Roadmap for Future Research Directions Current Status and What is New

WORK IN PROGRESS

Please provide feedback to francesco.mureddu@lisboncouncil.net and https://roadmap.bigpolicycanvas.eu/

Document Information

List of Contributors	
Name	Partner
Francesco Mureddu	Lisbon Council
David Osimo	Lisbon Council
Esther Garrido	ATOS
Ricard Munné	ATOS
Vittorio Loreto	Sony Computer Science Laboratories
Peter Parycek	Danube University Krems
Gianluca Misuraca	JRC Seville
Giuseppe Veltri	University of Trento

Table of Contents

List of Figures

1 Current Status and What is New

1.1 The Policy Cycles

Policy-making is typically carried out through a set of activities described as "policy-cycle" (Howard 2005). In this document we propose a new way of implementing policies, by first assessing their impacts in a virtual environment. While different versions of the cycle are proposed in literature, in this context, we adopt a simple version articulated in 4 phases:

- **Agenda setting** encompasses the basic analysis on the nature and size of problems at stakes are addressed, including the causal relationships between the different factors;
- **Policy design** includes the development of the possible solutions, the analysis of the potential impact of these solutions, the development and revision of a policy proposal;
- **Implementation** is often considered the most challenging phase, as it needs to translate the policy objectives in concrete activities, that have to deal with the complexity of the real world. It includes ensuring a broader understanding, the change of behaviour and the active collaboration of all stakeholders. This phase includes also adoption, where accountability and representativeness are needed. It is also the area most covered by existing research on e-democracy;
- **Monitoring and evaluation** make use of implementation data to assess whether the policy is being implemented as planned, and is achieving the expected objectives.

Figure 1 below (authors' elaboration based on Howard 2005) illustrates the main phases of the policy cycle (in the internal circle) and the typical concrete activities (external circle) that accompany this cycle. In particular, the identified activities are based on the Impact Assessment Guidelines of the European Commission[1].

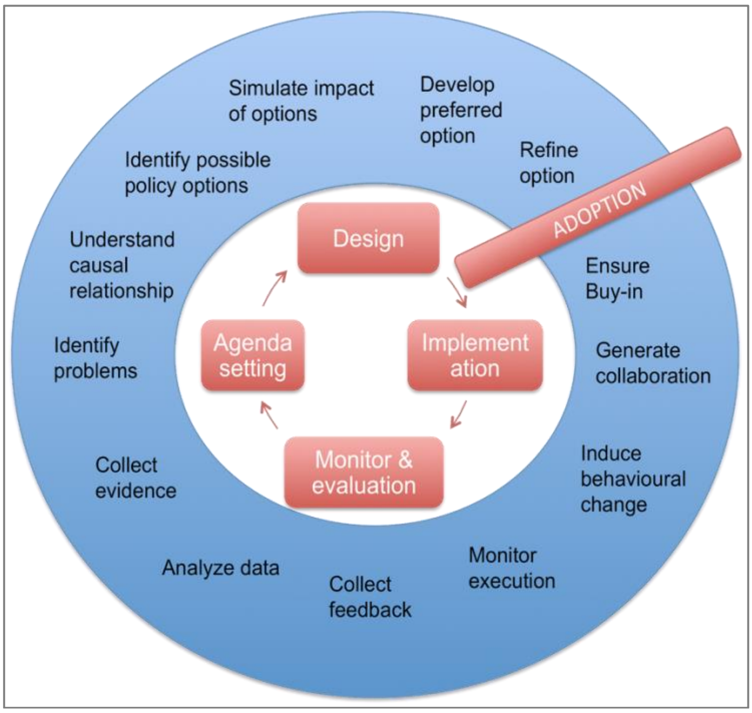

Figure 1 – Policy Cycle and Related Big Data Activities

[1] Impact assessment guidelines SEC(2009) 92. Key documents are on the IA website (http://ec.europa.eu/governance/impact/key_docs/key_docs_en.htm).

1.2 The Traditional Tools of Policy Making

Let us present now what are the methodologies and tools already traditionally adopted in policy-making. Typically, in the agenda-setting phase, statistics are analysed by government and experts contracted by government in order to understand the problems at stake and the underlying causes of the problems. Survey and consultations, including online ones, are frequently used to assess the stakeholders' priorities, and typically analysed in-house. General-equilibrium models are used as an assessment framework. Once the problems and its causes are defined, the policy design phase is typically articulated through an ex-ante impact assessment approach. A limited set of policy options are formulated in house with the involvement of experts and stakeholders. For each option, models are simulated in order to forecast possible sectoral and cross--sectoral impacts. These simulations are typically carried out by general-equilibrium models if the time frame is focused on short and medium term economic impacts of policy implementation. Based on the simulated impact, the best option is submitted for adoption. The adoption phase is typically carried out by the official authority, either legislative or executive (depending on the type of policy). In Some cases, decision is left to citizens through direct democracy, through a referendum or tools such as participatory budgeting; or to stakeholders through self-regulation. The Implementation phase typically is carried out directly by government, using incentives and coercion. It benefits from technology mainly in terms of monitoring and surveillance, in order to manage incentives and coercion, for example through the database used for social security or taxes revenues. The monitoring and evaluation phase is supported by mathematical simulation studies and analysis of government data, typically carried out in-house or by contractors. Moreover, as numbers aggregate the impacts of everything that happens, including policy, it is difficult to single out the impacts of one policy ex post. Final results are published in report format, and fed back to the agenda setting phase.

