The needle and the haystack: A literature review using Structural Topic Modeling in a Digital Government Corpus

Andres Aguilera Castillo

04 June 2022

Abstract

Digital Government is a growing and vibrant multidisciplinary field of research, the fast increase in research output has challenged researchers to explore and use novel computational ways and methods to perform evidence synthesis on the extant literature and be able to map a scientific discipline, explore the thematic evolution over time and identify potential avenues for further research. Topic modeling has emerged as a powerful technique from the computer science field that is contributing to the examination of large amounts of text data. This manuscript demonstrates the training a structural topic model aimed to assemble a 'smart literature review' on a subset of the Digital Government Reference Library (DGRL) version 17.5. Structural topic modeling is a conceptual and methodological evolution of 'vanilla' topic modeling that allow the estimation of covariates contained in the metadata of corpora to calculate topic prevalence in a corpus. To our best knowledge, this is the first attempt to use unsupervised machine learning techniques with this data set. This effort may contribute to creating a map of the field, identify the evolving themes in the literature and help to identify promising areas of research.

Findings?

Introduction

Recent trends in global scientific output demonstrate a rapid and sustained increase in the production of vast amounts of unstructured data in the form of digitized text. This bounty in content is challenging researchers to explore and pursue novel methodological approaches and techniques to examine massive volumes of scientific publications in a systematic, efficient and reproducible manner. The expanding amount of bibliographic information available is exceeding traditional approaches for processing research output making it necessary to apply computational-assisted approaches for science mapping and evidence synthesis.

Topic modeling is an iterative process, thus this manuscript explores the training set[1] of abstracts of journal articles contained in the Digital Government Research Library (DGRL) via a Structural Topic Model. Probabilistic topic models are a type of unsupervised machine learning processes that allow the exploration of a vast collection of documents (also known as corpus), perform the automated classification of large amounts of textual data and hence assist scholars in research tasks such as discovery, measurement, prediction and causal inference. Topic modeling enables the use of larger bibliographic data sets, and the extraction of relevant concepts from sizable corpora in a scalable way. To our best knowledge this is the first attempt to run a topic model for a corpus in the field of Digital Government Research.

The corpus used for this analysis is a subset of the journal articles in the version 17.5 of the Digital Government Reference Library. As argued by Grimmer, Roberts, and Stewart (2022a), texts are "expensive to

[1] Best practices found in the literature suggest the split of sample data for topic modeling between a training and test data set. The structural topic model used in this manuscript have been trained on 75% of the data. A subsequent product will make use of the held out data.

produce, gather and collate", the contents of previous versions of this data set have been used as primary or secondary source of data exploring the Digital Government field. The Digital Government Research Library is a collection of bibliographic references associated with Digital Government scholarship. In its 17.5 version, it contains more than 16500 references, including journal articles, book chapters and conference papers. Contributions to this research domain come from established disciplines such as information science, computer science, organization science, sociology, public administration, and political science (Scholl 2021a).

Previous explorations of this reference library have revealed the thematic evolution using bibliometric and scientometric approaches (AlcaideMuñoz et al. 2017), and identified the most influential journals, conferences and leading scholars in the field (Scholl 2021b). This data set has been used as well for conducting a systematic review on the impacts of e-Government using a public value perspective (MacLean and Titah 2021). According to (Webster and Watson 2002a), accumulating a "complete census" of the relevant literature and following a concept-centric framework are crucial in a literature review. Concept-centric approaches with topic modeling might be conducted via the use of "seed word dictionaries" in semi-supervised topic models (Watanabe and Zhou 2020), but this technique is out of the scope of this manuscript.

The study of the linkage between modern technologies and the quality and quantity labor has been on the research agenda of diverse disciplines and academic fields such as economics(Dosi et al. 2021; Fernández-Macías and Bisello 2022), industrial relations (Doellgast and Wagner 2022a), information systems (Klein and Watson-Manheim 2021), and organization studies (Stephen R. Barley 2020), primarily focused in the private sector from advanced democracies. Technological change is a very broad term that may include a wide array of ICT-enabled applications for automation, digitalization and robotization. Our attention is directed at the digitalization of government, but despite the momentum in digital government research, one aspect that remains under explored is the empirical assessment of the effects of digital technologies on the public sector workforce (Plesner, Justesen, and Glerup 2018).

The public sector "composition" can be very heterogeneous in terms of scale and scope among diverse jurisdictions. Public sector organizations rank high globally among the largest employers in the form of armies and other defense related operations, State-Owned Enterprises (SOE), and health care providers, to name a few. The 'industries' in which public organizations operate are very diversified, have distinct degrees of technological sophistication and mixed levels of interaction with citizens and firms. In general, the public sector commands a large, diverse, and highly educated workforce.

The public sector is also considered one of the largest adopters and users of ICT, and perform a key role in the creation and governance of enormous amounts of data (Guenduez, Mettler, and Schedler 2020; Lofgren and Webster 2020). Historically, governments have developed the required information infrastructure to manage data intensive operations such as population and property registries, tax collection, and medical records among others. The pervasive deployment and use of digital technologies, digital platforms, and digital infrastructures has accelerated the rate of new data creation thus transforming the operations of firms and public organizations with profound implications for the organization of work (Nambisan, Wright, and Feldman 2019).

The reported impact of digitalization on the organization work is diverse (Stephen R. Barley and Kunda 2001), it may automate work, create or eliminate jobs, deskill or reskill workers but also, little or negligible impact whatsoever. Digital government as a research field is in a phase of consolidation, allowing for the exploration of promising subfields for further inquiry. Digital technologies and the novel design of public services may facilitate a more intricate division of labor into smaller components (tasks), reconfiguring the workflow of public services, fostering new ways for multi-actor co-production (Bryson et al. 2016), promoting the implementation of self-service solutions and facilitating scenarios for the co-production of public services (Scupola and Mergel 2021), turning each citizen and user into "his or her own administrator, caseworker and bureaucrat" (Schou and Hjelholt 2018), and possibly creating detrimental effects such as administrative burden for citizens (Madsen, Lindgren, and Melin 2021).

These developments enabled by the implementation of digital technologies in public organizations are changing the interaction between citizens and public officials turning it into a technology-mediated public encounter (Lindgren et al. 2019), introducing changes in the organization of work in terms of task redundancies and the creation of new occupations, to cope with an increasingly digitalized public sector. The argumentative

arc presented above sparks the discussion of automation in a public sector context as an emerging topic of interest in the extant literature (Engin and Treleaven 2019; Andersson, Hallin, and Ivory 2021; Lloyd and Payne 2021).

Conceptual developments in Digital Government Research have considered the success factors of digital government initiatives from both the supply and demand sides. However, given the intrinsic complexity associated with the public sector, a more elaborate discussion is found in the design and use literature that incorporates analytic dimensions such as power, ideology, design, and institutional change in the study of how novel technologies affect the organization of work (Bailey and Barley 2020).

The relationship between digitalization and work is complex and multifaceted, its impacts are variegated among organizations, industries and employee groups (Doellgast and Wagner 2022b), however, it was the global health emergency in early 2020 that favored the newly gained awareness and increased the research interest in the subject matter (Nagel 2020; Dingel and Neiman 2020; Mazzucato and Kattel 2020; Leonardi, Woo, and Barley 2021; Faraj, Renno, and Bhardwaj 2021). Thus, it is deemed pertinent and timely to pursue the scholarly exploration of the effects of technological change in public organizations and its consequences for the public sector workforce.

