
Data for Policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking

Report about the State-of-the-Art

17 July 2015

Summary

Opportunities for policymakers

Big data and other innovative data-driven approaches for policymaking create opportunities for evidence-informed policy and modernization of public sectors. **More data, more relevant data, more timely data, etc.** The nature and magnitude of these opportunities are difficult to assess for a number of reasons. Opportunities may differ between policy areas and between stages of the policy cycle, depending on the relevance of data sources, data technologies (such as data analytics and visualisation), privacy issues, skills, enabling ICT infrastructures, etc. Gradually, information is becoming available about data-driven approaches for policymaking at local and regional level and, to some extent, at national and international level. Still, much more has been written about data-driven approaches in private sectors such as retail, media, insurance companies and manufacturing firms.

The objectives of this explorative study

The European Commission has commissioned the Technopolis Group, the Oxford Internet Institute (OII) and the Centre for European Policy Studies (CEPS) to conduct an international study on innovative data-driven approaches to inform policymaking.

The main objective of the study is to explore the opportunities that innovative data-driven approaches offer for evidence-informed policy making, including the relevant data sources and technologies.

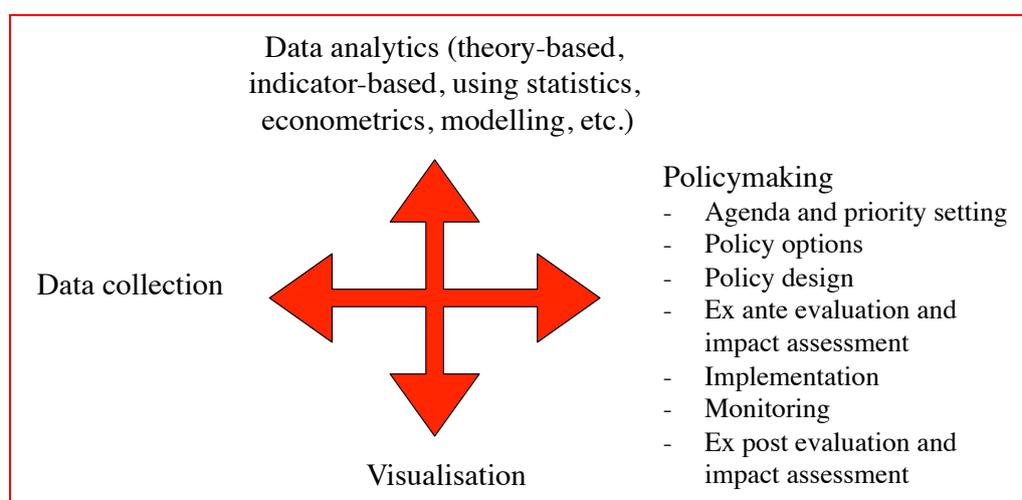
The study is action-driven and aims at the development of an Agenda for action for practitioners and other stakeholders (policymakers, public agencies, NGOs, companies that provide tools, collect data, etc.). To this end, it engages with interested parties and contributes to creating or linking relevant communities in the field.

Scoping the study...in a dynamic and broad field

This study explores a dynamic phenomenon. As a result, decisions were needed on the precise scope of the study in order to ensure its relevance and feasibility, as well as to build on its complementarity to existing studies.

The study covers four dimensions that constitute the focal areas for the analysis, i.e. **data collection and sources, data analytics**, the use of data in subsequent stages in the **policymaking cycle** and, to a lesser extent, **visualisation tools** (Figure 1).