1.3 The Key Challenges of the Policy Makers

Let us now briefly discuss the key challenges which are faced by policy makers. One first aspect to consider is the emergence of a distributed governance model. Traditionally, the policy cycle is designed as a set of activities belonging to government, from the agenda setting to the delivery and evaluation. However, in recent years it has been increasingly recognized that public governance involves a wide range of stakeholders, who are increasingly involved not only in agenda-setting but in designing the policies, adopting them (through the increasing role of self-regulation), implementing them (through collaboration, voluntary action, corporate social responsibility), and evaluating them (such as in the case of civil society as watchdog of government).

Detect and understand problems before they become unsolvable

The continuous struggle for evidence-based policy-making can have some important and potentially negative implications in terms of the capacity of prompt identification of problems. Policy-makers have to balance the need for prompt reaction with the need for justified action, by distinguishing signal from noise. Delayed actions are often ineffective; at the same time, short-term evidence can lead to opposite effects. In any case, government have scarce resources and need to prioritize interventions on the most important problems. For instance, the significant underestimation of the risks of the housing bubble in the late 2000s, and the systemic reaction that it would lead to, led to delayed reactions. Systemic changes do not happen gradually, but become visible only when it is too late to intervene or the cost of intervening is too high. For example, ICT is today recognized as a key driver of productivity and growth, but evidence to prove this became available at a distance of years from the initial investment. In fact, the initial lack of correlation between ICT investment and productivity growth was mostly due to incorrect measurement of ICT capital prices and quality. Subsequent methodologies found that computer hardware played an increasing role as a source of economic growth (see inter al. Colecchia and Schreyer 2002, Jorgenson and Stiroh 2000, Oliner and Sichel 2000). The problem is in this case is therefore twofold: to collect data more rapidly; and to analyse them with a wider variety of models that account for systemic, long term effects and that are able to detect and anticipate weak signals or unexpected wild cards.

Generate high involvement of citizens in policy-making

The involvement of citizens in policy-making remains too often associated with short-termism and populism. It is difficult to engage citizens in policy discussions in the first place: public policy issues are not generally appealing and interesting as citizens fail to understand the relevance of the issues and to see "what's in it for me". The decline in voters' turnout and the lack of trust in politicians reflects this. More importantly, there are innumerable cases where the "right" policies are not adopted because they are not politically acceptable. While the Internet has long promised an opportunity for widespread involvement, e-participation initiatives often struggle to generate participation. Participation is often limited to those that are already interested in politics, rather than involving those that are not. When participation occurs, online debates tend to focus on eye-catching issues and polarized positions, in part because of the limits of the technology available. It Is extremely difficult and time consuming to generate open, large scale and meaningful discussion.

Identify "good ideas" and innovative solutions to long-standing problems

Innovation in policy-making is a slow process. Because of the technical nature of issues at hand, the policy discussion is often limited to restricted circles. Innovative policies tend to be "imported" through "institutional isomorphism". Innovative ideas, from both civil servants and citizens, fail to surface to the top hierarchy and are often blocked for institutional resistance. Existing instruments for large-scale brainstorming remain limited in usage, and fail to surface the most innovative ideas. Crowdsourcing typically focus on the most "attractive" ideas, rather than the most insightful.

Reduce uncertainty on the possible impacts of policies

When policy options have been developed, simulations are carried out to anticipate the likely impact of policies. The option with the most positive impact is normally the one that is proposed for adoption. Most existing methodologies and tools for the simulation of policy impacts work decently with well known, linear phenomena. However, they are not effective in times of crisis and fast change, which unfortunately turn out to be exactly the situations where government intervention is most needed. This is especially true in case of economic crisis, as shown by the policies carried out to fight the financial crisis in 2008. But the need for new policy making tools is not limited to the economic realm: in the future it will become more and more important to anticipate non-linear potentially catastrophic impacts from phenomena such as: climate change (draught and global warming); threshold climate effects such as poles' sea-ice withdraw, out-gassing from melting permafrost, Indian monsoon, oceans acidification; social instability affecting economic well-being (social conflict, anarchy and mass people movements). The lack of understanding of systemic impact has driven to short term policies which failed in grasping long term, systemic consequences and side effects.

Ensure long-term thinking

In traditional economics, decisions are utility-maximising. Agents rationally evaluate the consequences of their actions, and take the decisions that maximize their utility. However, it is well known that this rationalistic view does not fully capture human nature. We tend to overestimate short-term impact and underestimate the long term. In policy-making, short-termism is a frequent issue. People are reluctant to accept short-term sacrifices for long-term benefits. Politicians have elections typically every 5 years, and often their decisions are taken to maximize the impact "before the elections". There is also the perception that laypeople are less sensitive to long term consequences, which are instead better understood by experts. Overall, long-term impact is less visible and easier to hide, due to lack of evidence and data. As a result, decisions are too often taken looking at short-term benefits, even though they might bring long term problems. This is especially true in a period in which populists movements are taking control.

Encourage behavioural change and uptake

Once policies are adopted, a key challenge is to make sure that all stakeholders comply with regulations or follow the recommendations. It is well known how the greatest resistance to a policy is not active opposition, but lack of application. For instance, several programmes to reduce alcohol dependency problems in the UK failed as they excessively relied on positive and negative incentives such as prohibition and taxes, but did not take into account peer-pressure and social relationships. They failed to leverage "the power of networks" (Ormerod 2010). For instance, any policy related to reduction of alcohol consumption through prohibitions and taxes is designed to fail as long as it does not take into account social networks. In another classical example (Christakis and Fowler 2007), a large scale longitudinal study showed that the chances of a person becoming obese rose by 57 per cent if he or she had a friend who became obese. The identification of social networks and the role of peer pressure in changing behavior is not considered in traditional policy-making tools.

