

Big data and policymaking: between real time management and the experimental dimension of policies

Grazia Concilio¹, Paola Pucci¹, Giovanni Vecchio², Giovanni Lanza¹

¹ Politecnico di Milano, Dastu, piazza da Vinci 32, 20133 Milano (Italy)

grazia.concilio@polimi.it; paola.pucci@polimi.it; giovanni.lanza@polimi.it

² Pontificia Universidad Católica de Chile, CEDEUS, Los Navegantes 1963, Providencia, Santiago de Chile (Chile) giovanni.vecchio@uc.cl

Abstract. The paper aims at exploring how big data can support decision making for and about cities at different strategic levels and temporal perspectives. Big data can improve the effectiveness of urban mobility policy, but such contribution heavily needs to consider the multiplicity of big data, as reflected by three elements: the different sources that produce data and the knowledge they provide; the many actors who produce, storage, manage and use of big data; the different roles that data may play in the different stages of a policy making process, and the many actors in the very production, storage, management and use of big data. Based on this, the paper presents a sound policy cycle focusing on the experimental dimension of policy making and provides a ground for the assessment of project implications for the ‘business of government’. The paper considers specifically mobility policies and, referring to the experience of the Polivisu research project, provides a policy cycle tested in relation to three pilot cases using big (open) data visualizations in a clear mobility policy context: Ghent (Belgium), Issy-les-Moulineaux (France), and Pilsen (Czechia). By considering the cycle of the policy process, the policy making activities the pilots are experiencing, and the data they are processing, the paper shows how the pilot cases are internalizing the policy experimentation opportunity, addressing the further pilots’ activities, into a continuous policy adaptation cycle.

Keywords: policy cycle, experimental dimension, pilot cases.

1. Introduction

Cities are more and more recognized and referred to as environments of data production. The digitalization of an enormous amount of existing services and the development of new smart ones are the evident causes of the city being a data producing engine. Data in fact represent the key ingredient for the services to be properly and effectively produced and supplied [1]: the use of services produce data that, at the same time, are used as basic resources for the services to be growingly effective and efficient. The more this is true, the more the produced data sets give rise to big-data, that is, “everything captured or recorded digitally by modern information and communications technologies such as networked sensors, ‘smart’ objects and devices, the web and social media” [2, p. 28]. Big data is both an opportunity and a resource for policy making: the size of information it provides, the epistemological value it assumes and its necessary interactions with other forms of knowledge provide “an important new source of insights into the management, governance and experience of urban life” [2, p.29], which nonetheless is still object of a huge academic and practice-oriented debate [3; 4; 5]. However, the opportunities ICTs provide in supporting urban policy making processes is a basic component of the discourse on “smart cities” and several public institutions, such as the EU, are devoting efforts to strategically promote their use towards a “smart” urban growth [6, p.3]. The aim of

these efforts is to experiment and establish processes of evidence-based policy making using the most advanced technologies and big data to “find patterns, linkages and relationships that shed fresh light on policy problems” [7].

Based on these premises, the paper aims at exploring how big data can support decision making for and about cities at different strategic levels and temporal perspectives. Drawing on an ongoing research project (Polivisu) in which three pilot cases are using big (open) data visualizations in relation to mobility issues, the paper presents a policy cycle focusing on the experimental dimension of policy making and provides a ground for assessing the project implications for the ‘business of government’ as far as mobility policies are concerned.

2. Urban decisions and the multiplicity of big data

The multiplicity intrinsic of big data explains the relevance of the knowledge it provides and its potential contribution to urban decision making, but also raises significant practical issues. Three elements, here explored in relation to urban mobility, are crucial for the multiplicity of big data: its sources, its possible contributions to policymaking, and the actors involved in its production, storage, management and usage.

