**Appendix 4**

**II. 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.[[1]](#footnote-1) The lead authors of this report are David Watts and Vanessa Teague.[[2]](#footnote-2) The members of the taskforce, many of whom also contributed to the text, include Christian d'Cunha (the European Data Protection Supervisor in Brussels), Alex Hubbard (the United Kingdom’s Information Commissioner's Office), Prof. Dr. Wolfgang Nejdl (Germany), Marty Abrams (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 Big Data and Open Data report is available in Supporting document VII[[3]](#footnote-3).

 **A. Framing the issues**

1. 21. One of the most significant challenges that twenty-first century information societies face 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 to respect, protect and fulfil their human rights obligations, but also risk undermining certain human rights, in particular the right to privacy.
2. 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 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.
3. 23. With the recognition by the Human Rights Council of privacy as an enabling right essential to the right to dignity and the free and unhindered development of one’s personality, the challenge posed by Big Data and Open Data broadens.[[4]](#footnote-4)
4. 24. Certain claims made about Big Data and Open Data have been labelled ‘utopian’[[5]](#footnote-5). 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.
5. 25. One of the major challenges encountered in developing 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 future implications for the future: paradoxically, that lack of data inhibits our understanding of the potential benefits and harms of Big Data and Open Data.

 **B. Data**

1. 26. Every day our digital activities produce about 2.5 quintillion bytes of data.[[6]](#footnote-6) This is 2.5 followed by eighteen zeros[[7]](#footnote-7) of bytes of data. To put this into perspective, an average three-hundred-page novel contains about 3 followed by five zeros bytes of data Ninety percent of all of the data in the world was created in the last two years[[8]](#footnote-8) and the rate at which it is being created keeps growing.
2. 27. In a connected world, data[[9]](#footnote-9) 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 then processed, stored and transmitted as electronic signals.
3. 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”.[[10]](#footnote-10)
4. 29. Some of the data created does not relate to individuals. It is data derived from activities like 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.[[11]](#footnote-11).
5. 30. Personal information captures our individuality as human beings. It is this ability to identify each individual which makes personal information so valuable.
6. 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 ecommerce transactions.
7. 32. But other data about us is generated in ways that are not obvious because it occurs, behind the scenes, in circumstances that are opaque and largely unknown – and unknowable – to us. It consists of ‘digital bread crumbs,’[[12]](#footnote-12) 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 GPS satellites, records of the websites we visit, or images collected by digital CCTV systems. These ‘digital breadcrumbs we leave behind and which are likely to remain in perpetuity on computer servers are clues to 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.’[[13]](#footnote-13)
8. 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:
9. 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.[[14]](#footnote-14)
10. 35. The way in which information and communications technologies permit individuals to become knowable through the analysis of their data involves ‘[l]ooking at the nature of a person as being constituted by that person’s information.’[[15]](#footnote-15) The phenomenon that enables this is widely known as Big Data.

 **C. Big Data**

1. 36. ‘Big Data’ is the term 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.
2. 37. There is no agreed definition of Big Data. The US National Institute of Science and Technology (NIST) describes it as:
3. ..the inability of traditional data architectures to efficiently handle the new datasets. Characteristics of Big Data that force new architectures are:

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

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

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

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.[[16]](#footnote-16)

39. The NIST description, as well as many other efforts to pinpoint the phenomenon of Big Data, such as the European Union’s statement that ‘[b]ig data refers to large amounts of data produced very quickly by a high number of diverse sources,’[[17]](#footnote-17) 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 generalisation these descriptions offer and their predominant focus on technologies does not sufficiently account for the phenomenon of Big Data.

40. A more exhaustive description of Big Data that extends further than the ‘V’s’ has been attempted by a variety of experts. A useful, and more detailed account is that Big Data is:

huge in volume, consisting of terabytes or petabytes of data;

high in velocity, being created in or near real-time;

diverse in variety, being structured and unstructured in nature;

exhaustive in scope, striving to capture entire populations or systems;

fine-grained in resolution and uniquely indexical in identification;

relational in nature, with common fields enabling the conjoining of different data sets;

flexible, adding new fields easily and able to expand in size rapidly.[[18]](#footnote-18)

41. Any particular instance of Big Data does not necessarily embody each 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:

 (1) Technology – maximizing computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets.