The use of text based techniques and topic models have gained traction among scholars exploring the nexus between novel technologies and labor markets. Among these novel approaches are (Montobbio et al. 2022) that explore robots and labor-saving technologies, and (Kogan et al. 2019) that analyze patent contents to estimate technological change and labor displacement.

Supervised, semi-supervised and unsupervised machine learning techniques for text analysis can be used in a wide range of disciplines to examine databases, repositories and corpora, hence expanding the methodological repertoire of researchers opening an opportunity to explore large troves of data. It is our opinion that this methodological innovation can be repurposed to explore the linkages between digitalization and organization of work in a public sector context.

This initial argumentation lead us to formulate the following research questions:

RQ1: What does topic modeling techniques applied to the Digital Government Research Library v17.5 reveal about the conceptual, intellectual and thematic evolution this academic field? –> Text mining and STM

RQ2: What structural changes can be interpreted from the topic model? –> (Covariates)

RQ3: What does the extant literature (corpus) on Digital Government reveal on the linkage between digitalization and the organization of work? – SeededLDA? (in-progress)

The objective of this exercise is to analyze and present the results of the application of a structural topic model, a novel method for evidence synthesis, in the exploration of the effects of digitalization in the organization of work in the public sector. The advent of computerization and digitalization has had broad impacts in most aspects of contemporary life, including scientific research. Digitalization has influenced how research is designed and conducted, allowing for the creation and increased availability of ever-growing data sets that require powerful computational methods and enhanced tools to handle abundant information (Meyer and Schroeder 2015). Therefore, this manuscript aims to offset the reported "excessive use" of qualitative methods in e-government research (Arduini and Zanfei 2014), and answering to calls in the extant literature towards the pursuit of quantitative and empirically oriented approaches (Wirtz and Daiser 2016).

Literature Review and Conceptual Framework

Studies in the history of science have identified a relatively sustained growth pattern in scientific publications over time, this exponential growth rate means a doubling in scientific output every 17 years approximately (Bornmann, Haunschild, and Mutz 2021). This level of growth might be attributed to the increased resources dedicated to the global scientific endeavor and consequently the communication of science via publications. However, it may also be due to what has been dubbed "salami sliced publishing" or the multiple publications of a single research study (Bornmann and Daniel 2007; Bornmann and Mutz 2015).

Research synthesis is part of the literature review process in which the extant scientific knowledge in each academic field is examined to help scholars understand the conceptual structure, themes, and debates to identify trends in the literature and potential areas for further research. This crucial task is labor-intensive, time-consuming, and restricted to a limited number of documents if conducted by traditional "manual" methods (Antons et al. 2020a; Asmussen and Møller 2019a). Still, computer-assisted text analysis does not substitute human intervention, instead it "augments our reading ability" (Grimmer, Roberts, and Stewart 2022b), human judgement is deemed necessary for the evaluation and validation of the outcome of these models (Barberá et al. 2021).

Quantitative research synthesis techniques like bibliometrics and computer-assisted text mining allow the analysis of a larger quantity of documents and may contribute to advancing the "research fronts" in inter-disciplinary fields such as Digital Government (Tanskanen et al. 2017). Computational tools and techniques developed originally in the computer science field have been repurposed in diverse disciplines but also have enabled social scientists to exploit Natural Language Processing (NLP) applications for classification tasks of large scientific corpora. Topic modeling techniques, a subset of machine learning and NLP allow for the automatic classification of vast amounts of text data.

Unstructured text has become one of the most prevalent types of data in the current "data deluge". In organization research, text is considered a key source of data as organizations publish content on their websites, social media and other searchable repositories (Kobayashi et al. 2017). The use of text analysis or text mining is not necessarily new; however, the digitalization of everyday life has facilitated the creation, storage and analysis of enormous quantities of data in text format. Nonetheless the usage of text mining techniques has remained "disconnected among fields" (Banks et al. 2018).

Probabilistic topic modeling is a method that extracts topics from a collection of text. According to the seminal work by (Blei, Ng, and Jordan 2003), Latent Dirichlet Allocation (LDA) applied to a corpus generates a probabilistic model in which documents are represented as the mixtures of latent topics, and topics are characterized by a distribution of words. LDA is considered the state-of-the-art, simplest and most used method to perform topic modeling (Asmussen and Møller 2019b).

LDA models are becoming widely used in social science, however these techniques are not infallible and require rigorous validation and human interpretability (Maier et al. 2018a), if not, it may be as factual as "reading tea leaves" as eloquently put it by (Chang et al. 2009). For a robust analysis it is advised to take an iterative approach for build, compute, critique, and rebuild topic models (Blei 2014).

Even though these techniques originated in the computer science field and at first sight may seem arcane to newcomers, there have been important progress in other research areas towards facilitating the adoption of this powerful computational tool by lowering the technical barriers, the creation of agreed-upon workflows for modeling and visualization, and the development of relatively accessible software packages in open source statistical software like R and Python (Rehurek and Sojka 2010; Benoit et al. 2018a; Roberts, Stewart, and Tingley 2019a).

Topic modeling techniques applied to bibliographic data have been explored in diverse scientific realms and academic disciplines such as statistics (De Battisti, Ferrara, and Salini 2015), economics (Ambrosino et al. 2018), cliometrics (Wehrheim 2019), innovation research (Antons et al. 2020b; Antons and Breidbach 2017), and management (Hannigan et al. 2019). The scope of these analyses can be very large, (Ambrosino et al. 2018) studied the evolution in the thematic structure of the economics discipline by applying LDA to the full texts of articles published in 188 journals in the JSTOR database from 1845 to 2013 (n= 250846). Other implementations have concentrated its attention and analysis, (Antons, Kleer, and Salge 2016) explored the full text corpus (n=1008) of a single top ranking journal in innovation research over a three decade span.

Structural Topic Modeling is a conceptual and technical evolution of the the typical topic modeling approach by incorporating the estimation of topic prevalence using covariates found in the metadata of the corpus (Roberts, Stewart, and Airoldi 2016). Applications of this method to bibliographic data have estimated the role of covariates such as temporal and geographic information in the analysis of the dissertation titles in economics and chemistry in East and West Germany before and after the German reunification (Rehs 2020).

As advised by (Barberá et al. 2021), there are "consequential decisions" in the methodological choices of automated text classification and the fact that human validation is a key component of text as data methods.

The selection of a corpus in itself is deemed a crucial decision that can be prone to four types of bias: resource bias, incentive bias, medium bias and retrieval bias, these selection biases are well acknowledged in the text as data literature (Grimmer, Roberts, and Stewart 2022b). It may be probable that the DGRL v17.5 has omitted important research that is not included in this collection. All decisions concerning text as data methodologies are "consequential", our aim is to make our workflow reproducible by documenting all the choices in the scripts associated with this document.