Manage crisis and the "unknown unknown"

The job of policy-makers is increasingly one of crisis management. There is robust evidence that the world is increasingly interconnected, and unstable (also because of climate change). Crises are by definition sudden and unpredictable. Dealing with unpredictability is therefore a key requirement of policy-making, but the present capacity to deal with crises is designed for a world where crises are exceptional, rather than the rule. Each crisis seems to find our decision-makers unprepared and unable to deal with it promptly. As Taleb (2007) puts it, we live in the age of "Extremistan": a world of "tipping points" (Schelling 1969), "cascades" and "power laws" (Barabasi 2003) where extreme events are "the new normal".

Detect non-compliance and mis-spending through better transparency

In times of budget constraints, it is ever more important for governments to ensure that financial resources are well spent and policies are duly implemented. But monitoring is a cost in itself, and a certain margin of inefficiency in resources deployment is somehow understandable. Yet the cost of this mismanagement is staggering: for instance, in 2010, 7.7% of all Structural Funds money was spent in error or against EU rules. The EU Commission has managed to bring the error rate down in recent years, achieving 2.4% in 2017 (3.1% in 2016, 3.8% in 2015 and 4.4% in 2014). This means more than €97 of every €100 spent by the EU was free from error. But the European Court of Auditors considers a 2% error rate as the level below which errors are not regarded as having a significant effect. Thereby it would be crucially important to be able to avoid the mismanagement with anticipatory corrective actions.

Moving from Conversation to Action

The collaborative action of people is able to achieve seemingly unachievable goals: experiences such as ZooGalaxy and Wikipedia show that mass collaboration can help achieve disruptive innovation. Yet too often web-based collaboration is confined to complaints and discussions, rather than action. A typical example is the electoral debate, in which before the elections we see an explosion of activity in social media discussing about the different candidates and their possible programs. Unfortunately, most of the time such energy then fails to translate into concrete action in the aftermath of the elections.

Understand the impact of policies

Measuring the impact of policies remains a challenge. Ideally, policy-makers would like to have real-time clear evidence on the direct impact of their choice. Instead, the effects of a policy are often delayed in time; the ultimate impact is affected by a multitude of factors in addition to the policy. Timely and robust evaluation remains an unsolvable puzzle. This is particularly true for research and innovation policy, where the results from investment are naturally expected at years of distance. As Kuhlmann and Meyer-Krahmer (1995) puts it, “the results of evaluations necessarily arrive too late to be incorporated into the policy-making process”.

1.4 Big Data Driven Policy Making

Let us now discuss the use of Big Data in policy making, and in particular in the policy cycle. First, we are going to introduce Big Data, their market and value chain, then we are discussing the application of big data to the policy cycle.

1.4.1 Big Data Value Chain

“Data-driven innovation is a key driver of growth and jobs that can significantly boost European competitiveness in the global market.”, was declared in the EC Strategy “Towards a common European data space”[2]. Not only is data produced, gathered and elaborated by an increasing set of stakeholders at growing rates both in the public and private sector, but a true knowledge economy can also only build on data understanding, data integration and data-driven predictions. The term big data is employed to stress the scale of the problem to be solved and is usually explained through the 4 V’s model:

- Scale of data (Volume) - 2.5 quintillion bytes created every day;
- Streaming/real-time data (Velocity) - 18.9 billion network connections in 2016;
- Heterogeneous formats (Variety) - 30 billion pieces of content are shared on Facebook every month;
- Data uncertainty (Veracity) - poor data quality costs the US economy $3.1 trillion every year.

Furthermore, as argued by Klievink et al. (2016), building on several studies (Adrian, 2011; Chen et al., 2014; Davenport et al., 2012; Gantz and Reinsel, 2011; Hota et al., 2015; Janssen and Kuk, 2016; Mayer-Schönberger and Cukier, 2013; OpenTracker, 2013; Simon, 2013), there are five differentiating characteristics of big data:

- Use and combining of multiple, large datasets, from various sources, both external and internal to the organization;
- Use and combining of structured (traditional) and less structured or unstructured (non-traditional) data in analysis activities;

[2] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “Towards a common European data space” (COM(2018) 232 final)

- Use of incoming data streams in real time or near real time;
- Development and application of advanced analytics and algorithms, distributed computing and/or advanced technology to handle very large and complex computing tasks;
- Innovative use of existing datasets and/or data sources for new and radically different applications than the data were gathered for or spring from.

Addressing a big data application means taking into account one or more of those four dimensions, which in turn require distinct technologies and approaches. The open data meme, on the other hand, emerged to highlight the transparency and legal issues related to data access and sharing. According to the Open Definition by the Open Data Institute, open data is "information that is available for anyone to use, for any purpose, at no cost" or, in other words, "data that can be freely used, modified, and shared by anyone for any purpose". On the pure technological side stands the concept of linked data that, coming from the academic research of the Semantic Web community, refers to a set of techniques and practices to structure, interlink and publish data. This is enabled by leveraging Web technologies (e.g. HTTP URI identifiers and hyperlinks) and machine-readable formats (e.g. RDF, related languages and relevant data models such as Data Cube, DCAT and ADMS) in order to foster interoperability. Certainly, to represent different perspectives on data, any combination of big, open and linked data is possible: the expression "big open linked data" thus refers to large datasets suffering from one or more of the four V's issues, released with an open license for both commercial and noncommercial purposes, and published on the Web in machine-readable format and interlinked with other data sources.
Considering all these characteristics, as reported by AGCOM(217/17CONS) it is possible to highlight this radical change in the approach to data analysis. In fact, in the time of data scarcity it was necessary to ask a research question and consequently collect data ("data-is-scarce-model"), or to acquire a sample, looking for the answer to a predetermined research question. In the time of data abundance, data are often collected regardless of specific research questions, which are then defined a posteriori after the analysis (Figure 2)[3].

