United Nations A/72/540

General Assembly

Distr.: General
19 October 2017

Original: English

**Seventy-second session**
Agenda item 72 (b)
**Promotion and protection of human rights: human rights questions, including alternative approaches for improving the effective enjoyment of human rights and fundamental freedoms**

# Right to privacy*

## Note by the Secretary-General

The Secretary-General has the honour to transmit to the General Assembly the report of the Special Rapporteur of the Human Rights Council on the right to privacy, Joseph A. Cannataci, submitted in accordance with Human Rights Council resolution 28/16.

* The present report was submitted after the deadline in order to reflect the most recent developments.

17-18460 (E) 021117

# Report of the Special Rapporteur of the Human Rights Council on the right to privacy

*Summary*

The present report is divided into two parts: the first provides an executive summary of activities undertaken during 2016 and 2017; the second contains the interim report on the work of the TASk Force on Big Data and Open Data established by the Special Rapporteur of the Human Rights Council on the right to privacy.

## Contents

# I. Overview of activities of the Special Rapporteur on the right to privacy 2016–2017

1. The 2016–2017 period has been a particularly hectic time for the mandate of the Special Rapporteur, involving engagements with civil society, Governments, law enforcement, intelligence services, data protection authorities, intelligence oversight authorities, academics, corporations and other stakeholders, involving 26 events in 15 countries and four continents. These engagements took the Special Rapporteur to over 30 different cities, some in Asia, North Africa and Central America, with 25 per cent of his engagements in the United States of America and over 50 per cent in Europe.

## A. Draft international legal instrument on surveillance and privacy

2. Security and surveillance were important issues leading to the creation of the mandate of the Special Rapporteur on the right to privacy by the Human Rights Council in 2015.

3. The mandate of the Special Rapporteur on the right to privacy, set out in Human Rights Council resolution 28/16, clearly states the duty: "To identify possible obstacles to the promotion and protection of the right to privacy, to identify, exchange and promote principles and best practices at the national, regional and international levels, and to submit proposals and recommendations to the Human Rights Council in that regard, including with a view to particular challenges arising in the digital age".[1]

4. The vacuum in international law with regard to surveillance and privacy in cyberspace, the core substance of the Snowden revelations, has been identified by the Special Rapporteur as a serious obstacle to privacy, and this subject is currently his primary concern. The Special Rapporteur is of the opinion that it is not only the lack of substantive rules that are an obstacle to the promotion and protection of privacy, but also the lack of adequate mechanisms.[2]

5. Under his mandate, the Special Rapporteur would strongly recommend that the Human Rights Council support the discussion and adoption, within the United Nations, of a legal instrument to achieve two main purposes:

(a) To provide Member States with a set of principles and model provisions that could be integrated into their national legislation, embodying and enforcing the highest principles of human rights law, and especially privacy when it comes to surveillance;

(b) To provide Member States with a number of options to help plug the gaps and fill the vacuum in international law, in particular those relating to privacy and surveillance in cyberspace.

6. While the need for such a legal instrument is clear, its precise scope and form are as yet unclear. Whereas the substance of its contents is emerging clearly from ongoing research and stakeholder consultations, the best vehicle for achieving these purposes has not yet been determined.

7. It has long been recognized that one of the few areas in which the right to privacy cannot be absolute is that of the detection, prevention, investigation and

[1] A/70/53, sect. III. part A, resolution 28/16, para. 4 (c).

[2] Report of the Special Rapporteur on the right to privacy to the Human Rights Council, March 2017 (advance unedited version available online, see A/HRC/34/60).

prosecution of crime, as well as in the area of national security. Preservation of democracies, however, requires checks and balances to ensure that any surveillance is undertaken to protect a free society. Prior authorization of surveillance and the subsequent oversight of surveillance activities are key parts of the rules, safeguards and remedies needed by a democratic society in order to preserve its defining freedoms.

8. The above-mentioned report of the Special Rapporteur to the Human Rights Council in March 2017 contained interim conclusions for a legal instrument regulating surveillance in cyberspace complementary to existing cyberlaw, such as the Convention on Cybercrime (the Budapest Convention) adopted by the Committee of Ministers of the Council of Europe in 2001. A pre-existing initiative, the European Union-supported Managing Alternatives for Privacy, Property and Internet Governance (MAPPING) project, is exploring options for a legal instrument regulating surveillance in cyberspace. A draft text, which is being debated by civil society and international corporations, will be made public before the spring of 2018.

9. The process is described in more detail in supporting document V.[3]

## B. Letters of allegation

10. Some of the letters of allegation sent by the Special Rapporteur to Governments related to surveillance will be published by the Office of the High Commissioner for Human Rights (OHCHR) in line with the communications reports of the special procedures mandate holders.

## C. Other letters: public domain; Japan

11. On 18 May 2017, the Special Rapporteur published a letter to the Government of Japan (see supporting document III).[4] In that letter, the Special Rapporteur expressed his concern about the shortcomings of proposed legislation, which allows surveillance without the necessary safeguards, ostensibly in order to permit Japan to ratify the 2000 United Nations Convention against Transnational Organized Crime. Attempts at engagement over this matter continue and will feature in the report of the Special Rapporteur to the Human Rights Council in March 2018.

## D. Other ongoing initiatives related to surveillance

12. There are other initiatives which the mandate is exploring on surveillance, security and privacy. If appropriate, details will be made public at a later stage.

## E. A better understanding of privacy

13. The Special Rapporteur is analysing privacy, inter alia, as an essential right, enabling an overarching fundamental right to the free, unhindered development of one's personality. The Chair of the Task Force on Privacy and Personality, Elizabeth Coombs, former Privacy Commissioner, New South Wales, Australia, has kindly accepted to undertake this work, with a special focus on gender and privacy.

[3] See www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.
[4] See www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

14. More information on the activities carried out by the Task Force is available in supporting document IV.[5]

## F. TASk Force on Health Data

15. The Special Rapporteur's TASk Force on Health Data has commenced its work under the leadership of Dr. Steve Steffensen of the United States of America. Consultations are expected to take place in the spring and summer of 2018.

## G. Use of personal data by corporations

16. The Special Rapporteur has continued to work on business models, privacy in the corporate use of personal data, both independently and within the MAPPING project, in the build-up to the launch of the Special Rapporteur's task force on the subject with time frames announced at the website of the Special Rapporteur (http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/ThematicReports.aspx).

## H. Official country visits

17. The following country visits have been undertaken or are planned: United States (19–28 June 2017),[6] France (confirmed dates, 13–17 November 2017); United Kingdom of Great Britain and Northern Ireland (confirmed dates, 11–17 December 2017); Germany (confirmed dates, 29 January to 2 February 2018); and the Republic of Korea (confirmed dates, 3–15 July 2018).

## I. Resourcing

18. Only the official country visit to the United States and the travel of the Special Rapporteur and other speakers to Hong Kong, China, for the International Conference of Data Protection and Privacy Commissioners and its discussions on personality and flows of information in Asia were financed under the budget for the mandate of the Special Rapporteur managed by OHCHR. The other visits have received extramural funding, largely from the hosts of the related events.