Figure 1 Focal points of the study



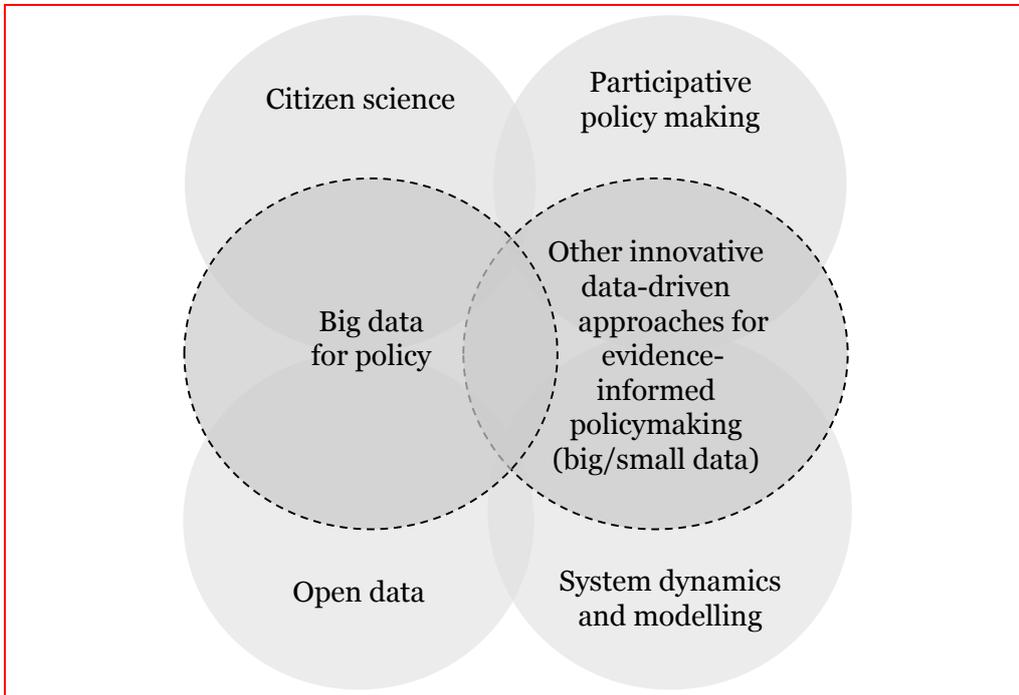
Technopolis, OII, CEPS (2015)

The study covers a broad range of policy areas because big data and other innovative data-driven approaches are relevant for all policy areas (transport, energy, education, security, health, research, taxation, etc.). Data-driven approaches also frequently cut across policy areas. For example, policymakers may combine datasets to explore the interactions between education, research, innovation, economic growth, etc.

Figure 2 illustrates the scope of the study. The study will focus on big data for policy but it also addresses innovative data-driven approaches that benefit from linking or scaling-up small data or that combine big data and small data. To some extent, the study

addresses four developments or building blocks that are related to the big data phenomenon: **open data**, **system dynamics and policy modelling** (e.g. complexity science, complex adaptive systems and agent-based modelling), **participative policymaking** and **citizen science**. The study only addressed these four developments when they are an important part of data-driven approaches for policymaking. For example, open data initiatives that start adding data analytics and where the results are used in policymaking can be relevant for the study. Along the same lines, some citizen science and participative policymaking can be data-driven, while system dynamics and modelling can be applied to policy-making.

Figure 2 Scope of the study



Technopolis, OII, CEPS (2015)

The study will address **policymaking at national and international level**, with particular attention for policymaking by the European Commission. As such, the study is complementary to studies on ‘big’ data used for or by smart cities, local police, fire departments, crowd control, etc.

Finally, the study covers **operational pilots, demonstrators and full implementations of data-driven approaches for policymaking**. 58 relevant examples are included in the inventory of relevant initiatives. However, it also looks at research projects in which new types of data are used and relevant tools are developed (the emphasis is on Europe’s Seventh Framework Programme).