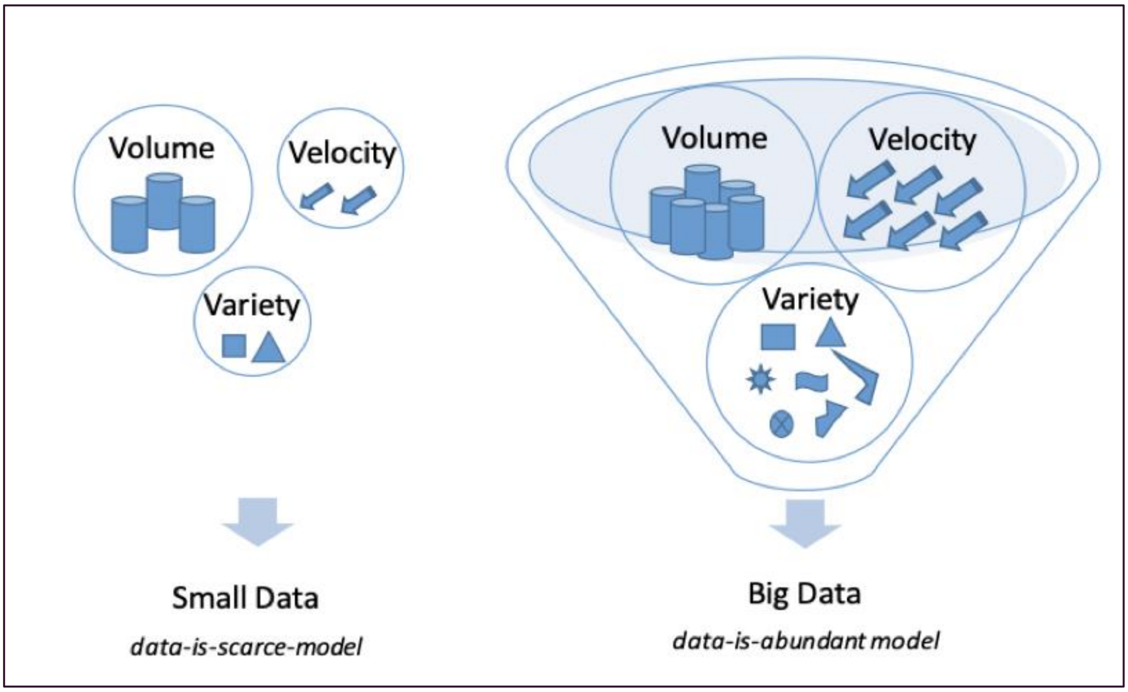

Figure 2 – Small and Big Data

The landscape of data analytics is very broad in terms of definitions, perspectives and actors. In the following, we give our definitions of the main topics – mainly in line with the definitions of the EU

[3] Source: AGCOM(217/17/CONS)

strategy – highlighting the current trends and technologies with special reference to evidence-based policy-making. By data analytics, we mean the set of approaches and methodologies to explore data with the purpose of drawing conclusions or taking decisions on top of that information. Data analytics as such is a complex process that encompasses collecting, organizing and processing of datasets, in order to discover hidden patterns and relations in data and to make predictions on future data. Data analytics usually aims at measuring business performance (KPIs), provide suggestions or support and inform decision-making. Thus, data analytics spans across different activities: from data analysis to data cleansing, from data processing to data modelling, and from data prediction to data visualization. To take the point of view of big industry players, IBM states that the purpose of analytics is to "discover what is happening, determine why it is happening, predict what is likely to happen and prescribe the best action to take". SAS interprets the current popularity of analytics technologies as a sign that "we are on the cusp of an analytics revolution that may well transform how organizations are managed, and also transform the economies and societies in which they operate". Ericsson has a vision of "Data-derived growth: creating innovative offerings and generating new revenue streams sparked by data analytics". Data analytics employs multiple types of data. Apart from the traditional distinction between structured data (e.g. databases) and unstructured information (e.g. text), different perspectives can be chosen to address data analytics.

Figure 3[4] illustrates a typical Big Data Value Chain and the respective technologies used in every step of the chain. While most of the techniques can be considered state-of-the-art (statistics, data mining, basic machine learning), the scientific-technical challenges – and the business opportunities – come from applying those and more sophisticated techniques to big data (in the sense of data fitting in the 4 V's model) and to contexts requiring data fusion and integration from multiple and heterogeneous sources (e.g. data that may require signal processing, NLP, spatiotemporal analysis and predictive modelling).

