digitising europe initiative 2016

Spurring Big Data-Driven Innovation and Promoting Responsible Data Governance in a Privacy-Centred Europe

Background

The big data revolution and social impact

As personal interactions with computers and capabilities in data storage have evolved in the last four decades in scale and use, the term "big data" has been used to characterise the phenomenon of rapid global changes in personal data use. Originally framed as the "3 V's" (volume, velocity and variety) in the early 2000s, big data has emerged as an ecosystem of "3 C's": digital "crumbs" (digital translations of human actions and interactions captured by digital devices); powerful capacities to collect, aggregate and analyse data; and communities involved in generating, governing and using data, including data generators, end users, policy-makers, experts, privacy advocates and civic hacker communities.[1] Much of the societal usefulness of big data comes from discoveries in secondary or alternative uses of passively collected data from companies, such as the use of call details records (CDR) or location data for humanitarian response and disease tracking. Enthusiasts of the "big data revolution" underline how big data can help society spot socially valuable insights, unlock new forms of economic value in data and uncover human and social dynamics.

Addressing privacy and innovation

The rules governing the big data ecosystem have been a source of constant debate in light of widespread corporate and government use of data that counter an individual's right to privacy. In today's increasingly connected, "big data" world, the emergence of big data problematises several of our governing principles around data collection, sharing and consent; individual users—largely accustomed to ubiquitous use of data-emitting digital devices—are largely unaware of how and through which channels their data is used and processed, and the mechanisms used by companies to provide notice and consent (e.g. terms and condition agreements, privacy policies, etc.) have often failed to provide meaningful choice. While the economic and societal benefits of the big data revolution have been underlined by the EU Horizon 2020 programme and forthcoming initiatives under the Digital Single Market, the recent passing of the General Data Protection Regulation (GDPR) has major implications for the ongoing conversation on maximising growth while minimising risks to privacy. How can we make full use of data analytics in a responsible and human-centred manner?

---

[1] Letouzé, Emmanuel. "Big Data and Development Overview Primer". Data-Pop Alliance, SciDev.Net and the World Bank (2015)

The digitising europe initiative
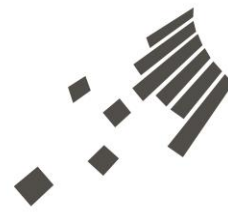
Through a series of regional stakeholder dialogues and debates in Berlin, Brussels and Madrid, the *digitising europe initiative* – organised by the Vodafone Institute for Society and Communications and Data-Pop Alliance – aims to provide insights for European companies and public sector institutions on initiating big data-driven projects and promoting responsible use and governance of big data for public good. The initiative convened academic experts, private companies, government civil society and the public from across Europe to discuss the applications and implications of big data for public good.

These efforts have culminated in this short position paper that both reflects the major takeaways from these events and provides insights for both the private and public sectors in unleashing big data-driven innovation and promoting responsible data governance for public good. This paper has been drafted for discussion during the final roundtable discussion in Dublin during a side event to the International Conference on Big Data for Official Statistics, organized by the CSO Ireland and the UN Statistics Division.

Highlights from the European roundtables

Each of the country roundtables highlight a specific dimension of big data-driven innovation: its potential, practice and promotion.

- **The potential of big data-driven innovation**: Bringing together German academic institutions, think tanks, businesses and other thought-leaders, the Berlin workshop and public forum focused on the possibilities and pitfalls of using big data analytics for economic growth and public good. The discussions in Berlin largely centred on issues of **anonymisation and data protection**; **objectivity and validity** of big data's claims; and the **diversity of expertise** needed to explore existing potential solutions and develop new tools and safeguards towards optimal data use.

  These events occurred against the backdrop of ongoing political discussion on the GDPR and EU legislation on data protection. Distinguished experts such as Alex 'Sandy' Pentland, MIT Professor and Co-Founder of the MIT Media Lab, and Andrew Keen, noted author and critic of the digital revolution, focused on the need for redefining the role of the consumer in the data revolution; exploring existing solutions and safeguards towards optimal data use; and improving society's understanding of the various risks and opportunities relating to big data and how to respond to them.