# II. TASk Force on Big Data and Open Data

19. The TASk Force on Big Data and Open Data established by the Special Rapporteur is led by David Watts.[7] The lead authors of the present report are David Watts and Vanessa Teague.[8] The members of the TASk Force, many of whom also contributed to this report, include Christian d'Cunha (European Data Protection Supervisor for the European Union), Alex Hubbard (of the office of the Information Commissioner of the United Kingdom), Professor Wolfgang Nejdl (University of

[5] See www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

[6] The final report on the official country visit to the United States of America is expected to be published in the spring of 2018: The end-of-mission statement is available at: http://www.ohchr.org/Documents/Issues/Privacy/SR_Privacy/VisitUSA_EndStatementJune2017.docx.

[7] David Watts is Adjunct Professor of Law at Latrobe University and at Deakin University. Until 31 August 2017 he was Commissioner for Privacy and Data Protection for the State of Victoria, Australia.

[8] Vanessa Teague is a Senior Lecturer in the Department of Computing and Information Systems at the University of Melbourne, Australia.

Hannover, Germany), Marty Abrams (Information Accountability Foundation, United States) and Marie Georges (France). Sean McLaughlan, Elizabeth Coombs and Joe Cannataci have also contributed to the report.

20. More information on the drafting process for the report on big data and open data is available in supporting document VII.[9]

## A. Framing the issues

21. One of the most significant challenges that information societies face in the twenty-first century is the task of reconciling the societal benefits offered by new information and communications technologies with the protection of fundamental rights such as the right to privacy. These new technologies have the potential to assist States in ensuring respect, protection and fulfilment of their human rights obligations, but also risk undermining certain human rights, in particular the right to privacy.

22. New methods of collecting and analysing data — the phenomenon of big data — and the increasing willingness of Governments across the world to publicly release the personal information they hold, albeit in de-identified form, in order to generate economic growth and stimulate scientific research — the phenomenon of open data — challenge many of the assumptions that underpin our notions about what privacy is, what it entails and how best to protect it.

23. With the recognition of privacy by the Human Rights Council as an enabling right, essential to the right to dignity and the free and unhindered development of one's personality (see Human Rights Council resolution 34/7 of 23 March 2017), the challenge posed by big data and open data has broadened.

24. Certain claims made about big data and open data have been labelled "utopian".[10] These claims argue that big data offers the means to develop new insights into intractable public policy issues such as climate change, the threat of terrorism and public health. At the other end of the spectrum are those who take a dystopian point of view, troubled by the increasing surveillance by State and non-State actors, unjustified intrusion into the private sphere and the breakdown of privacy protections.

25. One of the major challenges encountered in the development of this report has been navigating and evaluating the claims by these and other stakeholders involved in the complex debates surrounding big data and open data. Although both issues have generated significant commentary and scholarship, gaps exist in our understanding of the technologies and their implications for the future: paradoxically, that lack of data inhibits our understanding of the potential benefits and possible harm of big data and open data.

## B. Data

26. Every day our digital activities produce about 2.5 quintillion bytes of data.[11] This is 2.5 followed by 18 zeros[12] of bytes of data. To put this into perspective: an

[9] See www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

[10] Danah Boyd and Kate Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication and Society*, vol. 15, No. 5.

[11] See www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

[12] This is the calculation used in the United States. In the United Kingdom of Great Britain and Northern Ireland, a quintillion is 1 followed by 30 zeros.

average 300 page novel contains about 3 followed by five zeros bytes of data; 90 per cent of all of the data in the world was created in the last two years;[13] and the rate at which it is being created keeps growing.

27. In the present-day connected world, data is both pervasive and ubiquitous. Whenever we use a computer, a smartphone or even everyday devices that include sensors capable of recording information, data is created as a by-product. This takes the form of characters or symbols ultimately reduced by computing devices to binary code, which is then processed, stored and transmitted as electronic signals.

28. The sources of the data used for big data are as varied as the activities that take place using the Internet:

> "Data come from many disparate sources, including scientific instruments, medical devices, telescopes, microscopes, satellites; digital media including text, video, audio, email, weblogs, twitter feeds, image collections, click streams and financial transactions; dynamic sensor, social, and other types of networks; scientific simulations, models, and surveys; or computational analyses of observational data. Data can be temporal, spatial, or dynamic; structured or unstructured; information and knowledge derived from data can differ in representation, complexity, granularity, context, provenance, reliability, trustworthiness, and scope. Data can also differ in the rate at which they are generated and accessed".[14]

29. Some of the data created does not relate to individuals. It is data derived from activities such as the analysis of weather patterns, space exploration, scientific testing of materials or designs or the risks associated with securities trading in financial markets. But a large proportion is the data we create ourselves or that is created about us. The focus of this report is on this category of data — personal information — whether provided, observed, derived or inferred.[15]

30. Personal information captures our individuality as human beings. It is this ability to identify each individual which makes personal information so valuable.

31. The data we create ourselves involves our own agency. It includes our emails and text messages, as well as images and videos we create and share. Other data is created about us by third parties, but in circumstances where we have participated, at least to some extent, in its creation, for example electronic health records or e-commerce transactions.

32. But other data about us is generated in ways that are not obvious because it takes place behind the scenes, in circumstances that are opaque and largely unknown — and unknowable — to us. It consists of "digital bread crumbs",[16] electronic artefacts and other electronic trails left behind as a product of our online and offline activities. This data can encompass the times and locations when our mobile devices connect with mobile telephone towers or global positioning system (GPS) satellites, records of the websites we visit, or images collected by digital closed circuit television (CCTV) systems. These "digital breadcrumbs we leave behind and which are likely to remain in perpetuity on computer servers are clues to

[13] See www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

[14] United States, National Science Foundation, "Critical techniques and technologies for advancing big data science and engineering (BIGDATA)", Program Solicitation NSF 14-543, available from www.nsf.gov/pubs/2014/nsf14543/nsf14543.pdf at p3.

[15] Martin Abrams, "The origins of personal data and its implications for governance", available from http://informationaccountability.org/wp-content/uploads/Data-Origins-Abrams.pdf.

[16] Evan Schwartz, "Finding our way with digital bread crumbs", *MIT Technology Review*, 18 August 2010. Available from www.technologyreview.com/s/420277/finding-our-way-with-digital-bread-crumbs/.

who we are, what we do, and what we want. This makes personal data — data about individuals — immensely valuable, both for public good and for private companies".[17]

33. A world that is engulfed in data, computer processing and instant digital communication raises questions about how privacy rights can coexist with the new technologies that enable personal information to be collected, processed and analysed in ways that could not have been conceived when the 1948 Universal Declaration of Human Rights and the 1966 International Covenant on Civil and Political Rights were drafted.