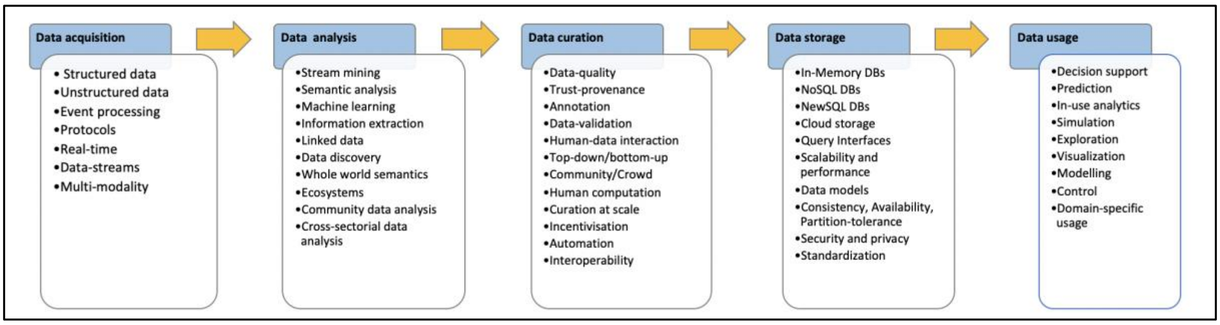

Figure 3 – Big Data Value Chain and Technologies

On the same line, in Figure 4[5], it is depicted the Big Data Ecosystem, which shows the interconnection among the following actors:

- Subjects generating data, i.e. data "providers";
- Technology providers, typically in the form of data management platforms;
- Users, i.e. who use the big data to create added value;
- Data brokers, which are organizations collecting data from a set of sources, both public and private, and that sell them to other organizations;
- Companies and research organizations, who develop new technologies, new algorithms through which to explore data and extract value;
- Public bodies who act as regulators and/or public administration providing products and services to the citizens based on data, or that use data in their processes.

[4] Source: Curry (2016)
[5] Source: AGCOM (217/17/CONS)

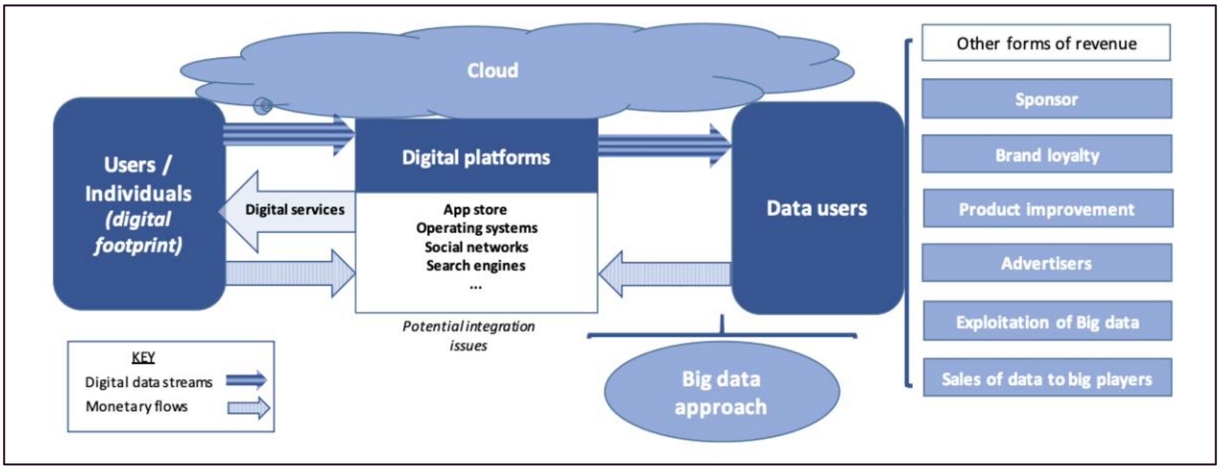

Figure 4 – Digital Data Market

The market of data analytics is continuously growing, even if marker research tends to focus on the big data aspect because it spans to a heterogeneous – and potentially expensive – set of solutions. Gartner affirms that 64% of organizations invested or planned to invest in big data in 2013[6]; 67% of the respondents to an MIT Sloan Management Review survey report that their companies are gaining a competitive edge from their use of analytics[7]; Forrester Research estimates that organizations effectively use less than 5% of their available data (Killmeyer et al. 2014). To take the point of view of a specific industry sector, McKinsey Global Institute calculates the potential value generated by big data solutions in US health care in $300 billion per year (Manyika et al. 2011). Big data and data analytics offer unquestionable added value to a diverse set of sectors; hereafter we give some short examples. Multimodal transport and logistics need the integration of sensor data with vast amounts of mobility and social data generated by an increasing set of devices and technologies; the opportunities are monitoring, controlling and managing transport and logistics processes, for example with goods delivery adaptation based on predictive monitoring or weather forecasts. In the environmental sector, the need is to understand planet and climate changes, also considering the role of human factors and interventions; the challenge is to improve the accuracy and availability of location data (e.g. those of the European programmes Copernicus and Galileo) to the benefit of the industry at large. The media and content market is another field in which big data and data analytics can improve the dynamic access and interaction patterns of users with media, also in a "prosumer" (producer and consumer) approach, to reach a seamless content experience. In the energy and manufacturing industries, the digitization of processes and the increasing supply of sensor networks require new approaches and technologies to (real-time) data management which in turn can lead to efficiency gains, predictive maintenance and better resource consumption.

