

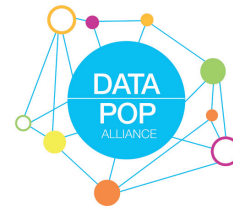


Vodafone Institute  
for Society  
and Communications



digitising europe  
initiative

In partnership with:



Digitising Europe Initiative  
Paper #1

# **BIG DATA AND PRIVACY**

## **UNDERSTANDING THE POSSIBILITIES AND PITFALLS OF THE DATA REVOLUTION IN GERMANY**

Key results from the Berlin Stakeholder  
Dialogue, November 2015

## Background

In the last decade, society has witnessed a digital revolution that has transformed the way in which we communicate, work, learn, and do business. As consumers participate within a technology-infused, sensor-laden society, their personal data follows along with them as well, often used and collected by businesses, public agencies and other entities. The Big Data Revolution helps us to spot socially valuable insights or unlock new forms of economic value in data. Numerous examples detail the impact that “Big Data” has had and will continue to have towards greater economic growth, innovation and public good.

Once 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 analyze data; and communities involved in generating, governing and using data, including data generators, end users, policy-makers, experts, privacy advocates and civic hacker communities.<sup>1</sup>

**This rapid data-driven development, however, raises some important political and ethical questions: How can we make full use of data analytics in a responsible and human-centered manner? Which forms of data use should be excluded, and who should set the rules?**

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.

The General Data Protection Regulation (GDPR), for which a political agreement was reached in December 2015, will apply to most or all digital interactions involving humans as a result of the broad definition of personal data, which also includes pseudonymous processing.

As the common standard for the EU, the GDPR aims to harmonize data protection in the member states and ease business activities for foreign companies. However, the GDPR remains vague across several elements, leaving room for interpretation and not always providing sufficient guidance for businesses and supervisory authorities. For instance, the rules applying to profiling based on Big Data are only vaguely articulated.

The underlying principles of the regulation involve informing the data subject and, at times, getting their consent, purpose limitation and data minimization. Critics note the limits of these policy tools in a Big Data world where the use cases often evolve after the actual collection of the data. Also, in view of the amount of data sources and networked nature of data processing, it seems inevitable that, while recognizing the need for transparency and reasonable choice, the emphasis will have to move from the actual collection of data into ensuring responsible and accountable use of data.

The passing of this legislation has major ramifications for the ongoing conversation on maximizing growth while minimizing risks to privacy. Given the EU’s strong market and political power, the outreach of the regulation will go far beyond Europe with implications for global data transfer. In addition, many data protection legislation around the world have been

---

<sup>1</sup> Letouzé, Emmanuel. “Big Data and Development Overview Primer”. Data-Pop Alliance, SciDev.Net and the World Bank (2015)

based on the old directive (D-95/46/EC) to allow data exchange with the EU. With the years to come it is to be expected that many legislations will be amended according to the new regulation.

Germany has a strong historical awareness and appreciation for data protection. In 1983, the German court ruling (later forming the base line of German data protection laws and also of the European Data Protection Guideline), stated that “in a networked world, there are no trivial data.” Given its historical awareness of data protection (stemming from the Nazi and the Stasi era) and its political and economic power, Germany constitutes an interesting case study for the discussion on Big Data and privacy.

As the first event paper in the digitising Europe’s series, this event paper captures the major key themes emerging from our events in Berlin in November 2015.

The focus of the Berlin events was on exploring the possibilities and pitfalls around the data revolution and identifying key insights and practical solutions for facilitating the use of data while protecting the privacy of citizens.

## Event themes and reflections

The Berlin workshop and public forum focused on the possibilities and pitfalls of using Big Data analytics for economic growth and public good. Bringing together German academic institutions, think tanks, businesses and other thought-leaders, the expert workshop focused on the ongoing political discourse in Berlin surrounding the elements framing the GDPR and EU legislation on data protection. During the evening public forum, Alex “Sandy” Pentland, MIT Professor and Co-Founder of the MIT Media Lab, and Andrew Keen, noted author and critic of the digital revolution, engaged in a debate on the motion: “Will the data revolution improve our lives?”

Participants in both the expert workshop and public forum highlighted the following key themes in the current discourse on the data revolution in both Germany and across Europe:

- 1. Redefining the role of the consumer in the data revolution;**
- 2. Exploring existing solutions and safeguards towards optimal data use; and**
- 3. Improving society’s understanding of the various risks and opportunities relating to Big Data and how to respond to them.**

## 1. Redefining the role of the consumer in the data revolution

Renegotiation has been essential in major social transformation and innovation processes and similarly, consumers need greater participation in the use and sharing of their data in the data revolution. During the debate, Pentland pointed to the U.S. introduction of global land-use satellites for agriculture, railroads, electricity and the internet as examples of historical cycles of technological innovation, public discourse and renegotiation on risks, and long term societal benefits. Following the deaths of over one million people in India as a result of famine, the U.S. invested in land-use satellites to observe global crops as a form of early warning system to prevent future famines. The implementation of this new technology was extremely controversial in Germany where policymakers and citizens saw these actions by the U.S. as threats to national security and a guise for global spying. However, this debate led to

much needed discussion on both how to control for the risks generated by the new satellites as well as the impact of their use: effective global decision-making for allocating resources for vulnerable populations. The impact of public innovation projects such as the railroad and electricity were both shaped by public demand and behaviour – forms of public renegotiation – and these originally controversial new technologies have created significant societal benefits.

