DELIVERABLE

Data Analytics in Policy-Making: The BIGPROD approach

Summary

This policy brief introduces the objectives of BIGPROD project and strives to spark a discussion by highlighting the preliminary achievements and ongoing actions in the project.

Deliverable Information

Deliverable number and name:	**Data Analytics in Policy-Making: The BIG-PROD approach**
Due date:	30th November 2020
Deliverable:	D11
Work Package:	WP5
Lead Partner for the Deliverable:	UNU-MERIT
Author:	Arho Suominen and Arash Hajikhani
Reviewers:	Scott Cunningham and Hugo Hollanders
Approved by:	Arho Suominen
Dissemination level:	Public
Version	v. 1.0 25th November 2020 v. 2.0 18th January 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

Disclaimer

This document contains a description of the **BIGPROD** project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium coordinator for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of BIGPROD consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu.int/)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

Introduction

Big data and data analytics have been seen as augmenting knowledge, ultimately leading to better decision- making. Arguments such as that the broad-based use of big data and data analytics will lead to the end-of-theory, speak volumes about our expectations of the transformative power of technologies. While industry has been leading the way to test big data and analytics, public actors have been slower to engage (Poel et al., 2018), despite an at least equal opportunity for big data and data analytics to augment the public policy process.

Utilizing big data and data analytics has become a near necessity due to our increasing capability for creating and collecting data at an extraordinary rate. The terms ”big data” and ”data analytics” have been among the buzzwords of recent years, leading to an upsurge of research, industry, and government applications (Zhou et al., 2014). Scholarly discourse has highlighted case studies and narratives on the implementation of big data and data analytics in the policy process, but the literature lacks a systematic view of the current state of big data and data analytics in public policy, and there are clearly identifiable research gaps (Desouza and Jacob, 2017)

Background

”Big data” is a general term used for the process of gathering massive amounts of data from different sources. Sources can include human-input data but also includes data from sensors or different types of monitoring systems that create process data while running. It is clear that we are accumulating data at a never before seen rate. Already, in 2014, the pace was staggering, with 90% of the worlds data being collected during the prior two years and 2.5 quintillion bytes of data added each day (Kim et al., 2014). Having access to massive amounts of data has enabled significant innovation in both the public and private domains. Looking at companies like Google and Amazon, with their innovation of new services for consumers, or at the recent ability for doctors to detect cancer cells more precisely thanks to massive training data about what a cancerous cell is, we can see that we are very much on the cusp of creating a broad utility of big data and analytics. This has been seen as a shift in the magnitude of the Industrial revolution (Richards and King, 2014) and has been widely hyped in business (Margetts and Sutcliffe, 2013). That said, public policy is not at the forefront of the use of big data and data analytics in decision making (Poel et al., 2018; Kaski et al., 2019). This non-adoption is due to multiple factors limiting the utility of these technologies (Malomo and Sena, 2017).

What we know is that approximately 90% of the data gathered is unstructured and in need of restructuring and cleaning prior to being used by existing machine learning methods (Kim et al., 2014). This and the ever- increasing amount of data offer possibilities for discovering new relationships and for inference on a multitude of problems. However, this comes with new challenges involving reproducibility, complexity, security, and risks to privacy, as well as a need for new technology and human skills. This is very much the case in public policy, where we need to clearly identify where big data can add value in an ethical and trustworthy manner. In a review, Giest (2017) highlighted three underlying factors to consider. First, institutional capacities have a significant role to play in the use of big data in public policy, being able to produce solutions that can enable users to easily interact with data, while also taking into account the siloed data structures in the public domain. However, what we know from previous research is that siloed structures are an important limiting factor for public policy utilization of big data (Malomo and Sena, 2017). Second, hand-in-hand with big

data comes the broader digitalization of public services. Digitalization allows for mediums to interact with big data but also enables the creation of new data. There is, however, evidence that digitalization changes the interactions between citizens and public officials and requires new skills for both parties. Third, big data information will have an impact on the policy cycle. Studies have found that there has been limited progress in taking advantage of big data and analytics (Poel et al., 2018) because it requires a significant change in the policy cycle (Höchtl et al., 2016).

Giest (2017) highlights two issues, the substantive role and the procedural role of big data in policy instruments. Procedural activities focus on regulatory activities, such as enabling open data, while the substantive actions relate to the collection of data for enhancing, for example, evidence-based policy making. Capacities, digitalization, and the role of big data in the (substantive and procedural) policy cycle are core to digital-era governance and evidence-based policy making. In this, it is important to note that policy-makers are not a homogeneous group, and policy cycles vary. Thus, the objectives of analytics throughout the policy cycle vary significantly (Daniell et al., 2016) whether or not we approach the policy cycle as separate discrete stages (Jann and Wegrich, 2007), and it has been shown that big data analytics, when used more in some policy stages than in others, notably improved government transparency, policy evaluation, foresight, and agenda setting (Poel et al., 2018). This should be reflected against findings that data analytics have been politically significant in all stages of the policy cycle (Van der Voort et al., 2019).