1.4.2 Big Data in the Policy Cycle

The utilization of Big Data technology and analytics for government is now in the early stages of a practical implementation (Moorthy et al. 2015). In fact, as reported by IDC[8], analytics alone will grow from $130.1 billion in 2016 to over $203 billion in 2020 among others driven by a shift towards a data-

[6] More info available at https://techcrunch.com/2013/09/23/64-of-organizations-have-invested-in-or-plan-to-invest-in-big-data-tech-but-only-8-have-started-using-it-says-gartner/
[7] More info available at http://ilp.mit.edu/media/news_articles/smr/2015/56320Wx.pdf
[8] Please refer to https://www.ciodive.com/news/big-data-business-analytics-revenues-to-hit-203b-in-2020-idc-says/427507/

driven mindset. Furthermore, according to the Tech America Foundation[9], 82% of public IT officials say the effective use of real-time Big Data is the way of the future. While being still immature, big data technology solutions are applicable throughout the different phases of policy cycle, from agenda setting to policy design, implementation and evaluation. To put it very simply, evidence-based policy-making is based on causality, and data-driven approaches appear to offer radical improvements in detecting causality links by cross-analysing very disparate datasets.

The digital transformation of our societies, industry and government has made them unrecognisable in many ways. Many jobs have been lost, many more have been created. Costs of production of goods, provision of services, transport, and communication have significantly decreased, while speed, quality and efficiency have dramatically increased. However, the opportunities that digital technologies offer – notably with regards to public services – are yet to be fully seized. Indeed, an increased take-up of digital tools and solutions has the potential to render public services faster, cheaper, as well as more efficient, transparent, and user-oriented. The indirect positive effects are manifold, whether on public finances, on productivity, on citizens' lives, and on the environment. Overall, digital government transformation translates in a more competitive and attractive society. Today's society faces complex challenges such as migration, poverty, and climate change, for which not one optimal solution exists (Millard, 2015; Janssen and Helbig (2015). In order to address such problems, governments aim to realize public sector innovation that gears them towards becoming platforms of open governance, making optimal use of information and communication technologies (ICTs) to create public value (Millard, 2015). In this regard, the role of European governments has increased in complexity and complication. To cope with such challenges, ICTs have been increasingly used for enhancing the process of policy making and therefore address societal problems by formulating and implementing laws, rules and guidelines. In practical terms, data-driven policy making aims to make use of new data sources and new techniques for processing these data and to realize co-creation of policies, involving citizens and other relevant stakeholders. Clearly it is related to the notion of evidence-based policy making, which considers relevant the inclusion of systematic research, program management experience and political judgement in the policy making process (Head 2018). However, data-driven policy making stresses the importance of big data and open data sources into policy making as well as with co-creation of policy by involving citizens to increase legitimacy (Bijlsma et al 2011) and decrease citizens' distrust in government (Davies 2017). In this regard, a reported by Höchtl et al. (2015), data analytics has significant potential to be used in the policy cycle by contributing to policy decision making, in particular for what concerns:

- Identifying underperforming areas of public services and help with reallocation of resources for optimisation of public service provision;
- Improving existing processes by providing solutions for the citizens faster and with less paperwork;
- Predictions and forecasts.

To achieve these benefits, it is however necessary to address the privacy and security issues arising from government handling vast amounts of citizen-related data (Höchtl et al., 2015).

Policy making is the process of creating and monitoring policies to solve societal challenges. In this respect, it is often conceptualized as a policy cycle, consisting of several different phases, such as agenda setting, policy formulation, decision- making, implementation and evaluation. Concerning the use of big data technologies in the policy cycle, according to Maciejewski (2017), big data supports better policy development and execution "by strengthening the information input for evidence-based decision-making and provides more immediate feedback on policy and its impacts" (p. 127). According to

[9] TechAmerica Foundation. Demystifying Big Data: A practical guide to transforming the business of government. Retrieved from http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf

Schintler and Kulkarni (2014), big data has great potential as a resource for helping to inform different points in the policy analysis process “from problem conceptualization to ongoing evaluation of existing policies, and even empowering and engaging citizens and stakeholders in the process” (p. 343).
Specifically, the most common usage is in the **agenda setting** phase. Here, the challenge addressed is to detect (or even predict) problems before they become too costly to face. One traditional problem of policy making is that statistics (official but also unofficial) become available only long time after the problems (or opportunities) have emerged, hence increasing the costs to solving the problems and grasping the opportunities. Alternative metrics and datasets can be used to identify early warning signs at an earlier stage and lower costs. Other applications aim at better understanding causal links behind a certain policy problems by identifying unexpected correlations and hence providing hypothesis to test. Specifically, according to Longo et al. (2017), big data can serve as an input for “framing a policy problem before it is apprehended as such, indicating where a need is being unmet or where an emerging problem might be countered early” (p. 83). Moreover, according to Höchtl et al. (2016, p. 159) governments can identify emergent topics early and to create relevant agenda points collecting data from social networks with high degrees of participation and identifying citizens’ policy preferences. In the **policy design** phase, big data and data analytics solutions can be used for providing evidence for the ex ante impact assessment of policy options, by helping to predict possible outcomes of the different options. In this regard, Giest (2017, pag. 376) argues that the increased use of big data is shaping policy instruments, as “The vast amount of administrative data collected at various governmental levels and in different domains, such as tax systems, social programs, health records and the like, can— with their digitization— be used for decision-making in areas of education, economics, health and social policy”. In the **policy implementation** phase, big data and data analytics can help identifying the key stakeholders to involve in policy or to be targeted by policies. One way in which big data can influence the implementation stage of the policy process is the real-time production of data. The execution of new policies immediately produces new data, which can be used to evaluate the effectiveness of policies and improving the future implementation processes. Testing a new policy in real time can provide insights whether it has the desired effect or requires modification. Furthermore, as shown by Dunleavy (2016, pp. 12-13) big data can be used for behavioral insights. In this respect, the production of new data can also stem from the involvement of citizens science experiments, aimed at collecting data from the real world in real time. In the pilots, we will present and execute several of those experiments. In the **policy evaluation** phase, big data and data analytics approaches can help detecting the impact of policies at an early stage (Höchtl et al., 2016, p. 149), before formal evaluation exercises are carried out, or detecting problems related to implementation, such as corruption in public spending. Most importantly, big data can be used for continuous evaluation of policies, to inform the policy analysis process, while even empowering and engage citizens and stakeholders in the process (Schintler and Kulkarni 2014, p. 343).
In summary, big data tools and technologies present interesting opportunities to address some of the aforementioned key challenges of data-based policy making:

