*Before the*
National Telecommunications & Information Administration
Department of Commerce
Washington, DC

| | |
|---|---|
| *In re*<br><br>Request for Comments on Big Data and and Consumer Privacy in the Internet Economy by the National Telecommunications and Information Administration | Dkt. No. 140514424-4424-01 |

**COMMENTS OF**
**THE INFORMATION ACCOUNTABILITY FOUNDATION**

Pursuant to the request for comments by the National Telecommunications & Information Administration (NTIA) published in the Federal Register at 79 Fed. Reg. 32,714 (6 June 2014), The Information and Accountability Foundation (Foundation) respectfully submits the following comments for review.[1]

**Introduction**

Big data and its necessary companion, analytics, promise to provide innovation to U.S. and global business, science, research and education. Powerful algorithms have already been used to help identify individuals in need of social services, detect fraudulent transactions, forecast the effects of natural disasters, support prognosis of medical conditions, recognise patterns in scientific research and discover trends in consumer

[1] The Foundation is a non-profit research and educational organisation founded to integrate accountability and data stewardship as key components of data governance to foster both data protection and information-driven innovation. A list of the Foundations supporters is available at http://informationaccountability.org/.

demand. Big data and analytics have begun to benefit all aspects of society—from understanding medicine to managing natural resources and improving education.

While big data and analytics in some instances may pose risks, the failure to use it to address significant issues in various contexts, such as healthcare, research, education and development will deny individuals and society of potential benefits. Thoughtful guidance that takes into account the realities of big data and analytics will allow organisations to use analytics in an effective and responsible manner to provide long-term solutions that can be adjusted over time. Developing a governance framework that incorporates individual interests in privacy with the potential of this processing power will make it possible to realise the significant and, in some cases, still unanticipated benefits of big data and analytics.

**I. Broad Questions Raised by the Big Data Report and the PCAST Report**

*(Questions 1-6)*

Modern information processing ranges from transactions to statistics to advanced analytics that we call big data. Moreover, big data may involve all sorts of data, including sensitive data, personal data, and varying levels of de-identified or aggregate data. Privacy must be respected in all forms of processing that affect individuals. Furthermore, the principles that govern privacy protections should remain the same. However, the more complex the processing, the less visible it is to the individual, the more the burden should shift from individual control to responsible stewardship. The governing principal of data stewardship is accountability.[2]

[2] A comprehensive description of accountability is best found in the following source: Centre for Information Policy Leadership (2009), "Data Protection Accountability: The Essential Elements" (The "Galway Paper"),

An accountable organisation demonstrates commitment to accountability, implements data privacy policies linked to recognised external criteria, and implements mechanisms to ensure responsible decision making about the management and protection of data. The essential elements as described in the "Galway Paper" are as follows:

1. Organisation commitment to accountability and adoption of internal policies consistent with external criteria.

2. Mechanisms to put privacy policies into effect, including tools, training and education.

3. Systems for internal, ongoing oversight and assurance reviews and external verification where appropriate.

4. Transparency and mechanisms for individual participation.

5. Means for remediation and external enforcement.

The Foundation believes that big data is best governed when it is seen as having at least two distinct phases. The first is a "discovery phase" in which focused research offers new insights. The second is an "application phase". While governance is applicable to both phases, the level of impact on the individual differs across the two. The discovery phase is where one finds correlations between data sets that would not be visible without the muscle of modern high speed computing and advanced analytic processes and technologies. In the discovery phase, one is not applying those insights but only conducting the research to illuminate them. Any implementation of the insights would occur in the application, not the discovery, phase. The discovery phase typically begins with a repurposing of data already in existence. The discovery phase is not usually personally impactful.

Research typically does not affect the individual, whereas the application phase may be more likely to affect the individual. Any law enacted to protect individuals in the

---

http://www.huntonfiles.com/files/webupload/CIPL_Galway_Accountability_Paper.pdf.

context of big data analytics should consider these different phases.[3]

Traditional privacy principles are based on the vast majority of data coming directly from consumers. Significant evidence exists that a great deal of data today does not come from individuals in a manner where those individuals are participatory. The Foundation issued a data taxonomy paper that categorises data into four classifications.[4] Those classifications are as follows:

- Provided;
- Observed;
- Derived; and
- Inferred.

Big data governance should take into consideration the attributes of how the data originates.

As stated in public documents, the Consumer Privacy Bill of Rights is based on the vast majority of data falling into the provided classification. (Below, in section II, we have discussed issues this creates for each principle.)

Analytics processes have been applied to personal data in the United States since at least the 1980s, when MDS and Fair Issacs first developed bankruptcy scores. However, analytics took a major leap forward when technicians developed the skill

---

[3] Abrams, Martin, Meg Leta Ambrose and Paula Bruening (February 2013), "Big Data and Analytics: Seeking Foundations for Effective Privacy Guidance A Discussion Document", Centre for Information Policy Leadership, Washington, DC, http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf.