34. As a result of pervasive computer mediation, nearly every aspect of the world is rendered in a new symbolic dimension as events, objects, processes and people become visible, knowable and shareable in a new way. The world is reborn as data and the electronic text is universal in scale and scope.[18]

35. The way in which information and communications technologies permit individuals to become knowable through the analysis of their data involves looking "at the nature of a person as being constituted by that person's information".[19] The phenomenon that enables this is widely known as big data.

## C. Big data

36. The term "big data" is commonly used to describe the large and increasing volume of data and the advanced analytic techniques used to search, correlate, analyse and draw conclusions from it.

37. There is no agreed definition of big data. The United States National Institute of Standards and Technology describes it as the inability of traditional data architectures to efficiently handle the new datasets. The characteristics of big data that force new architectures are:

(a) Volume (i.e., the size of the dataset);

(b) Variety (i.e., data from multiple repositories, domains or types);

(c) Velocity (i.e., rate of flow);

(d) Variability (i.e., the change in other characteristics).

38. These characteristics — volume, variety, velocity and variability — are known colloquially as the "Vs" of big data.[20]

39. The above description provided by the National Institute, as well as many other efforts to pinpoint the phenomenon of big data, such as the European Union's statement that "big data refers to large amounts of data produced very quickly by a high number of diverse sources",[21] direct attention to the technologies that are coalescing to make the collection, processing and analysis of large quantities of data a commonplace reality. However, the high level of generalization in these

[17] Julie Lane and others, eds., *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (New York, Cambridge University Press, 2014).

[18] Shoshana Zuboff, "Big other: surveillance capitalism and the prospects of an information civilization", *Journal of Information Technology*, vol. 30, No. 1 (March 2015).

[19] Luciano Floridi, "Four challenges for a theory of informational privacy", *Ethics and Information Technology*, vol. 8, No. 3 (July 2006).

[20] Other Vs are attributed, but these four are the key drivers. See http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf.

[21] See https://ec.europa.eu/digital-single-market/en/policies/big-data.

descriptions and their predominant focus on technologies do not sufficiently account for the phenomenon of big data.

40. A more exhaustive description of big data that extends further than the four "Vs" has been attempted by a variety of experts. A useful and more detailed account describes big data as:

(a) Huge in volume, consisting of terabytes or petabytes of data;

(b) High in velocity, being created in or near real-time;

(c) Diverse in variety, being structured and unstructured in nature;

(d) Exhaustive in scope, striving to capture entire populations or systems;

(e) Fine-grained in resolution and uniquely indexical in identification;

(f) Relational in nature, with common fields enabling the conjoining of different datasets;

(g) Flexible, adding new fields easily and able to expand in size rapidly.[22]

41. One particular instance of big data does not necessarily embody each and every one of these features.

42. Other approaches present big data as more than a technological phenomenon:

> "We define Big Data as a cultural, technological, and scholarly phenomenon that rest on the interplay of:
>
> "(a) Technology — maximizing computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets;
>
> "(b) Analysis — drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims;
>
> "(c) Mythological — the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy."[23]

43. A main claim made by proponents of big data is that it can provide a solution to the limits imposed on research from a lack of empirical evidence, i.e., a lack of data, and can provide us with the objective truth about circumstances or phenomena. These epistemological claims, which tend to elevate big data to a new form of scientific method, lie at the centre of the unease many have expressed about the limitations of, and risks posed by, big data.

44. There is broad agreement that big data can produce social benefits, including personalized services, increased access to services, better health outcomes, technological advancements and accessibility improvements.[24] The European

[22] Rob Kitchin, "Big data, new epistemologies and paradigm shifts", *Big Data and Society*, vol. 1, No. 1 (April–June 2014).

[23] Boyd and Crawford, "Critical questions for big data".

[24] There are also significantly contrary views. For example, see the statement of the European Union, Article 29 Data Protection Working Party on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the European Union, 16 September 2014: "Many individual and collective benefits are expected from the development of big data, despite the fact that the real value of big data still remains to be proven. The Working Party would naturally support genuine efforts at European Union or national levels which aim to make these benefits real for individuals in the European Union, whether individually or collectively". See http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf.

Commission states that "the need to make sense of 'big data' is leading to innovations in technology, development of new tools and new skills".[25]

45. The European Commission identifies information as being an economic asset, as important to society as labour and capital.[26] Significantly, this market is dominated by a small number of massive technology firms whose market share relies upon the use of data.

## D. Advanced analytics

46. The critical change is the tremendous use of data to inform the algorithm whose subsequent behaviour depends on the very data it accesses:

> "The term machine learning refers to automated detection of meaningful patterns in data. In the couple of decades, it has become a common tool in almost any task that requires information extraction from large data sets ...
>
> "One common feature of all of these applications is that, in contrast to more traditional uses of computers, in these cases, due to the complexity of the patterns that need to be detected, a human programmer cannot provide an explicit, fine-detailed specification of how such tasks should be executed ... Machine learning tools are concerned with endowing programs with the ability to learn and adapt."[27]

47. The key difference between "now" and "then" is the autonomous and semiautonomous nature of the new techniques.

48. One of the most commonly used analytic techniques is known as "data mining". This is a process whereby data is extracted from large data sets and subsequently analysed to determine whether patterns or correlations exist. Data mining facilitates the simplification and summarization of vast quantities of raw data[28] and the inference of knowledge from the patterns that appear.

49. The engine that drives these techniques and tools is the algorithm.

## E. Algorithms

50. Algorithms are nothing new. They "have been around since the beginning of time and existed well before a special word had been coined to describe them".[29]

51. Algorithms are not confined to mathematics. The Babylonians used them for deciding points of law, Latin teachers use them to ensure that grammar is correct and they have been used in all cultures for predicting the future, for deciding medical treatment and or for preparing food. Today, everybody uses algorithms of one sort or another, often unconsciously, when following a recipe, using a knitting pattern or operating household gadgets.[30]

---

[25] See https://ec.europa.eu/digital-single-market/en/making-big-data-work-europe.
[26] Ibid.
[27] Shai Shalev-Shwartz and Shai Ben-David, *Understanding Machine Learning* (New York, Cambridge University Press, 2014).
[28] Data that relates only to one individual.
[29] Jean-Luc Chabert, ed., *A History of Algorithms: From the Pebble to the Microchip* (Berlin, Springer-Verlag, Berlin, Heidelberg, 1999).
[30] Ibid.

52. In common with other elements of big data, "it is notoriously difficult to give a precise characterization of what an algorithm is".[31] For the purposes of this report, a useful working definition is:

> "a specific set of instructions for carrying out a procedure or solving a problem, usually with the requirement that the procedure terminate at some point. Specific algorithms sometimes also go by the name method, procedure, or technique … The process of applying an algorithm to an input to obtain an output is called a computation."[32]

53. What separates an algorithm used to bake a cake from an algorithm that assesses a person's creditworthiness is the degree of automation involved, its autonomous, non-linear, nature and the amount of data processed.