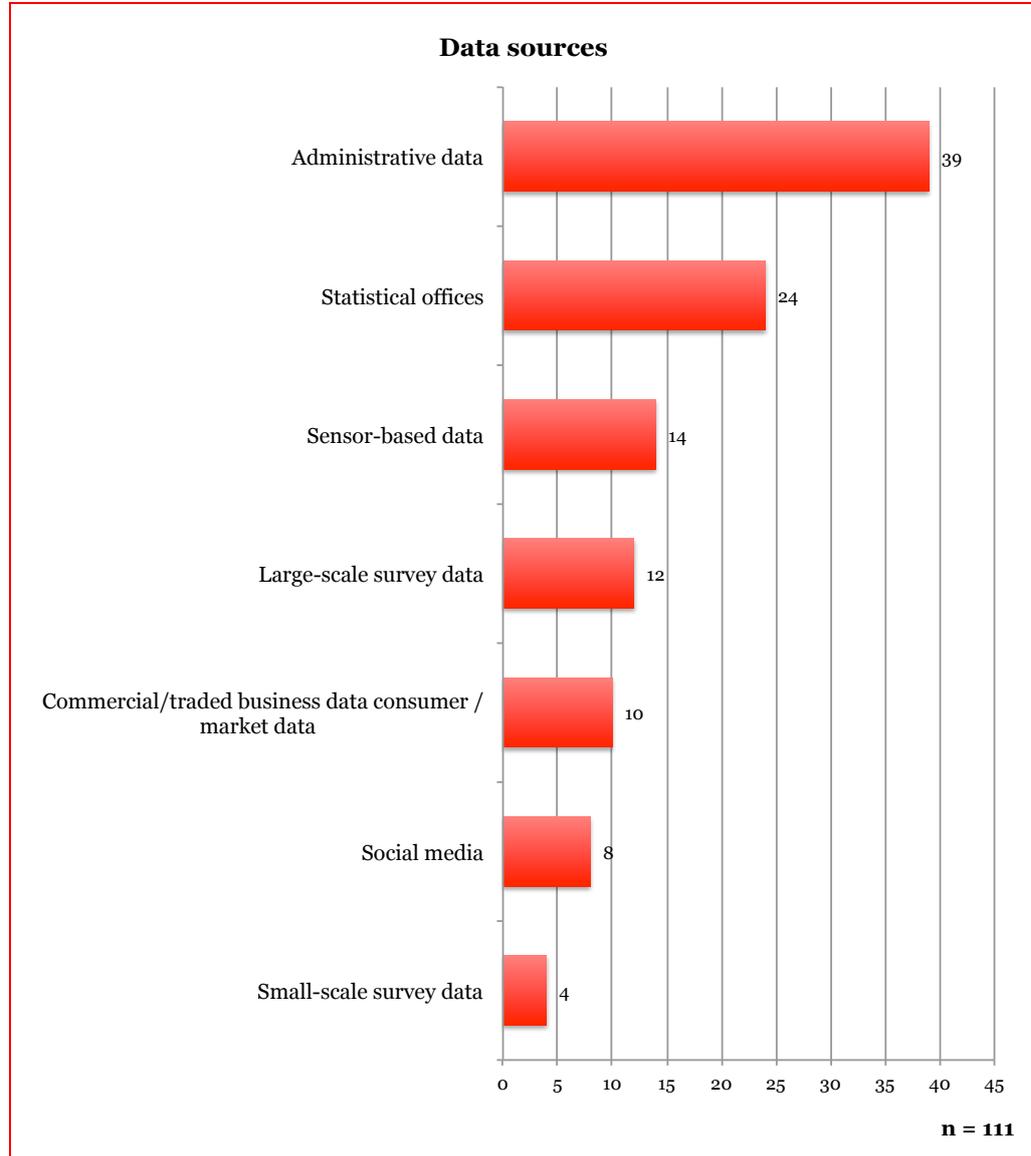
The next four pages present the main findings of the State-of-the-Art report.

Second half of 2015, the study proceeds with a number of case studies of data-driven approaches by the European Commission, development of one demonstrator and an international workshop.

Two main types of data

The **range of data sources** used for policymaking is increasing, while combining and linking data is becoming the norm. Figure 3 provides an overview of the data sources that were used in the 58 relevant initiatives that were identified during the study.