LDA is an unsupervised machine learning method which means the relationship between words and topics is ignored prior to the execution of the model. Thus is deemed good practice to split the data between a training set and a test set. Our approach is to train the model with 75% of the corpus, leaving the remaining proportion for testing purposes. The optimal number of topics (k) is unknown and the researcher should selected this parameter, there is technically no "right number of topics" and this choice might be specific to a corpus and research design (Grimmer and Stewart 2013). In general, a low number of topics is used for an overview, instead, a higher number of topics is used for more granular analysis of the corpus (Asmussen and Møller 2019b).

The evaluation of topic models can be performed through the calculation of goodness of fit statistics and the iterative calibration of the model to increase interpretability via "eye balling" the topics and their word-probability, and human judgement, meaning the implicit knowledge of the researcher on the subject matter of a corpus. A rule of thumb found in the documentation of the stm R package states that for small corpora, like the one used for this analysis containing "a few hundred to a few thousand" documents, 5 to 50 topics is "a good place to start, then an iterative calibration of the model is due. In addition, the stm R package includes functions for model selection, visualization and estimation of the effects of covariates in topic prevalence (Roberts, Stewart, and Tingley 2019b).

Four goodness of fit measures are usually considered when exploring the optimal number of topics to apply to a corpus: perplexity, coherence, residuals and lower bound. The held-out likelihood, also know as perplexity, measures how well the probability model predicts unseen data, a lower number in this measure implies a higher the accuracy of the model. Semantic coherence is maximized when the most probable words in a topic co-occur frequently (Roberts et al. 2014). The lower bound indicator explains the convergence in the iterations of the model, when there is small change among iterations the model is considered converged. As for residuals, this diagnostic measure calculates the sample dispersion, if the number for this value is greater than one (>1) it suggests that the number of topics are set too low (Taddy 2011).

Text is a type of unstructured data that requires intensive processing to be able to work with it. This means that before being able to create and analyze a corpus object containing the information deemed of interest, "consequential decisions" have to be made. It is considered a best practice to use version control systems in the all the phases of the analysis for efficiency but also for replicability and transparency purposes.

Text data is incredibly diverse in length and contents. Social media posts, political speeches, press releases and customer reviews are the usual targets of this kind analysis. For researchers exploring bibliographic data the unit of analysis can be the title of the document, the abstract or the whole text of the documents in the corpus. Text data can be coerced into a type of structure for processing using the bag of words approach. The bag of words assumption means that the order of words within each document is ignored and the thematic structure of the document can be inferred by the frequency distribution of words (Maier et al. 2018b).

The bag of words approach deliberately ignores the syntax or structure of the text, the creation of a bag of words is known as tokenizing. Additional treatment of text include the elimination of punctuation, transform each word to lowercase and in some cases stemming which is a way to reduce a word to its stem or root in order to reduce the sparsity of the resulting matrices. Even though these steps may seem difficult to understand at first, the publication of open software packages, the availability of vast documentation, tutorials and vibrant online knowledge communities have lowered the technical barriers of this powerful computational tool for research.

The next step in pre-processing is the creation of the document-feature matrix[2] containing all the documents

[2]In the quanteda R package the Document-Feature Matrix is equivalent to the Document-Term Matrix of alternative text mining software. Features in this context are the individual tokens (single words) from the documents in the corpus.

and the tokenized text, the usual result is a very sparse matrix. Best practices found in the literature suggest to perform dimensionality reduction to the matrix by dropping features with very low frequency of occurrence and the very common features, the most common words in the corpus given that it is assumed that these very common words will not contribute to the discovery of the latent structure of the corpus.

Computational tools like topic models are enabling researchers to explore and analyze larger data sets of bibliographic information to conduct evidence synthesis by facilitating the exploration of a vast corpora, perform the automated classification of textual data and assist scholars in research tasks such as discovery, measurement, prediction and causal inference.

Methods and Data

The Digital Government Research Library version 17.5 is a large curated repository of publications contributing to the field of Digital Government Research (DGR), it contains more than 16500 references among its records. The most prevalent types of documents are conference papers (33.2%) and journal articles (50%). The inclusion criteria of the DGRL are: to have passed academic peer review, to be published in an academic journal, to be published in English language (Scholl 2021a). The Library can be downloaded from the website DGRL.

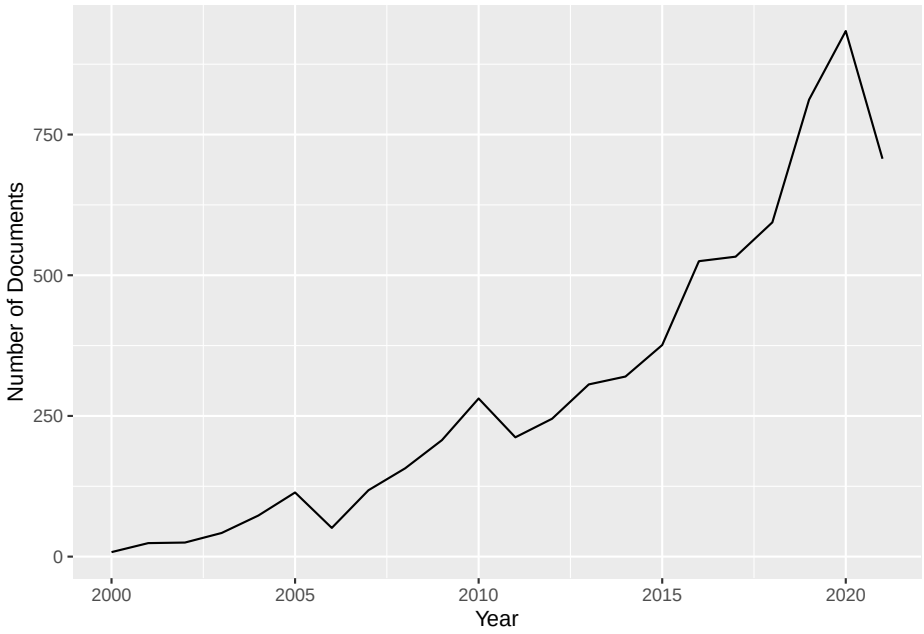
Table 1: Contents of DGRL v17.5 without processing

Document Type	Number of Documents
journalArticle	8278
conferencePaper	5492
bookSection	2083
book	636
report	33
thesis	3
magazineArticle	2
manuscript	1
webpage	1

The download package contains three types of bibliographic files BibTeX, RIS, and ENL (EndNote). In its raw and unprocessed form, the data has a large proportion of missing values, mostly clustered in metadata not considered relevant for the analysis. By exploring the different bibliographic formats, BiBTeX, RIS and ENL files, we noticed that the data sets had a large amount of missing data and that some information was available in a file type but not other. The script for initial data wrangling and data transformation documents the steps and choices made to the initial filtering and de-duplication. The unique digital object identifier (DOI) served as a exact key to merge the data sets, also as a "quality control" step to retain documents with valid DOIs.

The following variables have been deemed of interest for the analysis: type of reference (conference paper or journal article), year of publication, author, document title, publication title and the presence of an abstract. Text is a type of unstructured data that requires meticulous processing before using it. For replicability purposes, the script for the initial data cleaning and wrangling, including the R functions and packages used is available for revision, clarity, and replicability purposes and made publicly available in the scripts section of the GitHub repository for this project.