To overcome the challenges, Poel et al. (Poel et al., 2015) highlighted multiple topics that must be addressed to enable capacity building, digitalization, and the integration of data into the policy cycle. These are 1) a skills gap, 2) reduced transparency due to data analytics, 3) sources and tools, 4) standardization of methods and tools, 5) linking of policy experiments with impact assessments, and 6) enabling policy-makers to be informed about the tools that are developed and piloted. The highlighted themes give context to the issue of big data in policy. While we see the significant impacts being created by the use of big data in policy making, along with the subsequent adaptation of data analytics, we need to better explain and make transparent the utility and complementarity of big data driven analyses for the policy cycle (Vydra and Klievink, 2019).

The current body of literature lacks holistic research agendas (Desouza and Jacob, 2017) addressing the issues highlighted from practice by Giest et al. (Giest, 2017) and Poel et al. (Poel et al., 2015). While we can take note of emerging fields such as policy analytics (Tsoukias et al., 2013; De Marchi et al., 2016), there is a need to better understand the theoretical grounding and research gap of big data and data analytics in public policy making.

Research Agenda and BIGPROD

Suominen and Hajikhani (2020), in the extended working paper version of this policy brief, suggest a research agenda that builds cohesion among the communities. The core questions for this research agenda depart from the three broad communities of Big Data and Policy analytics research communities, supported by the smaller areas as follows:

1. Strategic capability: Focusing on the role of strategic capability (Chatfield et al., 2015) rather than technical capacity (Höchtl et al., 2016; Wang et al., 2015), the research should address the need for empirical evidence (Guenduez et al., 2020;

Vydra and Klievink, 2019) about how to reduce the skills gap and enable a balanced approach toward stakeholders (influence (Washington, 2014) and engagement (Bright and Margetts, 2016; Bertot and Choi, 2013)).

2. Data-based decision-making: Research should develop a systematic agenda (Desouza and Jacob, 2017) to take full advantage of big data, while in particular addressing transparency (Poel et al., 2018), inclusiveness (Longo et al., 2017; Mergel et al., 2016), and interaction (Bright and Margetts, 2016; Bertot and Choi, 2013).
3. Productivity increases: Research should create understanding of how big data and data analytics can address expectations for better service for the public (Maciejewski, 2017) while simultaneously increasing productivity (Chen and Zhang, 2014).
4. Policy analytics: Research should address new forms of policy analytics (Loukis et al., 2020) and empiricism (Starkbaum and Felt, 2019) by understanding facilitating conditions (Malawani et al., 2020; Loukis et al., 2020), impacts to the policy cycle (Bright and Margetts, 2016; Höchtl et al., 2016), and impact assessment (Scharaschkin and McBride, 2016).

The BIGPROD project addresses the research questions of productivity increases and policy analytics.The BIGPROD project has in its first year build the technical capability to do data based decision making in a large scale. With and objective of creating data on roughly 180 000 companies, BIGPROD has created the infrastructure to retrieve data and has now webscraped information from nearly 20 000 companies. The project has also completed a literature review[1] on how to extend the CDM model with novel big data metrics. The project has also engaged policymakers, economists as well as professionals in multiple stakeholder events and published policy briefs to spark discussion on the project objectives[2].

References

Bertot, J. C. and Choi, H. (2013). Big data and e-government: issues, policies, and recommendations. In Proceedings of the 14th annual international conference on digital government research, pages 1–10.

Bright, J. and Margetts, H. (2016). Big data and public policy: Can it succeed where e-participation has failed? Policy & Internet, 8(3):218–224.

Chatfield, A., Reddick, C., and Al-Zubaidi, W. (2015). Capability challenges in transforming government through open and big data: Tales of two cities.

Chen, C. P. and Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information sciences, 275:314–347.

Daniell, K. A., Morton, A., and Insua, D. R. (2016). Policy analysis and policy analytics. Annals of Operations Research, 236(1):1–13.

De Marchi, G., Lucertini, G., and Tsouki`as, A. (2016). From evidence-based policy making to policy analytics. Annals of Operations Research, 236(1):15–38.

[1] https://www.merit.unu.edu/publications/wppdf/2020/wp2020-050.pdf

[2] http://www.bigprod.eu/wp-content/uploads/2020/01/BIGPROD-Policy-Brief-1.2.pdf

Desouza, K. C. and Jacob, B. (2017). Big data in the public sector: Lessons for practitioners and scholars.Administration & society, 49(7):1043–1064.

Giest, S. (2017). Big data for policymaking: fad or fasttrack? Policy Sciences, 50(3):367–382.