- Anticipate detection of problems before they become intractable;
- Generate a fruitful involvement of citizens in the policy making activity;
- Making sense of thousand opinions from citizens;
- Uncover causal relationships behind policy problems;
- Identify cheaper and real-time proxies for official statistics;
- Identify key stakeholders to be involved in or target by specific policies;
- Anticipate or monitor in real time the impact of policies.

However, data-driven policy making raises also complex challenges related to the capturing, integration and reuse of data exist (Bertot and Choi 2013, Janssen et al. 2012), as well as to the involvement of citizens and other stakeholders in policy making (Janssen and Helbig 2015, Ferro et al. 2013, Linders 2012). Furthermore, the assumption that simply because of the emergence of new technologies bureaucracies and public administrations will quickly adapt does not necessarily lead to the expected transformational outcomes. Big data readiness is an important factor, in order to avoid breaches of privacy and security of personal data, unfair treatment of citizens through overly extensive and unethical datafication of decision-making processes, wrong or suboptimal decisions because of incorrect data handling, analyses and interpretation (Clarke 2016, Janssen and Van den Hoven 2015, Margetts and Sutcliffe 2013).

Summarizing, several important research and implementation questions still need to be addressed:

- What is the disruptive and transformation potential of big data technologies for public sector operations and policy making activities?
- Which other activities do we have to consider - upskilling of personnel, changes, and adaptation of potentially outdated regulations, and investments in infrastructure?
- What are potential pathways and roadmaps that are to be followed by European public administrations consider starting the transformative processes of their own policy making activity?
- At the same time, how can we ensure that privacy, ethical and legal considerations are not jeopardized by these new technologies?

References

- Adrian, M. (2011). Information management goes' Extreme': The biggest challenges for 21st century CIOs. *SAS: Business Analytics and Intelligence Software*.
- AGCOM(217/17/CONS) Big Data Interim Report
- Barabási, A. (2003). *How Everything is Connected to Everything Else and What It Means for Business, Science and Everyday Life*. New York: Plume Books, 294pp. ISBN 0452284392.
- Bertot, J.C., & Choi H. (2013). Big data and E-government: issues, policies, and recommendations. In: *Proceedings of the 14th Annual International Conference on Digital Government Research*. dg.o. 2013, 17–20 June, Quebec City, pp. 1–10.
- Bijlsma, R.M., Bots, P.W.G., Wolters, H.A., & Hoekstra, A.Y. (2011). An empirical analysis of stakeholders' influence on policy development. *Ecol. Soc*. 16(1), 51–66.
- Budäus, D., & Buchholtz K. (1997). Konzeptionelle Grundlagen des Controllings in öffentlichen Verwaltungen. *Die Betriebswirtschaft*, 57(3), 322–337.
- Buelens, B., Daas, P., Burger, J., Puts, M. and van den Brakel, J. (2014). Selectivity of Big Data.
- Discussion paper 201411, *Statistics Netherlands*, The Hague/Heerlen, The Netherlands.
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: a survey. *Mobile Networks and Applications*, 19(2), 171–209.
- Christakis N., & Fowler J. (2007). The Spread of Obesity in a Large Social Network Over 32 Years. *New England Journal of Medicine*, 357(4): 370-379.
- Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, 26(1), 77–90.
- Colecchia, A. & Schreyer P. (2002). ICT Investment and Economic Growth in the 1990s: Is the United States a Unique Case? A Comparative Study of Nine OECD Countries. *Review of Economic Dynamics*. April, 5:2, pp. 408–42.
- Curry, E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In: Cavanillas J., Curry E., Wahlster W. (eds) *New Horizons for a Data-Driven Economy*. Springer, Cham.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, 54(1), 43–46.
- Dunleavy, P. (2016). 'Big data' and policy learning. In G. Stoker, & M. Evans (Eds.), *Evidence-based Policy Making in the Social Sciences: Methods That Matter* (pp. 143-151). Bristol: Policy Press.
- Ferro, E., Loukis, E., Charalabidis, Y., & Osella, M. (2013). Policy making 2.0: From theory to practice. *Government Information Quarterly* 30(4):359–368.
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC Iview*, 1142, 1–12.
- Giest, S. (2017). Big data for policymaking: fad or fasttrack? *Policy Sciences*, 50(3), 367-382.
- Groppo, G., & Heck U. (2009). Strategische Neuausrichtung der öffentlichen Verwaltung. Ein Erfolgsfaktor zur Umsetzung der Verwaltungsmodernisierung. *Verwaltung und Management*, 15(5), 271-277.
- Grimmelikhuijsen, S., Porumbescu, G., Hong, B., & Im, T. (2013). The Effect of Transparency on Trust in Government: A Cross-National Comparative Experiment. *Public Administration Review*, 73(4), 575-586.
- Hanger. S., Pfenninger, S., Dreyfus, M., & Patt, A. (2013). Knowledge and information needs of adaptation policy-makers: a European study. *Regional Environmental Change*, 13(1), 91-101.
- Head, B. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration*, 67(1), 1–11.
- Höchtl, J., Parycek, P., & Schöllhammer, 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.
- Hota, C., Upadhyaya, S., & Al-Karaki, J. N. (2015). Advances in secure knowledge management in the big data era. *Information Systems Frontiers*, 17(5), 983–986.