After the initial data wrangling, the relevant data for 6682 journal articles or approximately 80.7% of the total number of articles in the DGRL v17.5 is further processed to create a corpus, the initial step towards a topic model. Documents published before year 2000 were dropped from further analysis due to their negligible quantity, also a single observation from year 2022, this reduced the corpus to be analyzed slightly to 6664 documents. A visualization in the publication trend demonstrate an incipient increase in number of journal articles after year 2000 and a steep increase in the beginning of the 2010 decade to present.



The subsequent step is the creation of a corpus object. A second script describes the phase of pre-processing related to preparing the unstructured text data into a format that is usable for analysis. Steps like tokenization, removal of stopwords, symbols, and special characters, and conversion to lowercase, are part of this phase (Maier et al. 2018b). There are several software packages for text analysis, mining, and visualization, our choice for pre-processing was conducted in R statistical software using the functions of the quanteda R package (Benoit et al. 2018b). After several iterations, we deemed pertinent the use of an stemming algorithm to aid to the dimensionality reduction in the matrices by cutting words to their root form.

As suggested by Webster and Watson (2002b), a complete review covers the relevant literature and it is not limited by a single research methodology, set of journals or geographic region, topic models contribute to expand the options available to researchers and the amplify the scope and reach of their inquiries. In this exercise, the top 10 publication titles (journal name) in the corpus represent almost a third of the documents in the sample. By making quick search in the Scimago Journal Rank website, it can be established that all publication titles on the table are listed in this database.

Table 2: Top 10 Journals in represented in the corpus

Publication Title	Number of Documents in Corpus
Government Information Quarterly	777
Transforming Government: People, Process and Policy	255
Journal of Information Technology & Politics	253
Electronic Government, an International Journal	202
International Journal of Electronic Governance	198
International Journal of Electronic Government Research	142
Information Technology for Development	123
Social Science Computer Review	107
Information Polity	105
International Journal of Public Administration	98

Text as data methods are inherently iterative thus requiring the adoption of suitable workflows and best practices for model calibration and version control systems of its operations, even though stop words are considered language specific and Natural Language Processing applications are advancing in sophistication,

some stop words are corpus specific. In the downstream of this process we found strings with no relevant meaning to the analysis but very prevalent in the corpus, thus the importance in the construction of the workflow in a programmatic manner in an R environment.

Quanteda pre-processing workflow includes functions that allow the finding of multi-word expressions via collocation analysis, this might be useful to identify proper names and other meaningful words. However, the "killer application" is the finding of n-grams, meaning the identification of single words (unigrams), or sequence of words (bigrams, trigrams) that tend to occur together with high frequency and may carry valuable information about the contents of the corpus.

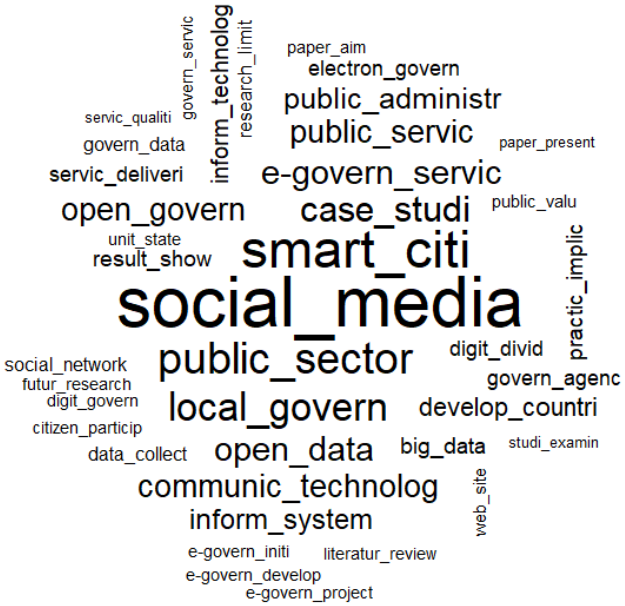


Figure 1: DGRL Corpus Bigrams Visualization

This visualization shows the top 40 bigrams found in the corpus, it can be seen that the words have been stemmed to their root form, the most salient bigram is "social_media" implying the centrality of these platforms for digital government scholarship, from adoption and use by public organizations (Mergel and Bretschneider 2013, 2013), to the role of social media in political campaigns (Karlsen 2010; Mascheroni and Mattoni 2013), the regulation of disinformation (Marsden, Meyer, and Brown 2020), and the provision of public services (Tursunbayeva, Franco, and Pagliari 2017; Criado and Villodre 2021).

Other salient bigrams include "smart_citi", "local_govern", and "open_data". A closer examination provides hints on methodological aspects, the bigram "case_studi" carries a lot of meaning informing about the frequency of this method in the sample. The ubiquity of the bigram "public_valu", for the public value theory shows the important evolution from New Public Management to alternative theoretical frameworks (Panagiotopoulos, Klievink, and Cordella 2019). For a summary of the most frequent theoretical frameworks used in Digital Government Research refer to the work of Bannister and Connolly (2015).

The ubiquity of the word service in bigrams such as "public_servic", "servic_deliveri", "servic_qualiti", "govern_servic" that provide a glimpse in the nature of government operations, the creation of public services but not necessarily with a service logic as argued by Cordella and Paletti (2018). This also suggests the influence of the work by Vargo and Lusch (2004) on service-dominant logic and its conceptual evolution including digital aspects (Barrett et al. 2015), and the adaptation to a public sector context by introducing a

"public service logic" (Osborne 2017), evidencing the rich conceptual roots from the field of service innovation studies that support digital government scholarship.

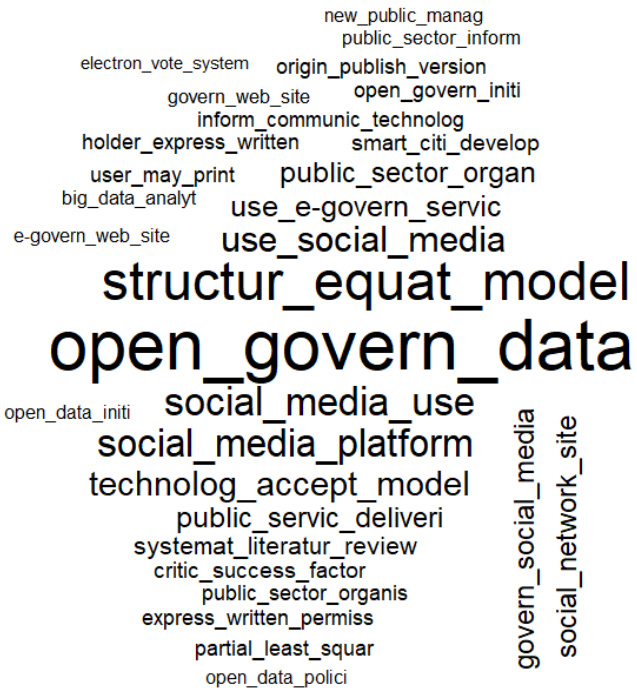


Figure 2: DGRL Corpus Trigrams Visualization

This visualization shows the top 30 trigrams found in the corpus, clearly dominant in the wordcloud is "open_govern_data", this is a concept usually associated with the public value theory. A literature review on the public value of e-government found that open government data contributes to values like openness, transparency, participation, communication and collaboration (Twizeyimana and Andersson 2019).