54. More and more how we understand ourselves and our relationship to the world takes place through the lens of an algorithm. Algorithms are now a crucial part of information societies, increasingly governing "operations, decisions and choices previously left to humans",[33] recommending matches on dating sites,[34] determining the best route to travel[35] and assessing whether people are good credit risks.[36] They are used for profiling — identifying personal characteristics and behaviour patterns to make personalized predictions, such as the goods or services we might be inclined to buy. They determine how data should be interpreted and what resulting actions should be taken. They "mediate social processes, business transactions, governmental decisions and how we perceive, understand and interact among ourselves and our environment".[37]

55. From an individual perspective, the recommendations and decisions that result from algorithmic processing appear to spring from an inscrutable and unknowable black box, a kind of twenty-first century Delphic oracle, which seemingly makes unchallengeable and authoritative pronouncements divorced from human agency. Unravelling the mechanisms of algorithmic processing, and thus assessing the risks that they pose, is complex and there is a multiplicity of issues that need to be considered. These complexities hinder our ability to understand how algorithms function and how they affect our lives.

56. There is a growing body of literature highlighting the problems algorithms can cause and urging caution before we run headlong into an algorithmic future without thinking about the safeguards we need to manage the risks.

### 1. Algorithms are value-laden

57. Contrary to their arithmetical construction, which gives them an appearance of objectivity, algorithms "are inescapably value-laden".[38] The values they embody

---

[31] Felicitas Kraemer, Kees van Overveld and Martin Peterson, "Is there an ethics of algorithms?", *Ethics and Information Technology*, vol. 13, No. 3 (September 2011).

[32] See http://mathworld.wolfram.com/Algorithm.html.

[33] Brent Mittelstadt and others, "The ethics of algorithms: mapping the debate", *Big Data and Society*, vol. 3, No. 2 (July-December 2016).

[34] See, for example, Rebecca Harrington, "Dating services tinker with the algorithms of love", *Scientific American*, 13 February 2015. Available from www.scientificamerican.com/article/dating-services-tinker-with-the-algorithms-of-love/.

[35] See, https://motherboard.vice.com/en_us/article/4x3pp9/the-simple-elegant-algorithm-that-makes-google-maps-possible.

[36] See Michael Byrne, "The simple, elegant algorithm that makes Google Maps possible", 22 March 2015. Available from http://mitsloan.mit.edu/media/Lo_ConsumerCreditRiskModels.pdf.

[37] Mittelstadt and others, "The ethics of algorithms".

[38] Ibid.

often reflect cultural or other assumptions of the software engineers who design them and embed them within the logical structure of algorithms as unstated opinions.

58. For example, a credit-scoring algorithm might be designed to inquire about a person's place of birth, where she or he went to school, where she or he resides and her or his employment status. The selection of these proxies involves a value judgment, which is that the answers to those questions are relevant to assessing whether credit should be offered and, if so, on what terms. Either way, the applicant for credit very often has no way of knowing the reason for any particular credit decision and cannot determine the value judgements that have been applied.

59. Although these data proxies might be relevant to credit assessments in some societies, they will be, at best, unhelpful distractions, or at worst damaging in others. For example, their deployment in some developing countries where much of the population might have no fixed address, may have had little formal education and may be self-employed would deny, in perpetuity, access to credit.

60. On the other hand, algorithms that analyse non-traditional forms of data could show that a person without a conventional credit history nevertheless could be a good risk — thus enabling human development.[39]

### 2. The problem of imperfect data

61. The raw material that fuels algorithms is data, but not all data is accurate, sufficiently comprehensive, up-to-date or reliable.[40] The provenance of some data, for example taxation records, can usually readily be established, but their accuracy may vary from taxation agency to taxation agency within one State and between States. Other data sources may have been drawn from antiquated databases never properly cleansed or from insecure sources or where there have been inappropriate data entry and recordkeeping standards.

62. The role of algorithms is to process data, and they "are therefore subject to a limitation shared by all types of data-processing, namely that the output can never exceed the input".[41] The "garbage in/garbage out" principle applies.

### 3. The choice of data

63. This risk regarding choice of data is similar to that noted in paragraph 62 above. Just as poor data produces poor outcomes, the selection of inappropriate or irrelevant data also produces outcomes that can be unreliable and/or misleading.

64. A significant amount of algorithmic processing involves inductive reasoning and identifying correlations between apparently disparate pieces of data. If the wrong data is used, any recommendations or decisions will be flawed.

### 4. Bias, discrimination and embedding disadvantage

65. Although some experts draw distinctions between bias and discrimination,[42] the risks they pose in the context of big data are sufficiently similar to warrant them being discussed together.

---

[39] United States, Federal Trade Commission, "Big data: a tool for inclusion or exclusion — understanding the issues" (2016). Available from www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

[40] For example, the interests of minority groups not well represented in a particular dataset may be affected by decisions and predictions subsequently taken on the basis of such information.

[41] Mittelstadt and others, "The ethics of algorithms".

66. Algorithms can be used for profiling, i.e., to identify correlations and make predictions about behaviour at a group-level, albeit with groups (or profiles) that are constantly changing and redefined by the algorithm using machine learning:

> "Whether dynamic or static, the individual is comprehended based on connections with others identified by the algorithm, rather than actual behaviour. Individuals' choices are structured according to information about the group. Profiling can inadvertently create an evidence-base that leads to discrimination."[43]

67. Some commentators have argued that advanced analytic techniques, such as profiling, intensify disadvantage. An example is predictive policing, which draws on the use of crime statistics and algorithmically based analysis to predict crime hotspots and make them the priorities for law-enforcement agencies.[44] As hotspots are more heavily policed, and are often located in socially disadvantaged areas rather than where white-collar crime occurs, more policing tends to localize arrests and convictions, which leads, in a vicious cycle, to the repeated and intensified identification of the same hotspot locations, thus exposing people who are disadvantaged to a higher risk of arrest and punishment under criminal law.

68. The possible use of such tools by Governments to control, target or otherwise harm certain communities has also raised concerns.[45]

### 5. Responsibility and accountability

69. Harm caused by algorithmic processing is broadly attributable to the difficulties associated with processing large volumes of disparate datasets and the design and implementation of the algorithms used for processing. As there are so many variables involved, it is difficult to pinpoint who is responsible for the harm caused. Oftentimes, big data analytics is based on discovery and exploration, as opposed to the testing of a particular hypothesis, so it is difficult to predict (and, for individuals, to articulate) what the ultimate purpose of the use of data will be at the outset.

70. Algorithm opaqueness is not necessarily "a given"; it is technically possible to retain the data used and the result of the application of the algorithm at each stage of its processing.

### 6. Challenges to privacy

71. The Organization for Economic Cooperation and Development (OECD) published its "Guidelines on the Protection of Privacy and Transborder Flows of Personal Data" in 1980.[46] The eight principles in the OECD Guidelines, together with the similar principles found in the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Data Protection Convention), adopted by the Council of Europe in 1981, and in the Guidelines for the regulation of computerized personal data files adopted by the General Assembly in its

---

[42] Bias is considered to be the consistent or repeated expression of a particular decision-making preference, value or belief. Discrimination is the adverse, disproportionate impact that can result from algorithmic decision-making.