Figure 3 Data sources



Technopolis, OII, CEPS (2015)

The analysis revealed two main types of data. The first is the use of **public datasets** (administrative (open) data and statistics about populations, economic indicators, education, etc.), which are now used on a larger scale, used more intensively, and linked. Open data is widely promoted in the public sector and among NGOs, and there are now analyses of the quality of freely available open government datasets that finds that heterogeneity is still an issue even if quality is often good. Half the datasets among the initiatives examined were open or semi-open, with some kind of registration needed. The rest of the datasets are restricted for private or non-commercial use, e.g. use by public authorities in a specific policy area rather than all public authorities, stakeholders or the general public.

The second main type of data is from **social media, sensors and mobile phones** that are typically new data sources for policymaking, analysed with novel methods such as sentiment analysis, location mapping or advanced social network analysis. The importance of these new types of data is increasing and they are linked to established data sources such as official statistics, surveys and commercial/business databases.

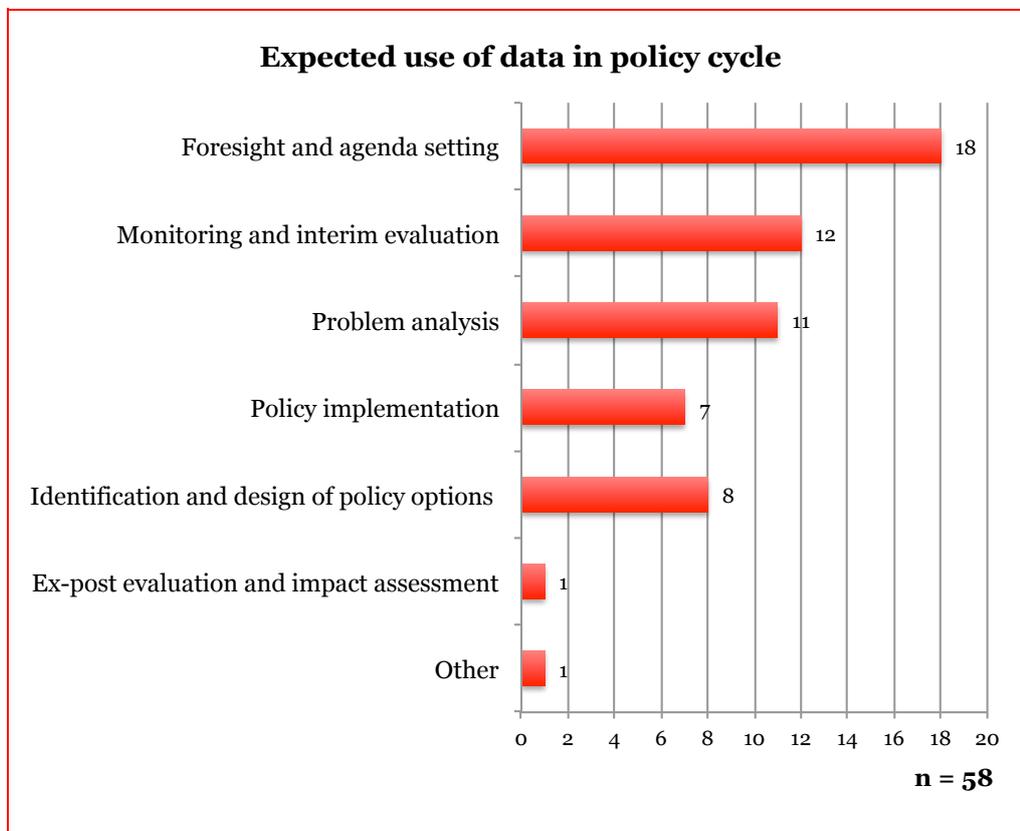
Privately held data is, currently, of less relevance. Private actors (market research, data brokers, data analytics, etc.) benefit from public datasets, while still investing in private datasets and bespoke data collection.