The next salient trigram is "structur_equat_model", the dominant position of this trigram in the visualization of this contrasts the reported over-reliance of qualitative methods in the digital government field, also the trigram "partial_least_squar" hint of the importance of these methodologies in the corpus. Definitely something worth exploring deeper.

The trigram "technolog_accept_model" refers to the technology acceptance model found in the seminal work of (Davis 1989) and adapted to the digital government field by (Hung, Chang, and Yu 2006), also the trigram "new_public_manag" provides a glimpse of the prevalence of this "paradigm" in the corpus (O'Flynn 2007).

The next step in the process is to create a Document-Feature Matrix, which is the method to provide a structure to the text and be able to conduct the quantitative analysis of the corpus. From this step in the process we gather that the corpus under study contains 6664 documents (abstracts) with their respective metadata and 18749 features (unique tokenized words). This is a very sparse matrix and the logical step is to conduct two process for dimensionality reduction: remove very common and very rare words. For the removal of rare words, the parameter was set to retain words with a minimum term frequency of 100, for the most common words, the criteria was to remove words that appears in more than 10% of the documents in the corpus. After these decisions the number of documents remain the same, the number of features, the vocabulary that will be used for the topic model, was reduced to 916. The documented code for these steps can be found in the GitHub page for this project for replicability purposes.

One crucial step before the initial calibration of the model is to split the data between a train and a test set, the model was trained with the 75% of the sample. The remaining 25% is used to apply the model to the unseen data with the calibrated parameters of the training of the model. The novelty of the structural topic model is possibility to include covariates found in the metadata to estimate topic prevalence, for our analysis the year of publication as covariate of interest.

The stm package includes the function searchK() that performs the estimation models with different K values to provide statistical analysis for goodness of fit measures in topic modeling. Perplexity, semantic coherence, residuals and lower bound can be estimated and visualized helping researchers to select the optimal number of topics in a data driven manner. However, statistical goodness of fit is not enough and it is widely advised to apply human validation and human judgement in the decision of the number of topics to model.

The following graphic shows the visualization of the results of the function searchK from the stm package, four goodness of fit measures are calculated for different values for K in a range from 25 to 60 topics. The values for held-out likelihood, or perplexity seem to be optimal in this range at k=53, semantic coherence is higher in k=44, k=53, and k=60, residuals values above 1 indicate sample dispersion meaning that the number for k is set too low, after several iterations with the training set, k=53 is deemed optimal for further analysis. The lower bound value indicate model convergence, small changes between the compared values are preferred.

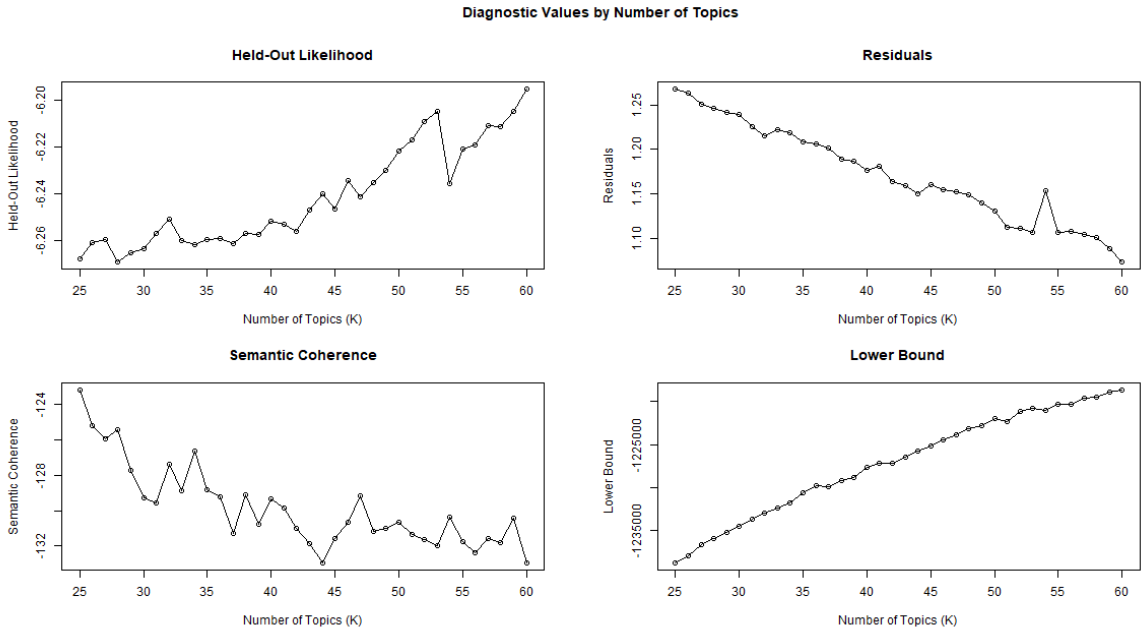


Figure 3: Comparative Goodness of Fit Measures for Different values for K

The different indicators for goodness of fit and the iterative revision of these parameters with the training set led us to choose k=53 as an optimal number of topics. The following graph presents the estimated topic proportions found in the training set. Topic 47, 27 and 51 are estimated to be more prevalent in the analyzed set.

Topic modeling estimation is not necessarily difficult, the real work for researchers comes in the interpretation and visualization of the model output. The visualization settings of the stm package have clear strengths like the inclusion of covariates in calculating topic prevalence but also crippling limitations regarding the visualization of the models, however there have been important developments in associated software packages that enhance interpretability such as LDAvis (Sievert and Shirley 2014).

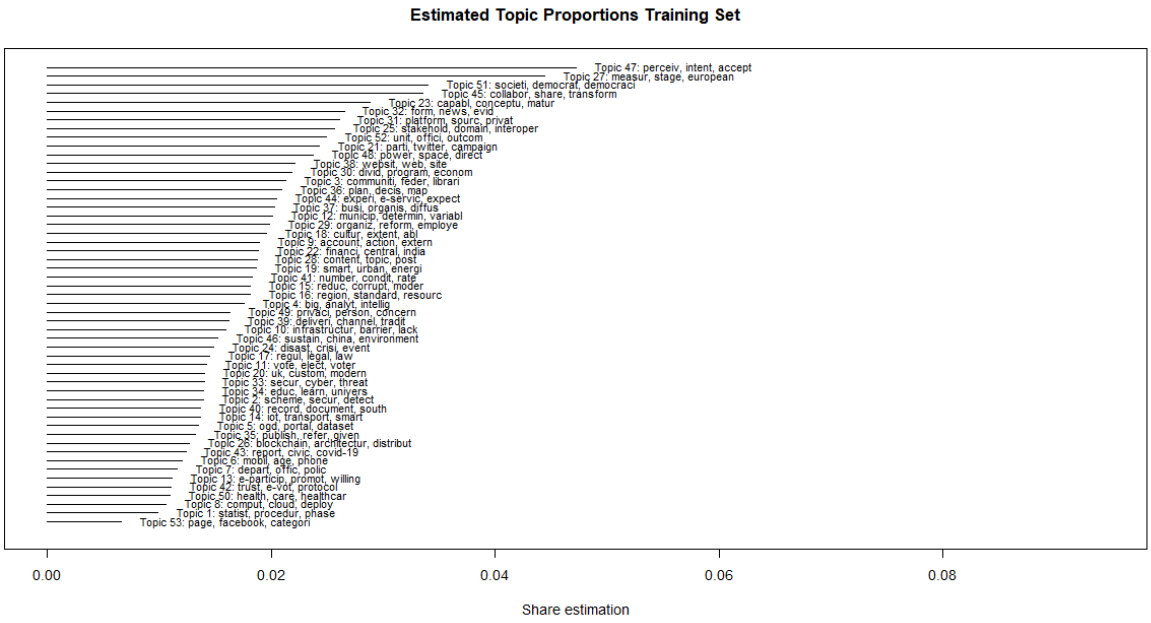


Figure 4: Topic Proportions K=53

Discussion

The identification and visualization of bigrams and trigrams enrich the researcher's ability to have a quick overview of the co-ocurrence of words, this contribute to the quick detection of meaning word combinations contributing to the interpretation of the contents of the corpus.