[43] Mittelstadt and others, "The ethics of algorithms".

[44] See, for example, www.predpol.com/how-predictive-policing-works/.

[45] Lee Rainie and Janna Anderson, "Code-dependent: pros and cons of the algorithm age", Pew Research Center, 8 February 2017. Available from www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

[46] Organization for Economic Cooperation and Development (OECD), "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data". Available from www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

resolution 45/95 of 14 December 1990 have informed information privacy laws across the world.

72. The foundational principle found in both OECD Guidelines and the Data Protection Convention, the "collection limitation principle", is that personal information should only be collected lawfully and fairly and, where appropriate, with the knowledge and consent of the individual concerned.[47] The "purpose specification principle" requires that the purpose of the collection of personal information should be specified at the time of collection and that the subsequent use of the information should be limited to the purpose of collection or a compatible purpose and that these should be specified whenever there is a change of purpose.[48] The "use limitation principle" restricts the disclosure of personal information for incompatible purposes except with the individual's consent or by legal authority.[49] The "data quality principle" is challenged by the collection of vast quantities of data and the requirement to only process personal information that is adequate, relevant and not excessive. The 1990 United Nations Guidelines for the regulation of computerized personal data files posit the principle of proportionality in data retention for the purpose of data processing.

73. Big data challenges these principles while posing ethical issues and social dilemmas arising from the poorly considered use of algorithms. Rather than solving public policy problems, there is a risk of unintended consequences that undermine human rights, such as freedom from all forms of discrimination, including against women, persons with disabilities and others.

74. At the same time, there are signs of a change of mindset in algorithm design leading to better algorithmic solutions for big data algorithms with, for example, the initiative of Institute of Electrical and Electronics Engineers Standards Association on ethically aligned design.[50]

75. In terms of privacy, relevant international instruments extend the meaning of the right to privacy beyond the information privacy rights that are the focus of the principles of the OECD Guidelines and the Data Protection Convention. Given the recognition of privacy as an enabling right, which is important to the enjoyment of other human rights, and as a right strongly linked to concepts of human dignity and the free and unhindered development of one's personality (see Human Rights Council resolution 34/7), the challenges posed by big data to privacy broaden towards the inclusion of a diversity of human rights. The tendency of big data to intrude into the lives of people by making their informational selves known in granular detail to those who collect and analyse their data trails is fundamentally at odds with the right to privacy and the principles endorsed to protect that right.

76. The regulatory implications are as profound as the changes evident in evolving industry and government practices.

## F. Open data

77. Open data is a concept that has gained popularity in parallel to the development of advanced analytics. It seeks to encourage the private and public

[47] See OECD Privacy Principles. Available from http://oecdprivacy.org/.
[48] Ibid.
[49] Ibid.
[50] Institute of Electrical and Electronics Engineers (IEEE), IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, *Ethically Aligned Design: A Vision for Prioritizing Wellbeing with Artificial Intelligence and Autonomous Systems*, ver. 1 (IEEE Press, 2016). Available from http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf.

sectors to release data into the public domain to encourage transparency and openness, particularly in government.

78. Open data is defined as:

> "… data that can be freely used, reused and redistributed by anyone — subject only, at most, to the requirement to attribute and share alike".[51]

79. Open data can consist of practically any category of data. The Open Knowledge Foundation summarizes these as follows:

(a) Culture: data about cultural works and artefacts — for example titles and authors — generally collected and held by galleries, libraries, archives and museums;

(b) Science: data that is produced as part of scientific research from astronomy to zoology;

(c) Finance: data such as government accounts (expenditure and revenue) and information on financial markets (stocks, shares, bonds, etc.);

(d) Statistics: data produced by statistical offices such as the census and key socioeconomic indicators;

(e) Weather: the many types of information used to understand and predict the weather and climate;

(f) Environment: information related to the natural environment such as presence and level of pollutants, the quality, and rivers and seas.[52]

80. In order to satisfy the requirements of the definition, open data is often released under Creative Commons' licenses. Creative Commons license CC BY 4.0 permits the unrestricted copying, redistribution and adaptation (including for commercial purposes) of the licensed material, provided that attribution requirements are met.[53]

81. Government-held data about its citizens would not fall under any of these categories. Open data and open government were intended to provide access to data about government itself and the world we live in. It was not intended to include data that governments collect on citizens. In recognition of this, some jurisdictions explicitly exclude "personal" and other categories of information, such as commercial or "cabinet-in-confidence" information, from open data.[54] We should not lose sight of the fact that, amidst terminology such as "sharing" and "connecting", a reversal has occurred, that is, rather than releasing data about how government works and which the public can use to hold government to account, governments are releasing data about their citizens.

[51] See http://opendatahandbook.org/guide/en/what-is-open-data/.
[52] See https://okfn.org/opendata/.
[53] See https://creativecommons.org/licenses/by/4.0/.
[54] Australia, New South Wales government, "Open data policy", Department of Finance and Services, 2013.

## G. Open government

82. One of the first acts of the Obama administration was to issue an executive order to encourage the release of Government-held information to enable public trust and to promote transparency, participation and collaboration.[55]

83. Following this, the Open Government Partnership was formed. In September 2011 the Partnership issued the Open Government Declaration.[56] The Declaration focuses on providing individuals with more information about the activities of government and emphasizes the need for greater civic participation and government transparency, fighting corruption, empowering citizens and harnessing "the power of new technologies to make government more effective and accountable."

84. The Open Government Declaration commits its members, on a non-binding and voluntary basis, to:

(a) Increase the availability of information about governmental activities;

(b) Support civic participation;

(c) Implement the highest standards of professional integrity throughout the administration;

(d) Increase access to new technologies for openness and accountability.[57]

85. This first executive order of the Obama administration was followed by another executive order, issued on 9 May 2013, which sought to make all United States Government information open and machine-readable by default.[58] The emphasis had changed from the earlier, 2009, order. Open government data, it stated: "promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation".[59]

86. In the following years, open data has evolved to a point where, in 2017, its ambitions lie beyond the release into the public domain of data that has never been or is not derived from personal information to the release of de-identified personal information. Proponents of this approach assert that much "value" is locked away in government databases or other information repositories and that making this information available publicly will encourage research and stimulate the growth of the information economy.

87. Open data that is derived from personal information thus wholly relies on the efficacy of "de-identification" processes in order to prevent re-identification and thus its linkage back to the individual from whom it was derived. Debates about whether or not de-identification delivers both privacy protection and "research-useful" data have proven to be highly contentious.

---

[55] President Obama, "Transparency and Open Government", 21 January 2009, memorandum for the Heads of Executive Departments and Agencies. Available from https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

[56] See https://www.opengovpartnership.org/open-government-declaration.

[57] https://www.opengovpartnership.org/open-government-declaration.

[58] President Obama, executive order of 9 May 2013, on "Making open and machine readable the new default for Government information". Available from https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government.

[59] Ibid.

