Big Data, Analytics, and the Human in the Middle

Donald M. Allen

Technical Services
Cisco Systems, Inc.
Colorado Springs, CO 80921-3548, USA

ABSTRACT

Big data and the analysis of that data is being marketed as the next big opportunity for "innovation, competition, and productivity" according to McKinsey and Company. The primary focus of the Big Data opportunity has been on the value of Big Data to businesses that can is achieved through augmenting businesses and government's ability to better understand their customers and the product and services "needs" of those customers. This understanding provides the opportunity to springboard the development of new products and to deliver focused advertising to specific groups of consumers and businesses. The analysis of Big Data is seen as the mechanism that is the catalyst in the development of new, robust customer behavioral models and the discovery of as yet hidden behavioral trends. This paper, however, will focus on the issues related to the human factor in this new data frontier specifically providing a review of the psychological and legal challenges created as a function of the consolidation and analysis of what have historically been disparate data sets. Data storage, data processing, data security, and privacy will be discussed in the context of their impact on corporate and government policy making and the present and future impacts on the human who is both the marketing focus and owner and creator of the resources vital to Big Data. Finally the role that service science can play in shaping the services that are and will be created by Big Data.

Keywords: Big Data, User Experience, Service Science, Security, Privacy

INTRODUCTION

The term Big Data is used to describe the collection of large and potentially complex data sets containing both structured and unstructured data into commonly accessible data sets. It has been described to have the potential to be as important to business and society as internet is today. Big data represents the technology that can support the aggregation of a data that was traditionally stored in siloes, organized around specific applications (business transactions) or tasks (content search) and typically stored in relational data bases or file systems located within the companies data center. Access to these traditional data sets was optimized around specific transactional uses such as business transactions (tasks). Big data has to potential to aggregate a large number of these transactional or task oriented data sets into a single ubiquitous access point where behavioral and statistical analysis techniques can be applied uncover new behavioral patterns and market segments. With the advent of cloud technologies, specifically cloud storage, it has become cost effective for companies, governments, and other organizations to store large volumes of unstructured and semi-structured data "the cloud", typically in off-premise and often times off-country, data centers. The extension of the storage of this data outside of the tradition corporate IT sphere of control and has introduced a number of significant security and privacy issues for consumers (humans), companies, governments,

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2091-6

and service providers (Wood, 2014).

The change Big Data could introduce can be explored using traditional banking systems as an example. A bank typically stores customer's transaction information in a relational data base in their data center. The bank uses this information to maintain customer identification data and track financial transactions. This information can be used to provide its customers with transaction history and to ensure compliance with governmental banking regulations. This data was proprietary to both the customer and to the bank and therefore not accessible outside the specific use that it was intended and banks did not have access to a customer's data or activities in other financial institutions. Using the available information a bank is able to assess a customer's current accounts and determine if there other products the bank has that it could offer. The bank is limited in their ability to develop a complete picture of a customer's current financial positions or banking activities through other financial institutions. The application of Big Data provides the potential for financial institutions to begin to share customer information in a cost effective way. Current laws, as well as technologies, prevent institutions from these activities without customer consent, however, these laws as well as technologies are changing. The complexities and as yet unanswered questions in the areas of privacy and data ownership leave open the opportunity for the sharing of information to become a reality. Consider how many internet users do not consider the privacy and terms of use statements that they click through when signing up for services. Without careful scrutiny on the part of the user or strict government laws and policies to limit such data aggregation there is the potential for people to unwittingly giving consent to activities and they would never consent to if they understood the ramifications. Enabling business to understand the motivation of a person's actions, financial or other, through analysis of aggregated behavior puts that person at a disadvantage if the person enters into a relationship with an unscrupulous company. Clearly there is a need to protect against this behavior through education, law, and governmental policies.

The marketing of Big Data to large companies and the potential for financial windfall and competitive advantage is demonstrated in an excerpt from an invitation to a corporate breakfast briefing sent out by a Big Data consulting firm:

> "Big Data sources are compelling organizations to create their competitive advantage by uncovering new ways to engage with customers, new revenue streams and go to market channels. Explosion of new data sources is enabling organizations to deliver personalized insights to customers that were not previously available."

The corporate marketing of Big Data and the analysis is growing in intensity and has been touted as the next big opportunity for "innovation, competition, and productivity" in business according to McKinsey and Company (Manyika, 2011). The initial focus on Big Data has been primarily in two areas: 1) Efficiency and 2) Security. Efficiency in the Big Data context is manifested through enabling businesses and governments to become more effective in acquiring, retaining, and servicing customers. The issue of security encompasses concerns with data protection, organizational and personal privacy, and data sovereignty. The primary perspective of the ongoing work has been that of the corporation, associations, and government focusing on legal and ethical issues as they pertain to them. This paper will explore Big Data from the perspective of the human in terms of the issues that impact them and how service design and service science can shape the future of how users and service providers navigate the complexities of Big Data.