2.1 Big data and its different sources

Big data improves our ability to understand human mobility in many respects: first, it makes possible to overcome the limitations in the detection of latency typical of traditional data sources; second, it can exploit the pervasiveness guaranteed by the ubiquity of digital devices and mobile phone networks; third, a ‘longitudinal perspective’ becomes available on the variability in human travel activities [8], thus validly complementing traditional research methods. These conditions depend strongly from the type of sources of big data: whether data is produced voluntarily or not, aggregated or not, anonymous or not, its possible uses in urban planning and in other policy domains changes, and determines different challenges in terms of use and interoperability with other data relevant for evidence-based policymaking processes.

Different sources produce three main categories of big data [9, p. 4]: directed, automated and volunteered. *Directed data*, monitoring the movement of individuals or physical objects moved by humans, is generated by traditional forms of surveillance. *Automated data* is produced by automatic digital devices such as mobile phones, capture systems, clickstreams and sensors, but also by image data like aerial and satellite images. *Volunteered data* is generated by users who interact with social media or produce crowdsourced data contributing to a common system, such as OpenStreetMap. Such multiplicity of sources raises the issue of organizing and structuring the data itself as well as reducing its size by trying to keep the information (signal) alive and by eliminating everything that is not needed for the analysis (noise). This issue is crucial for the policy use of big data: the availability of a large amount of data improves the accuracy and completeness of the measurements to capture phenomena that were previously difficult to investigate, but at the same time increases the level of complexity in the approaches finalized to process, integrate and analyze this data [10]. Central in this sense is also the expediency to improve/validate/integrate big data with existing ones. If the main features characterizing big data are velocity (data that can be obtained in real time), volume (large amount of data) and variety (data coming from different sources), it follows that big data is not ‘big’ for its size, but for the possibility of being linked to other data from different sources [11]. Data integration would be crucial also to gain new insights on existing and potential policy issues involving multiple actors.

2.2 Big data and actors' expertise

According to Kitchin [9, p.136], “data are not useful in and of themselves. They only have utility if meaning and value can be extracted from them”. The use of big data to face mobility issues is therefore deployed in an environment in which different expertise are involved, from the wide range of actors who have a deep knowledge on software and hardware management, statistics, machine learning and other technical specialists [9, p.202], to the subjects with the political and managerial competencies required to define a policy problem that the use of big data can contribute to face, design a policy solution to it and mediate between all the actors involved in the process. This last skill is fundamental, considering the crucial role of private companies in relation to big data [12, p.7]. Four categories of expertise should be involved as actors to work effectively with data in a policy making process [13, pp.15-21]:

Domain expertise, with a profound knowledge and understanding of the humanities and social science tradition connected to the policy issue. It can guide the process of data selection and help in defining the policy problem;

Data expertise, to select, gather, assess the quality and validity of data and to understand the relation between them and the material objects and dynamics they describe in a digital form;

Analytical expertise, to help obtaining relevant insights from big data through specific technical actions and digital tools for research and analysis purposes;

Project management expertise, relevant to make the whole process and the interactions between different expertise running smoothly.

The multiplicity of big data is thus a challenge for policy makers and has relevant consequences in an operative perspective.

2.3 Big data and their role in mobility policy making

Even if the increasing use of big data in the policy domain poses relevant challenges and is subject to limitations, networked technologies are important drivers of knowledge for mobility and urban practices: from a policy perspective, big data provides a huge amount of information, offering a question-oriented knowledge. Depending on the question or issue faced by the subject using them, big data both supports a real-time knowledge of mobility needs/practices and offers valuable support for ‘re-scaling’ urban policy, contributing to multi-scalar maps of heterogeneous urban phenomena. In both perspectives, big data becomes a valuable tool for (mobility) policy makers: the use of data along the policy making process in fact can contribute to create evidences for developing a broader knowledge of complex issues and for the design of more effective policy.

