Wilton Park

Image: Kts | Dreamstime.com

Report

Safeguarding rights in the big data revolution

Monday 13 – Wednesday 15 June 2016 | WP1478

In partnership with:

Institute for Human Rights and Business

Shaping Policy • Advancing Practice • Strengthening Accountability

With support from:

ERICSSON

Wilton Park

Report

Safeguarding rights in the big data revolution

Monday 13 – Wednesday 15 June 2016 | WP1478

The Wilton Park conference brought together international experts from a range of countries including Finland, Germany, Italy, India, Netherlands, Sweden, Norway, UK and US, with backgrounds in international organisations, government, business, civil society and academia to discuss the challenges of safeguarding rights in the use of big data. The conference explored the opportunities and risks of big data in the ICT sector and beyond and considered practical steps business can take to ensure rights are respected.

Discussion included an exploration of the complexities of big data, the opportunities to help realise human rights and the challenges data controllers face regarding ownership, consent, transparency, accountability and trust. The conference also explored the prospect of developing a set of human rights based principles in order to embed privacy considerations into company practice, particularly in the way data is collected, stored, processed and shared.

Reliance on technology has resulted in a data explosion. We create and release data about ourselves and our activities every minute of every day. Big data is a great enabler and has the potential to advance human rights. But there are also poignant questions to consider. What do privacy rights mean in the age of big data, the internet of things and increased surveillance around the world? How to protect other rights such as non-discrimination, freedom of association and freedom of expression?

Billions of devices are connected to the internet and producing data from all over the world. In the current environment only a very small percentage of data is analysed but this is set to increase. In this fast moving environment, there is an urgent need for discussion about ways in which to implement policies and processes to safeguard rights.

Key points

- It is tempting to be distracted by the infinite possibilities and complexities of technology, big data and the internet of things. However, the positive and negative impacts of these developments on people should continue to be a key focus. The focus should be on inclusion and dignity, ensuring individuals are free from discrimination and are able to enjoy the right to privacy.
- A small number of high profile companies are actively engaged in discussion about the implications of big data and associated responsibilities. However, there are many more companies actively mining and selling data which are not yet part of this debate. Their business practices will have an increasingly significant

impact on people's rights.

- Privacy discussions tend to focus on the individual. These should expand to include collective rights, as information about individuals is frequently used to make decisions about groups or communities.
- It is unclear how the majority of algorithms work, so decisions about their application are frequently unchallenged. People may not know they have been treated unfairly, or been discriminated against and on what grounds. This allows limited access to remedy.
- Rather than asking 'who owns the data?', a more valuable question could be: 'who owns the insights into individual data and the value it holds'?
- It is not desirable to put all the onus on an individual to make decisions regarding their data. While user control and consent is very important, data controllers need to mitigate risk and protect the environment which the user inhabits.
- Any future rights based principles relating to the behaviour of companies should first address the relationship between business and user and could be based on the framework of the UN Guiding Principles on Business and Human Rights.

"The commitment of the UN Sustainable Development Goals to "leave no-one behind", can be supported by technological advancements"

Understanding the opportunities and risks of Big Data

1. The enormous potential for big data to help people and save lives places a responsibility on both government and business. The commitment of the UN Sustainable Development Goals to "leave no-one behind", can be supported by technological advancements[1].
2. The advent of big data has generated many insights and conclusions that could not be realised from smaller quantities of data. The increase in data collection and better methodologies, along with advances in processing, allows computers to can gain more insights than was previously possible.
3. There are multiple discussions about the opportunities and risks of big data and many examples to illustrate both aspects. For example, sensors could be embedded into a car seat to record the posture of the driver. On the positive side, the data could be used as an anti- theft device or could identify driver fatigue. However, this ability to identify individuals with a high degree of certainty, also gives companies access to an unprecedented level of personal information.
4. Similarly, 'smart' meters provide information which allows better control of energy use, thereby managing scarce resources. However, this information about power use, domestic habits and occupancy, could also be used for criminal gain.
5. Large datasets can be created not just for the individual but also on a societal level. For example, an analysis of online search queries could assist in identifying people who are looking up early symptoms of particular diseases, leading to early and more effective diagnosis. However, this same methodology could also be used by governments to identify the browsing history and interests of people they may regard as dissidents. Whilst data can be used to protect people it can also be deployed to invade privacy with a negative impact on other rights.