Guenduez, A. A., Mettler, T., and Schedler, K. (2020). Technological frames in public administration: What do public managers think of big data? Government Information Quarterly, 37(1):101406.

Höchtl, J., Parycek, P., and Sch¨ollhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. Journal of Organizational Computing and Electronic Commerce, 26(1-2):147–169.

Jann, W. and Wegrich, K. (2007). Theories of the policy cycle. Handbook of public policy analysis: Theory, politics, and methods, 125:43–62.

Kaski, S., Ailisto, H., and Suominen, A. (2019). International ai experts: Towards the third wave of artificial intelligence. In Leading the way into the age of artificial intelligence: Final report of Finlands Artificial Intelligence Programme 2019, pages 28–42. Ministry of Economic Affairs and Employment.

Kim, G.-H., Trimi, S., and Chung, J.-H. (2014). Big-data applications in the government sector. Communications of the ACM, 57(3):78–85.

Longo, J., Kuras, E., Smith, H., Hondula, D. M., and Johnston, E. (2017). Technology use, exposure to natural hazards, and being digitally invisible: Implications for policy analytics. Policy & Internet, 9(1):76–108.

Loukis, E. N., Maragoudakis, M., and Kyriakou, N. (2020). Artificial intelligence-based public sector data analytics for economic crisis policymaking. Transforming Government: People, Process and Policy.

Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. International Review of Administrative Sciences, 83(1 suppl):120–135.

Malawani, A. D., Nurmandi, A., Purnomo, E. P., and Rahman, T. (2020). Social media in aid of post disaster management. Transforming Government: People, Process and Policy.

Malomo, F. and Sena, V. (2017). Data intelligence for local government? assessing the benefits and barriers to use of big data in the public sector. Policy & Internet, 9(1):7–27.

Margetts, H. and Sutcliffe, D. (2013). Addressing the policy challenges and opportunities of big data. Policy & Internet, 5(2):139–146.

Mergel, I., Rethemeyer, R. K., and Isett, K. (2016). Big data in public affairs. Public Administration Review, 76(6):928–937.

Poel, M., Meyer, E. T., and Schroeder, R. (2018). Big data for policymaking: Great expectations, but with limited progress? Policy & Internet, 10(3):347–367.

Poel, M., Schroeder, R., Treperman, J., Rubinstein, M., Meyer, E., Mahieu, B., Scholten, C., and Svetachova,M. (2015). Data for policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking. Report about the State-of-the-Art. Amsterdam: technopolis, OxfordInternet Institute, Center for European Policy Studies.

Richards, N. M. and King, J. H. (2014). Big data ethics. Wake Forest L. Rev., 49:393.

Scharaschkin, A. and McBride, T. (2016). Policy analytics and accountability mechanisms: judging the value for moneyof policy implementation. Annals of Operations Research, 236(1):39–56.

Starkbaum, J. and Felt, U. (2019). Negotiating the reuse of health-data: Research, big data, and the european general data protection regulation. Big Data & Society, 6(2):2053951719862594.

Tsoukias, A., Montibeller, G., Lucertini, G., and Belton, V. (2013). Policy analytics: an agenda for research and practice. EURO Journal on Decision Processes, 1(1-2):115–134.

Van der Voort, H., Klievink, A., Arnaboldi, M., and Meijer, A. J. (2019). Rationality and politics ofalgorithms. will the promise of big data survive the dynamics of public decision making? Government Information Quarterly, 36(1):27–38.

Vydra, S. and Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. Government Information Quarterly, 36(4):101383.

Wang, X., White, L., Chen, X., and Amankwah-Amoah, J. (2015). Safety or no safety in numbers? governments, big data and public policy formulation. Industrial Management & Data Systems.

Washington, A. L. (2014). Government information policy in the era of big data. Review of Policy Research, 31(4):319–325.

Zhou, Z.-H., Chawla, N. V., Jin, Y., and Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]. IEEE Computational intelligence magazine, 9(4):62–74.

For more information, please contact

Dr. Arho Suominen (Consortium leader)
Tel. +358 50 5050 354
arho.suominen@vtt.fi

About BIGPROD

BIFPROD is a research project focusing on Big Data based analysis of productivity using webscraped data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

The project partners in the project are Quantitative Science and Technology Studies team, Foresight-driven Business Strategies, 1) VTT Technical Research Centre of Finland, Competence Center Innovation and Knowledge Economy (Coordinator), 2) Fraunhofer ISI, Economics of Knowledge and Innovation team, 3) UNU-MERIT, Maastricht University, 4) Public Policy and Management Institute, 5) Economics of Technology and Innovations, Faculty of Technology, Policy and Management, 6) Delft University of Technology, Economics of Technology and Innovations, 7) Faculty of Technology, Policy and Management, Delft University of Technology

www.bigprod.eu