Key term extractions from the corpus - Stemming

Collect names (Buffat et al)

Insert bigrams and trigrams graphics

Conclusions

References

AlcaideMuñoz, Laura, Manuel Pedro RodríguezBolívar, Manuel Jesús Cobo, and Enrique HerreraViedma. 2017. "Analysing the Scientific Evolution of e-Government Using a Science Mapping Approach." *Government Information Quarterly* 34 (3): 545–55. https://doi.org/10.1016/j.giq.2017.05.002.

Ambrosino, Angela, Mario Cedrini, John B. Davis, Stefano Fiori, Marco Guerzoni, and Massimiliano Nuccio. 2018. "What Topic Modeling Could Reveal about the Evolution of Economics." *Journal of Economic Methodology* 25 (4): 329–48. https://doi.org/10.1080/1350178X.2018.1529215.

Andersson, Christoffer, Anette Hallin, and Chris Ivory. 2021. "Unpacking the Digitalisation of Public Services: Configuring Work During Automation in Local Government." *Government Information Quarterly*, December, 101662. https://doi.org/10.1016/J.GIQ.2021.101662.

Antons, David, and Christoph F. Breidbach. 2017. "Big Data, Big Insights? Advancing Service Innovation and Design with Machine Learning:" *Journal of Service Research* 21 (1): 17–39. https://doi.org/10.1177/1094670517738373.

Antons, David, Eduard Grünwald, Patrick Cichy, and Torsten Oliver Salge. 2020a. "The Application of Text Mining Methods in Innovation Research: Current State, Evolution Patterns, and Development Priorities." *R&D Management* 50 (3): 329–51. https://doi.org/10.1111/RADM.12408.

———. 2020b. "The Application of Text Mining Methods in Innovation Research: Current State, Evolution Patterns, and Development Priorities." *R&D Management* 50 (3): 329–51. https://doi.org/10.1111/RADM.12408.

Antons, David, Robin Kleer, and Torsten Oliver Salge. 2016. "Mapping the Topic Landscape of JPIM, 1984–2013: In Search of Hidden Structures and Development Trajectories." *Journal of Product Innovation Management* 33 (6): 726–49. https://doi.org/10.1111/jpim.12300.

Arduini, Davide, and Antonello Zanfei. 2014. "An Overview of Scholarly Research on Public e-Services ? A Meta-Analysis of the Literature." *Telecommunications Policy* 38 (5-6): 476–95. https://doi.org/10.1016/j.telpol.2013.10.007.

Asmussen, Claus Boye, and Charles Møller. 2019a. "Smart Literature Review: A Practical Topic Modelling Approach to Exploratory Literature Review." *Journal of Big Data* 6 (1): 1–18. https://doi.org/10.1186/S40537-019-0255-7/TABLES/6.

———. 2019b. "Smart Literature Review: A Practical Topic Modelling Approach to Exploratory Literature Review." *Journal of Big Data* 6 (1): 1–18. https://doi.org/10.1186/S40537-019-0255-7/TABLES/6.

Bailey, Diane E., and Stephen R. Barley. 2020. "Beyond Design and Use: How Scholars Should Study Intelligent Technologies." *Information and Organization* 30 (2): 100286. https://doi.org/10.1016/J.INFOANDORG.2019.100286.

Banks, George C., Haley M. Woznyj, Ryan S. Wesslen, and Roxanne L. Ross. 2018. "A Review of Best Practice Recommendations for Text Analysis in r (and a User-Friendly App)." *Journal of Business and Psychology 2018 33:4* 33 (4): 445–59. https://doi.org/10.1007/S10869-017-9528-3.

Bannister, Frank, and Regina Connolly. 2015. "The Great Theory Hunt: Does e-Government Really Have a Problem?" *Government Information Quarterly* 32 (1): 1–11. https://doi.org/10.1016/j.giq.2014.10.003.

Barberá, Pablo, Amber E. Boydstun, Suzanna Linn, Ryan McMahon, and Jonathan Nagler. 2021. "Automated Text Classification of News Articles: A Practical Guide." *Political Analysis* 29 (1): 19–42. https://doi.org/10.1017/PAN.2020.8.

Barley, Stephen R. 2020. *Work and Technological Change*. Oxford University Press.

Barley, Stephen R, and Gideon Kunda. 2001. "Bringing Work Back In." *Organization Science* 12 (1): 75–95. https://about.jstor.org/terms.

Barrett, Michael, Elizabeth Davidson, Jaideep Prabhu, and Stephen L. Vargo. 2015. "Service Innovation in the Digital Age." *MIS Quarterly* 39 (1): 135–54. https://doi.org/10.25300/MISQ/2015/39:1.03.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018b. "Quanteda: An r Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3 (30): 774. https://doi.org/10.21105/JOSS.00774.

———. 2018a. "Quanteda: An r Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3 (30): 774. https://doi.org/10.21105/JOSS.00774.

Blei, David M. 2014. "Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models." *Annual Review of Statistics and Its Application* 1 (1): 203–32. https://doi.org/10.1146/annurev-statistics-022513-115657.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *The Journal of Machine Learning Research* 3 (null): 9931022.

Bornmann, Lutz, and Hans-Dieter Daniel. 2007. "Multiple Publication on a Single Research Study: Does It Pay? The Influence of Number of Research Articles on Total Citation Counts in Biomedicine." *Journal of the American Society for Information Science and Technology* 58 (8): 1100–1107. https://doi.org/10.1002/ASI.20531.

Bornmann, Lutz, Robin Haunschild, and Rüdiger Mutz. 2021. "Growth Rates of Modern Science: A Latent Piecewise Growth Curve Approach to Model Publication Numbers from Established and New Literature Databases." *Humanities and Social Sciences Communications 2021 8:1* 8 (1): 1–15. https://doi.org/10.1057/s41599-021-00903-w.

Bornmann, Lutz, and Rüdiger Mutz. 2015. "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References." *Journal of the Association for Information Science and Technology* 66 (11): 2215–22. https://doi.org/10.1002/asi.23329.

Bryson, John, Alessandro Sancino, John Benington, and Eva Sørensen. 2016. "Towards a Multi-Actor Theory of Public Value Co-Creation." *Public Management Review* 19 (5): 640–54. https://doi.org/10.1080/14719037.2016.1192164.