"Technology has always been used in ways other than intended and, to some extent, this 'dual use' catalyses innovation"

6. This medical example also raises ethical questions: is there a moral obligation on data analysts to notify people of their concerns regarding searches for early symptoms? In known projects, researchers did not notify as the data was anonymised.
7. Technology has always been used in ways other than intended and, to some extent, this 'dual use' catalyses innovation. But it can also be the door to misuse. A unique aspect of big data is that it is compiled and analysed from many different sources to provide insights. Once data is joined together, it will provide new insights which are

[1] UN Sustainable Development Goals http://www.un.org/sustainabledevelopment/

likely to be used for reasons not originally imagined. And once data is joined up, it can be hard to unwind it.

8. Big data promises much, and there are some assumptions that traditional research methodologies are now redundant. However, it is a mistake to believe that inputting a large amount of data into a "black box", and applying an algorithm will produce the "right" answer without a human check on the analysis. A large amount of data doesn't automatically produce valuable insights; the relationship between causation and correlation should continue to be scrutinised. Robust research methodologies are still necessary, dependent on well framed questions at the beginning of the process. Interrogation of data in a meaningful way requires the investment of time, collaboration and hard work.

"it is easy to be distracted by the infinite possibilities and complexities of technology"

9. Big data and the internet of things can be overwhelming and it is easy to be distracted by the infinite possibilities and complexities of technology. However, an understanding of the human impact should be at the core of discussion, risks mitigated, and benefits explored and enabled. There is a concern that amidst experimentation people are being reduced to numbers, statistics and subjects rather than individuals with agency.

10. The focus should be on inclusion and dignity, ensuring people are free from discrimination and are able to enjoy the right to privacy. In order to understand possible negative impacts on people, it is useful to think the unthinkable and explore uncomfortable possibilities in order to imagine worst case scenarios.

Understanding personal and non-personal data

"all non-personal or machine data is potentially personal data"

11. The majority of data is generated by and handled at machine level. Most data will never be seen by a human eye. However, all non-personal or machine data is potentially personal data. For example, if a pacemaker communicates with an insulin pump, it is machine and also personal data. Similarly, a fingerprint used to start a car, connects non-personal and personal data to the activity. An oil company gathering global data is non personal, but the resultant discovery of new oil fields, is likely to have a high impact on people. These examples illustrate that data not considered sensitive or personal could be used to extract sensitive information which may later be deployed with high impact on individuals.

12. It has been said that the distinction between personal and non-personal is misleading. It is, therefore, unhelpful to regulate by types of content, rather it should be applied to the purpose for which the data is used.

13. In a human rights context, accountability rests on three pillars: protect, respect and remedy. When machines make decisions, it difficult to know who is responsible or accountable and for what. Regulated industries will need to consider these implications, particularly with regard to eg machine-learning in medicine.

Privacy in the age of big data

14. There are distinctions between the concept of online and offline privacy. For example, observable facts such as height, hair colour and race, can be private or undisclosed on the internet. Whilst personal privacy on and offline may differ, the same rights based principles should apply.

"Anonymisation is a useful tool but not a silver bullet"

15. Anonymisation is a useful tool but not a silver bullet with regards to privacy protection. Anonymised information about groups can be used to make decisions about individuals. For example, in the US, many people use 'Uber' taxis rather than calling more expensive ambulance services. However, this information could be sold to insurance companies which may result in higher premiums based on the assumption that the area or user is high risk.

