Transforming Government: People, Process and

Research Challenges for the use of Big Data in Policy-Making

Journal:	*Transforming Government: People, Process and Policy*
Manuscript ID	TG-08-2019-0082.R2
Manuscript Type:	Research Paper
Keywords:	Big Data, Public Sector, Governance, evidence based practice, data informed policy making, Roadmap
File Type:	

Research challenges for the use of Big Data in Policy-Making

Keywords: Big Data, Public Sector, data-informed Policy-making

1 Abstract

1.1 Purpose

The manuscript aims at presenting pertinent research challenges in the field of (Big) data-informed policy-making based on the research, undertaken within the course of the EU-funded project Big Policy Canvas (BPC). Technological advancements, especially in the last decade, have revolutionised the way that both every day and complex activities are conducted. It is thus expected that a particularly important actor, such as the public sector, should constitute a successful disruption paradigm through the adoption of novel approaches and state-of-the-art Information and Communication Technologies (ICT).

1.2 Design

The research challenges stem from a need, trend and asset assessment based on qualitative and quantitative research as well as from the identification of gaps and external framework factors that hinder the rapid and effective uptake of data-driven policy-making approaches.

1.3 Findings

The current paper presents a set of research challenges categorised in six main clusters:

1. Public Governance Framework
2. Privacy, Transparency, Trust
3. Data acquisition, cleaning and representativeness
4. Data clustering, integration and fusion
5. Modelling and analysis with big data
6. Data visualisation

1.4 Originality/value

The paper provides a holistic overview of the interdisciplinary research challenges in the field of data-informed policy-making at a glance and shall serve as a foundation for the discussion of future research directions in a broader scientific community. It furthermore underlines the necessity to overcome isolated scientific views and treatments due to a high complex multi-layered environment.

2 Introduction

The EU funded project BPC aims at renovating the public sector at a cross-border level by mapping the needs of public administrations to trends, methods, technologies, tools and applications available both in the public & the private sector, stepping upon the power of open innovation and the rich opportunities for analysis and informed policy-making generated by big data.[i] The project delivered a live roadmap that proposes short and midterm milestones and relevant actions needed towards achieving the expected impacts for the public sector and the society (Mureddu et al., 2019). The current paper presents a set of research challenges categorised in six main research clusters. The description of these clusters and corresponding research challenges precede a more detailed explanation of the general approach of the road mapping exercise in the following section.

3 General approach of the road-mapping exercise

The aim of the BPC Roadmap for Future Research Directions in Data-Driven Policy-making is to put forward the different research and innovation directions that should be followed in order to reach Europe's anticipated vision for making the public sector a key player in tackling societal challenges through new data-driven policy-making approaches (European Commission, 2016, p. 3).

Within the course of the project, methods, technologies and tools were mapped to the public sector needs they address, trends they exploit and possible challenges/barriers they either meet or overcome as a preparation for the evidence-based gap analysis in the public administration sector (Mureddu et al., 2019, p. 13). The identification of needs, trends and assets encompassed an in-depth literature review and qualitative interviews as well as a quantitative analysis of twitter and the web of science database (Schmeling and Marx, 2018). The developed BPC assessment framework contributed to the prioritisation and mapping process (Markaki, 2018). These actions allowed defining the implementation as well as the research challenges and their transformation into recommendations for the next EU working programmes.

The first version of research clusters and challenges has been elaborated building on the preparatory phase including the aforementioned steps and the additional input from BPC experts (see section 6) provided via a set of memos and the input provided in BPC community workshops. The further development of the research challenges rested on a highly collaborative and multidisciplinary approach putting forward the different research and innovation directions. The second version of research clusters and challenges was available in a commentable format in MakingSpeechesTalk, a proprietary tool of the Lisbon Council, and promoted a broader discussion in the scientific community.

The relations among the clusters can be found in **Error! Reference source not found.**.