 (2) Analysis – drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.

 (3) 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.’[[19]](#footnote-19)

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 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 personalised services, increased access to services, better health outcomes, technological advancements and accessibility improvements.[[20]](#footnote-20) The European Commission states that “the need to make sense of ‘Big data’ is leading to innovations in technology, development of new tools and new skills.”[[21]](#footnote-21)

45. It identifies information as being an economic asset, as important to society as labour and capital.[[22]](#footnote-22) 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."[[23]](#footnote-23)

47. The key difference between ‘now’ and ‘then’ is the autonomous and semi-autonomous 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 summarisation of vast quantities of raw data[[24]](#footnote-24) and to infer 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.’[[25]](#footnote-25)

51. Algorithms are not confined to mathematics… The Babylonians used them for deciding points of law, Latin teachers used them to get the grammar right, and they have been used in all cultures for predicting the future, for deciding medical treatment, or for preparing food. Everybody today uses algorithms of one sort or another, often unconsciously, when following a recipe, using a knitting pattern, or operating household gadgets.[[26]](#footnote-26)

52. In common with other elements of Big Data, ‘it is notoriously difficult to give a precise characterisation of what an algorithm is.’[[27]](#footnote-27) 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.[[28]](#footnote-28)

53. What separates an algorithm used to bake a cake from an algorithm that assesses a person’s credit worthiness 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 lenses of algorithms. Algorithms are now a crucial part of information societies, increasingly governing ‘operations, decisions and choices previously left to humans.’[[29]](#footnote-29) They recommend matches on dating sites,[[30]](#footnote-30) determine the best route to travel[[31]](#footnote-31) and assess whether we are a good credit risk[[32]](#footnote-32). They are used for profiling – identifying personal characteristics and behaviour patterns to make personalised predictions, such as 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.’[[33]](#footnote-33)

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 that 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 they can cause and which urge 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**

1. 57. Contrary to their arithmetical construction that gives them an appearance of objectivity, algorithms ‘are inescapably value-laden.’[[34]](#footnote-34) The values they embody 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 judgement 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.[[35]](#footnote-35)

 **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.[[36]](#footnote-36) 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 record-keeping 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.[[37]](#footnote-37) The ‘garbage in/garbage out’ principle applies.

 **3. The choice of data**

63. This risk is similar to that noted in the previous paragraph. Just as poor data produces poor outcomes, the selection of inappropriate or irrelevant data also produces outcomes that can be unreliable and 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,[[38]](#footnote-38) the risks they pose in the context of Big Data are sufficiently similar to warrant them being discussed together.

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.[[39]](#footnote-39)

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 these priorities for law-enforcement agencies.[[40]](#footnote-40) As the hotspots are more heavily policed and often located in socially disadvantaged areas rather than where white-collar crime occurs, more policing tends to localise arrests and convictions. This leads to hotspot locations persisting and intensifying in a repeating cycle, exposing those 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.[[41]](#footnote-41)

 **5. Responsibility and accountability**

69. Harm caused by algorithmic processing is broadly attributable to the difficulties associated with processing large volumes of disparate data sets and the design and implementation of the algorithms used for the processing. As there are so many variables involved, it is difficult to pinpoint who is responsible for any harm caused. Often, Big Data analytics is based on discovery and exploration, as opposed to testing 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 Organisation for Economic Co-operation and Development (OECD) published its Guidelines on the Protection of Privacy and Transborder Flows of Personal Data in 1980[[42]](#footnote-42). The eight principles in the OECD Guidelines, together with the similar principles found in the 1981 Council of Europe’s (CoE) Data Protection Convention and the 1990 Guidelines for the regulation of computerized personal data files[[43]](#footnote-43) have informed information privacy laws across the world.