16. There are concerns that the EU General Data Protection Regulation (GDPR) does not fully address big data and the internet of things. For example, there is dissatisfaction

regarding the issue of consent and the definition of "legitimate interest" in the processing of personal data[2] and anonymity.

"The tendency to 'hoard' data can constitute a challenge to privacy"

17. The tendency to 'hoard' data can constitute a challenge to privacy. Data controllers should exercise caution against the desire to collect and keep as much data as possible. There may be a tendency to overestimate the benefits of retaining data and to underestimate the negative. While there may be beneficial secondary uses, data minimisation is a worthwhile principle to be preserved.

Consent

"What does free, prior and informed consent mean in big data terms?"

18. It is difficult to determine the ultimate direction of the debate over consent, but it is an important component in the overall discourse on protection and respect for privacy. What does free, prior and informed consent mean in big data terms?

19. The debate could be advanced by the principle question- 'consent to what and with whom?' Companies are increasingly collecting data as a core part of their business. This includes sectors which may not have institutional knowledge or experience of appropriate collection, storage and use. For example, some healthcare apps may not have appropriate restrictions on data collected without user knowledge, compared to pharmaceutical companies, which have well established restrictions and protocols.

20. Company terms and conditions are notoriously lengthy and opaque and the majority of users do not read them prior to acceptance. This could be considered a broken consent model and it is apparent that new models are needed. Is it realistic or worthwhile to have privacy policies for every collection of data and use? Will this simply lead to a scenario whereby users will click 'accept' on policies without reading them, thereby maintaining a flawed status quo? People want fast services: how can they be convinced to take more time and consider what they are consenting to? Is there a digital equivalent to speed bumps which would introduce some caution by slowing down the transaction?

"Is there a digital equivalent to speed bumps"

21. Consent is often bundled into one issue, which can encompass use of a lifestyle website, to iris scanning of refugees in Syria. As there are different levels of personal information and use, should there also be different levels and safeguards on consent? Just because data is sensitive does not mean it cannot be used, but there should be higher levels of concern and related scrutiny.

22. The key to consent may lie elsewhere than in vague privacy policies or user permissions for each of the many services and uses of personal data. An alternative could be an obligation on companies to inform users what has been done with the data, for example in the last month. This would demonstrate trustworthiness and may do much to promote customer loyalty to a service or brand. However, for it to be effective, users would need a channel or instrument whereby inappropriate or unauthorised usage could be challenged.

Ownership of data and data insights

23. The concept of privacy is associated with maintenance and preservation of the self. But ownership is ambiguous. For example, a person with a prosthetic limb (which can be detached, bought and sold) will have a different sense of 'ownership' than that taken towards a real limb. This can help to illustrate the nuances which may be applied to the cross cutting issues of ownership.

"Does ownership rest with the individual, or the organisation that collected the data"

24. Does ownership rest with the individual, or the organisation that collected the data about the individual? The prevailing view is that the individual should own personal data, however this may not always be the case.

25. It is undesirable to assume that the trade off for efficient access to innovative products

[2] EU General Data Protection Directive (GDPR), Article 6 http://www.eugdpr.org/

and services should be the surrender of fundamental rights and associated protections.

26. An individual may not give information directly, but inferences and insights about that person can be made when data is compiled and analysed from different sources. Data portability is one of the central components of the GDPR, but a data download in itself, doesn't provide the insights and the decisions made about the person based on that data. A more valuable question about data ownership is: 'who owns the insights into personal data and the associated value?'.

27. Data is not solely about the individual. Individual data is used in a collective way in order to make decisions about other people. Community ownership of data and privacy as a communal right should also be addressed.

28. It has been recognised that people want to control their own data. Small companies have responded by developing products and services which allow people to take control. This can foster a mutually reinforcing opportunity for growth and innovation[3].

29. There is an imbalance between data controller and data subject. The user needs to be empowered to have control over their own data. However, it is not desirable to put all the responsibility on the user. The user should be protected not because of the nature of their decisions, but because the overall environment is safe.