## H. Complexity of big data

88. In 2015, Australian journalist Will Ockenden published his telecommunications metadata online and asked people to tell him what they could infer about his life. The metadata included the exact times of all telephone calls and SMS messages, along with the nearest phone tower. Although he replaced phone numbers with pseudonyms, questions like "where does my mother live?" were easily and correctly answered based on communication and location patterns alone. It wasn't complicated — viewers simply guessed (correctly) that his mother lived in the place he visited on Christmas Day.

89. This is a key theme of privacy research: that patterns in the data, without the names, phone numbers or other obvious identifiers, can be used to identify a person and hence to extract more information about them from the data. This is particularly powerful when those patterns can be used to link many different datasets together to build up a complex portrait of a person.

90. Some data inevitably must be exposed. Phone companies know what numbers each customer is dialing, and doctors know their patients' test results. Controversies therefore arise on the disclosure of that data to others, such as corporations or researchers, and on the ways governments can use information and impact the exercise of the human rights of their citizens.

91. Other data is deliberately harvested, often without the individual's knowledge or consent. Researchers at the Electronic Frontier Foundation published the results of "Panopticlick", an experiment that showed it was possible to fingerprint a person's web browser based on simple characteristics such as plugins and fonts.[60] They warned that web browsing privacy was at risk unless limits were set on the storage of these fingerprints and their links with browsing history. No significant policy changes were made. Today, in 2017, web browsing privacy is gone. Many companies routinely and deliberately track people, generally for commercial reasons. Web tracking is now almost ubiquitous and evaded only with great effort.

92. Much of the economy of the modern Internet depends on harvesting complex data about potential customers in order to sell them things, a practice known as "surveillance capitalism".[61] However, surveillance does not seem any more justifiable to data-driven efficiency than child-labour is to an industrial economy. It is only the most convenient and easiest way to exploit information. It is not remotely to be considered as a fundamental right, as is the right to privacy. Indeed, the data-driven economy would survive and prosper if minimal standards and improved technologies forced corporations and governments into a world in which ordinary people had much greater control over their own data.[62]

93. Governments would also be able to innovate with a more legitimate license. The community's level of trust in government strongly shapes how they view the possible impact of open data and open government initiatives. Those who trust

[60] Peter Eckersley, "How unique is your web browser?" in *Privacy Enhancing Technologies*, Mikhail Atallah and Nicholas Hopper, eds. (Berlin, Springer-Verlag, 2010).

[61] Shoshana Zuboff, "Big other: surveillance capitalism and the prospects of an information civilization", *Journal of Information Technology*, vol. 30, No. 1 (March 2015).

[62] Corporations and Governments do not necessarily need to be forced to provide privacy protections. For examples of ethical approaches adopted by companies see Information Commissioner's Office, "Big data, artificial intelligence, machine learning and data protection", ver. 2.2 (2017). Available from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.

government are far more likely to think that there are benefits to open data.[63] Research shows that people are for the most part comfortable with their government providing online data about their communities, although they sound cautionary notes when the data hits close to home. Citizen comfort levels vary depending on what area of data collection is being discussed.[64]

94. Most information privacy laws regulate the collection and processing of personal information: if information is not personal information it is not regulated by information privacy laws. Many such laws recognize that personal information may be de-identified so that it can be used or processed for purposes such as public interest research in a way that does not interfere with individuals' information privacy rights. Governments and others have sought to maintain the trust of those whose data they collect by assurances of de-identification.

95. This leads to the important consideration "do de-identification processes deliver data that does not interfere with individuals' information privacy rights"?

96. Simple kinds of data, such as aggregate statistics, are amenable to genuinely privacy-preserving treatment such as differential privacy. Differential privacy algorithms work best at large scales, and are being incorporated into commercial data analysis. Randomized algorithms achieving differential privacy are a valuable tool in the privacy arsenal, but they do not provide a way of blanket de-identification of highly complex datasets of unit-record[65] level data about individuals. The use of these techniques by the Apple corporation in 2016 is an example of how differential privacy is used on a large scale.[66]

97. High-dimensional unit-record level data cannot be securely de-identified without substantially reducing its utility. This is the sort of data produced by a longitudinal trace of one person's data for health, mobility, web searching and so on. Supporting document I[67] provides a summarized account of de-identification tools and controversies.

### Open government data

98. There are numerous examples of successful re-identification of individuals in data published by governments.[68] This "public re-identification" is public in two senses: the results are made public; and re-identification uses only public auxiliary information.

---

[63] John Horrigan and Lee Rainie, "Americans' views on open Government data", Pew Research Center, 21 April 2015.

[64] Ibid.

[65] Relates only to one individual.

[66] Andy Greenberg, "Apple's 'differential privacy' is about collecting your data — but not your data", 13 June 2016. Available from https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/https://techcrunch.com/2016/06/14/differential-privacy/ https://arxiv.org/abs/1709.02753.

[67] See www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

[68] In testimony to the Privacy and Integrity Advisory Committee of the Department of Homeland Security on 15 June 2005, Sweeney stated it was in 1997 that she "was able to show how the medical record of William Weld, the Governor of Massachusetts at the time could be re-identified using only his date of birth, gender and ZIP. In fact, 87 per cent of the population of the United States is uniquely identified by date of birth (e.g., month, day and year), gender, and their 5-digit ZIP codes. The point is that data that may look anonymous is not necessarily anonymous". See www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_06-2005_testimony_sweeney.pdf; see also Latanya Sweeney, "Matching known patients to health records in Washington state data", Harvard University, 2012. Available from http://dataprivacylab.org/projects/wa/1089-1.pdf and http://dataprivacylab.org/index.html; Latanya Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, No. 5 (2002).

99. The more auxiliary information is available, the easier it becomes to re-identify a larger number of individuals. As more datasets are linked, there is a reduction in the auxiliary information necessary for re-identification. The public disclosure and linking of datasets gathers vast auxiliary information about individuals in the same place, making it much easier to re-identify any data related to them.

100. The re-identifiability of open data is a small indication of a much larger problem — the re-identifiability of "de-identified" commercial datasets that are routinely sold, shared and traded.

101. Arrayed against the right to privacy in the big data and open data era are powerful forces. The weakest possible de-identification permitted is likely to be the most financially preferred by all who deal in data, whether for commercial or other purposes, and governments come under pressure not just in relation to opening up access to data about individuals, but also in relation to the regulation of this access.

102. Non-government organizations have voiced concerns about the growth of big data without due consideration for the involvement of the individual, the ethical and legal issues arising from inadequate management of the personal information of individuals or adequate regulation.[69] Such organizations will continue to advocate for adequate protection and appropriate action.