72. The foundational principle found in both the OECD and CoE rules, 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.[[44]](#footnote-44) The purpose limitation 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.[[45]](#footnote-45) The use limitation principle restricts the disclosure of personal information for incompatible purposes except with the individual’s consent or by legal authority.[[46]](#footnote-46) The data minimization 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 to the purpose of the data processing.[[47]](#footnote-47)

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 mind-set in algorithm design leading to better algorithmic solutions for big data algorithms with, for example, the IEEE Standards Association initiative on ethically aligned design.[[48]](#footnote-48)

75. In terms of privacy, relevant international international instruments extend the meaning of the right to privacy beyond the information privacy rights that are the focus of the OECD Principles and the COE Convention 108. Given the recognition of privacy as an enabling right that 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, the challenges posed by Big Data to privacy broaden towards a diversity of human rights.[[49]](#footnote-49) 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 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, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike”.[[50]](#footnote-50)

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

Culture: Data about cultural works and artefacts — for example titles and authors — generally collected and held by galleries, libraries, archives and museums.

Science: Data that is produced as part of scientific research from astronomy to zoology.

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

Statistics: Data produced by statistical offices such as the census and key socioeconomic indicators.

Weather: The many types of information used to understand and predict the weather and climate.

Environment: Information related to the natural environment, such as presence and level of pollutants, the quality of water in rivers and seas.[[51]](#footnote-51)

80. In order to satisfy the requirements of the Open Data 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 attribution requirements are met.[[52]](#footnote-52)

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 the 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.[[53]](#footnote-53) We should not lose sight amidst terminology such as ‘sharing’ and ‘connecting’, that 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.

 **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 information to enable public trust and to promote transparency, participation and collaboration.[[54]](#footnote-54)

83. Following this, the Open Government Partnership was formed. It issued an Open Government Declaration in September 2011 (OGD). The recitals to the OGD focus on providing individuals with more information about the activities of government and emphasise 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.’[[55]](#footnote-55)

84. The OGD[[56]](#footnote-56) commits its members to:

Increase the availability of information about governmental activities.

Support civic participation.

Implement the highest standards of professional integrity throughout the administration.

Increase access to new technologies for openness and accountability.[[57]](#footnote-57)

85. This was later followed by a further executive order on 9 May 2013 that sought to make all United States Government information open and machine-readable by default.[[58]](#footnote-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]](#footnote-59)

86. Over the succeeding 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 much ‘value’ is locked away in government databases or other information repositories and 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 to prevent the 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.

 **H. The 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 which showed it was possible to fingerprint a person's web browser based on simple characteristics such as plugins and fonts.[[60]](#footnote-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. 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]](#footnote-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 the information. It is not 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]](#footnote-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 government are far more likely to think that there are benefits to Open Data.[[63]](#footnote-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 across topics.[[64]](#footnote-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 recognise that personal information may be ‘de-identified’ so that the de-identified data 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. Randomised 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]](#footnote-65) level data about individuals. Apple's use of these techniques in 2016 is an example of how differential privacy is used on a large scale.[[66]](#footnote-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]](#footnote-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]](#footnote-68) This ‘public re-identification’ is public in two senses: the results are made public, and re-identification uses only public auxiliary information.

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 to the involvement of the individual, the ethical and legal issues arising from inadequate management of the personal information of individuals, or adequate regulation.[[69]](#footnote-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 sown. For example, the notion of highly complex “anonymized” data is cultivated by an industry that benefits from users’ mistaken feeling of anonymity.[[70]](#footnote-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 a dataset unintentionally exposed by a political organisation from the United States showed.[[71]](#footnote-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, the Social Credit Project aims to score not only the financial creditworthiness of citizens, but also their social and possibly political behaviour. It relies upon data from a variety of sources, primarily online sources, over time [[72]](#footnote-72)

105. Data brokers —companies that collect consumers’ personal information and resell or share that information with others—are important participants in the Big 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]](#footnote-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 consumers’ knowledge.[[74]](#footnote-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 this data for eligibility determinations or unlawful discriminatory purposes.[[75]](#footnote-75)

107. The European Parliament's recent draft report on European Privacy regulation recommends that "Users should be offered, by default, a set of privacy setting options, ranging from higher (for example, ‘never accept tracker and cookies’) to lower (for example, ‘always accept trackers and cookies’) and intermediate."[[76]](#footnote-76)

108. The need to increase individuals’ control is being raised. This approach sees individuals use their own devices and their data, to get the information they require, such as maps and directions, and which advertisements they want. 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]](#footnote-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 for data the person does 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, however there are technologies for obstructing the collection of data.[[78]](#footnote-78) The highly successful “TLS 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 from other profiles they choose to present.