[4] See appendix of Abrams, Martin (2014), "The Origins of Personal Data and its Implications for Governance", OECD, http://informationaccountability.org/wp-content/uploads/Data-Origins-Abrams.pdf.

necessary to use unstructured data in models sparked the new big data age.[5] Big data constitutes a fundamental paradigm shift from earlier analytic processes. It is not just more data. It is complete and diverse data sets that involve making decisions without the necessity of the causation link. The processing generates new hypotheses rather than just validating existing hypotheses. These differences challenge traditional privacy principles but do not break basic objectives. When discovery is conducted, big data analytical processes should not utilise data that is inappropriate for purpose. The data should be stored, accessed and processed in a secure manner. In addition, the data should be understood and potential risks to individuals should be mitigated. A recent article in "International Data Privacy Law" by K. Krasnow Waterman and Paula Bruening[6] clearly articulates the risks in the discovery phase. The basic objectives of privacy protection apply to application as well and may be applied in a more traditional manner without restricting the innovation that comes from big data.

Big data governance relies the most on the last Consumer Privacy Bill of Rights principle, accountability. The data management programme that best describes what accountability in practice might look like was issued in Canada[7] and modified for Hong Kong.[8] Canadian private sector privacy law includes accountability as the first principle.

---

[5] McCabe, Bruce (2007), "The Future of Business Analytics", S2 Intelligence.
[6] Waterman, K. Krasnow and Paula Bruening (2014), "Big Data Analytics: Risks and Responsibilities", International Data Privacy Law, Volume 4, issue 2, pp. 89-95.
[7] Office of the Privacy Commissioner of Canada, Office of the Information and Privacy Commissioner of Alberta, Office of the Information and Privacy Commissioner of British Colombia (2013), "Getting Accountability Right with a Privacy Management Program", https://www.priv.gc.ca/information/guide/2012/gl_acc_201204_e.asp.
[8] Hong Kong Privacy Commissioner (2014), "Best Practice Guide on Privacy Management Programmes", http://www.pcpd.org.hk/engindex.html.

Any law governing big data should create the legal obligation for comprehensive programmes and grant authority to enforce that obligation.

The Foundation is conducting a big data ethics project. Our initial report will be presented on or around 15 September 2014. The first part of the project includes the creation of a common ethical frame that takes into consideration the interests of both stakeholders within the organisation, as well as outside the organisation. The external stakeholders include agencies charged with privacy enforcement, government agencies charged with economic growth, individuals both as economic and social beings, and individuals aggregated into groups. An approximation of the common ethical frame as it links to big data would be based on the following characteristics:

- Valuable;
- Progressive;
- Sustainable;
- Respectful; and
- Fair.

Privacy issues are wrapped into each of the values above.

**II. Specific Questions Raised by the Big Data PCAST Report** *(Questions 7-12)*

The Foundation supports the principles articulated in the Consumer Privacy Bill of Rights with specific areas of emphasis.

1. **INDIVIDUAL CONTROL: Consumers have a right to exercise control over what personal data companies collect from them and how they use it.** The principle speaks to "data collected from them" [consumers]. This is data that the

Foundation's analysis refers to as provided data.[9] Notice and consent is traditionally addressed by this data class. However, this class of data only covers a portion of the data related to individuals. Individual opportunities for notice and consent and control becomes more problematic when one tries to address data classifications suggest as observed, derived and inferred.

2. **TRANSPARENCY**: **Consumers have a right to easily understandable and accessible information about privacy and security practices.** Transparency is a significant challenge in the big data environment. Privacy notices serve the dual purpose of defining the breadth of uses for privacy enforcement agencies and providing an understanding of data use for individuals. Regulators require a full description of uses—which leads to very long notices—while consumers need an understanding at a glance. This creates a challenge in devising a complete notice that is also simple, easy to understand and conveys the key issues for individuals. Ten years ago, the concept of multi-layered notices was considered a best practice to make privacy notices clearer. Since then, data use—particularly big data—has challenged an organisation's ability to communicate the possible uses in a concise manner. This principle requires a concerted effort to create new communications concepts.
3. **RESPECT FOR CONTEXT: Consumers have a right to expect that companies will collect, use and disclose personal data in ways that are**

[9] Abrams, Martin (2014), "The Origins of Personal Data and its Implications for Governance", OECD, http://informationaccountability.org/wp-content/uploads/Data-Origins-Abrams.pdf.

