

Big Data for Public Policy – The State of Play

Author: Charlotte van Oirsouw

Big Data for Public Policy – The State of Play In an increasingly fast-paced, complex and digitized society, public administrations need to develop policies that address problems efficiently and effectively and in order for them to remain relevant. Data-driven policy making is a form of policymaking that uses information and communication technologies (ICTs) to capture the benefits of new data sources such as Big- and Open Data to create better policies. Data-driven policymaking holds promises of innovating public sector and has captured the interest of the European Commission. This concept has grasped our attention due to the prominent role of Big Data, hence the BDVe organized the webinar 'Big Data 4 Policy – The State of Play: How to bring data-driven approaches into the policy cycle, and to which benefits?' on May 14. This blog reports what was discussed.

Data-driven policy making builds upon the concept of evidence-based policy making. Its purpose is to create better policies through usage of several data sources, but adds the aim of creating legitimacy by involving citizens and stakeholders throughout different phases in the policy cycle. A very important aspect of data-driven policymaking is a methodological approach that involves co-creation and multi-stakeholder approaches. The policy cycle consists of different phases, such as agenda setting, formulation, decision-making, implementation and evaluation. For example, using social media data for agenda setting or evaluation of a policy. Data used for policy making can be gathered either by usage of new data sources or by digitizing an existing service.

Anne Fleur van Veenstra from TNO kicked off the webinar session by presenting her findings of the Policy Lab project she carried out in Rotterdam (NL). The Policy Lab is a methodology developed by TNO for conducting experiments with new data sources and new technologies, with the aim of exploring data-driven policy instruments. During the webinar, she elaborated the Policy Lab Methodology and the Rotterdam project.

The Policy Lab idea came up due to the gap between present attention for data-driven services in research and its lacking implementation in policy making. At the same time, the European Commission also stated that they saw the increasing need to enhance the legitimacy of policy making. The purpose of the Policy Lab is not only to help governments out in creating better policy, but also to gain insights in what it means to take up such a mission. One of the most important aspects in their methodology was involving stakeholders, including governments, municipalities, public- or semi-public organizations, citizens and sometimes private organizations. Some of the challenges they faced throughout the project were the level of organizational readiness and the willingness of policy makers and politicians to accept such methodologies. They also found that role of doing experiments is crucial, as it allows for an understanding of what issues come to the surface during the process of implementing data-driven policy making. A clear example of such an issue are legal requirements, because with non-compliance, outcomes end up being unusable in practice.

The case she presented was the experiment in Rotterdam on Youth Policy. Its focus was on getting insight in factors that influence social-emotional skills of youngsters. The municipality of Rotterdam believed that these insights could improve their policy model, among others improving education levels and decreasing government costs. As Anne Fleur expected, they did not find one particular variable that influenced social-emotional skills, several factors turned out to play a role. The strongest factor that they found to be of influence was the value of houses that youngsters lived in, but they believe this to be a proxy for income rather than an actual factor itself.

Throughout the experiment, they experienced that it was hard to obtain data because it was scattered, resulting in the pre-process taking lot of time. In addition, they found that joint development of data-protection agreements is of major importance when it comes to establishing

trust between parties. When using many different data sources, there are many factors that have (potential) influence and the model becomes complex as a result. With such complex issues touching upon multiple domains, you need domain specific knowledge to interpret the results of the data analysis. A multidisciplinary approach is therefore important for getting meaningful outcomes. It also remains important to differentiate between the experiment and its application in practice, because the gathered data can include biases when they have been developed in other settings. A challenge that presented itself is the explainability of machine learning, as explanation of the underlying reasoning is a requisite for implementation of the policy. Openness also proves to be a challenge, and raises the question to what degree the policy making process should be open. Should algorithms or gathered data be revealed? It is also important to be conscious that with increased openness comes increased scrutiny and accountability as well. They found that the policy cycle process accelerates when using data-driven policy making. This experiment took 1,5 year and even though it proved to be challenging, it is definitely also very promising for future research.

Francesco Mureddu from the Lisbon Council then proceeded by presenting their Roadmap for a Fair Data Economy. Their Roadmap analyzes the current situation, underlying trends, a new vision for the data economy and what needs to be done to make it happen.

The Roadmap carried out an analysis of the gaps and external factors in public administration, with a focus on Big Data. The findings are that Big Data tools and technologies present interesting opportunities to address some of the challenges faced by public policy. These include anticipatory detection of policy problems before they become intractable, fruitful involvement of citizens in the policy making process and uncovering causal relationships behind policy problems. They defined five research clusters related to Big Data and policy making. These are privacy, transparency and trust; data acquisition, cleaning and storing; data clustering, integration and fusion; modelling and analysis with Big Data and data visualization. The Roadmap also identified several current research challenges. Francesco firstly touched upon the challenge of privacy, algorithmic bias and transparency. As algorithms are developed by humans and trained by data, there is room for the possibility that algorithmic decision-making reinforces biases. In this respect, there is a need for common practices and tools to explain what happens throughout the decision-making processes. Another research challenge he mentioned was the integration and interoperability of public administration databases. Governments hold massive amounts of data. The once only-principle of e-government requires that we exploit it as much as possible, but here data protection and integration of data prove to be very challenging. Currently there are frameworks under development that aim to solve this matter on the public administration level, for instance the data analytics framework. The Roadmap will be presented at the Co-VAL Summit in Brussels, the Data for Policy Conference in London and at the Big Data Value PPP Summit in Riga.

Interesting links

- The webinar: big-data-value.eu/resources/webinar
- Policy lab: tno.nl/en/focus-areas/strategic-analysis-policy/expertise-groups/strategy-policy/policy-lab-developing-data-driven-policies
- Big data policy canvas project: bigpolicycanvas.eu