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

112. A simplistic approach to Big Data – Open Data blind to the complex interaction between perceived privacy management business practices, trust in respect for privacy and behaviours of individuals will not facilitate ‘Big Data’, but 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 address privacy risks 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 seven principles of data sharing:[[81]](#footnote-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]](#footnote-82) In a medical context, the explicit granting withdrawal of consent has been put into practice in the Dynamic Consent interface.[[83]](#footnote-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]](#footnote-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, to experiment with the refineries, amongst others.[[85]](#footnote-85) The underlying dynamic is the empowerment of individuals and 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 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, so as 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 values, such as fairness, justice, freedom, autonomy, welfare, and others more specific to the context in question.”[[86]](#footnote-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 this report are available at the Special Rapporteur’s website[[87]](#footnote-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, and VII. Procedural clarifications on the thematic report on Big Data and Open Data.

 **IV. Conclusion**

**122. The issues identified in this 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 at risk their potential value.**

**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 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 on-going investigations and letters of allegation to Governments, the Special Rapporteur is considering the following recommendations for a more final version of this report to be published in or after 2018:**

**126. Open Data policies require clear statements of the limits to using personal information based on international standards and principles, including 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 the 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. Establish frameworks 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 be taken to prevent data inputs and algorithms from being interfered with without authorisation**

**f. develop 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 to be commensurate with the new challenges posed by big data for example, the ability for regulators to be able to scrutinise the analytic process and its outcomes**

**j. Examination of privacy laws to ensure these 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. Formalise consultation mechanisms, including ethics committees, with professional, community and other organisations and citizens to protect against the erosion of rights and identify sound practices;**

**l. Undertake a broadbased consultation on the recommendations and issues raised by this report such as the appetite, for example, for prohibition 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, uses of personal information including for research, technological mechanisms to enhance individual control of their data and to increase their ability to utilise it for their needs.**

1. 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. [↑](#footnote-ref-1)
2. Dr Vanessa Teague is a Senior Lecturer in the Department of Computing and Information Systems at The University of Melbourne, Australia. [↑](#footnote-ref-2)
3. Available at http://www.ohchr.org/Documents/Issues/Privacy/SR\_Privacy/ReportSR\_SupportingDocuments.pdf. [↑](#footnote-ref-3)
4. United Nations, Human Rights Council 34th Session, A/HRC/34/L.7/ Rev.1, Agenda Item 3 Protection of all Human Rights Civil, Political Economic, Social and Cultural Rights including the Right to Development, 22 March 2017. [↑](#footnote-ref-4)
5. Danah Boyd and Kate Crawford, *Critical questions for Big Data*, Information, Communication and Society, Vol 15, No. 5, p662 at p663. [↑](#footnote-ref-5)
6. IBM, https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html. [↑](#footnote-ref-6)
7. This is the calculation used in the USA. In the UK, a quintillion is 1 followed by 30 zeros. [↑](#footnote-ref-7)
8. IBM, ibid. [↑](#footnote-ref-8)
9. In this report, ‘data’ is used as a collective singular as well as a plural and as a mass noun. [↑](#footnote-ref-9)
10. US National Science Foundation, Critical Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA), Program Solicitation NSF 14-543. See <https://www.nsf.gov/pubs/2014/nsf14543/nsf14543.pdf> at p3. [↑](#footnote-ref-10)
11. The Information Accountability Foundation, Origins of Personal Data and its Implications for Governance” see <http://informationaccountability.org/wp-content/uploads/Data-Origins-Abrams.pdf>. [↑](#footnote-ref-11)
12. Evan Schwartz*, Finding our way with digital bread crumbs*, MIT Technology Review, 18 August 2010. See https://www.technologyreview.com/s/420277/finding-our-way-with-digital-bread-crumbs/.

