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

Andres Aguilera Castillo

31 May 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 of a topic model aimed to dissect and perform 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 prelavence. 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.

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 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).

[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 in this manuscript has been trained on 75% of the data.

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 might be tackled 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 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 2022), information systems (Klein and Watson-Manheim 2021), and organization studies (Stephen R. Barley 2020), primarily focused in the private sector of advanced democracies. Technological change is a very broad term that may include a wide array of information and communication technologies (ICT) enabled applications for automation, digitalization and robotization. Our attention is located in the digitalization of government, but despite the momentum in digital government research, one aspect that remains under explored is 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. Nonetheless, public sector organizations rank high among the largest sole employers globally 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. The public sector, in general, commands a large, diverse, and highly educated workforce.

Governments are also considered one of the largest adopters and users of ICT. (insert Guenduez, Lofgren) and it is considered a Recent technological developments have allowed the deployment of digital technologies, digital platforms, and digital infrastructures by both private corporations and public organizations with deep 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.

Digitalization, in general, allows a more specific division of labor into the smallest possible tasks (Cherry 2015), and in a public service context, opens more opportunities for 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 generating administrative burden of citizens (Madsen, Lindgren, and Melin 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 analytical 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).

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), probably generating changes in the organization of work as well as tasks and job redundancies. On the other hand, it can also create new occupations to cope with an increasingly digitalized public sector. Thus, it is deemed pertinent and timely to extend the scholarly exploration of the effects of technological change in the organization of work in public organizations and the potential consequences for the public sector workforce.

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

RQ1: What does topic modeling techniques applied to DGRL v17.5 reveal about the conceptual, intellectual and thematic evolution this academic field? –> Run LDA

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?

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 chapter aims to offset the "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 // 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 2022a), 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 interdisciplinary 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 and have been used for the analysis of bibliographic content in diverse fields of research and academic disciplines including 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).

Digital Government Research (DGR) as a multidisciplinary research field is experiencing rapid growth in its research output. The diversity in scope and methods of these disciplines converge in the field of Digital Government enriching it, but at the same time, raising questions as regards the lack of native theoretical developments, thus relying upon frameworks, theories, and conceptualizations from related disciplines (Bannister and Connolly 2015).

On occasions the scope of analysis 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 of these techniques (Antons, Kleer, and Salge 2016), have explored the full text corpus of a single top ranking journal in innovation research over a three decade span (n=1008), alternative uses of this technique have considered

the titles of dissertations 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 2022a). 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.

In addition, 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. In our opinion, this methodological innovation can be repurposed to explore the linkages between digitalization and organization of work in a public sector context.

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. 2018), 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).

The implementation of LDA to bibliographic data has a relatively agreed-upon workflow. The script for the initial data cleaning and wrangling, including the R functions and packages used, is available for revision, clarity, and replicability. 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. The pre-processing was conducted in R statistical software using the functions of the quanteda R package (Benoit et al. 2018).

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 was to train with 75% of the corpus, leaving the remaining proportion for testing purposes. In addition, LDA topic modeling typically means that the topic structure is unknown and the number of topics (k) should be selected by the researcher. 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).

Model optimization measures: predictive likelihood * perplexity * coherence * exclusivity * cutoff criteria loadings Antons et al 2016.

Computational tools like topic models are enabling researchers to explore and analyze larger data sets of bibliographic information to conduct evidence synthesis.

The scholarly exploration of the linkages between digitalization and the organization of work remains relatively unexplored in a public sector context.

Unsupervised machine learning techniques for text analysis, this research technology can be used in a wide range of disciplines to examine databases, repositories and corpora thus expanding the methodological repertoire of scholars and opening an opportunity to explore large troves of data.

Using a text as data approach and an unsupervised machine learning technique known as topic modeling. 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).

Maier

Evidence synthesis –>

What is digital government?

Collect names (Buffat et al)

STM

reproducibility

human interpretable

Insert list of top journals.