**consistent with the context in which consumers provide the data.** Respect for context is very similar to the OECD concept of limiting purposes consistent for which the data was originally collected. By further comparison, the Opinion by the European Union's Article 29 Working Party on purpose limitation states research as always a compatible purpose.[10] The Foundation agrees. The Foundation also supports a two-phase approach in which research and application are both governed—but governed serially. However, the Foundation does not believe that consent works for governing the research phase. Instead, the Foundation believes that governance for this phase should be governed by the accountability principle. The utility of consent-based governance in the application phase is dependent on the particular application. For example, consent has greater applicability to marketing than it does for cyber security or fraud prevention.

4. **SECURITY: Consumers have a right to secure and responsible handling of personal data.** Foundation believes this principle works for big data without the need to make significant adjustments.
5. **ACCESS AND ACCURACY: Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data is inaccurate.** The Foundation believes that more robust access to data is

[10] Article 29 Data Protection Working Party (2013), "Opinion 03/2013 on purpose limitation", European Commission, Directorate General Justice, Brussels, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

necessary in an age of big data. While individuals may be aware of the data that is provided by them, they also have a strong interest in seeing observed, derived and inferred data as well. Access is not an absolute right. For example, there are some instances where access to data used for fraud prevention would put those applications at risk.

6. **FOCUSED COLLECTION: Consumers have a right to reasonable limits on the personal data that companies collect and retain.** As stated in Section I, collection is not the best way to characterize the means in which data originates. The Foundation agrees that there should be no secret aggregations of data. But this principle needs to be amended to deal with the fact that it really speaks to an era when data was provided by individuals in a manner in which they were aware. It needs to be updated to be relevant to data observed, derived and inferred as well.
7. **ACCOUNTABILITY: Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights.** Likewise, as mentioned above, the Foundation believes that accountability is the overarching principle that operates to implement effective privacy protections. We believe that the "Responsible Data Use Model" as it is described in the big data report is a synonymous with the concept of accountability.
8. **DE-IDENTIFICATION:** The Foundation believes de-identification has value even if advanced technologies make re-identification a risk. The PCAST report creates doubts about the ability of de-identifying technologies alone to obscure

perfectly the identity of individuals in a big data research project. If other individuals or organisations have the resources and motivation, including access to other data sets, they can re-identify the subjects of the research. The Foundation agrees with that assessment.

However, depending on the type of big data being deployed, technological de-identification can be combined with policy requirements (e.g., internal access controls, separation of duties, contract restrictions, in addition to management oversight, policies, procedures, and training) to provide adequate protection for individuals. This is particularly the case where de-identified data is used internally or shared with third parties under appropriate contract restrictions. The use of de-identification when combined with effective policy has been used in analytics for forty years to ensure appropriate protections. For example, marketing of credit cards has made use of simple de-identified files to analyse which credit offers should go to which consumers. In such cases, names, addresses and other identifiers were replaced with sequence codes. The use of sequence codes protected the identity of the millions of individuals whose data was being processed while allowing the analysis to go forward. The de-identification process also allowed for compliance with the Federal Fair Credit Reporting Act (FCRA). The penalties in the law are sufficient to assure effectiveness of de-identification. When combined with algorithmic techniques to obscure personal data and large data sets offer promising countermeasures to prevent the potential re-identification of data.

The Netflix study is an example where contracts (a policy proscription)

could have added protection to the technological anonymization. In this case, Netflix shared anonymized files with researchers in the quest for better predictive algorithms. University of Texas researchers, operating without any restrictions, used other data to re-identify a number of consumers. If contracts had been in place with researchers that prohibited re-identification, the penalties for breach of contract and would likely have prevented the re-identification and certainly the publication of the information.

By analogy, as prudent individuals, we lock the doors to our houses, engage security systems, and put security signs in our front lawns. We do so with the knowledge that thieves can pick locks and overcome security systems. But we also know that thieves will pick the easiest mark, and each security measure contributes to the decision to move on to an easier target. As a mitigator of risk technologies, de-identification has to be better but not perfect. The EU Working Party papers on anonymization and legitimate interests point to obscuring technologies as useful risk mitigation strategies. The Foundation agrees.

9. **CONCERN FOR POETENTIAL SOCIAL HARMS.** Both the PCAST report and the Federal Trade Commission have raised the issue of big data analytics increasing the potential for discrimination based on consumer and civil rights laws in areas such as fair lending, employment and housing. The same issues were raised pursuant to the adoption of credit scores and scorecards in the late 1980s and early 1990s. As described in this paper, the Foundation believes big data governance is applicable to both a discovery and application phase. In the application phase, one consideration is to determine if the processing will be fair.