- **The practice of big data-driven innovation**: During the Brussels workshop and public forum, experts such as Kenneth Cukier, Data Editor of The Economist, and Dr. Linnet Taylor, Marie Curie Research Fellow at the University of Amsterdam, Giovanni Buttarelli, European Data Protection Supervisor (EDPS), and several others discussed how society can unleash innovation in the big data revolution while preserving privacy. These discussions focused on existing infrastructure and frameworks in place to address issues of **access, governance, and ethics**.

  With experts across government, academia and the private sector, this roundtable specifically highlighted the new role of the private sector in the responsible collection, governance and use of big data for societal purposes. Panelists discussed the need for all stakeholders to "do no harm" by co-creating responsible governance structures
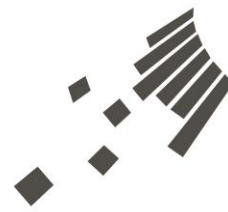
and test cases on data use; understanding current notions of privacy; and assessing both individual and group privacy implications of big data-driven innovation.

- **The promotion of big data-driven innovation**: The most recent events in Madrid focused squarely on forming new kinds of public-private-people partnerships involving big data, exploring the potential for these new partnerships to both promote innovation and collaboration, as well as foster collusion and surveillance. Organised in partnership with El País, experts such as Esteban Moro, Associate Professor at Universidad Carlos III de Madrid, and Marco Bressan, Chief Analytics Officer at BBVA, focused on the lessons learned from years of public-private-people data partnerships and projects – their formation, structure, maintenance and impact – and insights for companies and governments initiating these projects. Overall, these discussions focused largely on issues of **transparency, accountability, consent and literacy.**

  Panelists emphasised the need to redefine partnerships between the public and private sector around personal data; the need to set up clear and transparent internal governance rules, standards and procedures for companies (who gets access to which data and under which conditions); create shared language around big data use and communities with infomediaries; and invest in data literacy efforts and developing language for the public through media.

Across each of the country workshops, participants identified the following **knowledge gaps in the potential, practice and promotion of big data-driven innovation**:

1. **Unpacking new discoveries and possibilities of big data.** The need to help both companies and citizens understand the possibilities of the data revolution, through capacity-building and case studies of companies using data for social impact;

2. **Protecting consumer data.** The need to recognize the relationship between consumers and their own data, and help promote civic understanding of the data revolution through legal safeguards as well as the promotion of transparency-enhancing, privacy-preserving tools and platforms;

3. **Incentivising the private sector.** The recognition of the private sector as part of larger societal aspect of the big data revolution and consider criteria towards responsible data governance for social purposes;

4. **Encouraging greater collaboration towards understanding the complexity of big data ethics and operationalising ethical frameworks.** The need for multi-stakeholder collaboration and infrastructure to deepen conversation and understanding around "do no harm";

5. **Understanding "lessons learned" from existing data partnerships.** The need for further exploration and analysis of the emergence of public-private-people data partnerships, and their role in facilitating the evolution of principles and modalities around responsible data use.
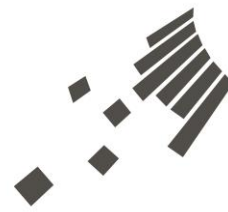
Action points and guidelines: towards a roadmap for initiating big data-driven innovation projects and partnerships

As echoed by the EU Horizon 2020, big data has increasingly been viewed as a critical lever of innovation for society. In the last seven years, several big data-driven innovation research projects have emerged involving companies sharing "private sector data" for public good through various cross-sector public-private partnerships. Though public-private partnerships are in no way a new phenomenon in governance innovation, the nature through which data (and its value) is shared, as well as the norms and practices around data use, will require setting new rules for collaboration and public-private business relations that emphasise societal benefits and lead towards scalable innovation in the public interest.

The following roadmap reflects the numerous insights from existing initiatives and participant discussions, and highlights the requirements and milestones for initiating big data-driven innovation projects.

