POLICIES FOR BIG DATA

In the second part of their article, **ANDREW HAIRE** and **VIKTOR MAYER-SCHÖNBERGER** discuss controls for adverse uses of big data and list the current forums for action

It is clear given the powerful qualities of big data and the likelihood that it will shape all sectors of the economy, and considering its significant dark sides, that policymakers at all levels will want to play a role in influencing its trajectory. But what should they focus on? We suggest four areas of regulatory involvement.

ENSURING PROTECTION

The most obvious is to ensure effective protection of individuals' privacy. As we have discussed, current mechanisms of privacy protection will become increasingly ineffective in the context of big data. This is not only problematic because it potentially harms affected individuals; it is also detrimental to the acceptance and use of big data, because without sufficient societal trust users will not be able to operate. It is not just in the interest of society, but in the interest of all responsible users of big data to ensure that effective mechanisms protecting privacy are in place.

What is needed is an additional and more effective protection mechanism. Recent work undertaken by a group of privacy experts from around the world points towards a regulatory mechanism that would shift the focus of privacy protection from informed consent at the point of collecting personal data to accountable and responsible uses of personal data. The core idea is that with such a mechanism in place users of personal data would have to evaluate the privacy harms and implications of a potential new use of such data and what safeguards would need to be put in place to reduce the privacy harms before this use could commence.

And while this assumes that big data users would have to evaluate their intended applications, incorrect evaluations and insufficient implementation of safeguards would not only lead to civil and criminal liability. The mechanism also foresees a well-resourced privacy regulator with the expertise and power to enforce such a use-based privacy protection mechanism.

The advantages of such an additional mechanism are clear: privacy protection would not rely on the mystical ability of individuals to fully comprehend the complex uses of their personal data at the moment of collection, and data users could not forego the implementation of stringent privacy safeguards by pointing towards rather formal 'consent of the data subjects'. And enforcement would not depend on individuals suing data users (which we know very rarely if ever happens), but rely on much more powerful privacy regulatory agencies with sufficient resources and stamina to regulate and enforce even against the most powerful data users. In return, data users would be permitted to reuse personal data for novel purposes (and thus unleash its power) as long as a comprehensive privacy assessment had shown that it would produce minimal privacy risks.

Of course, such an additional mechanism would not solve all privacy challenges, but we suggest that a focus on responsible and accountable data use will go a long way in addressing some of the most troubling privacy challenges.

TAMING PROBABILISTIC PREDICTIONS

Probabilistic predictions, the operational outgrowth of big data analyses, can be tremendously useful. They reduce uncertainty and risk and help big data users and by extension society to prepare for the future by better decision-making in the present. But probabilistic predictions also pose unique policy challenges, especially when they are used to decide who to punish or hold responsible, based only on predictions. For instance, if a government uses predictions to decide which individual to put under surveillance or police heavily, not because of past behaviour but based on prediction, such a policy would rightly be viewed as infringing dangerously on personal freedom.

Regulatory authorities are therefore well advised to put in place clear restrictions on how and for what purpose government agencies can use predictions. There must be a bright red line that interdicts such misuses.

Uses by government agencies as well as commercial entities of predictions of future behaviour that result in negative treatment, quasi-punishment or the withholding of benefits granted to others, while not prohibited per se, must (we suggest) meet strict scrutiny. This includes providing transparency into the data analysis, as well as the guaranteed right to affected individuals to disprove the prediction.

PROVIDING BIG DATA EXPERTISE

Transparency and the right to disprove predictions will only be usable for the public if individuals

do not have to confront complex analysis themselves, but can avail themselves of big data experts who are obliged to help. We envision a new cadre of such experts – the 'algorithmists'. Specially trained, they would take vows of impartiality, confidentiality and professionalism, much like civic engineers or doctors.

Individuals who believe they have been mistreated because of false big data predictions could contact algorithmists, who in turn would investigate and render a decision. They would also help individuals in disproving predictions if an individual believes such a prediction is wrong.

Algorithmists could also advise data users on how to best implement transparent, disprovable predictive decision-making, and how to ensure responsibility and accountability in their predictions.