## I. Considering the present: big commercial data and privacy

103. The exponential increase in data collection and the rush to connect seemingly every object to the Internet, with insufficient regard for data security, has created risks for individuals and groups. In efforts to assure consumers and individuals of the security of information identifying them, a number of notions have been spread in the public domain. For example, the notion of highly complex "anonymized" data is cultivated by an industry that benefits from users' mistaken feeling of anonymity.[70]

104. A great deal of data is gathered from ordinary users without their knowledge or consent. This data can be sold and linked with data from other sources to produce a complex record of many aspects of a person's life. This information serves many purposes, including political control, as has been demonstrated by a dataset unintentionally exposed by a political organization in the United States.[71] The dataset included personal details of almost 200 million United States voters, along with astonishing detail gathered (or guessed) about their political beliefs. In China, there is a "social credit" project that aims to rate not only the financial creditworthiness of citizens, but also to track their social and possibly political behaviour. It relies upon data from a variety of sources, primarily online sources, over time.[72]

105. Data brokers —companies that collect consumers' personal information and resell or share that information with others—are important participants in the big

[69] See www.privacyinternational.org/node/8.

[70] Even if anonymized, this does not remove the relevance of privacy principles and considerations such as "consent".

[71] Sam Biddle, "Republican data-mining firm exposed personal information for virtually every American voter", *The Intercept*, 19 June 2017 Available from https://theintercept.com/2017/06/19/republican-data-mining-firm-exposed-personal-information-for-virtually-every-american-voter/.

[72] "China invents the digital totalitarian state", *The Economist*, 17 December 2016. Available from https://www.economist.com/news/briefing/21711902-worrying-implications-its-social-credit-project-china-invents-digital-totalitarian; Lucy Hornby, "China changes tack on 'social credit' scheme plan", *Financial Times*, 4 July 2017. Available from www.ft.com/content/f772a9ce-60c4-11e7-91a7-502f7ee26895.

data economy. In developing their products, data brokers acquire a vast array of detailed and specific information about consumers from a variety of sources;[73] analyse it to make inferences about consumers, some of which may be sensitive; and share the information with clients in a range of industries. All of this activity takes place without the consumers' knowledge.[74]

106. While data broker products help to prevent fraud, improve product offerings and deliver personalized services, many purposes for which data brokers collect and use data pose risks to consumers. Concerns exist about the lack of transparency, the collection of data about young people, the indefinite retention of data and the use of such data for the determination of eligibility or for unlawful discriminatory purposes.[75]

107. The European Parliament's recent draft report on European Privacy regulation recommends that "end-users should be offered a set of privacy setting options, ranging from higher (for example, 'never accept cookies') to lower (for example, 'always accept cookies') and intermediate…".[76]

108. The need to increase the control of individuals over their Internet privacy is being widely discussed. Individuals use their own devices and their data to obtain the information they require, such as maps and directions, and to view the advertisements they are interested in. In this regard, it is vital to ask, while technologies facilitating end-user control are important, to what extent can individuals exert sufficiently comprehensive protective control? The adoption of these tools conflicts with the economic forces currently shaping the Internet.[77] Do governments have a role in the development and adoption of these tools?

**Technologies for controlling data collection**

109. Controlling (including stopping) data collection is relevant to the data people do not want to share. With "old" technology this was not a consideration, as the user was inevitably in control because technology did not enable anything other than user determination: devices had physical covers on cameras or ethernet-only Internet connections that could be manually unplugged. Now there are internal Wi-Fi and coverless cameras. Television sets have microphones that cannot be turned off. Manual disabling features have disappeared, although there are

[73] There are many reported illustrations of large-scale commercial data acquisition from smart devices such as televisions, "intimate appliances", children's toys and, ride sharing apps to "connected cars".

[74] United States Senate Committee on Commerce, Science and Transportation, "A review of the data broker industry: collection, use, and sale of consumer data for marketing purposes", staff report, 18 December 2013. Available from http://educationnewyork.com/files/rockefeller_databroker.pdf.

[75] United States Federal Trade Commission, "Data brokers: a call for transparency and accountability", May 2014. Available from www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

[76] Marju Lauristin, "Draft report on the proposal for a regulation of the European Parliament and of the European Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC", European Parliament, Committee on Civil Liberties, Justice and Home Affairs, 2017.

[77] For example, AdNauseum defeats tracking by automatically clicking on all the advertisements presented to a user in order to obscure which ones the user truly reads. This has been blocked by Google Chrome. Other sites detect and block individuals who visit with ad blockers installed. See Daniel Howe and Helen Nissenbaum, "Engineering privacy and protest: a case study of AdNauseam". Available from https://adnauseam.io/.

technologies for obstructing the collection of data.[78] The highly successful "TLS [Transport Layer Security] Everywhere" campaign means that most Internet traffic is now encrypted and much less likely to be collected in transit by an entity unknown to the user. Such technologies have benefits that need to be further explored and supported.

110. The idea of obfuscating who you are and what you do is also not new — consider the battle between some social networks' "real names" policies and the efforts of those who defend their right to register under pseudonyms. To obfuscate requires tools that allow users to present a "reserved" profile and separate it from other profiles they choose to present.

111. Research shows consistently that if individuals are concerned about the personal information practices of organizations they deal with, they are more likely to provide inaccurate or incomplete information.[79] Because privacy and data protection generate trust, they have a beneficial effect on data quality and also on data analytics. The confidence of users in their privacy is also important for the stability and accuracy of the machine-learning algorithms. Ordinary machine learning can be highly susceptible to deliberately contrived confusing inputs.[80] What would happen if a large number of people deliberately adopted tools for obfuscating themselves due to privacy concerns?

112. A simplistic approach to big data — open data that is blind to the complex interaction between perceived privacy management business practices, trust in the respect for privacy and the behaviours of individuals will not facilitate "big data", but will rather lead to potentially inaccurate and poor-quality decision-making.

## J. Principles for the future: controlling data disclosure

113. Privacy law tends to be based on principles that enable sufficient flexibility to allow privacy risks to be addressed as they evolve. There is value in considering whether additional principles are required to complement existing privacy principles in order to protect personal data from technologically-based privacy incursions.

114. One formulation proposes the following seven principles of data sharing:[81]

1. Moving the algorithm to the data: sharing outcomes rather than sharing the data directly.

2. Open algorithms: open review and public scrutiny of all algorithms for data-sharing and privacy protection, so that errors or weaknesses can be identified and corrected.

3. Permissible use: respect for the (explicit or implicit) permission for uses of the data or "contextual integrity".[82] In a medical context, the explicit

[78] The Tor (anonymity network) router obscures who communicates with whom (i.e., telecommunications metadata), but it is not widely used. Some browsers (such as Firefox and Brave) include a "private browsing" mode that obstructs data collection. The Electronic Frontier Foundation's "Privacy Badger" and New York University's "TrackMeNot" are highly effective, but are not widely adopted.

[79] Office of the Australian Information Commissioner, Australian community attitudes to privacy survey, 2017 and 2013; Deloitte, "Trust starts from within: Deloitte Australian privacy index 2017", 2017.

[80] Ian Goodfellow, Jonathon Shlens and Christian Szegedy, "Explaining and harnessing adversarial examples", ArXiv preprint, 2014.