As such, a nuanced picture emerges with respect to ‘data as the new oil.’ Only some data are scarce. A more accurate metaphor would be to say that data offers new fields for harvesting, but some will bear better fruit than others, especially since when for so many datasets, the main task is cleaning and making data available.

Users and uses in the policy cycle

National governments and national and international agencies are becoming common users of innovative data-driven approaches. Based on our inventory of 58 initiatives, the most common uses include utilising data for **agenda setting** and **problem analysis** (e.g. measuring global priorities via Twitter and tracking traffic via sensors and mobile phone data), the use of open data for **transparency, accountability and enhancing participation** (with initiatives by policy makers and NGOs) and using administrative data and statistical data for **implementation of policy** and **monitoring** the output policy (Figure 4).

Figure 4 Expected use of the data in the policy cycle



Technopolis, OII, CEPS (2015)

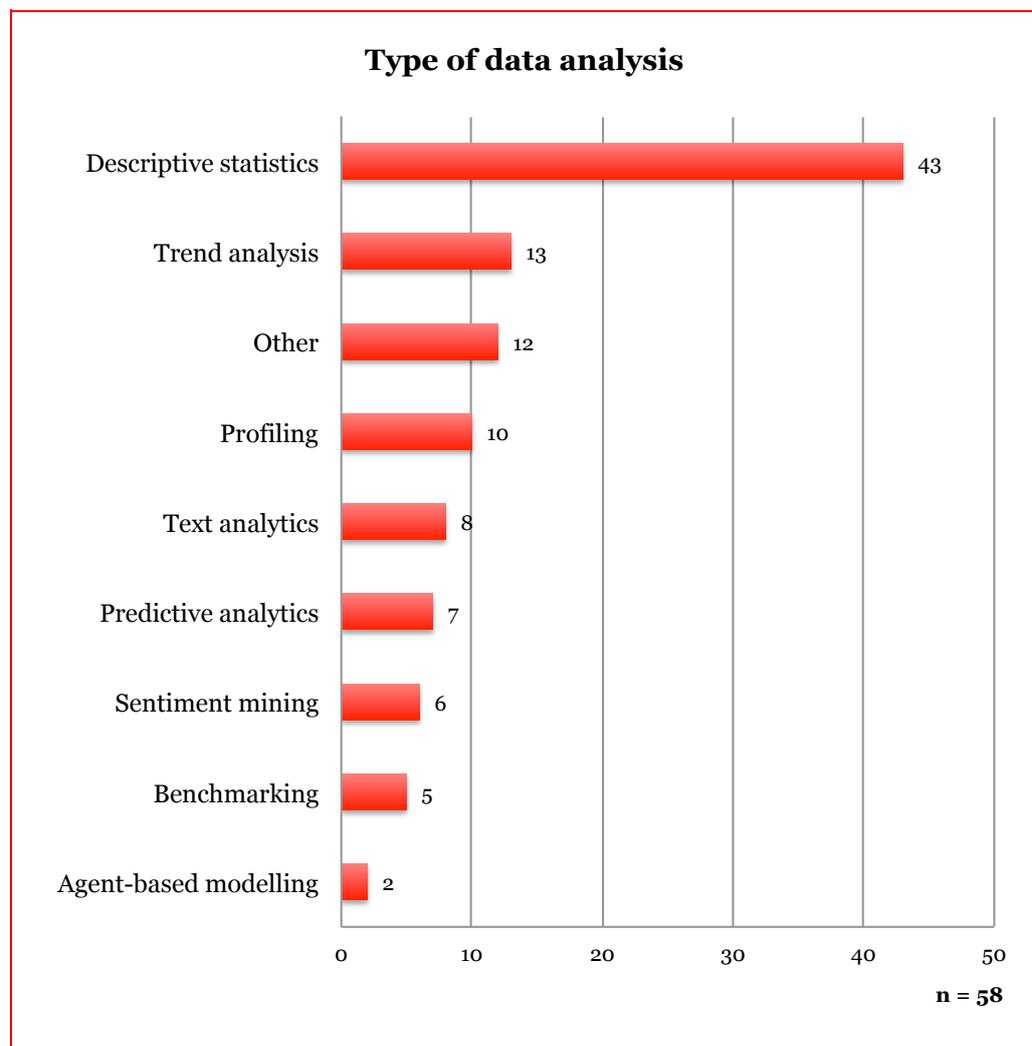
The inventory of 58 initiatives and a quick scan of relevant European research projects (and UN's Global Pulse) illustrate the opportunities for implementation of data-driven approaches across all stages of the policy cycle, including policy design, evaluation and impact assessment (ex ante and ex post).