Algorithmists would have special big data expertise, which includes statistical and technical training, but would also be well versed in ethical considerations and the legal and regulatory constraints in place.

KEEPING DATA MARKETS FLUID

We have focused on the role of regulatory authorities to defend and enforce the rights of the individuals but there is another, equally important, dimension that is not directly related to these rights. As some data markets are becoming more concentrated over time, and more data is held by fewer and fewer commercial entities, ensuring competition in the data economy becomes paramount. Otherwise big data may face the same fate as steel manufacturing and railways in the late 19th century in the US. The concentration of power of these industries in few hands gave rise to the first effective antitrust and competition legislation in the world, and to the recognition that government plays a role in ensuring powerful, market-stifling trusts do not form, and where they have formed they are busted.

Ensuring competition in data markets can take a variety of forms. The most obvious is for data holders to be forced to let others access their data holdings under fair and reasonable terms. Such 'fair, reasonable and non-discriminatory' (FRAND) licensing (as it is popularly termed) has been routinely used in certain areas of patent protection, and shown to be effective. Moreover, the US government has in recent years in a number of cases already used a FRAND licensing mandate to constrain data holders' power after they had acquired large datasets.

The advantage of this approach is not only that the mechanism has already been tested and found to be effective, but that it is well known to competition authorities and so easier to get it employed. Moreover, such a mechanism is using market competition to reduce the power of large data holders, which is much preferable to more limiting restrictions or market interventions.

Some experts have gone one step further and suggested that for data markets to truly function well, one needs to put in place a legal exclusion right for data, much like we already have in place for intellectual property. Whether such a right is truly needed, and what its features and limitations would be, we cannot answer. It is important, however, to note these experts' opinion in this context.

> **"Ensuring competition in the data economy becomes paramount."**

ORGANISATIONS AND REPORTS

BIG – Big Data Public Private Forum

In Europe, the Big Data Public Private Forum (BIG – see big-project.eu) is working towards the definition and implementation of a strategy for research and innovation, and it is giving a major boost for technology adoption and supporting actions for the successful implementation of the big data economy.

In addition, each year various government and private sector entities meet to exchange their views on projects of importance at the European Data Forum (2014.data-forum.eu). The forum is designed to capture a larger umbrella of views by examining open data, linked data and big data. This year's forum included work on open data in the transport and communications sectors in Finland, public sector information at the European Commission, the European single digital market, and predicting parking supply to satisfy demand in a smart city.

World Economic Forum

The WEF acknowledges that a new approach to handle data is necessary to protect the rights and wellbeing of individuals. A report published in 2013 lays out three subthemes:[1]

- From transparency to understanding: people need to understand how data is being collected, whether with their consent or without – through observations and tracking mechanisms given the low cost of gathering and analysing data.
- From passive consent to engaged individuals: too often the organisations collecting and using data see their role as a yes-no/on-off degree of consent. New ways are needed to allow individuals to exercise more choice and control over this data that affects their lives.
- From black to white to shades of gray: the context by which data is collected and used matters significantly. How is the data used; much like money, it means little until it is used.

To achieve a level of trust during the flow of data at least five issues were discovered about the data: protection, accountability, empowerment, transparency and respect.

The WEF has also released 'Risk and responsibility in a hyperconnected world', which focuses on the malicious intent to disrupt or capture information (data) in both the private and public sectors.[2]

European Union

The European Commission (EC) adopted a communication on the data-driven economy as a response to the European Council's conclusions of

October 2013, which focused on the digital economy, innovation and services as drivers for growth and jobs and called for EU action to provide the right framework conditions for a single market for big data and cloud computing.

As part of its big data strategy, the EC and Europe's data industry will invest €2.5 billion in a public-private partnership that aims to "strengthen the data sector and put Europe at the forefront of the global data race". Data protection and pseudonymisation mechanisms are said to be among the priorities, to strengthen data privacy. In March 2014, the European Data Protection Supervisor published 'Privacy and competitiveness in the age of big data: the interplay between data protection, competition law and consumer protection in the digital economy'.[3]

International Telecommunication Union (ITU)

The big data approach taken by ITU so far focuses on the following areas and questions.[4] To address these increasingly important issues, reports are being prepared as well as workshops and dedicated sessions in ITU events:

Standardisation[5]

- Which standards are required to facilitate interoperability and allow technology integration in the big data value chain?
- Which definitions, taxonomies, secure architectures and technology roadmaps need to be developed for big data analytics and technology infrastructures?
- What is the relationship between cloud computing and big data in view of security frameworks?
- Which techniques are needed for data anonymisation for aggregated datasets such as mobile phone records?