Processing that has an effect of discriminating based on prohibited grounds would surely be found to be inappropriate and unfair. Processing that has the effect of discriminating even if not targeted to prohibited grounds is a more nuanced question. However, the issue is not the processing but, rather, the individuals and organisations that use insights in a prohibited discriminatory fashion. The fact is that big data has the potential to isolate underserved populations and suggest methods of serving those populations more effectively. For example, the Mobile, Alabama, school district is already using big data to reduce dropout rates, enhancing opportunities for thousands of children. To the Foundation, the solution lies in effective internal governance and robust and fair oversight and enforcement.

## III. Possible Approaches to Big Data Suggested by the Reports and the Big Data Workshops (Questions 13-20)

### A Framework for the Ethical Uses of Big Data

The Foundation is currently conducting a big data ethics project with assistance from HP Labs in Bristol, England, and a team from business and academia to draft ethical guidance for responsible and answerable organisations. Initial results will be released in time for the FTC big data workshop during 15 September 2014 for the purpose of receiving multi-stakeholder input. The results will also be discussed at the 36th International Conference of Data Protection and Privacy Commissioners in Mauritius on a big data ethics panel moderated by the Foundation's Executive Director Martin Abrams.

Big data ethics do not only address data protection and privacy.[11] The UN Charter of Human Rights reminds us that privacy is only one of a number of rights affected by the growing use of information. These rights include standard of living, education, healthcare and sharing in the benefits of scientific advances as well as freedom of religion, expression, family life and association. All of these rights are affected by big data analytics and must be part of the ethical considerations to be mapped and established by this project.

**The Project's Domain is Big Data**

The project starts with the assumption that there are risks associated with most forms of data processing. This is particularly true when data is used in mathematical modelling of human behaviour. However, the domain of the project is focused on big data analytics. There are many definitions of big data that have been articulated in different fields over the past few years. Cukier's and Mayer-Schoenberger's definition is probably the most applicable one. They say "big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organisations, the relationship between citizens and government, and more." They focus on the ability of big data to change the manner in which key questions are confronted by looking for interesting correlations between data sets that would not have been visible using legacy systems and intuition. Advanced analytic processes that make it possible to use unstructured data to conduct legacy forms of analysis with greater and more diverse data is included in the definition

[11] For a full discussion of big data ethics, see Richards, Neil and Jonathan King (2014), "Big Data Ethics", *Wake Forest Law Review, 2014,* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174.

of big data, but the most interesting questions are those focused on the processing of data that makes what would have been considered impossible insights possible. This class of big data projects begin with a set of questions to be explored but, unlike other forms of analytics or processing, do not always begin with a narrow hypothesis to be verified. First and foremost, it looks for correlations in data sets that would not have been obvious or clear from human intuition or more traditional processes.

Some advanced analytic processes follow more traditional statistical techniques that use huge data sets to verify narrow hypothesis. As stated above we are more interested in the more cutting-edge forms of big data. The outcomes from this project may be helpful in these traditional domains, but these forms of processing fall outside the scope of this project.

**Our Aim is Effective Information Policy not Just Privacy Policy**

There are numerous ongoing debates about whether privacy is more limited than data protection, or whether data protection should be seen as a fundamental right that supports other fundamental rights. The Foundation will not be limited by the debate over what is and is not a privacy interest. For us, inappropriate processing and the absence of beneficial processing because of reticence are both a problem.

**A Unified Ethical Framework**

The project's first phase will focus on developing a unified ethical framework that encompasses the varied interests of internal and external stakeholders of an organisation in a responsible manner. That will mean consideration of a data scientist's desire for progressive results as well as a privacy officer's desire for appropriate, legal and moral data use.

**Five General Principles of a Common Ethical Framework**

The project is currently defining five characteristics that will serve as the outline of the framework:

- Valuable;
- Progressive;
- Sustainable;
- Respectful; and
- Fair.

Once defined, we will begin developing interrogation questions in order to test their utility.

**Next Steps and Timing**

The Foundation expects to vet these principles and interrogation questions with stakeholders of the project in early August. We will then begin vetting externally with the following events in mind:

- Late August – Session with the policy staff at the Office of the Federal Privacy Commissioner of Canada;
- 15 September – Ideally, the Foundation will have an opportunity to discuss the ethical frame at the FTC workshop on big data;
- Late September – Discussions with various authorities in Europe; and
- 10 October – Big data ethics session at the 36th International Conference of Data Protection and Privacy Commissioners.

**Second Phase**

Once the ethical frame and interrogation questions are fully vetted and updated,

they will form the basis for developing a draft code of conduct.

The Foundation looks forward to sharing the common ethical frame in September and submitting the draft code of conduct at the completion of the project.

Respectfully Submitted,

Martin Abrams
John Kropf
The Information Accountability Foundation
A Non-Profit Charitable 501(c)(3) Organisation
Tax ID: #46-1416947
1811 River Heights Dr.
Little Rock, AR 72202
(972) 781-6667

4 August 2014