Data chains

30. Discussion on the "supply chain" of data, included: the need to respond to a lack of visibility; the roles and responsibilities of companies and the consumer; what opportunities there may be to give consumers more ownership of their data; and what is realistic and achievable.

31. There are opportunities to raise awareness and educate people about what happens to their data. Automated processes could be deployed to increase user control. For example, a user could be alerted when an organisation wants to use their data (via a "data angel") or metadata could be tagged so a user could follow where it goes.

32. The solution should not put the onus on the individual- a trace on user data could be used by intermediaries to hold data controllers to account.

33. While these solutions may be realistic, there is a danger they may only work for those least at risk and not the most vulnerable, and therefore may not create systemic change. These solutions tend to focus on individual/consumer rights rather than collective rights.

Transparency

34. It is a concern that big data analytics and the use of algorithms can lead to decisions being made about people of which they have no knowledge.

35. It is unclear how the majority of algorithms work, and decisions may not be challenged. People may not know they have been treated unfairly, or been discriminated against and on what grounds.

36. For example, a social networking site with a 'real name' policy reportedly used an algorithm to identify users using fake names. As a result, Native American names were flagged and user accounts suspended because the algorithm did not recognise the formation of Native American names (noun + adjective).

"The use of algorithms in predictive policing is controversial"

37. Algorithms focus on propensity, from an individual's shopping habits to the potential of an individual to commit a crime. The use of algorithms in predictive policing is controversial. On the one hand, police need quick and available information to enable them to take action and respond. However, information from algorithms may trigger

[3] Facebook (2016) A New Paradigm for Personal Data? https://www.ctrl-shift.co.uk/news/2016/06/21/facebook-a-new-paradigm-for-personal-data/

deployment of police to patrol areas considered crime 'hotspots'. This may create further hostility in areas with a troubled relationship between police and community, thereby hindering efforts to build trust.

38. Big data often infers that the 'real answer' will materialise, without needing a human to oversee. However, these examples illustrate the need for a feedback loop which allows human intervention to improve or correct the system. Without this human input, the algorithm may perpetuate the underlying problem.

"information which is incomprehensible or impenetrable does not necessarily constitute transparency"

39. There is a need for transparency in order to ensure informed consent, but this could result in a flood of information. Furthermore, information which is incomprehensible or impenetrable does not necessarily constitute transparency.

Transparency of algorithms

40. One of the transparency challenges is that multiple diverse parties may be contributing to an algorithm. Elements may change as developers seek to improve the algorithm and there may be intellectual property issues. There are also many different kinds of algorithms eg machines that deal with nuclear plants will differ from those designed for traffic management.

"Artificial Intelligence (AI) is not monolithic"

41. Artificial Intelligence (AI) is not monolithic: there are different styles that can be applied to identify mistakes. Projects should have cyber proofing built in to the concept: simulations on how the machine might behave and what could go wrong. Smart monitoring capabilities are needed to warn if an algorithm is crossing a line.

42. There is limited protection for researchers who try to probe algorithms, leading to legal challenges. For example, a diabetic security researcher in the US hacked his own insulin pump and was subsequently prosecuted under the DMA because he circumvented private software.

43. The GDPR is a recognised example of European regulation and illustrates the need for regulators to have the competence to analyse algorithms.

Ethics and trust

44. Without user trust, the full potential of big data cannot be realised. Ethics and trust are fundamental for companies' relationships with customers/users. It would be useful to focus on voluntary reciprocal relationships rather than compulsive elements of rights and regulations.

45. Ethics can be defined as an accepted set of values across a community, relevant to cultures and communities. Trust is often based on confidence in adherence to a shared view of 'doing the right thing'.

46. Terms and conditions should be fair and consistent and seen by companies as the basis of a trust relationship with the service user. In practice, will companies be nervous about liability and regard them as legally limiting, or will they interpret them in a reasonable and open manner?