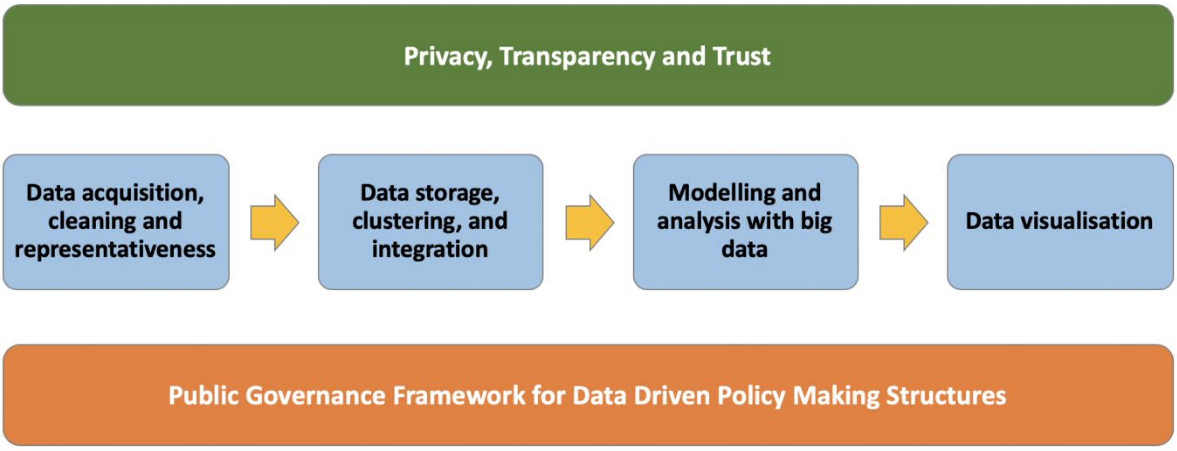
Privacy, Transparency and Trust
Data acquisition, cleaning and representativeness
Data storage, clustering, and integration
Modelling and analysis with big data
Data visualisation
Public Governance Framework for Data Driven Policy Making Structures

Figure 1 – Structure of Research Clusters

4 Research Clusters and Challenges

4.1 Research Cluster: Public Governance Framework

Generally, the governance notion is associated with shaping and designing areas of life and particularly the way that rules are set and managed in order to guide policy-making and policy implementation (Lucke and Reinermann, 2002). Core dimensions of governance are efficiency, transparency, participation and accountability (United Nations, 2007). Corresponding to the definition of electronic governance, evidence-based and data-informed policy-making in the information age applies agile technologies to efficiently transform governments and their relationship with citizens, businesses, and other stakeholders while creating a positive impact on society (Estevez and Janowski, 2013). The smart governance approach includes an intelligent use of digital technologies to improve decision-making (Pereira *et al.*, 2018). In this frame, governance needs to focus on how to leverage data for more effective, efficient, rational, participative and transparent policy-making processes.

4.1.1 Forming and monitoring of societal and political will

European governments have been trying to establish data platforms and the current development in the open data movement contributes to data-driven decisions in the public sector. However, it is ambiguous whether the status quo is sufficient as well as what is needed to leverage data for advanced data-based decision support processes in the public sector. The legislative and political objectives are often neither clear nor discussed in advance, leading to a huge amount of data available without the right data sets to assess specific political problems. In that sense, governance structures and frameworks as well as outcome and target oriented approaches are needed, so that the right data can be on the one hand available and on the other interpreted bearing in mind societal and legislative goals (Schmeling *et al.,* 2019). Objectives in the public sector can be multifarious since they aim at the common good and not only on profit maximisation. Therefore, shared targets have the potential to transform common policies and legislative intentions on a horizontal and a vertical level into public organisations (James and Nakamura, 2015). Research is needed to investigate how political and societal will can be operationalised in order to be able to design monitoring systems and performance measurement systems based not simply on financial information but rather on outcome and performance-oriented indicators.

4.1.2 Dataprovider-oriented governance models

To enhance databased decisions on policy-making, data must be gathered from different sources and various stakeholders including enterprise data, citizens' data, third sector data or public administrations' data. Every stakeholder group requires different approaches to provide and exchange data. Since a plurality of independent stakeholder groups is involved in the fragmented process of data collection, the governance mode requires negotiation-based interactions (Sørensen and Torfing, 2007). The public administration is in its origin an important advisor of the political system and is not to be underestimated in this context. If company data like traffic data from navigation device providers or social media data from social network providers is necessary to assess political questions, guidance and governance models to purchase or exchange this data is needed. For all aforementioned cases, ICT standards and ICT architecture frameworks for processing data stored in different infrastructures, constituting so called data spaces, are required. Data spaces coordinate respective rules and access rights on a meta-level and have to be established through intelligent data exchange standards, data connectors or controlled harvesting methods (Cuno *et al.*, 2019). In this frame, robust, modular, scalable anonymisation algorithms that guarantee anonymity are needed and it is important to ensure adequate forms of consent management across organisation and symmetric transparency, allowing stakeholders to see how their data is being used, by whom and for what purpose. Blockchain of things for example can provide authentication for machine to machine transactions (Reyna et al. 2018). Anonymisation algorithms and secure multiparty mining algorithms over distributed datasets allow guaranteeing anonymity (Selva Rathna and Karthikeyan 2015).

