# Innovation Policy-Making in the Big Data Era

There has been an explosion of interest in the potential of big data as a driver of better decision-making in many policy areas, including innovation policy. In this paper, we draw on the theory of innovation, and on Nesta's own experience[1] as an innovation agency in order to address the following research questions:

- *What are the characteristics of innovation policy that make it a suitable domain for the application of big data sources and analytical methodologies?*
- *What is the state of the art, and what are the emerging opportunities for the application of big data for innovation policy?*

In doing this, we seek to provide a firmer conceptual grounding for work in this area, and to set a vision for the development of big data applications addressing the needs of innovation policymakers.

We begin by identifying the main rationales for innovation policy: *market failures* linked to the fact that innovators often fail to fully capture the benefits of their investments, *systems failure* caused by gaps in the "system of innovation" that ought to connect innovation agents, and *inhibited emergence*, where a state of uncertainty about the future configuration of a market or technology field hinders its development [Gustafsson and Autio, 2011].

There are many policy options to remove these barriers to innovation, ranging from direct public investments on R&D to regulation and procurement [Edler et al., 2013]. Their design, implementation and evaluation has traditionally been based on data sources such as business and innovation surveys, administrative data, and metrics of scientific and technological output (academic publications and patents) [Fagerberg et al., 2006]. Three defining characteristics of innovation do however limit the usefulness of these data sources and outputs for innovation policymaking:

1. **Innovation involves novelty in inputs, processes and outputs:** it is associated with new capabilities, forms of organisation and industries which, by definition, are not captured by existing classifications of economic activity such as Standard Occupational Classifications (SOC) and Standard Industrial Classifications (SIC).
2. **Innovation is not confined to science and technology:** it may reflect changes in, say, business model, marketing or aesthetic: as a result, it is not always captured by traditional metrics such as academic papers and patents.
3. **Innovation is a complex, networked process:** it reflects a dynamic combination of resources and capabilities of many different agents and institutions. Measuring it requires combining data from a multitude of these sources. In turn, those who stand to benefit from access to data on innovation goes beyond policymakers, and encompasses investors, entrepreneurs and corporates, to name a few. However, in practice, most (aggregated, lagging) innovation data outputs are of limited relevance for these agents.

Big data can help overcome some of these challenges for innovation policymaking using conventional data inputs and outputs. Following Schroeder and Cowls [2014], we define big data as datasets of a volume, variety (complexity) and velocity unprecedented in the innovation policy domain, together with new analytical techniques and data outputs (such as data visualisations and interactive data platforms) used to analyse and create value from these data.

Big datasets (e.g. information provided by businesses on their websites) are often unstructured and closer to real-time than official data. This means that they can be used to identify new innovation areas as they emerge, even when these do not respect traditional occupational or industry boundaries. Some metrics that policymakers use to measure innovation are steeped in scientific and technological understandings of innovation; big datasets, by expanding possible sources of data, need not be so constrained. Big data is high-resolution (when it is based on publicly available data, it is often possible to

[1] e.g. http://www.nesta.org.uk/blog/big-data-better-innovation-policy

identify individual agents like businesses or investors in it in ways that official data, which is subject to non-disclosure constraints, is not). This makes it easier to republish it in interactive formats, say, that can be queried and exploited by a variety of innovation agents in addition to policymakers (this is manifest in the recent creation of a variety of online platforms to map innovative industries, clusters and ecosystems using, for example, publicly available data from Companies House).

Big data is however no panacea for policymakers, and its use in innovation policy presents serious methodological challenges. Nevertheless, there exists significant policy work done in the open science space in which to analyse and leverage [Royal Society, 2012]. The innovation landscape is constantly shifting, leading to the arrival of new data sources to study, and structural changes that can impact the reliability of algorithmic data collection and analysis. There is a risk that online data sources might offer a biased representation of innovation activity that privileges digitised industries at the expense of those trading in physical goods and services, and consumer facing industries at the expense of business-to-business sectors. Privacy is of course another critical consideration, especially where personal information is involved [House of Commons Science & Technology Committee, 2014], but there is an imperative for open and sharable data to provide the platform for effective (and transparent) policy-making.

Notwithstanding these important issues, we consider that there is significant scope for innovation in the use of big data for innovation policy – we conclude this paper by outlining some of the main opportunities, and setting out Nesta's future programme of research and platform development in this space:

1. Go beyond innovation maps based on SOC and SIC coding and official geographies, and make more use of unstructured data collection and (supervised and unsupervised) classification methods.
2. Complement descriptive analyses and sample-based inference with predictive modelling.
3. Explore the opportunities of real-time data and interactive data visualisation for innovation policy.
4. Combine big data sources with official and policy activity data in order to evaluate innovation policy impacts.
5. Develop standards for data-sharing so as to minimise the risk of fragmentation into incompatible platforms capturing disparate aspects of innovation activity.
6. Develop open datasets and transparent (as compared to "black box") methodologies for big data analysis.
7. Creatively explore the potential use of big data sources and methods for the study of industries which may not currently be data science-literate – and therefore less well catered for by online data – but are of great importance for policymakers, such as manufacturing.

## References

Robin Gustafsson and Erkko Autio. A Failure Trichotomy in Knowledge Exploration and Exploitation. *Research Policy*, 40(6):819–831, 2011.

Jakob Edler, Paul Cunningham, Abdullah Gök, and Philip Shapira. Impacts of Innovation Policy: Synthesis and Conclusion. 2013. Compendium of Evidence on the Effectiveness of Innovation Policy Intervention Project, Manchester Institute of Innovation Research, University of Manchester.

Jan Fagerberg, David C. Mowery, and Richard R. Nelson, editors. *The Oxford Handbook of Innovation*. Oxford University Press, 2006.

Ralph Schroeder and Josh Cowls. Big Data, Ethics, and the Social Implications of Knowledge Production. Paper presented at Data Ethics Workshop, KDD@Bloomberg, 2014.

Royal Society. Science as an open enterprise. 2012.

House of Commons Science & Technology Committee. Responsible Use of Data. 2014. Fourth Report of Session 2014-2015.