Methods and Data

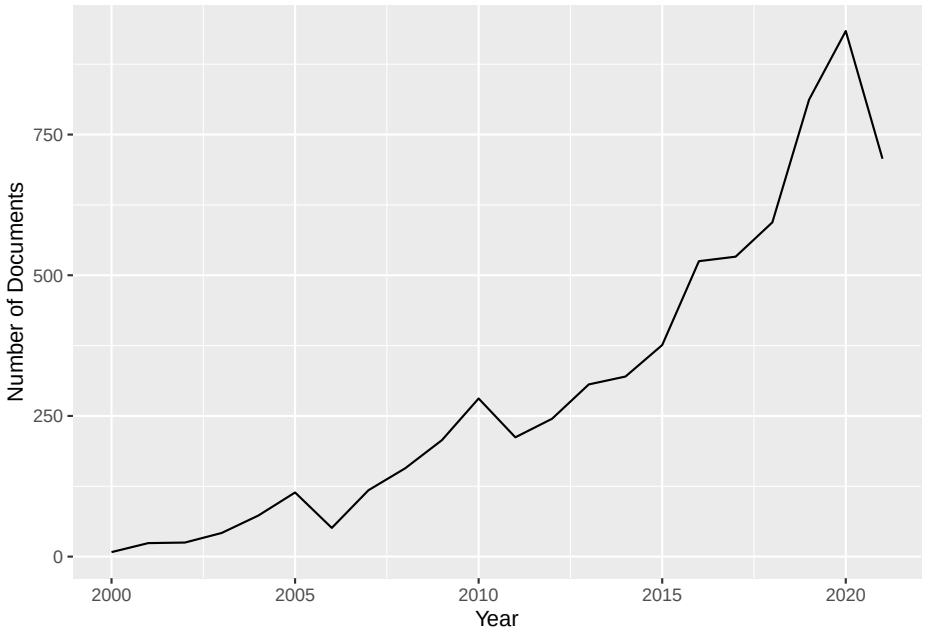
The Digital Government Research Library version 17.5 is a large curated repository of publications contributing to the field of Digital Government Research (DGR), it contains more than 16500 references among its records. The most prevalent types of documents are conference papers (33.2%) and journal 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. 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 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. For this exercise, 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 data wrangling, cleaning and overall processing of the topic model is available in the scripts section of the GitHub repository for this project.

Table 1: Table 1. Contents of DGRL v17.5 without pre-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 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 (2022b), texts are "expensive to 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 approach of this chapter is not as ambitious as Ambrosino *et. al.*, nor focalized on a single journal publication as Antons *et. al.*, instead the aim is to analyze the abstracts of 6682 journal articles in the version 17.5 of the DGRL via the application of a topic model.

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. 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 workflow for topic modeling includes text pre-processing, meaning further data cleaning and data transformation. This means that before creating a corpus object with the available information from the DGRL v17.5, the text should be prepared before running the initial explorations. Our main unit of analysis is the abstract of the journal articles contained in the processed data set of the DGRL v17.5. Text as a type of unstructured data can be structured for processing using the bag of words approach or the splitting of abstracts into separate word units or terms and every occurrence of a term is defined as a token. The creation of a bag of words is known as tokenizing. The bag of words approach deliberately ignores the syntax or structure of the text, 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 Document Frequency Matrix.

Stop words are context specific

Split into train and test data set Define K perplexity Asmussen (more or less topics) granularity.

LDA? Unsupervised learning... latent topics.

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. 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 title search in the Scimago Journal Rank website, it can be established that all names on the table are listed in this database.

Table 2: 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

Publication Title	Number of Documents in Corpus
Information Polity	105
International Journal of Public Administration	98

Dictionary based Quantitative Text Analysis Based on the work of Montobbio et al. 2022

Truncated words

Theory based Watanabe.

Results

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

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.

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.

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.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "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.

Cherry, Miriam A. 2015. "Beyond Misclassification: The Digital Transformation of Work." *Comparative Labor Law & Policy Journal* 37.

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.

Doellgast, Virginia, and Ines Wagner. 2022. "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.

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

———. 2022b. *Text as data: a new framework for machine learning and the social sciences*. Princeton,

New Jersey Oxford: Princeton University Press.

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.

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.

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.

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. 2018. "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.

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.

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.

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.

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.

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.

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.