4.1.3 Administrative boundaries and jurisdictional silos

Decisions in the political environment are often facing trans-boundary problems on different administrative levels, in different jurisdictions and different organisations. Thus, the data collection to understand these problems and to investigate possible solutions causes manifold barriers and constraints and can be ouvercome through modern governance approaches and models. Solutions that can balance the need for data with safeguards on data protection need to be developed. This development needs to take place at a European level to ensure the achievement of the goals of the Tallinn declaration (European Commission, 2017b). What is more, transparency and full consent by citizens needs to be considered.

Data integration has long been a priority for policy-making, but with the new European Interoperability Framework and the objective of the once only principle (European Commission, 2017a) it has become an unavoidable priority. Data integration and integrity are the basic building blocks for ensuring sufficient data quality for decision-makers.

4.1.4 Data science-oriented education and personnel development

Governance plays an important role in all questions of education and personnel development[ii] in order to ensure that the right capabilities are available in terms of data literacy, data management and interpretation. The need to develop these skills has to be managed and governed as a frame to design qualification strategies, trainings and employee developments (OECD, 2019).

Governance in personnel development promotes effective and efficient fulfillment of public duties like evidence-based policy-making. This includes research focusing on standards making the assessment criteria of education policies transparent, giving incentives to motivate specific types of behavior, clear definitions of outputs and outcomes and accountability to examine that given outcomes and outputs can be delivered (Lewis and Pettersson, 2009).

4.2 Research Cluster: Privacy, Transparency and Trust

This research cluster deals with core elements such as data ownership, security and privacy from one side, and transparency of the policy-making on the other side. The overall aim is to increase trust in the government, especially in the public services, and a fair policy-making activity and public service provisioning. An initial exploration on views and opinions of practitioners has been undertaken within the course of the project BPC which brought forward a controversial debate regarding the use of data analytics in the public sector that is closely linked to the emerging lack of trust in the public sector and personal privacy (Schmeling and Marx, 2018). Thus, big data requires systems for determining and maintaining data ownership, data definitions, and data flows.

4.2.1 Big Data Nudging

Nudging has long been recognised as a powerful tool to achieve policy goals by inducing changes in citizens' behaviour, while at the same time presenting risks in terms of respect of individual freedom. Nudging can help governments, for instance, reducing carbon emissions by changing how citizens commute, using data from public and private sources. However, it is not clear to what extent governments can use these methods without infringing citizens' freedom of choice. Research on ethical implications and the transparency of algorithms is requested. The recent case of Cambridge Analytica acts as a powerful reminder of the threats deriving from the combination of big data with behavioural science (Venturini and Rogers, 2019). These benefits and risks are multiplied by the combination of nudging with big data analytics. When nudging can exploit thousands of data points on any individual, based on data held by governments but also from private sources, the effectiveness of such measures is exponentially higher (Thaler and Sunstein, 2009).

4.2.2 Algorithmic bias and transparency

Many decisions, in the public as well as in the private sector, are today automated and performed by algorithms. Algorithms are designed by humans, and increasingly learn by observing human behaviour through data. Therefore they tend to adopt the biases that exist in the analysed data. Thus, it is particularly difficult to detect bias, and can be done only through ex-post auditing and simulation rather than ex-ante analysis of the code. There is a need for common practice and tools controlling data quality, bias and transparency in algorithms (World Wide Web Foundation, 2017). Furthermore, it is required to explain machine decisions in a human format. The risk of manipulation of data should also be considered, which might lead to ethical misconduct.

4.2.3 Open Government Datasets

Open Data are defined as data accessible with minimal or no cost, without limitations as to user identity or intent. Meaning that data should be available online in a digital, machine readable format. The notion of Open Government Data (OGD) concerns all the information that governmental bodies produce, collect or pay for, including geographical data, statistics, meteorological data, data from publicly funded research projects, traffic and health data. In this respect, the definition of Open Public Data is applicable when that data can be readily and easily consulted and re-used by anyone with access to a computer.[iii] Clearly opening government data can help in displaying the full economic and social impact of information, and create services based on all the information available. Nevertheless, transparency does not directly imply accountability. A taxonomy of Open Government Data (OGD) research areas and topics is provided by Charalabidis et al. (2016) and distinguishes between four main OGD research areas: Management and Policies, Infrastructures, Interoperability, Usage and Value.