THE BIG-AGGREGATED DATA DIFFERENCE

One of the core tenants of Big Data is the aggregation of data from various large data sources consisting of both structured and unstructured data into a single repository that can be used to identify trends and patterns in behavior, behavioral models, and data visualizations. This potential value of data aggregation is not new to science; Charles Minard (1869) designed charts that demonstrated the value in the mid 1800's that are still discussed and admired today. Minard's graph of the decreasing size of Napoleons army during its march to Moscow and back (Figure 1) is a canonical example of his work. In his graph Minard overlaid the size of Napoleon's army on a map of the route and included a temperature table to provide a visualization of the interrelatedness of the different measures. To construct this visualization Minard had to manually access this data from a number of different unstructured data sources and combine them into a single graphical representation.

The aggregation and consolidation of different data sets continues to the creating interesting and enlightening data

visualizations. Von Worley's (2009) depiction of the distance to nearest McDonalds in the contiguous United States demonstrates the ease of which large data sets can be obtained and visualized. Von Worley obtained physical location data from AggData, a data service provider that specializes in collecting data from various public resources and consolidating it into a single data set (AggData lists ~2,500 location data sets on their website for businesses ranging from banks to hospitals to movie theaters). In his graph Von Worley augmented the location data with the distance that a reasonable person might drive to dine at a McDonales. This visualization may give comfort to those addicted to McDonald's fast food, an insight into America's problem with obesity, and serves as a demonstration of the ease in which large data sets are available and the techniques that can be applied to the data to demonstrate a theme.

The availability and breadth of Big Data sets will enable the application of new or non-traditional analysis methods. One sure method, Mosaic Theory which holds that individually innocuous pieces of information, when combined with other such pieces of information can result in a composite - a mosaic – producing a clearer concept that when the information is viewed separately. Used for some time in the intelligence, legal, and investment fields, this approach is being applied to address public policy issues. Mathew Connelly of Columbia University is directing a project to identify the inherent bias in declassified information from the United States government. The project called "The Declassification Engine" employs data mining techniques to evaluate public records of publicly release, form secret government consolidated from multiple sources into a single document repository. The approach attempts to identify what the government is trying to keep secret from the public by evaluating redacted content in these documents (Garfield, 2013). The application of Mosaic Theory to the document attempts not to identify trends in government policy not simply from the information present in the data set but also from what has been removed from it with the goal of identifying what the government is attempting to be kept secret.

Technology-enabled humans produce a significant amount of data in a day, week, or year. On average it is estimated that a young adult age 18-24 sends 67 text messages a day or over 2,000 per month. In 2013 it was estimated that over 100 billion emails are sent and received each day worldwide (Radicati, 2013). Facebook, Twitter, browser histories and search engine questions, television viewing behaviors, bank and trading transactions, medical histories, voter and vehicle registrations, driving histories, etc. all add to the data that is available for capture and storage in large data repositories. Understanding the issues with what data can be stored, by whom, with what permissions, with what security, and for how long are critical to not only the development of laws and policies to provide governance but also in protecting and educating the people producing the data.

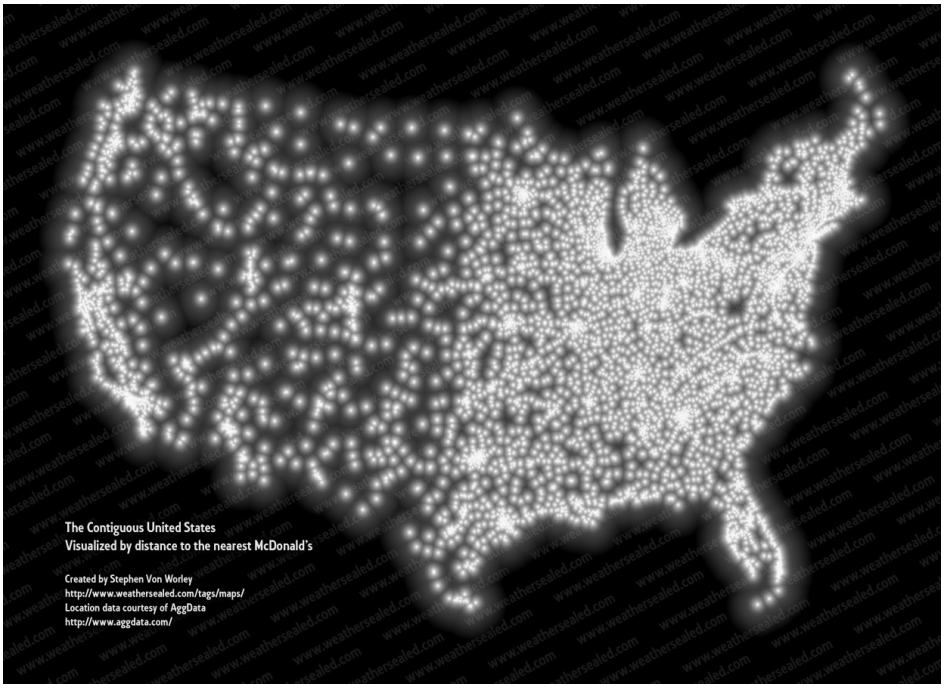

Figure 2: The Contiguous United States Visualized by distance to the nearest McDonald's

THE HUMAN ELEMENT IN BIG DATA

There are a number of factors that combine to form a human's view of Big Data and define the role human play in the Big Data universe. Privacy and Security are common topics are the focus of most Big Data laws, policies, and standards being developed. There are however other topics, with a human focus, that need to be considered in the Big Data discussion such as the monetary value and ownership of the data. This section contains information on Big Data issues as they relate to the human side of Big Data.