"companies need incentives"

47. What might be entailed in an ethical risk assessment? There are high costs associated with effective assessments, so companies need incentives in the form of a clear business case. Short-term costs would be compensated by the long-term benefit of increased user trust, based on transparent processes.

48. It is important for companies to realise that customers do not expect their data to be used out of context. There are many examples of companies using customer data contrary to human dignity, eg the dating website OK Cupid purposefully sent users on dates with people that were 'bad' matches as a behavioural experiment[4]. Facebook

[4] http://www.bbc.co.uk/news/technology-28542642

allowed academic researchers to manipulate news feeds in order to alter moods[5]. Uber assumed that customers who ordered cars between 7-9am at the weekend had been on a one-night stand[6].

Remedy

49. There is currently limited remedy for mistakes made in big data analytics which have impacted a person's rights resulting in discriminatory practices such as refusal of mortgage or problems in accessing services eg healthcare. It is also extremely difficult to find out if an incorrect decision has been made, particularly in an environment where it is increasingly easy to 'experiment' on people with regard to data analytics with no liability structure in place when things go wrong. How to get to a position where individuals know about violations of their rights and what to do to ensure there is a remedy?

Big data challenges in humanitarian work

50. Humanitarian activity provides many examples of the complexities of big data. In a crisis situation, humanitarian agencies often have difficulty obtaining data, particularly in places that do not have a population census or a functional statistic office. In many cases, data collected or aggregated may not be used due to slow analysis.

51. Agencies work with small, often messy, datasets. A simple standard for aligning data would ensure that robust small data contributes to qualitative insights subsequently drawn from big data.

"Humanitarian experts are generally not data scientists"

52. Humanitarian experts are generally not data scientists and it would be useful to address this skills gap to ensure more effective handling of data in an accountable way. The humanitarian community could learn much from the private sector regarding data protection and it would be useful to explore ways in which corporates could assist further.

53. The elements of a data policy or data-risk framework could be[7]:

    - Assessment: Where is the data controller located?
    - Data inventory: What type of data is held and where is it being stored? What is in the data?
    - Risks and harms- Are lives at risk? Could organisational reputation be put at risk following a data breach?
    - Counter-measures- What is needed to improve the model?

"questions asked of the data are under-developed"

54. Organisations and agencies regularly request information from telecommunications operators. However, many of these requests are generated by people who are not data scientists: the inherent challenges are often not fully understood and the questions asked of the data are under-developed. There are currently no specific frameworks for operators to share insights about humanitarian needs. How can public and private sector better collaborate?

55. Telecommunications operators need to ensure compliance with license agreements on the use of customer data, including for humanitarian purposes. This regulatory framework needs to be understood by NGOs and academics seeking the data. In contrast, internet companies are less regulated and have permissions to use data for

[5] http://www.bbc.co.uk/news/technology-28051930

[6] http://www.marketplace.org/2014/11/18/business/final-note/ubers-data-makes-creepy-point-about-company

[7] See also OCHA (2016) Building Data Responsibility into Humanitarian Action https://docs.unocha.org/sites/dms/Documents/TB18_Data%20Responsibility_Online.pdf

multiple purposes written in to their extensive privacy policies.

56. Before operators can carry out a research project they need: proof of concept; approval from the national regulator and boundaries set on whether data can be taken out of the country or shared with partners; to address privacy, regulation and national security issues; and arrange peer reviews from researchers and the scientific community so that research can be published. From start to finish, a project can take 2-3 years.

"The outcome of analytics is only as good as the data source"

57. Data analytics are hard to do and the source of information is important. The outcome of analytics is only as good as the data source. For example, an analysis of the spread of disease through call data records is meaningless in the absence of information on: the disease; how it is spreading; and reliable intelligence on incidences of outbreaks. Multiple data points are needed in order to attain the impacts. These can be achieved through a collaborative approach, but all parties need to understand and appreciate the challenges faced by other partners.