4.3 Research Cluster: Data acquisition, cleaning and representativeness

Data to be used in policy-making come from various sources, such as official statistics, government administrative data, user-generated web content, tracking data, etc. All these data of different size that could span across time series, are digital data (their enhanced version is big data), and they are well equipped to capture behavioural information.

4.3.1 Real time big data collection and production

The rapid development of web technologies enables ordinary users to generate vast amounts of data on a daily basis. In the Internet of Things (IoT)[iv] paradigm data can be produced real-time or near real-time from sensors embedded in numerous devices.

Research is needed to investigate how real time big data could be leveraged for efficient policy-making, for example to evaluate policies, to monitor the effects of policy implementations, to collect data that can be used for agenda setting (e.g. traffic data) and to analyse the sentiment and behaviour of the citizens, as well as to monitor and evaluate the government communication and engagement process (Venturini and Rogers, 2019).

4.3.2 Quality assessment, data cleaning and formatting

Assessing the quality is an important phase integrated within data pre-processing. It is a phase where the data is prepared according to the user or application requirements. After the assessment of data quality follows data cleaning. This is the process of correcting (or removing) corrupt or inaccurate records from a record set, table, or database. This research challenge also deals with formatting, as it is ambiguous whether the format of downloaded sets will be suitable for further analysis and integration in the existing platforms (Neumaier, S. et al., 2016).

Apart from the elimination of systematic errors, the quality of data is to be assessed, before used in the policy-making process. For this assessment, new frameworks need to be developed. These new frameworks should include big data quality dimensions, quality characteristics, and quality indexes.

4.3.3 Representativeness of data collected

A key concern with many Big Data sources is the selectivity, (or conversely the representativeness) of the dataset. Buelens et al. (2014) have used metrics from statistics to define the representativeness of a dataset. Some indicators have been developed and used to measure how information available in the data source differ from the information for the in-scope population. One of the indicators is using the covariates as a way of profiling the units of the dataset.

In the policy-making process, the representativeness of the data used is crucial, specifically when used in sentiment analysis, in studying characteristics of the population. Attention should be given to bias, when using data sets to formulate policies.

The appropriate sampling process could be a way to ensure the representativeness of data and limit bias. Kim et al. (Kim *et al.*, 2016) propose a new method of survey data integration using fractional imputation under the instrumental variable assumption, and Park et al. (Park *et al.*, 2017) use a measurement error model to combine information from two independent surveys.

4.4 Research Cluster: Data clustering, integration and fusion

The research cluster deals with information extraction from unstructured, multimodal, heterogeneous, complex, or dynamic data. The produced data from various sources are generated for different purposes and there is no unanimous point of reference on how they have to be structured.

4.4.1 Identification of patterns, trends and relevant observables

This research challenge deals with technologies and methodologies that allow businesses and policy makers to visualise patterns and trends of data, both structured and unstructured that may have not been previously visible. This ability can be very useful when developing a policy agenda. For example, an interesting application is anomaly detection, which is most commonly used in detecting fraud.

Following Gullo (2015), one of the most used big data methodologies for identification of pattern and trends is data mining. A combination of database management, statistics and machine learning methods are useful for extracting patterns from large datasets. Some examples include mining human resources data in order to assess some employee characteristics or consumer bundle analysis to model the behavior of customers. It has also to be considered that most of the data are not structured and have a huge quantity of text. In this regard, text mining is another technique that can be adopted to identify trends and patterns.

4.4.2 Extraction of relevant information and feature extraction

Prior to the analysis data need to be structured in a homogeneous way, otherwise nuances cannot be detected. Most computer systems work better if multiple items are stored in an identical size and structure. Efficient representation, access and analysis of semi-structured data is necessary.

While feature extraction may seem irrelevant to policy-making, the veracity of the information obtained needs to be ensured, to assure the widest reuse of the data for a variety of purposes. The data must be adapted according to the use and analysis that they are destined for. There are available Bayesian techniques for meaning extraction; extraction and integration of knowledge from massive, complex, multi-modal, or dynamic data; data mining; scalable machine learning (Mureddu et al., 2019, pp. 86-87).

4.5 Research Cluster: Modelling and analysis with big data

The intrinsic complexity of the emerging challenges, human beings collectively face, requires a deep comprehension of the underlying phenomena in order to plan effective strategies and sustainable solutions. In this regard, a main challenge in the use of big data for applications related to policy-making is coping with unanticipated knowledge. The term “Unanticipated Knowledge” refers precisely to the observation of events whose existence cannot even been foreseen. One typical solution are predictions which have to be based on modelling (Franke *et al.*, 2016).