The contribution of big data changes according to the stages of the policy making process (fig. 1), which can be considered a recurring sequence of three main cycles: design, implementation, and evaluation [14]. These cycles are strongly interdependent, and this is at the core of the experimental dimension of policy making that our proposal intends to implement. The *policy design* cycle is focused on highlighting the existence of a collective problem, mobilizing a set of goals and objectives in relation to it, and defining policy strategies and actions as attempts to contribute to solving the problem; in this phase, the usage of big data could be relevant to promote forms of problem crowdsourcing [15]. The *policy implementation* cycle gives form to the policy defined in the previous stages, delivering it to the public; data can contribute to putting the policy into practice and reshaping it, if necessary, depending on the reactions of the setting or on the achieved results. In this complex stage, data contributes to manage urban mobility issues in real time; to support private initiatives

addressing emerging urban mobility needs [16]; to improve the ‘infostructure’ of mobility by providing information that is updated, complete and personalized [17]; to involve citizens in the implementation of policy measures; to provide accountable information. Finally, the *policy evaluation* cycle examines the desired and undesired impacts achieved with the implementation of a policy; data can contribute to estimating the obtained outcomes and if the desired changes were obtained.

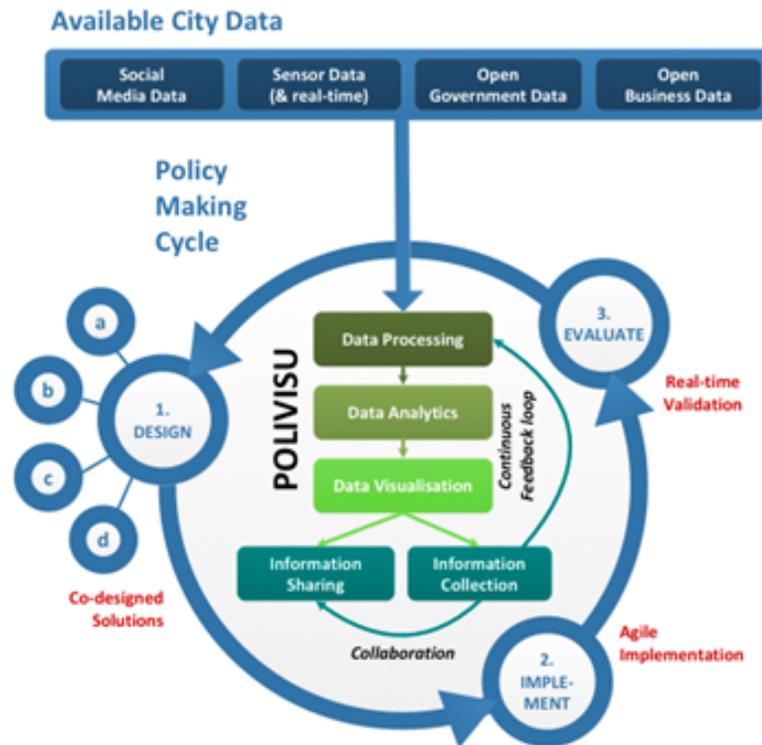


Fig. 1. Data integration in the Policy Experimentation Cycle [18].

3. Data relevance at the intersection between data production systems and policy making

3.1 Data for short- and long-term decisions and reasoning

The three issues discussed in the previous section show how big data introduces unprecedented complexity in policymaking, using new technologies but, more importantly, questioning the very basis of the policy making process towards new interpretative models. The growing data availability increasingly affects the way we analyze complex urban problems and make decisions for cities: data are a promising resource for more effective decisions as well as for better interacting with the context where decisions are implemented.

Decisions for and about cities are made at different urban scales, refer to different strategic levels and have different time perspectives, with reciprocal interdependencies that are changing due to the data availability. Here we mainly focus on the interplay between the different steps of decisions in policy making (those introducing the longest time perspective) and those necessary for the daily management of the city (connected to the shortest, real-time perspective), an intersection at which data can play a key role. Short-term management is embedded in the smart sphere of decisions for/about cities: here decisions are less analytic and more routinary. Routines may depend on data-driven learning mechanisms (also using data series) supporting smart systems to recognize situations and apply

solutions/decisions that have already proved to work. The decision has a temporary value related to the specific condition detected in a precise moment by the smart system. Opposite to real-time decisions, policy making works in a long-time perspective. Anticipatory is the prevailing mode for reasoning in this case and data-driven models are often adopted as supporting means to deal with the impacts of the policy measures, representing thus a relevant source for exploring decision options mainly having a strategic nature (since they consider recurring issues and aim at more systemic changes).

