwijk, A., & Janssen, M. (2014). The negative effects of open government data-investigating the dark side of open data. *Proceedings of the 15th ACM annual international conference on digital government research*, 147–152.