4.5.1 Identification and validation of suitable modelling schemes inferred from existing data

The traditional way of modelling starts with a hypothesis about how a system acts and data to test the model. The amount of data collected is rather small since it rarely existed already and had to be generated with surveys, or perhaps imputed through analogies. Finally, statistical methods established enough causality to arrive at enough truth to represent the system.

Nowadays deductive models are forward running and they end up representing a system not observed before. With the current huge availability of data, it is possible to identify and create new suitable modelling schemes that build on existing data. These are inductive models that start by observing a system already in place and one that is putting out data as a by-product of its operation.

Therefore, the real challenge is being able to identify, accept and validate from existing data models that are valid and suitable to cope with complexity and unanticipated knowledge. According to Mcintosh et al. (2008) there is a “different perceptions of model users and model developers on what a model should look like.” Furthermore, Van Delden et al. (2011) argues that model acceptance and validation is hindered by a lack of transparency, inflexibility and a focus on technical capabilities. Even the concept of model acceptance is not clear. In that regard McIntosh et al. (2011) clarified four levels of acceptance: first, when a model development has been completed and presented to its intended users; secondly, when the users have been trained in the use of a model, but there is limited evidence of actual use; further when the model has been used on a one-off basis, and finally in case the model is adopted routinely used in the daily business.

4.5.1 Collaborative model simulations and scenario generation

Collaborative model simulation should encompass participation of all stakeholders in the policy-making process through the implementation of online-based easy-to-use tools for all skills levels. Scenario simulators and decision support tools are needed, that allow a realistic forecast of how a change in the current conditions will affect and modify the future scenario. In this framework it is highly important to launch new research directions aimed at developing effective infrastructures merging the science of data with the development of highly predictive models, to come up with engaging and meaningful visualisations and friendly scenario simulation engines (Ranjan, 2014).

The weakest form of involvement is feedback to the session facilitator, similar to the conventional way of modelling. Stronger forms are proposals for changes or (partial) model proposals. In this particular approach the modelling process should be supported by a combination of narrative scenarios, modelling rules, and e-Participation tools (all Integrated via an ICT e-Governance platform).

4.5.2 *Integration and re-use of modelling schemes*

This research challenge seeks to find the way to model a system by using already existing models or composing more comprehensive models by using smaller building blocks, either by reusing existing objects/models or by generating/building them from the very beginning. Therefore, the most important issue is the definition/identification of proper (or most apt) modelling standards, procedures and methodologies by using existing ones or by defining new ones.

Furthermore, the present challenge calls for establishing the formal mechanisms by which models might be integrated in order to build bigger models or to simply exchange data and valuable information between the models. Finally, the issue of model interoperability as well as the availability of interoperable modelling environments should be tackled, as well as the need for feedback-rich models that are transparent and easy for the public and decision makers to understand.

Concerning current practices, the CEF BDTI[v] building block provides virtual environments that are built on a mix of mature open source and off-the-shelf tools and technologies. Specifically, the Big Data Test Infrastructure (BDTI) provides a set of data and analytics services, from infrastructure, tools and stakeholder onboarding services, allowing European public organisations to experiment with big data technologies in the view of developing data-driven decision making. Applications of the BDTI include descriptive analysis, Social Media Analysis, Time-series Analysis, Predictive analysis, Network Analysis, and Text Analysis. In this respect, BDTI allows public organisations to share data sources across policy domains and organisations, experiment with big data methods and tools, launch pilot projects on big data and data analytics. An interesting application is the European Big Data Hackathon 2019 carried out by EUROSTAT, aimed to modernise statistics through automated data collection and more accurate indicators to better support policy decisions.[vi]

4.6 Research Cluster: Data visualisation

Making sense and extracting meaning of data can be achieved by placing them in a visual context: patterns, trends and correlations that might go undetected in text-based data can be exposed and recognised easier with data visualisation software.

4.6.1 *Automated visualisation of dynamic data in real time*

Since most analysis and visualisation methods focus on static data sets, adding a dynamic component to the data source results in major challenges for both the automated and visual analysis methods. Besides typical technical challenges such as unpredictable data volumes, unexpected data features and unforeseen extreme values, a major challenge is the capability of analysis methods to work incrementally. Furthermore, scalability of visualisation in face of big data availability is a permanent challenge, since visualisation requires additional performances with respect to traditional analytics in order to allow for real time interaction and reduce latency. Finally, visualisation is largely a demand- and design-driven research area.