action	reaction	adaptation	planning
reasoning	rule-based	reflective	anticipatory
decision	temporary	reversible	strategic

Fig. 2. Decision/reasoning along diverse timeframes

Between short-term and long-term decisions a variety of situations is possible, which may be considered as characterized by decisions having a reversible nature: they are neither strategic in value (like those oriented to long term perspective for systemic changes), nor aiming at dealing with temporary, contingency situations asking decisions which are known as having the same (short) duration of the phenomenon to be managed. For such decisions, reasoning is not (fully) anticipatory and its temporariness allows reflection as embedded in action. Within the three different time-frames, actions are different in nature and show different use and role of data:

in the *short term*, action (the smart action) is mainly reactive; real time data are used as reference info to interpret situations;

in the *medium term*, action is mainly adaptive; data series, including current data, are used to detect impacts of the action itself and to improve it along time;

in the *long term*, action has a planning nature; data series become crucial to detect problems and to develop scenarios for long lasting changes.

3.2 Opportunities for an experimental approach to policy design

The interdependency between policy design, implementation and evaluation is strictly related to two factors, especially when considering the role (big) data can play. Design and implementation can be clearly and sequentially distinguished when a systematic, impact-oriented analysis is possible at the stage of design as it allows a clear costs and benefits assessment of different action opportunities (see the planning mode of strategy making in [19]).

Comprehensive analyses have the value to drive long range, strategic actions, and consequently show a clear dependency and distinction of the implementation from the design cycle. At the same time, the bigger is the uncertainty (not only related to the possible lack of data, rather also consequent to the high complexity of the problem/phenomenon to be handled), the smaller is the chance to carry out a comprehensive analysis. Therefore, goals and objectives cannot be defined clearly and the policy making shifts from a planning towards an adaptive mode [19]. Inevitably, this shift reduces the distance between design and implementation, transforming policy design into a more experimental activity that uses learning from implementation into food for design within adaptive dynamics.

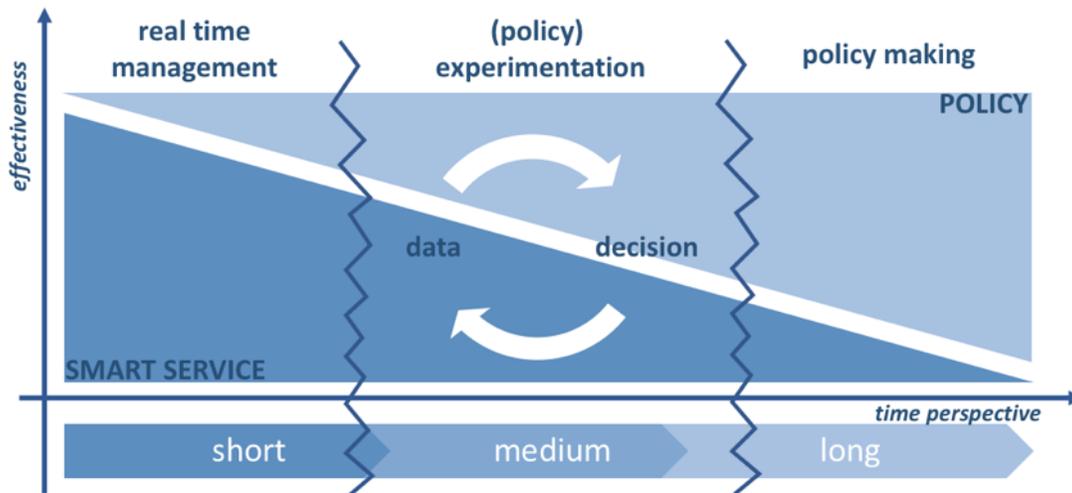


Fig. 3. Real time management vs policy making.