[81] Alex Pentland and others, "Towards an Internet of trusted data: a new framework for identity and data sharing", 2016.

granting withdrawal of consent has been put into practice in the dynamic consent interface.[83]

4. Always return “safe answers”: differential privacy in practice.

5. Data always in encrypted state: encrypted data can be read only by those who know the decryption key.[84]

6. Networked collaboration environments and block chains for audit and accountability.

7. Social and economic incentives.

115. These principles are not necessarily complete solutions in themselves as they in turn raise more questions. For example, transparency is particularly challenging when the techniques used to protect privacy are so sophisticated that only a handful of people have the capacity to understand them. The “open algorithms” principle is a vital first step, but the exact algorithms being used and their implication will still be challenging in practice.

116. Other “principle” approaches have been proposed, such as “agency” and “transparency”, with agency including the right to amend data, to blur your data or to experiment with the refineries.[85] The underlying dynamic is the empowerment of individuals and the introduction of a levelling of power between the data companies/holders and the users. Others raise the principles of the opportunity to obfuscate, prevent or opt out of data collection.

117. Overall, the principles of transparency and user control are important so that users can choose what data they reveal without unreasonable loss of facility or services.

118. Above all, attempts to produce big data-open data principles that respect privacy provide a useful starting point for discussion. Whatever principles are adopted, there should be adequate consultation across stakeholders, including civil society organizations, to ensure the fitness of any such principles.

119. Implementing these principles raises questions of the role of government and the type of incentives and regulation that will facilitate the protection of privacy and other human rights and assessing “their comparative impacts on ethical and political

[82] Privacy is defined as “the requirement that information about people (‘personal information’) flows appropriately, where appropriateness means in accordance with informational norms … Social contexts form the backdrop for this approach to privacy …”. See Solon Barocas and Helen Nissenbaum, “Big data’s end run around anonymity and consent”, in *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Julia Lane and others, eds. (Cambridge University Press, 2014).

[83] Jane Kaye and others, “Dynamic consent: a patient interface for twenty-first century research networks”, *European Journal of Human Genetics*, vol. 23, No. 2 (2014).

[84] Recent advances in cryptography allow multiple parties to jointly compute a function of their private inputs, subsequently revealing only the well-defined outcome. There are very general tools, based on multiparty computation (see, for example, Ivan Damgård and others, “Multiparty computation from somewhat homomorphic encryption”, *Advances in Cryptology — CRYPTO*, vol. 7417 (2012); and homomorphic encryption, available from www.microsoft.com/en-us/research/project/homomorphic-encryption/#). Most tools do not run sufficiently fast for big datasets, but simpler variants may in the future. There are many specific protocols that solve specialized problems on large datasets. The general notion of computing on encrypted data works very well for simple computations on one dataset, but may not be feasible for complex computations or datasets distributed over several locations.

[85] Andreas Weigend, *Data for the People: How to Make our Post-Privacy Economy Work for You* (New York, Basic Books, 2017).

values, such as fairness, justice, freedom, autonomy, welfare, and others more specific to the context in question".[86]

120. An innovative information economy would probably achieve greater community support if there was observable adherence by Governments and corporations to strong regulation around the acquisition, sharing and control of people's data.

# III. Supporting documents

121. The following documents supporting the present report are available at the website of the Special Rapporteur:[87]

I. Understanding history: de-identification tools and controversies;

II. Engagements by the Special Rapporteur in Africa, America, Asia and Europe;

III. Background on the open letter to the Government of Japan;

IV. Activities of the Task Force Privacy and Personality;

V. Description of the process for the draft legal instrument on surveillance;

VI. Acknowledging assistance;

VII. Procedural clarifications on the thematic report on big data and open data.

# IV. Conclusion

122. **The issues identified in the present report are not confined to a few countries. The availability of vast new collections of data allows more and better reasoned decision-making by individuals, corporations and States around the globe, but poor management of privacy puts their potential value at risk.**

123. **Careful understanding and successful mitigation of risks to privacy, other related human rights and ethical and political values of autonomy and fairness are required.**

124. **Data is and will remain a key economic asset, like capital and labour. Privacy and innovation can and do go together. Understanding how to use big data efficiently and how to share its benefits fairly without eroding the protection of human rights will be hard, but ultimately worthwhile.**

# V. Recommendations

125. **Pending feedback during the consultation period to March 2018 and the results of ongoing investigations and letters of allegation to Governments, the Special Rapporteur is considering the following recommendations for inclusion an updated version of the present report, which is to be published in or after 2018.**

126. **Open data policies require clear statements on the limits to the use of personal information, based on international standards and principles, including**

[86] Solon Barocas and Helen Nissenbaum, "Big data's end run around anonymity and consent", in *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Julia Lane and others, eds. (Cambridge University Press, 2014).

[87] See http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/SRPrivacyIndex.aspx; see also www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

**an exempt category for personal information with a binding requirement to ensure the reliability of de-identification processes to render this information appropriate for release as open data, and robust enforcement mechanisms.**

127. **Any open government initiative involving personal information, whether de-identified or not, requires a rigorous, public, scientific analysis of data privacy protections, including a privacy impact assessment.**

128. **Sensitive high-dimensional unit-record level data about individuals should not be published online or exchanged unless there is sound evidence that secure de-identification has occurred and will be robust against future re-identification.**

129. **Frameworks should be established to manage the risk of sensitive data being made available to researchers.**

130. **Governments and corporations should actively support the creation and use of privacy-enhancing technologies.**

131. **The following options are to be considered when dealing with big data:**

**Governance**

(a) **Responsibility: identification of accountabilities, decision-making process and, as appropriate, identification of decision makers;**

(b) **Transparency: what occurs, when and how, to personal data prior to it being publicly available, and its use, including "open algorithms";**

(c) **Quality: minimum guarantees of data and processing quality;**

(d) **Predictability: when machine learning is involved the outcomes should be predictable;**

(e) **Security: appropriate steps to prevent data inputs and algorithms from being interfered with without authorization;**

(f) **Development of new tools to identify risks and specify risk mitigation;**

(g) **Support: train (and, as appropriate, accredit) employees on legal, policy and administrative requirements relating to personal information;**

**Regulatory environment**

(h) **Ensure arrangements to establish an unambiguous focus, responsibility and powers for regulators charged with protecting citizens' data;**

(i) **Regulatory powers should be commensurate with the new challenges posed by big data, for example, the ability for regulators to be able to scrutinize the analytic process and its outcomes;**

(j) **Examination of privacy laws to ensure they are "fit for purpose" in relation to the challenges arising from technology advances, such as machine-generated personal information, and data analytics, such as de-identification;**

**Inclusion of feedback mechanisms**

(k) **Formalize consultation mechanisms, including ethics committees, with professional, community and other organizations and with citizens to protect against the erosion of rights and to identify sound practices;**

(l) **Undertake a broad-based consultation on the recommendations and issues raised in the present report, for example the appetite for prohibitions on the provision of government datasets;**

**Research**

(m) **Technical: investigate relatively new techniques such as differential privacy and homomorphic encryption to assess if they provide adequate privacy processes and outputs;**

(n) **Examine citizens' awareness of the data activities of governments and businesses, the uses of personal information, including for research, and technological mechanisms to enhance individual control of their data and to increase their ability to utilize it for their needs.**