Concerning relevance and applications in policy-making**,** visualisation of dynamic data in real time allows policy makers to react timely with respect to issues they face. Following Toasa et al. (2018), there is a set of visualisation techniques, which are suitable for real time data, including autocharting, correlation matrix, network and Sankey diagrams. Moreover, an important feature to set up an actual automatic dashboard that reacts when some data is entered in real time can be achieved with technologies such as Node.js and Socket.io.

A final interesting application is provided by Buschmann et al. (2015) that developed a technique for visualising massive 3D movement trajectories. Their technique allows to visualise real-time simulated movement data by individual attributed trajectories or by aggregated density maps, facilitating spatial reasoning with respect to different application fields such as urban development, environmental analysis and simulation, as well as risk and disaster management.

4.6.2 Interactive data visualisation

Traditionally, visualisations were performed as post-processing steps after an analysis or simulation had been completed. As simulations increased in size, this task became increasingly difficult, often requiring significant computation, high-performance machines, high capacity storage, and high bandwidth networks. In this regard, there is the need of emerging technologies that address this problem by "closing the loop" and providing a mechanism for integrating modelling, simulation, data analysis and visualisation. This integration allows to interactively perform data analysis while avoiding many of the pitfalls associated with the traditional batch / post processing cycle. It also plays a crucial role in making the analysis process more extensive and, at the same time, comprehensible (Keim *et al.*, 2006).

Concerning relevance and applications in policy-making**,** policy makers should be able to independently visualise results of analyses.

Regarding future developments, intuitive interfaces and devices are needed to interact with data results through clear visualisations and meaningful representations. User acceptability is a challenge in this sense, and clear comparisons with previous systems to assess its adequacy. An interesting approach would be to investigate two, or even three, tiers of visualisation tools for different types of users: experts and analysts, decision makers (which are usually not technical experts but must understand the results, make informed decisions and communicate their rationale), and the general public (Vornhagen et al. 2019). Visualisation for the general public will support buy-in for the resulting policies as well as the practice of data-driven policy-making in general. According to Wang et al. 2015, the step of interacting visualisation are the following: A) interactive selection of data entities or subset or part of whole data or whole data set according to the user interest; B) Linking of relating information among multiple views; C) Filtering the amount of information for display; D) Rearranging the spatial layout of the information.

5 Conclusions

The article presents six research clusters related to the use of big data in policy-making. Four of them are built on the big data cycle and value chain, while the first two are transversal at each phase of the cycle. The roadmap serves as a basis for further research and has been transferred into policy recommendations in order to be considered in future European working and research programmes and aims at promoting public sector transformation through evidence based practices.

The overview underlines the need to investigate how and at what point to apply new data science methodologies to design big data analyses as efficient as possible within the policy-making phases. The combination of 'sensors data' with behavioural/usage ones and self-reported (e.g. user-generated content) offers a unique possibility for tracing dynamic unfolding at different levels of measurements. New modelling schemes, possibly data-driven, have to be conceived to better grasp the complexity of todays' and tomorrow's challenges aimed at fostering a renewed dialogue between scientific research and policy-making. It is crucial to provide analysis techniques and metaphors that are capable of analysing large real time data streams in time, and to present the results in a meaningful and intuitive way. Therefore, the integration of visual analytics with opinion mining and participatory sensing, and the research on visualisation as a way to provide (persuasive) feedback and change in attitudes, opinions, behaviours, as well as a medium for grassroots/crowd-sourced participation, collaboration on data-related issues have to be evolved.

It emphasises furthermore the importance of ethical implications and transparency of algorithms, which encompasses e.g. the development of impact assessment frameworks for algorithms, privacy enhancing technologies, techniques to contrast manipulation of statements and misinformation as well as techniques to ensure representativeness of data collected.

Frames and ICT architecture reference models for public and private data spaces have to be improved to establish new governance models, which include collection, storing and sharing of data through multi-stakeholder data sharing agreements, commons-based data crowdsourcing or city data commons. In this context, standards and techniques to assess data quality need to be innovated, given the complexity of the dataset now available.

Finally, education can help to de-mystify technology and data analysis. There is a pressing need also for ethical education, both for developers and policy makers, and for the public at large.

6 Acknowledgement

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 769623.

We thank our experts Gianluca Misuraca (JRC Seville), Giuseppe Veltri (University of Trento), Vittorio Loreto (Sony Computer Science Laboratories) and Peter Parycek (Danube University Krems) who provided insight and expertise that greatly assisted the research.

7 References

Badham, J, Chattoe-Brown, E, Gilbert, N, Chalabi, Z, Kee, F & Hunter, R 2018, "Developing Agent-Based Models of Complex Health Behaviour", *Health and Place*, Vol. 54, pp. 170-177.