Coherently with the discussion on the time frame perspective adopted in the previous paragraph, it may be clear that a merge between policy design and implementation is consistent with the situation described in the medium term: within an adaptive mode for decisions, policy making can clearly become experimental. Experiments may refer to both the policy strategies and measures. They can reduce the risk of trial-errors approaches while considering the learning in action opportunity to improve, adapt, adjust the policy while making it in order to increase its capacity to affect the context in an evolving manner. Differently from the Mintzberg's considerations [19], the merge between policy design and implementation does not represent a sort of inevitable but not preferred option when a comprehensive analysis is not possible. In the era of (big) data availability, this merge can be looked at as an opportunity to create policies while verifying the policies themselves throughout their interactions with the contexts.

4. The experimental dimension at work in the Polivisu pilots

The experimental dimension of policy making is at the core of the concept of the Polivisu project. The experimental dimension puts emphasis on the learning dynamics of policy making through a strong relation with iterations cycle. This section reconsiders the elements discussed in the previous sections and observes them in action, discussing three pilot cases that differently face the limits and opportunities related to data sourcing, mobility policy-related data use and collaboration among different actors: these use big (open) data visualizations in a clear mobility policy context, to deal with different transport and mobility issues. The pilot cases are Ghent (Belgium), Issy-les-Moulinaux (France), and Pilsen (Czechia). This process works at two scales. From one side, it relies on small scale experiments, exploiting the real time opportunities offered by big data, in each of the pilot. From the other, it concerns also large-scale experiments as an entire policy cycle taking into account the in-depth throughout monitoring enabled by big data. For each pilot, the next subsections describe the problem to be faced, the imagined solutions, the actions taken, the following steps and a brief discussion on the short- and long-term decisions and reasoning allowed by data as experimented by the pilots.

4.1. Ghent

Ghent is the second biggest city in the Flanders with 260.000 inhabitants and it is an industrial and commercial focal point for the region. Moreover, it is an important cultural center due to the presence

of the well-known Ghent University. The participation of Ghent in the Polivisu project is related to the increasing number of students and commuters attracted by the local university and other important research centers located within the municipality: around 440.000 city users, not registered as residents, are estimated to reach or reside in Ghent on a daily base. This trend generates wide impacts on many policy domains, including mobility, and is combined with other urban issues, such as uncontrolled increase in housing prices and crisis of commercial activities.

The interest of Ghent in the opportunities provided by the Polivisu toolbox are mainly related to the potential role that big data can play in measuring the extent of the city user population, with a specific focus on students' housing. Big data in fact may help policy makers in adapting city services to changing populations, assessing the impact of policy proposals on previously "hidden" inhabitants but also estimating the effects determined by students on city services. The process followed by Ghent started in 2016 with an evaluation on local policies that assessed the necessity to use data to understand the extent of the student population. Subsequently, the aim of the following design phase was to find available microscale data sources to build informative and intuitive visualization for the different stages of the policy making process related to student housing. The pilot worked in the implementation phase assessing the availability and quality of data related to student housing to build a visualization tool, defining relevant local stakeholders to collaborate with.

Ghent uses big data to react to an emerging issue, the growing presence of city users and the impact they generate. The city intends to use visualisations as a tool for anticipatory reasoning, contributing to represent the presence of city users before defining evidence-based strategic measures to deal with their growing number. Within this ongoing process, until now an evidence-based policy making process involving the use of (big) data has proved to be challenging due to the difficulties in obtaining information that is privately owned or with different degrees of quality and completeness. For this reason, the main action that has been taken by the pilot is a negotiation that is both internal (across different administrative entities) and external (with communication providers) to facilitate data gathering. In parallel, promoters have started considering data scraping from social media (Twitter and Facebook) as a potential source of information. In future steps, the intention is to increase the amount of data, experimenting new ways to find information that were not considered before and could help to produce more insightful visualizations.