Chang, Jonathan, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." In, 288296. NIPS'09. Red Hook, NY, USA: Curran Associates Inc.

Cordella, Antonio, and Andrea Paletti. 2018. "ICTs and Value Creation in Public Sector: Manufacturing Logic Vs Service Logic." *Information Polity* 23 (2): 125–41. https://doi.org/10.3233/IP-170061.

Criado, J. Ignacio, and Julian Villodre. 2021. "Delivering Public Services Through Social Media in European Local Governments. An Interpretative Framework Using Semantic Algorithms." *Local Government Studies* 47 (2): 253–75. https://doi.org/10.1080/03003930.2020.1729750.

Davis, Fred D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly* 13 (3): 319. https://doi.org/10.2307/249008.

De Battisti, Francesca, Alfio Ferrara, and Silvia Salini. 2015. "A Decade of Research in Statistics: A Topic Model Approach." *Scientometrics 2015 103:2* 103 (2): 413–33. https://doi.org/10.1007/S11192-015-1554-1.

Dingel, Jonathan I., and Brent Neiman. 2020. "How Many Jobs Can Be Done at Home?" *Journal of Public Economics* 189 (September): 104235. https://doi.org/10.1016/J.JPUBECO.2020.104235.

Doellgast, Virginia, and Ines Wagner. 2022a. "Collective Regulation and the Future of Work in the Digital Economy: Insights from Comparative Employment Relations." *Journal of Industrial Relations*, May, 002218562211011. https://doi.org/10.1177/00221856221101165.

———. 2022b. "Collective Regulation and the Future of Work in the Digital Economy: Insights from Comparative Employment Relations." *Journal of Industrial Relations*, May, 002218562211011. https://doi.org/10.1177/00221856221101165.

Dosi, G., M. Piva, M. E. Virgillito, and M. Vivarelli. 2021. "Embodied and Disembodied Technological Change: The Sectoral Patterns of Job-Creation and Job-Destruction." *Research Policy* 50 (4): 104199. https://doi.org/10.1016/j.respol.2021.104199.

Engin, Zeynep, and Philip Treleaven. 2019. "Algorithmic Government: Automating Public Services and Supporting Civil Servants in Using Data Science Technologies." *The Computer Journal* 62 (3): 448–60. https://doi.org/10.1093/COMJNL/BXY082.

Faraj, Samer, Wadih Renno, and Anand Bhardwaj. 2021. "Unto the Breach: What the COVID-19 Pandemic Exposes about Digitalization." *Information and Organization* 31 (1): 100337. https://doi.org/10.1016/J.INFOANDORG.2021.100337.

Fernández-Macías, Enrique, and Martina Bisello. 2022. "A Comprehensive Taxonomy of Tasks for Assessing the Impact of New Technologies on Work." *Social Indicators Research* 159 (2): 821–41. https://doi.org/10.1007/s11205-021-02768-7.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022b. *Text as data: a new framework for machine learning and the social sciences*. Princeton, New Jersey Oxford: Princeton University Press.

———. 2022a. *Text as data: a new framework for machine learning and the social sciences*. Princeton, New Jersey Oxford: Princeton University Press.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97. https://doi.org/10.1093/PAN/MPS028.

Guenduez, Ali A., Tobias Mettler, and Kuno Schedler. 2020. "Technological Frames in Public Administration: What Do Public Managers Think of Big Data?" *Government Information Quarterly* 37 (1): 101406 [1–12]. https://doi.org/10.1016/j.giq.2019.101406.

Hannigan, Timothy R., Richard F. J. Haan, Keyvan Vakili, Hovig Tchalian, Vern L. Glaser, Milo Shaoqing Wang, Sarah Kaplan, and P. Devereaux Jennings. 2019. "Topic Modeling in Management Research: Rendering New Theory from Textual Data." *Academy of Management Annals* 13 (2): 586–632. https://doi.org/10.5465/ANNALS.2017.0099.

Hung, Shin-Yuan, Chia-Ming Chang, and Ting-Jing Yu. 2006. "Determinants of User Acceptance of the e-Government Services: The Case of Online Tax Filing and Payment System." *Government Information Quarterly* 23 (1): 97–122. https://doi.org/10.1016/j.giq.2005.11.005.

Karlsen, Rune. 2010. "Does New Media Technology Drive Election Campaign Change?" *Information Polity* 15 (3): 215–25. https://doi.org/10.3233/IP-2010-0208.

Klein, Stefan, and Mary Beth Watson-Manheim. 2021. "The (Re-)Configuration of Digital Work in the Wake of Profound Technological Innovation: Constellations and Hidden Work." *Information and Organization*

31 (4): 100377. https://doi.org/10.1016/J.INFOANDORG.2021.100377.

Kobayashi, Vladimer B., Stefan T. Mol, Hannah A. Berkers, Gábor Kismihók, and Deanne N. Den Hartog. 2017. "Text Mining in Organizational Research:" *Organizational Research Methods* 21 (3): 733–65. https://doi.org/10.1177/1094428117722619.

Kogan, Leonid, Dimitris Papanikolaou, Lawrence Schmidt, and Bryan Seegmiller. 2019. "Technology-Skill Complementarity and Labor Displacement: Evidence from Linking Two Centuries of Patents with Occupations," December. https://doi.org/10.2139/ssrn.3585676.

Leonardi, Paul M., Da Jung Woo, and William C. Barley. 2021. "On the Making of Crystal Balls: Five Lessons about Simulation Modeling and the Organization of Work." *Information and Organization* 31 (1): 100339. https://doi.org/10.1016/J.INFOANDORG.2021.100339.

Lindgren, Ida, Christian Østergaard Madsen, Sara Hofmann, and Ulf Melin. 2019. "Close Encounters of the Digital Kind: A Research Agenda for the Digitalization of Public Services." *Government Information Quarterly* 36 (3): 427–36. https://doi.org/10.1016/j.giq.2019.03.002.

Lloyd, Caroline, and Jonathan Payne. 2021. "Fewer Jobs, Better Jobs? An International Comparative Study of Robots and 'Routine' Work in the Public Sector." *Industrial Relations Journal* 52 (2): 109–24. https://doi.org/10.1111/IRJ.12323.

Lofgren, Karl, and C. William R. Webster. 2020. "The Value of Big Data in Government: The Case of 'Smart Cities'." *Big Data & Society* 7 (1): [1–14]. https://doi.org/10.1177/2053951720912775.

MacLean, Don, and Ryad Titah. 2021. "A Systematic Literature Review of Empirical Research on the Impacts of e-Government: A Public Value Perspective." *Public Administration Review*, August. https://doi.org/10.1111/PUAR.13413.

Madsen, Christian Østergaard, Ida Lindgren, and Ulf Melin. 2021. "The Accidental Caseworker – How Digital Self-Service Influences Citizens' Administrative Burden." *Government Information Quarterly*, November, 101653. https://doi.org/10.1016/J.GIQ.2021.101653.

Maier, Daniel, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, et al. 2018a. "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology." *Communication Methods and Measures* 12 (2-3): 93–118. https://doi.org/10.1080/19312458.2018.1430754.

———, et al. 2018b. "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology." *Communication Methods and Measures* 12 (2-3): 93–118. https://doi.org/10.1080/19312458.2018.1430754.