- Howard, C. (2005). The Policy Cycle: A Model of Post-Machiavellian Policy Making? *Australian Journal of Public Administration*, 64: 3-13.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag*. 29(4), 258–268.
- Janssen, M., & Helbig N. (2015). Innovating and changing the policy-cycle: policy-makers be prepared! *Gov. Inf. Q*.
- Janssen, M., & Kuk, G. (2016). Big and open linked data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, 26(1–2), 3–13.
- Janssen, M., & Van den Hoven, J. (2015). Big and open linked data (BOLD) in government: a challenge to transparency and privacy? *Government Information Quarterly*, 32(4), 363–368.
- Jorgenson, D., & Stiroh K. (2000). Raising the Speed Limit: U.S. Economic Growth in the Information Age. *Brookings Papers on Economic Activity*, (1), 125–211.
- Kaplan, R., & Norton, D. (1992). The Balanced Scorecard: Measures that Drive Performance. *Harvard Business Review*, 70(1), 71-79.
- Killmeyer J., Crandall R., Crandall, W. (2014). *Principles of Supply Chain Management*. CRC Press.
- Kim, J.K. (2011). Parametric fractional imputation for missing data analysis, *Biometrika*, 98, 119–132.
- Kim, J.K., Berg, E. & Park, T. (2016). Statistical matching using fractional imputation. *Surv. Methodol*., 42, 19–40.
- Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2016). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers: a journal of research and innovation*, 1-17.
- Kuhlmann S., & Meyer-Krahmer F. (1995). Practice of Technology Policy Evaluation in Germany: Introduction and Overview. In: Becher G., Kuhlmann S. (eds) Evaluation of Technology Policy Programmes in Germany. Economics of Science, Technology and Innovation, vol 4. Springer, Dordrecht.
- Linders, D. (2012). From E-government to we-government: defining a typology for citizen coproduction in the age of social media. *Gov. Inf. Q*. 29(4), 446–454.
- Longo, J., Kuras, E., Smith, H., Hondula, D. M., & Johnston, E. (2017). Technology use, exposure to natural hazards, and being digitally invisible: implications for policy analytics. *Policy & Internet*, 9(1), 76-108.
- 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.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work and think*. London: John Murray Publishers.
- Manyika, J. et al. (2011). Big Data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, McKinsey & Co.
- Margetts, H., & Sutcliffe, D. (2013). Addressing the policy challenges and opportunities of Big data. *Policy & Internet*, 5(2), 139–146.
- Millard, J. (2015). Open governance systems: doing more with more. *Gov. Inf. Q*.
- Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big Data: Prospects and Challenges. *Vikalpa*, *40*(1), 74–96.
- Olabe, P. B. (2017). Responding to citizen´s need: Public services and trust, in: OECD, Trust and Public Policy. How Better Governance Can Help Rebuild Public Trust, *OECD Publishing*, Paris, 47-65.
- OECD (2013). Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by "big-data", in: OECD: Supporting Investment in Knowledge Capital, Growth and Innovation. *OECD Publishing*, Paris, 319-356.
- OECD (2017a) OECD Digital Government Toolkit. 12 Principles.
- OECD (2017b). Knowledge management in the public and private sectors: similarities and differences in the challenges created by the knowledge-intensive economy.

- Oliner, S., & Sichel, D. (2000). The Resurgence of Growth in the Late 1990s: Is Information Technology the Story? *Mimeo, Federal Reserve Board*, February.
- OpenTracker (2013). Definitions of big data. *OpenTracker*.
- Ormerod, P. (2010). N squared: public policy and the power of networks. *RSA Pamphlets*, p36.
- Park, S., Kim, J.K. & Stukel, D. (2017). A measurement error model for survey data integration: combining information from two surveys, *Metron,* 75, 345–357.
- Pollanen, R. M. (2016). Linking Strategic Planning and Performance. Measurement in Canadian Public organizations: Does it improve Performance? *Financial Management Institute of Canada*.
- Schelling, T. (1969). Models of Segregation. *American Economic Review,* 59 (May), pp. 488-493.
- Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research*, 31(4), 343-348.
- Simon, P. (2013). *Too big to ignore: The business case for big data*. Hoboken: Wiley.
- Taleb, N. (2007) *The Black Swan: The Impact of the Highly Improbable*. Random House, ISBN 978-1400063512.
- Thomas, J. C. (2013), Citizen, Customer, Partner: Rethinking the Place of the Public in Public. *Public Administration Review*, 73(6), 786-796.
- Wang, X., & Van Wart, M. (2007), When Public Participation in Administration Leads to Trust: An Empirical Assessment of Managers´ Perceptions. *Public Administration Review*, 67(2), 265-278.
- Weber, M. (1980). *Wirtschaft und Gesellschaft: Grundriss der verstehenden So*ziologie. Mohr, Tübingen.
- Wigg, K. M. (2002). Knowledge management in public administration. *Journal of Knowledge Management*, 6(3), 224-239.