In terms of policy areas, the inventory includes several initiatives in transport and mobility, environment, information society, budget, financial and economic affairs. 11 out of 58 initiatives concern the cross-cutting theme of transparency and accountability.

Data analytics and visualisation

The most novel analytics tools are in the area of social media analysis, network analysis and visualisation tools for mapping mobile phone locations. More important than the tools are ways in which they can be combined and refined and applied, and more important than new visualisations are the skills needed to use them (and not be misled). The opportunities for policymakers are substantial, as **a range of analytical tools has become available**. Still, most of the 58 initiatives **focus on descriptive statistics**. There is little use of advanced analytics or visualization techniques, which is mostly confined to academic research, public-private partnerships and NGO initiatives.

Figure 5 Data analysis



Technopolis, OII, CEPS (2015)

Privacy and inclusion

The concerns over privacy from large-scale data analysis have been much discussed, while **differentiating between datasets** (e.g. official statistics vs. social media data). To be sure, new mechanisms for more transparency in data collection are needed, as are new mechanisms for obtaining consent. The European Union's new proposed legislation will be a much-watched model as to whether data protection that is adequate to the novel big data environment can be provided. The new data protection proposals are aimed preventing harms and at the same time promote the Digital Single Market, and the main impact in relation to data for policy making can be anticipated to be reassuring the public that data are being handled appropriately. In the initiatives examined for this project, privacy and data protection played a role in only a small number of initiatives.

In the meantime, **inclusion concerns arise**. Certain parts of the population with no digital traces or uncaptured environments (e.g. less sensors) might be left out of data collection procedures. If so, their data and interests are not taken into account when using data across the policy cycle. The initiatives examined for this project appeared to provide little attention to inclusion concerns and mitigation strategies.

Known unknowns

There are advanced uses of data-driven approaches in the areas of **crime, fraud and terrorism** (although there are concerns that hierarchical, centralised organisations may not explore decentralised and innovative data-driven approaches). However, little is knowable about these largely secret initiatives, at least by means of a study of this type.

Another 'known unknown' is whether the data collection strategies and data analytics that are developed in research projects can be **implemented** successfully, taking into account costs, benefits, privacy, etc. Many tools are at the stage of technical tests and operational pilots. For example, there are a number of initiatives that gauge prices, monitor public opinion, and try to detect emerging crises.

Outlook

While consulting experts and discussing the draft report, a number of topics emerged, as the field of data-driven approaches for policymaking is developing rapidly:

- Concerns about the availability of relevant skills in public sector organisations, e.g. skills related to data collection, data analytics and interpretation of (visual) data.
- The risk that data-driven approaches can reduce transparency of the policy process when data collection and data analytics (e.g. algorithms, machine learning) are not fully understood and explained by policymakers and other stakeholders.
- The balance between collecting the data that is most crucial (given a policy issue, relevant factors and indicators) and using data that is readily available.
- Acknowledge that stakeholders influence the selection of data sources and tools. Big data doesn't reduce the incentives for stakeholders to present policy-based evidence.
- The continuous or even increasing importance of international collaboration on data standardisation/harmonisation related to skills, company data, air quality, etc.
- Opportunities to combine policy experiments (e.g. in specific countries or for specific target groups) and data-driven approaches for impact assessment.
- The need for (better) strategies to ensure that policymakers are informed about the tools that are being developed and piloted successfully in research projects.