Marsden, Chris, Trisha Meyer, and Ian Brown. 2020. "Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?" *Computer Law & Security Review* 36 (April): 105373. https://doi.org/10.1016/j.clsr.2019.105373.

Mascheroni, Giovanna, and Alice Mattoni. 2013. "Electoral Campaigning 2.0—The Case of Italian Regional Elections." *Journal of Information Technology & Politics* 10 (2): 223–40. https://doi.org/10.1080/19331681.2012.758073.

Mazzucato, Mariana, and Rainer Kattel. 2020. "COVID-19 and Public-Sector Capacity." *Oxford Review of Economic Policy* 36 (Supplement_1): S256–69. https://doi.org/10.1093/oxrep/graa031.

Mergel, Ines, and Stuart I. Bretschneider. 2013. "A Three-Stage Adoption Process for Social Media Use in Government." *Public Administration Review* 73 (3): 390–400. https://doi.org/10.1111/puar.12021.

Meyer, Eric T., and Ralph Schroeder. 2015. *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Infrastructures. Cambridge, Massachusetts: The MIT Press.

Montobbio, Fabio, Jacopo Staccioli, Maria Enrica Virgillito, and Marco Vivarelli. 2022. "Robots and the Origin of Their Labour-Saving Impact." *Technological Forecasting and Social Change* 174 (January): 121122. https://doi.org/10.1016/J.TECHFORE.2021.121122.

Nagel, Lisa. 2020. "The Influence of the COVID-19 Pandemic on the Digital Transformation of Work." *International Journal of Sociology and Social Policy* 40 (9-10): 861–75. https://doi.org/10.1108/IJSSP-07-2020-0323/FULL/PDF.

Nambisan, Satish, Mike Wright, and Maryann Feldman. 2019. "The Digital Transformation of Innovation and Entrepreneurship: Progress, Challenges and Key Themes." *Research Policy* 48 (8): 103773. https://doi.org/10.1016/J.RESPOL.2019.03.018.

O'Flynn, Janine. 2007. "From New Public Management to Public Value: Paradigmatic Change and Managerial Implications." *Australian Journal of Public Administration* 66 (3): 353–66. https://doi.org/10.

1111/J.1467-8500.2007.00545.X.

Osborne, Stephen P. 2017. "From Public Service-Dominant Logic to Public Service Logic: Are Public Service Organizations Capable of Co-Production and Value Co-Creation?" *Public Management Review* 20 (2): 225–31. https://doi.org/10.1080/14719037.2017.1350461.

Panagiotopoulos, Panos, Bram Klievink, and Antonio Cordella. 2019. "Public Value Creation in Digital Government." *Government Information Quarterly* 36 (4): 101421 [1–8]. https://doi.org/10.1016/j.giq.2019.101421.

Plesner, Ursula, Lise Justesen, and Cecilie Glerup. 2018. "The Transformation of Work in Digitized Public Sector Organizations." *Journal of Organizational Change Management* 31 (5): 1176–90. https://doi.org/10.1108/JOCM-06-2017-0257.

Rehs, Andreas. 2020. "A Structural Topic Model Approach to Scientific Reorientation of Economics and Chemistry After German Reunification." *Scientometrics* 125 (2): 1229–51. https://doi.org/10.1007/S11192-020-03640-0/TABLES/4.

Rehurek, Radim, and Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." In, 4550.

Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi. 2016. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association* 111 (515): 988–1003. https://doi.org/10.1080/01621459.2016.1141684.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019a. "Stm: An r Package for Structural Topic Models." *Journal of Statistical Software* 91 (October): 1–40. https://doi.org/10.18637/JSS.V091.I02.

———. 2019b. "Stm: An r Package for Structural Topic Models." *Journal of Statistical Software* 91 (October): 1–40. https://doi.org/10.18637/JSS.V091.I02.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–82. https://doi.org/10.1111/AJPS.12103.

Scholl, Hans J. 2021a. "The Digital Government Reference Library (DGRL) and Its Potential Formative Impact on Digital Government Research (DGR)." *Government Information Quarterly*, July, 101613. https://doi.org/10.1016/J.GIQ.2021.101613.

———. 2021b. "The Digital Government Reference Library (DGRL) and Its Potential Formative Impact on Digital Government Research (DGR)." *Government Information Quarterly*, July, 101613. https://doi.org/10.1016/J.GIQ.2021.101613.

Schou, Jannick, and Morten Hjelholt. 2018. *Digitalization and Public Sector Transformations*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-76291-3.

Scupola, Ada, and Ines Mergel. 2021. "Co-Production in Digital Transformation of Public Administration and Public Value Creation: The Case of Denmark." *Government Information Quarterly*, November, 101650. https://doi.org/10.1016/J.GIQ.2021.101650.

Sievert, Carson, and Kenneth Shirley. 2014. "Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces." In, 63–70. Baltimore, Maryland, USA: Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-3110.

Taddy, Matthew A. 2011. "(AISTATS) 2012." In. La Palma, Canary Islands: arXiv. https://doi.org/10.48550/arXiv.1109.4518.

Tanskanen, Kari, Tuomas Ahola, Anna Aminoff, Johanna Bragge, Riikka Kaipia, and Katri Kauppi. 2017. "Towards Evidence-Based Management of External Resources: Developing Design Propositions and Future Research Avenues Through Research Synthesis." *Research Policy* 46 (6): 1087–1105. https://doi.org/10.1016/j.respol.2017.04.002.

Tursunbayeva, Aizhan, Massimo Franco, and Claudia Pagliari. 2017. "Use of Social Media for e-Government in the Public Health Sector: A Systematic Review of Published Studies." *Government Information Quarterly* 34 (2): 270–82. https://doi.org/10.1016/j.giq.2017.04.001.

Twizeyimana, Jean Damascene, and Annika Andersson. 2019. "The Public Value of e-Government – a Literature Review." *Government Information Quarterly* 36 (2): 167–78. https://doi.org/10.1016/j.giq.2019.01.001.

Vargo, Stephen L., and Robert F. Lusch. 2004. "Evolving to a New Dominant Logic for Marketing." *Journal*

of Marketing 68 (1): 1–17. https://doi.org/10.1509/jmkg.68.1.1.24036.

Watanabe, Kohei, and Yuan Zhou. 2020. "Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches:" *Social Science Computer Review*, February. https://doi.org/10.1177/0894439320907027.

Webster, Jane, and Richard T. Watson. 2002a. "Analyzing the Past to Prepare for the Future: Writing a Literature Review." *MIS Quarterly* 26 (2): xiii–xxiii. https://www.jstor.org/stable/4132319.

———. 2002b. "Analyzing the Past to Prepare for the Future: Writing a Literature Review." *MIS Quarterly* 26 (2): xiii–xxiii. https://www.jstor.org/stable/4132319.

Wehrheim, Lino. 2019. "Economic History Goes Digital: Topic Modeling the Journal of Economic History." *Cliometrica* 13 (1): 83–125. https://doi.org/10.1007/S11698-018-0171-7/TABLES/4.

Wirtz, Bernd W., and Peter Daiser. 2016. "A Meta-Analysis of Empirical e-Government Research and Its Future Research Implications:" *International Review of Administrative Sciences* 84 (1): 144–63. https://doi.org/10.1177/0020852315599047.