Regulation

- What are the key regulatory issues at stake and how can and should big data be regulated?
- How does big data impact the regulation of privacy, copyright and intellectual property rights (IPR), transparency and digital security issues?
- What is the link between big data and open data?
- Is there a need to regulate data management and service providers?

ICT data collection and analysis

- How can big data complement existing ICT statistics to better monitor information society developments?
- Which type of data from ICT companies are most useful and for which purposes?
- What are key issues that need to be addressed, and by whom, in terms of collecting and disseminating big data in telecoms?
- What is the role of national statistical offices and how can big data complement official ICT data?

UN Global Pulse

This was launched by the UN secretary-general to respond to the need for more timely information to track and monitor the impacts of global and local socioeconomic crises. It explores how new digital data sources and real-time analytics technologies can help policymakers understand wellbeing and emerging vulnerabilities in real-time, to better protect populations from shocks:[6]

"The initiative was established based on a recognition that digital data offers the opportunity to gain a better understanding of changes in human wellbeing, and to get real-time feedback on how well policy responses are working. The overarching objective of Global Pulse is to mainstream the use of data mining and real-time data analytics into development organisations and communities of practice."

US: big data initiative

Initiatives to fund research on six big data themes were announced by the president two years ago. They aim to:

- Advance technologies needed to collect, store, preserve, manage, analyse and share huge quantities of data
- Harness these technologies to accelerate the pace of discovery in science and engineering, strengthen national security, and transform teaching and learning
- Expand the workforce needed to develop and use big data technologies.

This year, the Executive Office released two reports. The first, 'Big data: seizing opportunities, preserving values', is a comprehensive treatment.[7] Among its scope is preserving privacy values (both in the US and through interoperable global privacy frameworks); recognising schools as an important sphere for using big data; and preventing new modes of discrimination that some uses of big data may enable. The second report, 'Big data and privacy: a technological perspective', examines the nature and evolution of technology and the challenges for protecting individual privacy.[8] It concludes with the notion that technology alone cannot protect privacy – policy needs to play a strong role and reflect what is technologically feasible. There are five policy recommendations:

- Focus on the use of big data, and less on collection and analysis
- Avoid embedding technological solutions into policy
- Strengthen US research in privacy-related technologies and social science
- Encourage education and training opportunities in privacy protection
- The US to take the lead both internationally and at home by adopting policies that stimulate the use of practical privacy protecting technologies.

This report was debated by privacy advocates, who said it relies on policy aimed at the use of data, and not as much on its collection. Critics thought there could be discrimination against the poor, older people, minorities and even children – those not in a position to protect themselves – by placing the burden of protection on the individual.

ANDREW HAIRE *is a policy and strategy advisor based in Washington, DC and* ***VIKTOR MAYER-SCHÖNBERGER*** *professor of internet governance and regulation at the Oxford Internet Institute. This article and the first part in the autumn edition are based on a longer paper presented at the 2014 ITU Global Symposium for Regulators.*

REFERENCES

1 World Economic Forum (2013). Unlocking the value of personal data: from collection to usage. bit.ly/19yopil

2 World Economic Forum (2014). Risk and responsibility in a hyperconnected world. bit.ly/1dYeLMZ

3 European Data Protection Supervisor (2014). Privacy and competitiveness in the age of big data: bit.ly/1o0JvIH

4 ITU (2014). Measuring the information society.

5 For further information on the work on big data carried out by the ITU Telecommunication Standardization Bureau (TSB), see: bit.ly/1qiVvAP

6 UN Global Pulse. unglobalpulse.org

7 Executive Office of the President (2014). Big data: seizing opportunities, preserving values. 1.usa.gov/1kyOreK

8 Executive Office of the President (2014). Big data and privacy: a technological perspective. 1.usa.gov/1rTipM2