Buschmann, S.; Trapp, M.; Döllner, J., "Real-time visualization of massive movement data in digital landscapes", in *Proceedings of the 16th Conference on Digital Landscape Architecture (DLA 2015), Dessau*, *Germany, 4–6 June 2015*, pp. 213–220.

Buelens, B., Daas, P., Burger, J., Puts, M. and van den Brakel, J. (2014), *Selectivity of Big Data*.

Charalabidis, Y.Alexopoulos, C. and Loukis, E. (2016), "A Taxonomy of Open Government Data Research Areas and Topics", *Journal of Organizational Computing and Electronic Commerce*. Vol. 26.

Cuno, S., Bruhns, L., Tcholtchev, N., Lämmel, P. and Schieferdecker, I. (2019), "Data Governance and Sovereignty in Urban Data Spaces Based on Standardized ICT Reference Architectures", *MDPI*, Vol 4 No.1.

Estevez, E. and Janowski, T. (2013), "Electronic Governance for Sustainable Development—Conceptual framework and state of research", *GOVERNMENT INFORMATION QUARTERLY,* 2013, 94-109.

European Commission (2017a), T*he New European Interoperability Framework*, available at: https://ec.europa.eu/isa2/eif_en (accessed 28 August 2019).

European Commission (2017b), *The Tallinn Declaration*, available at: https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration (accessed 25 August 2019).

European Commission (2016), *European Cloud Initiative -Building a competitive data and knowledge economy in Europe*, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0178&from=EN (accessed 20 September 2019)

Franke, B., Plante, J.-F., Roscher, R., Lee, E.-s.A., Smyth, C., Hatefi, A., Chen, F., Gil, E., Schwing, A. and Alessandro (2016), "Statistical Inference, Learning and Models in Big Data", *International Statistical Review*, No. 84 Issue 3, pp. 371–389.

Gullo, F. (2015), "From Patterns in Data to Knowledge Discovery: What Data Mining Can Do", *Physics Procedia*, Vol. 62, pp. 18–22.

James, O. and Nakamura, A. (2015), "Shared performance targets for the horizontal coordination of public organizations: control theory and departmentalism in the United Kingdom's Public Service Agreement system", *International Review of Administrative Sciences*, No. 81, Issue 2, pp. 392–411.

Keim, D.A., Mansmann, F., Schneidewind, J. and Ziegler, H. (2006), "Challenges in Visual Data Analysis", Tenth International Conference on Information Visualization (IV'06), 5-7 July 2006, *London, IEEE*.

Kim, J.K., Berg, E. and Park, T. (2016), "Statistical matching using fractional imputation", *Surv. Methodol*, No. 42, pp. 19–40.

Lewis, M. and Pettersson, G. (2009), *Governance in Education: Raising Performance*, available at: https://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1239047988859/Governance-in-education-master-22Dec09-GP.pdf (accessed 22 August 2019).

Lucke, J. von and Reinermann, H. (2002), *Speyerer Definition von Electronic Governance*, available at: http://www.joernvonlucke.de/ruvii/Sp-EGvce.pdf (accessed 24 August 2019).

Markaki, O. (2018), *The Big Policy Canvas Needs and Trends Assessment Framework*, available at: https://www.bigpolicycanvas.eu/updates/news-events/big-policy-canvas-needs-and-trends-assessment-framework (accessed 23 August 2019).

Mcintosh, B. S., Alexandrov, G., Matthews, K., Mysiak, J., & van Ittersum, M. (2011). Preface: Thematic issue on the assessment and evaluation of environmental models and software. *Environmental Modelling and Software*, 26(3), 245–246.

Mcintosh, B. S., Giupponi, C., Voinov, A. A., Smith, C., Matthews, K. B., Monticino, M., et al. (2008). "Bridging the gaps between design and use: Developing tools to support environmental management and policy", in A. J. Jakeman, A. A. Voinov, A. E. Rizzoli, & S. H. Chen (Eds.), *Environmental Modelling, Software and Decision Support: State of the art and new perspective.* Amsterdam: Elsevier.

Mureddu, F., Osimo, D., Garrido, E., Munné, R., Schmeling, J., Loreto, P., Parycek, P., Misuraca, G., Veltri, G. (2019), *Roadmap for Future Research Directions*, D5.2, available at: https://www.bigpolicycanvas.eu/sites/default/files/roadmap/BPC_D5.2_Roadmap_for_Future_Research_Directions-Final_Version.pdf (accessed 15 February 2020).