4.2 Issy-les-Molinaux

Issy-les-Molinaux is a large town of 70.000 inhabitants in the Ile de France Region, bordering the city of Paris, particularly vital both from the demographic and economic points of view. It hosts a relevant amount of economic activities (tertiary industry) and it is affected by traffic congestion, even if residents have a limited use of car and public transport system is well developed. The main leverage to find innovative solution for mobility and traffic management is the upcoming opening of massive work sites for the construction of Grand Paris Express, a new system of metro lines that will serve Issy and several other municipalities surrounding Paris. In forecast of these massive work that will have a relevant impact on already existing issues of traffic congestion, local authorities had set three main targets to be reached by using big data and visualization: understand in detail traffic trends, evaluate the impact of mobility-related policies and upcoming construction works, help people improve mobility behaviors. The main expected outcome is an easy to use visualization toolkit able to register and to predict mobility behaviors in relation to forecasted road works, useful not only for policymakers but also for citizens that could rely on real time monitoring and predictions applications to make more sustainable mobility choices.

The pilot worked, in a first step, creating three focus groups involving public and private actors in the definition of a project to be presented to various local stakeholders. In parallel, the pilot worked

matching data and policy making by analyzing existing data, defining other valuable digital information not yet available and analyzing the role that data could play in policy making processes. Finally, an effort has been made to stimulate potential users of the toolkit by analyzing existing barriers in the use of data and data literacy. This preliminary process highlighted challenges related to data availability (public bodies have access to limited resources and need to cooperate with private entities), institutional fragmentation (big data-measurable trends are related to administrative scopes that are wider than the single municipality) and interoperability. In the future steps, the pilot intends to deepen the availability and use of other datasets to provide more relevant visualization prototypes, developing a traffic app and involving other stakeholders in the process so to build a wider set of visualization tools.

In the pilot, the smart actions taken act as answers to local and metropolitan infrastructural and economic changes. The experience of Issy-les-Molinaux intercepts two main decisional timeframes. Real time data are used to guide actions as exploratory reactions to contingent events (roadworks, traffic jams, closures...). These reactions, under the form of data series, are then used to detect the outcomes of the action itself and to understand the impact on travelers' behaviors and may contribute to policy measures promoting modal shift, moving thus the contribution of data to a strategic level. Nonetheless, until now Issy-les-Molinaux is working assessing the role of data to guide decisions that are inbetween the short and the medium terms.

4.3 Pilsen

Pilsen is a medium sized city of 170.000 inhabitants, one of the most important industrial centers in the Czech Republic. The city is affected by huge traffic congestion exacerbated by the presence of important productive sites, the weakness of the road network and the lack of diagnostic tools for traffic and road work management. Aware of these issues, the local public administration drafted in 2016/17 a plan for sustainable mobility, promoting relevant strategies such as new bypass roads, reconstruction of important arteries and bridges and a new tram line. In forecast of these interventions, local authorities have considered the opportunity to introduce big data-related tools and visualization in the governance of urban mobility at the implementation stage of the policy cycle, to better plan and coordinate upcoming road works, identify the most congested areas and simulate the impact of new infrastructures on traffic trends.

The pilot aims at developing, with the support of different stakeholders, a web application that will visualize approved road closures and will provide a recalculation of traffic volumes based on real time measurement of the actual mobility trends, allowing a better coordination in presence of contingent and long planned interventions. The tool will be also useful for citizens, who will obtain via the web application all the relevant information related to major traffic problems and the schedule of planned works, towards a minimization of negative impacts of road works in their daily displacements.

The pilot worked promoting a direct cooperation among public and private actors in defining and take a set of relevant actions to test the contribution of big and open data in traffic and infrastructure management, starting with the identification of significant datasets (public open data, police data, sensor recording...). In these first steps, pilot promoters faced relevant challenges related to the multiplicity of actors involved in infrastructure planning and construction, which generates problems of coordination and a limited possibility to know information related to closures in sufficient advance. As for the future, the city of Pilsen intends to furtherly develop the web application by creating a reliable list of planned roadwork and cooperating with the police to guarantee a continuous flow of information from the city sensors.