 [↑](#footnote-ref-12)
13. Julie Lane and Ors (eds), *Privacy, Big Data, and the Public Good*, Cambridge, 2014 at p193. [↑](#footnote-ref-13)
14. Shoshana Zuboff, *Big Other: Surveillance capitalism and the prospects of an information civilization*, (2015) 30 Journal of Information Technology, *30* (1), 75-89, p75 - p77, 2015. [↑](#footnote-ref-14)
15. Luciano Floridi, *Four challenges for a theory of informational privacy*, (2006) 8 Ethics and Information Technology, p109, p111. [↑](#footnote-ref-15)
16. Other V’s are attributed but these four are key drivers. See <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf> . [↑](#footnote-ref-16)
17. See https://ec.europa.eu/digital-single-market/en/policies/big-data. [↑](#footnote-ref-17)
18. Rob Kitchin, *Big Data, new epistemologies and paradigm shifts*, Big Data & Society, April-June 2014, p1. [↑](#footnote-ref-18)
19. Danah Boyd and Kate Crawford, *Critical questions for Big Data*, Information, Communication and Society, Vol 15, No. 5, p662 at p663. [↑](#footnote-ref-19)
20. There are also significantly contrary views. For example, the EU Article 29 Data Protection Working Party Statement on the impact of the development of big data on the protection of individuals on processing of their personal data in the EU, 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 EU or national levels which aim to make these benefits real for individuals in the EU, whether individually or collectively.’ See http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221\_en.pdf . [↑](#footnote-ref-20)
21. <https://ec.europa.eu/digital-single-market/en/making-big-data-work-europe> [↑](#footnote-ref-21)
22. Ibid. [↑](#footnote-ref-22)
23. <https://books.google.com.au/books?h=en&lr=&id=Gf6QAwAAQBAJ&oi>=.