58. The majority of data is retained for commercial purposes, rather than in response to humanitarian or crisis needs. There is, therefore, caution and concern about companies collecting data for one purpose, such as call data records, and then converting to another, such as monitoring outbreak of disease. Whilst a company whose core business is not the tracking and treatment of disease, may justify data collection for this purpose, there is a risk that data protection principles will be forgotten.

"what are fundamental rights in the digital space?"

Developing a set of human rights big data principles relating to the behaviour of companies

59. Given the speed and scale of developments, the development of and commitment to human rights big data principles needs to be addressed as a matter of urgency. The key questions include: what are fundamental rights in the digital space?; and what principles to focus on? The principles need to be understood by a range of stakeholders, including start-ups and 'millennials' (those born between 1980-2000).

60. In this context a human rights based approach should be grounded in the fundamental question: how to enable positive usage and outcomes while ensuring that people are protected, particularly the vulnerable, those who are most marginalised, excluded or experiencing discrimination?

"Principles should incorporate the UN Guiding Principles on Business and Human Rights"

61. Principles should incorporate the UN Guiding Principles on Business and Human Rights[8] (UNGPs) as these constitute a recognised common standard. The UNGP approach to risk provides a practical starting point to address mitigation of risk to individuals. This approach could help drive a global standard. It is important that principles consider risks to people, not companies. Mitigate the risk to an individual, and mitigation of corporate risk will follow.

62. The UNGPs refer to risks in a general sense without defining their nature. It would be useful to specify risks as they relate to particular scenarios, in order to design appropriate controls. Principles should be tested against specific scenarios.

"Principles should not relate solely to the ICT sector"

63. Principles should not relate solely to the ICT sector. Big data impacts many different sectors, so principles should be developed to ensure relevance and resonance with a broad cross section.

64. A set of Principles would focus on the relationship between business and user. Other relationships should be taken into account (eg business to government and government to business) but as a starting point, and in order for them to work, data principles must be user focused. In the short term, a set of principles could help to reach the longer-term goal of good data governance.

[8] UN Guiding Principles on Business and Human Rights
http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

65. The UNGPs are not a perfect fit, in part because the concept of consumers and individuals is not at their core. However, in developing a new set of business-customer principles, it is important that the onus should not be on the user. The design should not place the overall responsibility on individuals to opt out and consent. Companies should take affirmative action to mitigate risk.

66. There is value in a set of principles which are relevant not only for companies but also for regulators and other organisations that collect and process data. This approach would also serve to educate users of their rights.

"consider ways of avoiding 'principle fatigue'"

67. There should be incentives for companies to sign up to and apply principles. These are lacking from the current system. It is also important to recognise that there are already numerous principles on the global policy scene and to consider ways of avoiding "principle fatigue".

Proposed policy recommendations

Integrate privacy from the beginning. Don't make privacy a last minute check-point, design/integrate from the beginning, not just into the product but the process. A company needs to act before something negative happens.

Conduct due diligence: In order to understand possible impacts on people, think the unthinkable and be put in an uncomfortable position to imagine what could happen.

Practice good data governance: It is important to understand the underlying aim. A knowledge of the landscape and the associated risks will guide appropriate development of effective data governance. It is good practice to: define environments and constraints; write relevant policies, apply them, and check them on a regular basis. There should be no data governance without audit. An organisation needs to measure and assess governance, ideally through an independent body.

Test new technologies: Consider if testing methods are fair. For example, is it ethical to offer lower insurance premiums if a user allows access to personal data?

Transparency: If a company is willing to tell customers what data it holds, then it should do so. Respect for privacy should be recognised as a competitive advantage over companies who cannot or will not tell customers how their data is used.

Create a feedback loop for human insights into algorithms.

Context: It is important for companies to realise that customers do not expect their data to be used out of context If companies are using information out of context, users need to know and additional consent sought.

Accountability: Be clear about who is accountable, responsible and informed about risk. Challenge complacency by ensuring Board level commitment and involve the whole organisation, from the engineering stage all the way through to product delivery.

Lucy Purdon
Wilton Park | July 2016