The experience of Pilsen clearly shows how data use in policy making intercepts multiple decisional dimensions. The smart model for traffic management that has been developed can guide reasoning and

decision making at any timeframe (short-, medium- and long-term). However, the model is particularly relevant in a long-time perspective, because it allows to develop anticipatory reasoning on how to implement the local plan for sustainable mobility and to define an experimental path to orient related strategic decisions.

5. Discussion and conclusions

Considering the cycle of the policy process (design, implementation, evaluation), the policy making activities the pilots are experiencing, and the data they are processing, the paper shows that it is possible to internalize the policy experimentation opportunity, addressing the further pilots' activities, into a continuous policy adaptation cycle.

Mapping the pilots' activities clearly shows that:

- the policy making process is not a linear one and each pilot bridges more than one activity related to mobility problems (among those listed by the model). This does not affect the correctness of the model and rather suggests that it is more effective to observe and describe their work at the scale of single activities rather than referring to cycles or steps;
- open and big data processed are mainly used to monitor the current conditions of a setting and provide evidence for making decisions accordingly;
- a clear usage of big data within the policy making cycle emerges only from a limited share of actors, with a focus on the formulation of policy measures;
- one of the pilots has opened new smart city services to the public, and they are now producing the most of the available big data which are not considered for future initiatives regarding policy making by the policymakers and politicians in the city.

The type of action that results from big and open data visualization is not often attributable to a policy, which by its nature affects the medium-long term perspective. In many cases in fact, big data improves the *management* of a policy. This suggests considering the time dimension of a policy cycle distinguishing short, medium and long term, exploring thus the tension between real-time management and the experimental dimension that the use of big data enables.

Acknowledgments

The authors acknowledge the funding received from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769608 "Policy Development based on Advanced Geospatial Data Analytics and Visualisation". Polivisu Project H2020 -SC6-CO-CREATION-2016-2017.

References

- Abella, A., Ortiz-de-Urbina, Criado M., De-Pablos-Heredero, C.: A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. *Cities*, 64: 47-53 (2017).
- Rabari, C., Storper, M.: The digital skin of cities: urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society*, 8(1), 27–42 (2015).
- Batty, M.: Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274–279 (2013).
- Kitchin, R.: Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1) (2014).
- Kitchin, R.: The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1–14. (2014).

- Caragliu, A., Del Bo, C., Nijkamp, P.: Smart cities in Europe. Research Memoranda VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics, 48 (2009).
- European Commission: Quality of Public Administration: a Toolbox for Practitioners (2017).
- Järv, O., Ahas, R., Witlox, F.: Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies*, 38, 122–135 (2014).
- Kitchin, R.: The data revolution. Big Data, Open Data, Data Infrastructures and Their Consequences. Sage, Singapore (2014).
- Einav, L., Levin, J.D.: The Data Revolution and Economic Analysis. *NBER Working Paper Series*, 19035 (2013).
- Boyd, D., Crawford, K.: Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679 (2012).
- Gantz, J., Reinsel, D.: Extracting value from chaos. IDC IView, 1-12 (2011).
- Williford, C., Henry, C.: One culture. Computationally intensive research in the humanities and social science. A report on the experiences of first respondents to the digging into data challenge. Council on Library and Information Resources, 151 (2012).
- Pucci, P., Vecchio, G., Concilio, G.: Big data and urban mobility: a policy making perspective. *Transportation Research Procedia* (forthcoming).
- Brabham, D.C.: Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory*, 8(3), 242–262 (2009).
- Vecchio, G., Tricarico, L.: “May the force move you”: Roles and actors of information sharing devices in urban mobility. *Cities* (forthcoming).
- Schwanen, T.: Beyond instrument: Smartphone app and sustainable mobility. *European Journal of Transport and Infrastructure Research*, 15(4), 675–690 (2015).
- Polivisu: The PoliVisu Policy Making Model (DRAFT) (2018).
- Mintzberg, H.: Strategy-Making in Three Modes. *California Management Review*, 16(2), 44-53 (1973).