In a 2008 – 2009 World Economic Forum report on mobility in a networked world, Pentland argues for what has become known as the “New Deal on Data,” a renegotiation of the relationships between companies, citizens and their own data that involves gaining “key rights over data that are about them.”<sup>2</sup> Similar to the essential human right to govern your own body, citizens should be in control of their digital identities as well.

**“Greater citizen participation in the use and sharing of their own data not only helps protect against privacy violations, but as discovered by recent research, encourages greater citizen sharing of their own data towards social causes.”<sup>3</sup>**

This renegotiation highlights the need for legally and institutionally enshrining a system of specific individual control over each piece of personal data. Such a proposal flows naturally from both existing policies on data rights and privacy protection which echo a need for a kind of user consent that is more fine-grained and more informed.

## 2. Exploring existing solutions and engineering new tools and safeguards towards optimal data use

Big Data – through new data sources, capacities and communities – helps give evidence for answering key questions on the social experience. In the use and analysis of “data crumbs” on how people actually behave, Big Data helps provide resources towards a scientific and systematic exploration of the human social experience. Pentland’s book *Social Physics* describes how data from communication patterns can predict the creativity and productivity of a group; through Big Data analysis, the book emphasizes that creativity does not emerge in the siloed, individual activity of an intelligent individual, but rather through the combination of multiple, diverse actors. Pentland’s work helps recognize the role of social activities in financial decision-making as well, highlighting the role and influence of the social interactions in addition to individual choice assumptions of behavioural economics.

A diversity of expertise is needed to explore existing potential solutions and develop new tools and safeguards towards optimal data use. Europe faces a multitude of socio-political and environmental crises; the use of Big Data in the data revolution provides opportunities towards greater efficiency and better decisions with scarce resources. Experts have already begun discovering how data can technically, safely and legally be used.

<sup>2</sup> See also Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Eds.). (2014). *Privacy, big data, and the public good: Frameworks for engagement*. Cambridge University Press.

<sup>3</sup> Vodafone Institute for Society and Communications (2016): *Big Data*. <http://www.vodafone-institut.de/bigdata>

The value of consent remains a debated topic, particularly given the development of other transparency-enhancing, privacy-preserving solutions and cited over-use of “notice and consent” as a tool for effective policy. Below are two recent examples of privacy engineering solutions emerging from the MIT Media Lab that represent advancements in using and mining data in parallel with privacy preservation.

### **Protecting metadata: OpenPDS2.0 SafeAnswers (SA).**

OpenPDS2.0/SA. is a platform for individuals and organizations to manage existing metadata in their own secure Personal Data Stores (PDS), allowing users to “collect, store and give fine-grained access to their data all while protecting their privacy.”<sup>4</sup> In terms of data use, anonymization of high-dimensional data has proved to be rather difficult to achieve while retaining data quality.

SafeAnswers represents the key privacy and use innovation in the platform, in which applications run code against the data and receive answers calculated against the metadata, rather than attempt to anonymize the data itself.

### **Using homomorphic encryption to analyse data without sharing: MIT’s Enigma Project.**

MIT’s Enigma is a privacy-preserving platform allows for entities to run computations on private data while keeping the data completely private.

Enigma “removes the need for a trusted third party,” cryptographically guaranteeing that users can safely share data without risks to their privacy. The technology has the potential to help “companies analyse consumer data without giving employees access to individual customers’ personal information or letting loan applicants submit data for automated underwriting without ever sharing their information with a human being.”<sup>5</sup>

## **3. Improving society’s understanding of the various risks and opportunities relating to Big Data and how to respond to them**

Pentland and Keen both acknowledge the failure of the data revolution in fostering an environment conducive for collective public discussion. As Pentland noted, social media can be helpful in generating panics and spurts of raised concerns, but largely has driven conversations towards extreme, polarized positions. Keen also described social media culture as narcissistic and self-focused, failing to live up to the potential of the data revolution as a communications revolution fostering consensus-building for rational decision-making. It is significant “that even if data might provide clear evidence for better policies or actions, the impact of Big Data may be diminished by the meaning (or lack of meaning) created by society.”

<sup>4</sup> <http://openpds.media.mit.edu/>

<sup>5</sup> <http://www.fastcompany.com/3054976/elasticity/mit-new-blockchain-project-enigma-wants-to-let-you-share-your-data-on-your-terms>

Both data literacy and transparency are needed for driving and enhancing social discourse. Ongoing debate persists on how much onus should be on either data collectors and their transparency, or consumers and their understanding of the implications behind their digital interactions and activities.