Invest in big data internal knowledge and support

1. **Focus projects and resources on solving public problems with clearest ethical imperative and problem definition.** Identify issue areas and public problems in which your data can play an important role towards decision-making and/or resource optimisation. Kenneth Cukier during the Brussels public event described "not using data [as] the moral equivalent of burning books," and described several case studies in which big data could have been leveraged by the right actors. Additionally, look for "low hanging fruit" — data from your company could even play a small but critical role in a specific issue area rather than solve a much larger problem.

2. **Analyse current risks and tradeoffs in various forms of data sharing to make the case for internal support.** Evaluate existing case studies in which companies have previously shared similar data. What modalities did they use to share data? How have others handled privacy issues in using data towards this particular issue? What kinds of risks and tradeoffs would exist in the creation of this specific project? How can this project both make social impact as well as support the company's overall objectives? This could take the form of sponsoring or hosting a series of expert workshops, participating in larger forums on big data and social impact or soliciting help from external consultants or researchers who would be embedded into your organisation and teams. Ultimately, answering these questions internally is critically both for gaining support as well as building internal momentum and enthusiasm.

3. **Recognise organisational challenges and constraints in initiating innovation.** What resource constraints exist for your company? How much funding or support is available and what kind of projects will be possible?

4. **Invest in knowledge management and implementation of privacy engineering solutions.** Evaluate the existing legal framework around privacy and learn more through existing solutions how to incorporate privacy-preserving elements into your project.
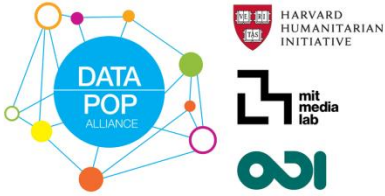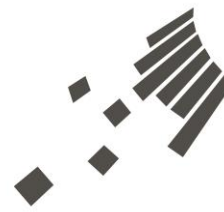
Invest in developing mutually beneficial big data partnerships
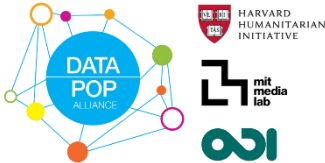
5. **Catalyse mutually beneficial partnerships with the right stakeholders with diversity of expertise and capacity for thought leadership.** Several participants stressed that many of the existing big data projects would not exist outside of years of building trust and taking steps (and risks) with the right stakeholders. Identifying the right stakeholders involved both evaluating expertise and thought leadership, as well as willingness to take on risks and process alignment.

6. **Focus partnerships on determining best use of private sector data to enhance existing traditional data.** Using data sources such as call detail records, transaction data and other new data sources will require some form of ground truth data from existing traditional sources. Identify partners that can also provide data or capacity to analyse data from public or open sources.

7. **Adapt features of existing ethical frameworks and guiding principles toward initial "do no harm" governance model.** While companies may aim to "do no harm" in the development of their projects, what guiding principles or ethical frameworks govern their use, and what happens when problems arise? While researchers continue to develop and assess ethical frameworks for data use (e.g. the Menlo Report), several companies have experimented with various models to incorporate ethical standards into their projects. This has involved the development of codes of conduct (in the case of BBVA, for example), organising ethical roundtables, and incorporating an ombudsman to oversee data sharing projects.

8. **Create and encourage mechanisms for public feedback and consultation.** Consult stakeholder groups to help inform the project design and provide opportunities for the public to give feedback.

Invest in meaningful civic engagement through multi-stakeholder big data literacy promotion

9. **Communicate project results through data visualisations.** Evaluate how to best communicate project results for both intended beneficiaries and the public, using infographics and data visualisations for example.

10. **Enable meaningful mechanisms for opt out.** Assess applicable notice and consent solutions and provide option for users to remove their data from use in the project.

11. **Promote opportunities for shared language and principles among stakeholders** through participation in cross-sector, cross-disciplinary collaboration, platforms and events led by intermediaries.

12. **Invest in long-term civic engagement and public understanding** of how data is being used through data literacy efforts through media, government and civic outreach.