 fnd& pg=PR15&dq=machine+learning+definition&ots=2HpfNhnHJ0&sig=LWPG20hBU4OiF4JDZ8OtHx13Y0#v=onepage&q=machine%20learning%20definition&f=false. [↑](#footnote-ref-23)
24. Data that relates only to one individual. [↑](#footnote-ref-24)
25. Jean-Luc Chabert (ed), *A History of Algorithms: from the Pebble to the Microchip*, Berlin, 1999 at p1. [↑](#footnote-ref-25)
26. Ibid. [↑](#footnote-ref-26)
27. Felicitas Kraemer and Ors*, Is there an Ethics of Algorithms?* (2011) Ethics and Information Technology 13(3), p251 [↑](#footnote-ref-27)
28. See http://mathworld.wolfram.com/Algorithm.html. [↑](#footnote-ref-28)
29. Brent Mittelstadt and Ors, *The Ethics of Algorithms: Mapping the Debate*, 2016, Big Data and Society, p1. [↑](#footnote-ref-29)
30. See, for example, https://www.scientificamerican.com/article/dating-services-tinker-with-the-algorithms-of-love/. [↑](#footnote-ref-30)
31. See, for example, https://motherboard.vice.com/en\_us/article/4x3pp9/the-simple-elegant-algorithm-that-makes-google-maps-possible. [↑](#footnote-ref-31)
32. See, for example, http://mitsloan.mit.edu/media/Lo\_ConsumerCreditRiskModels.pdf. [↑](#footnote-ref-32)
33. Brent Mittelstadt and Ors, *op cit*, p1. [↑](#footnote-ref-33)
34. Ibid. [↑](#footnote-ref-34)
35. US Federal Trade Commission, ‘Big Data: A Tool for Inclusion or Exclusion’ 2016. <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>. [↑](#footnote-ref-35)
36. For example, minority groups not well represented in a particular dataset, may be subject to decisions and predictions subsequently taken. [↑](#footnote-ref-36)
37. Brent Mittelstadt and Ors, *op cit*, p4. [↑](#footnote-ref-37)
38. 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. [↑](#footnote-ref-38)
39. Brent Mittelstadt and Ors, *op cit*, p8. [↑](#footnote-ref-39)
40. See, for example, http://www.predpol.com/how-predpol-works/. [↑](#footnote-ref-40)
41. http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/. [↑](#footnote-ref-41)
42. http:// www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm. [↑](#footnote-ref-42)
43. Adopted by General Assembly resolution 45/95, 14 December 1990. [↑](#footnote-ref-43)
44. See Article 7. [↑](#footnote-ref-44)
45. See Article 9. [↑](#footnote-ref-45)
46. See Article 10. [↑](#footnote-ref-46)
47. UN document E/CN.4/1990/72, 1990. [↑](#footnote-ref-47)
48. <http://standards.ieee.org/news/2016/ethically_aligned_design.html>; <http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf>. [↑](#footnote-ref-48)
49. United Nations, Human Rights Council 34th Session, A/HRC/34/L.7/ Rev.1, Agenda Item 3 Protection of all Human Rights Civil, Political Economic, Social and Cultural Rights including the Right to Development, 22 March 2017. [↑](#footnote-ref-49)
50. See <http://opendatahandbook.org/guide/en/what-is-open-data/> The full definition can be located at http://opendefinition.org/od/2.1/en/. [↑](#footnote-ref-50)
51. See https://okfn.org/opendata/. [↑](#footnote-ref-51)
52. See https://creativecommons.org/licenses/by/4.0/. [↑](#footnote-ref-52)
53. Australia, New South Wales Government, *Open Data Policy,* Department of Finance & Services*,* 2013. [↑](#footnote-ref-53)
54. See https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government. [↑](#footnote-ref-54)
55. See https://www.opengovpartnership.org/open-government-declaration. [↑](#footnote-ref-55)
56. On a non-binding, voluntary basis. [↑](#footnote-ref-56)
57. Op cit, n17. [↑](#footnote-ref-57)
58. See https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-. [↑](#footnote-ref-58)
59. Ibid. [↑](#footnote-ref-59)
60. Eckersley, P. (2010). How unique is your web browser? *Privacy Enhancing Technologies,* 2010. [↑](#footnote-ref-60)
61. Zuboff, S. *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization,* Journal of Information Technology (2015) 30, 75–89*. doi:10.1057/jit.2015.5,* 17 Apr 2015  . [↑](#footnote-ref-61)
62. Corporations and governments do not necessarily need to be forced to provide privacy protections. For examples of ethical approaches adopted by companies see <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> [↑](#footnote-ref-62)
63. Pew Research Center, Americans’ Views on Open Government Data, report by J. B. Horrigan and L. Rainie April 21, 2015. [↑](#footnote-ref-63)
64. Ibid. [↑](#footnote-ref-64)
65. Relates only to one individual. [↑](#footnote-ref-65)
66. <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>. [↑](#footnote-ref-66)
67. Available at http://www.ohchr.org/Documents/Issues/Privacy/SR\_Privacy/ReportSR\_SupportingDocuments.pdf. [↑](#footnote-ref-67)
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 of the time could be re-identified using only his date of birth, gender and ZIP. In fact, 87% 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”. <http://www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_06-2005_testimony_sweeney.pdf>. See also Sweeney L, Matching Known Patients to Health Records in Washington State Data, 2012 at <http://dataprivacylab.org/projects/wa/1089-1.pdf> and <http://dataprivacylab.org/index.html>; Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems , 10* (05 ). [↑](#footnote-ref-68)
69. Privacy International, *Related Privacy 101s*: [*Data Protection*](https://www.privacyinternational.org/node/44)*,*https://www.privacyinternational.org/node/8.