Keen largely criticized the role of companies and their positive impact in the data revolution, underlining how several companies are entrenched in business models that disincentivize them from taking responsibility in increasing transparency and making privacy relevant. However, strong case examples such as the Orange Data for Development Challenge and BBVA's Innova Challenge have emerged where companies have experimented in new "social responsibilities," particularly in sharing data with the public sector for social purposes. Vodafone has also initiated similar activities with the public sector. Most recently, Vodafone has helped local government in the UK with their transport infrastructure planning by sharing anonymous and aggregated statistics about movements of groups of people, derived from network data, and an opt-out for those who wished not to participate. All users were informed about the activity.

As companies continue to navigate their role in the data revolution, transparency in their efforts remains essential. A recent study from the Vodafone Institute highlighted European user uncertainty and skepticism of the benefits derived from the use of their information as a part of Big Data initiatives. Surveying more than 8,000 European digital users, the Vodafone Institute study suggests that users are "skeptical of the Big Data phenomenon because public and private organizations are failing to explain clearly how and why their data is analysed, and do not give them adequate control over how their data is being used."<sup>6</sup>

## The key themes from the Berlin event highlight:

**1.**

the need to improve the relationship between consumers and their own data through both legal safeguards as well as the development of transparency-enhancing privacy-preserving tools and platforms;

**2.**

the need to evaluate the ethical and social roles of the European corporate sector in participating in the European Big Data revolution and consider criteria towards responsible data governance for social purposes; and finally,

**3.**

the need to help both companies and citizens realise their roles and responsibilities in the data revolution through capacity-building and toolkit solutions.

The second event in the series took place in Brussels and focused on the role and participation of the corporate sector in the data revolution, in which Kenneth Cukier, Data Editor of The Economist, debated Dr Linnet Taylor, Marie Curie Research Fellow at the University of Amsterdam, on the motion: "Key to the future or the end of privacy – how will Big Data impact society?"

---

<sup>6</sup> Vodafone Institute for Society and Communications (2016): Big Data. <http://www.vodafone-institut.de/bigdata>

## About the project

This event paper was commissioned by the Vodafone Institute for Society and Communications as a part of its digitising Europe initiative, a high-quality forum for cross-sector dialogue on Europe's digital transformation. Launched in December 2014 with an inaugural keynote address by German Chancellor Angela Merkel, the initiative brought together business leaders, startup founders, and policymakers from across the continent to discuss data-driven transformation of education and the world of work.

In the second phase of this initiative, the Vodafone Institute for Society and Communications has partnered with the Data-Pop Alliance to launch a comprehensive series of European stakeholder debates and dialogues in order to analyse the impact and implications of the 'data revolution' for European companies and consumers in the digital economy.

This new series aims to investigate the social, economic and ethical dimensions of the data revolution with its stakeholders – businesses, governments and people – and identify both practical solutions and policy recommendations for shaping a people-centered, growth-driven data revolution for European companies and citizens.

Through hosting cross-sector expert workshops and public dialogues across the continent, this series invites experts across Europe to evaluate the impact of Big Data analytics for society, with the intention to further deepen the level of discourse and broaden public participation on the benefits, risks and responsibilities characterizing the European data revolution.

In 2016 digitising Europe will host a series of workshops and public debates in European capitals. Each of the digitizing

Europe events will be summarized as event papers from the initiative's flagship cities, followed by a comprehensive publication on the broader European context, which will be presented in summer 2016. Please note that this paper, as well as future publications, in this series do not necessarily represent the official position of the Vodafone Group plc.

## About Data-Pop Alliance

Data-Pop Alliance is a global coalition on Big Data and development created by the Harvard Humanitarian Initiative (HHI), MIT Media Lab and Overseas Development Institute (ODI) that brings together researchers, experts, practitioners and activists to promote a people-centered Big Data revolution through collaborative research, capacity building, and community engagement.

[www.datapopalliance.org](http://www.datapopalliance.org)

## About the Vodafone Institute

The Vodafone Institute for Society and Communications explores the potential of mobile and digital technologies to improve political, social and economic participation and to give better access to education. The Institute is a think and do tank that fosters dialogue between science, business and politics. It initiates projects and research, and publishes reports as a source of practical recommendations for decision makers. Through events and social media communications the Vodafone Institute provides a platform for debate.

[www.vodafone-institut.de](http://www.vodafone-institut.de)

## Imprint

**Authors** David Sangokoya and Emmanuel Letouzé, Data-Pop Alliance **Editor**  
Vodafone Institute for Society and Communications, Behrenstraße 18, 10117 Berlin, Germany  
**Chairman of the Advisory Board** Matthew Kirk **Board of Directors** Ingrid M. Haas,  
Thomas Holtmanns, Dr Mark Speich **Project Lead** Dr David Deissner **Editorial Management**  
Friedrich Pohl **Research Support** Vivian Weitzl **Layout** Interpol<sup>+</sup> Studios, Hamburg, Berlin

🏠 [www.vodafone-institut.de](http://www.vodafone-institut.de)

f [www.facebook.com/VodafoneInstitute](https://www.facebook.com/VodafoneInstitute)

🐦 @vf\_institute

©Vodafone Institute for Society and Communications, January 2016