Neumaier, S.; Umbrich, J.; Polleres, A. (2016): "Automated Quality Assessment of Metadata across Open Data Portals", *Journal of Data and Information Quality*, Vol. 8 (1), pp. 1-29. DOI: 10.1145/2964909.

OECD (2019): OECD Skills Strategy 2019. Skills to Shape a Better Future, Paris 2019.

Park, S., Kim, J.K. and Stukel, D. (2017), "A measurement error model for survey data integration: combining information from two surveys", *Metron*, No. 75, pp. 345–357.

Pereira, G., Parycek, P., Falco, E. and Kleinhans, R. (2018), "Smart governance in the context of smart cities: A literature review", *Information Polity*, Vol. 2018 No. 23(2), pp. 1–20.

Ranjan, R. (2014), “Modeling and Simulation in Performance Optimization of Big Data Processing Frameworks”, *IEEE Cloud Computing*, No. 1 Issue 4.

Reyna, A.; Martín, C.; Chen, J.; Soler, E.; Díaz, M. On blockchain and its integration with IoT Challenges and opportunities. Future Gener. Comput. Syst. 2018, 88, 173–190.

Schmeling, J. and Marx, A. (2018), *Needs and Trends in Public Administrations* No. D3.1, available at: https://www.bigpolicycanvas.eu/sites/default/files/misc/deliverables/D3.1_Needs_and_Trends_in_Public_Administrations_v1.1.pdf (accessed 22 August 2019).

Schmeling, J., Marx, A. and Kurrek, H. (2019), *Evidenzbasiert steuern. Die integrierte Nutzung von Verwaltungsdaten*, available at: http://publica.fraunhofer.de/dokumente/N-538072.html (accessed 20 August 2019).

Selva Rathna and T. Karthikeyan, “Survey on Recent Algorithms for Privacy Preserving Data mining”, IJCSIT, Vol. 6 (2) , 2015, 1835-1840, ISSN: 0975-9646.

Sørensen, E. and Torfing, J. (2007), *Theories of Democratic Network Governance*, palgrave macmillan.

Thaler, R.H. and Sunstein, C.R. (2009), *Nudge. Improving Decisions About Health, Wealth and Happiness*, Yale University Press, New Haven.

Toasa, R.; M. Maximiano, C. Reis and D. Guevara, "Data visualization techniques for real-time information — A custom and dynamic dashboard for analyzing surveys' results," *2018 13th Iberian Conference on Information Systems and Technologies (CISTI), Caceres, 2018*, pp. 1-7.

United Nations (2007), *Public Governance Indicators: A Literature Review*, available at: https://publicadministration.un.org/publications/content/PDFs/E-Library%20Archives/2007%20Public%20Governance%20Indicators_a%20Literature%20Review.pdf (accessed 24 June 2019).

van Delden, H., Seppelt, R., White, R., & Jakeman, A. J. (2011). "A methodology for the design and development of integrated models for policy support", *Environmental Modelling and Software*, Vol. 26 (3), pp. 266–279.

Venturini, T., Rogers, R. (2019), " "API-Rased Research" or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach", Digital Journalism, Vol. 7 No (4), pp. 532-540.

Vornhagen, H., Young, K. and Zarrouk, M. (2019) "Understanding My City through Dashboards. How Hard Can It Be?", in Virkar, S., Glassey, O., Janssen, M., Parycek, P., Polini, A, Re, B., Reichstädter, P., Scholl, H.J., Tambouris, E. (eds.) E*GOV-CeDEM-ePart 2019: Proceedings of Ongoing Research, Practitioners, Posters, Workshops, and Projects of the International Conference EGOV-CeDEM-ePart 2019. 2-4 September 2019, San Benedetto del Tronto,* Italy, pp. 21-30.

Wang, Lidong, Guanghui Wang, and Cheryl Ann Alexander. "Big Data and Visualization: Methods, Challenges and Technology Progress", *Digital Technologies,* 1.1 (2015), pp. 33-38.

World Wide Web Foundation (2017), *Algorithmic Accountability. Appliying the concept to different country contexts*, available at: http://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf (accessed 29 August 2019).

[i] https://www.bigpolicycanvas.eu/

[ii] https://www.bigpolicycanvas.eu/community/kb/deeper-understanding-it-potential-and-it-processes

[iii] https://ec.europa.eu/digital-single-market/en/open-data; https://www.bigpolicycanvas.eu/community/kb/open-data

[iv] https://www.bigpolicycanvas.eu/community/kb/internet-things

[v] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Big+Data+Test+Infrastructure

[vi] https://ec.europa.eu/eurostat/cros/EU-BD-Hackathon_en