 [↑](#footnote-ref-69)
70. Even if anonymised, this does not remove relevancy of privacy principles and considerations such as ‘consent’. [↑](#footnote-ref-70)
71. Biddle, S. (2017, June 20). *The Intercept.* https://theintercept.com/2017/06/19/republican-data-mining-firm-exposed-personal-information-for-virtually-every-american-voter/. [↑](#footnote-ref-71)
72. <https://www.economist.com/news/briefing/21711902-worrying-implications-its-social-credit-project-china-invents-digital-totalitarian>; Financial Times Beijing, July 2017 https://www.ft.com/content/f772a9ce-60c4-11e7-91a7-502f7ee26895 [↑](#footnote-ref-72)
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’. [↑](#footnote-ref-73)
74. US Senate Committee on Commerce, Science, and Transportation, *A Review of the Data Broker Industry: collection, use, and sale of consumer data for marketing purposes,* December 18, 2013. <http://educationnewyork.com/files/rockefeller_databroker.pdf>. [↑](#footnote-ref-74)
75. United StatesFederal Trade Commission, *Data Brokers**- A Call for Transparency and Accountability* May 2014 at <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>. [↑](#footnote-ref-75)
76. Lauristin, M. *Draft Report on the proposal for a regulation of the European Parliament and of the 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. [↑](#footnote-ref-76)
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 Chrome. Other sites detect and block individuals who visit with ad blockers installed. (Howe, D., Zer-Aviv, M., & Nissenbaum, H. (n.d.). *AdNauseam*. From https://adnauseam.io/). [↑](#footnote-ref-77)
78. The TOR router obscures who communicates with whom (*i.e.* telecommunications metadata), but isn't widely used. Some browsers (such as Firefox and Brave) include a "private browsing" mode which obstructs data collection. EFF's Privacy Badger and NYU’s TrackMeNot are highly effective, but are not widely adopted. [↑](#footnote-ref-78)
79. Office of the Australian Privacy Commissioner *Community Attitudes to Privacy*, May 2017; Office of the Australian Privacy Commissioner *Community Attitudes to Privacy*, October 2013; Deloitte, *Trust starts from within - Deloitte Australian Privacy index 2017,* May 2017. [↑](#footnote-ref-79)
80. Goodfellow, I. J., Shlens, J., & Szegedy, C. *Explaining and Harnessing Adversarial Examples.* ArXiv preprint, 2014. [↑](#footnote-ref-80)
81. Pentland, A., Shrier, D., Hardjono, T., & Wladawsky-Berger, I. *Towards an Internet of Trusted Data: A new Framework for Identity and Data Sharing.* Input to the Commission on Enhancing National Cybersecurity, 2016. [↑](#footnote-ref-81)
82. Privacy is defined as "the requirement that information about people ('personal information') flows appropriately, here appropriateness means in accordance with informational norms … Social contexts form the backdrop for this approach to privacy…." Barocas, S., & Nissenbaum, H., Big data's end run around anonymity and consent. In *Privacy, big data, and the public good: Frameworks for engagement* (pp. 44-75). Cambridge University Press, 2014. [↑](#footnote-ref-82)
83. Kaye, J., Whitley, E.A. Lund, D., Morrison, M., Teare, H. and Melham, K., *Dynamic consent: a patient interface for twenty-first century research networks* EJHGOpen, European Journal of Human Genetics (2015) 23, 141–146; doi:10.1038/ejhg.2014.71; published online 7 May 2014. [↑](#footnote-ref-83)
84. Recent advances in cryptography allow multiple parties to compute together a function of their private inputs, then reveal only the well-defined outcome. There are very general tools, based on multiparty computation (such as Damgård, I., Pastro, V., Smart, N., & Zakarias, S. (2012). Multiparty computation from somewhat homomorphic encryption *Advances in Cryptology–CRYPTO*) and homomorphic encryption (such as Lauter, K., Laine, K., Gilad-Bachrach, R., & Chen, H. (2016, March). *Simple Encrypted Arithmetic Library (SEAL)*. From https://www.microsoft.com/en-us/research/project/homomorphic-encryption/#). Most do not run sufficiently fast for big datasets, but simpler variants may in the future. There are many specific protocols that solve specialised problems on large datasets. The general notion of computing on encrypted data works very well for simple computations on one dataset, but can be infeasible for complex computations or datasets distributed over several locations. [↑](#footnote-ref-84)
85. Weigend, A., *Data for the people: how to make our post-privacy economy work for you.* New York: Basic Books 2017. [↑](#footnote-ref-85)
86. Barocas, S., & Nissenbaum, H., Big data's end run around anonymity and consent. In *Privacy, big data, and the public good: Frameworks for engagement* (pp. 44-75). Cambridge University Press, 2014. [↑](#footnote-ref-86)
87. Available at http://www.ohchr.org/Documents/Issues/Privacy/SR\_Privacy/ReportSR\_SupportingDocuments.pdf. [↑](#footnote-ref-87)