Privacy

Personal privacy is a first principal in addressing Big Data people-focused issues. The loss of personally identifiable information (see PII section below) through a breach in security has been sensationalized in the media which has raised awareness in the public that other Big Data issues do not have. Breaches of corporate data have resulted in significant monetary, brand, and consumer confidence losses. Traditional and social medial has stress the importance of properly protecting personal information and government and standards organizations are diligently establishing and updating policies of securing personal information. Governments are increasing the enforcement of privacy laws through the creation and funding of departments focused privacy protection. Standards and government policy makes must continue to refine and update their policies as Big Data analytics evolves. The identification of what is considered private personal data is core to developing the laws and policies required to govern it.

PII: Personally Identifiable Information

The guidance provided by the U.S. National Institute of Standards and Technology (NIST), the United States agency that works with private industry to develop and apply technology, measurements, and standards, definition of privacy is purposefully vague and open to interpretation. The intent of their guidance is to establish boundary conditions within which decisions on what constitutes private information can be discussed. NIST used the term Personally Identifiable Information to identify this category of personal data. NIST identifies Personally Identifiable Information (PII) as "(1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information." (McCallister, 2010). NIST provides an extensive set of examples such as names, personal

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2091-6

identification numbers (social security numbers, passport numbers, etc.), contact information, and also less obvious onse such as a person's place of birth, religion, weight, activities, etc. The less obvious "linked or linkable" information are included to avoid using these characteristics to decrease the probability of using this type of information in aggregate to identify an individual.

In a research project investigating the relationship between a person's ability to manage their own anonymity effectively, Professor Sweeney of the Harvard College demonstrated the need to extend privacy protection to less obvious personal information. Sweeney and her team used data that was publically available in the Personal Genome Project (PGB) records and was able to "re-identify" 84-97% of the profiles in the PGB data base to names and contact information by linking the participants publically available profile information to public records data bases such as voter lists (Sweeney, 2013). Participants in the PGB study were able to withhold personal information from their record in order to maintain their public anonymity. Participants were cautioned that the PGP could not guarantee the confidentiality of any of the information uploaded by the participants and it was the responsibility of participant to maintain their anonymity. The researchers found that participants regularly uploaded information containing personally identifiable information including their names and addresses demonstrating that even though participants were advised of the risks associated with using the PGB data base to store their information they were unable to properly do so on their own. Sweeney has proposed additional methods to enable participants to more effectively determine their vulnerability through the use of programs that determine the re-identification exposure when postal code, data of birth, and gender are made public. The program utilizes the three demographics and searches through lists readily accessible in public records and demonstrates this exposure to participants.

Anonymity

Anonymity, the act of remaining publically unknown, and control over it is an important aspect of a person's life particularly when it comes to navigating the internet. Anonymity is becoming increasing more difficult to maintain as more internet sites are requiring PII information from users in order to utilize a provided server or to obtain "registered" user access to services they provide. When presented with the decision to provide this information people must weigh the risk to benefits of this information. Site sponsors are producing extensive Terms of Use and Privacy policies that users must accept during registration. Design to address the potential users concerns about anonymity protection and security it is uncertain how this content corresponds to those concerns. Issues of trust, site ownership, data security, and concerns over how the provided information will be used as well as the lack of oversight and penalty if it is inappropriately used are common. Article 17 of the new European Union General Data Protection Regulation includes a provision that will give people the "right of erasure" (renamed from the previous term 'right to be forgotten') that has been proposed to address some of these concerns. Under the article people are granted the right to request the "erasure" of personal data relating to them, including links to, copies and any replications of private data provided to a third party. Ensure people have the ability to erase their personal information is a significant first step, the next logical issue is assisting them in tracking the businesses they have provided it to in support of them removing it when they deem appropriate.

The core expectation of erasure is to assist in people in maintain the degree of anonymity the wish to maintain. The identification of what is personally identifiable information to support this removal is challenging as was demonstrated in the Personal Genome Project (Sweeney, 2013). The removable of PII data increases in complexity when data is aggregated from multiple data sources. During the data consolidation process the source of the information along with the governance of its use specified during its collection is likely to be lost. The loss of "context" decreases the ability for entities using the information in the aggregated data store to abide by its original intended use. Technologies have been proposed to address the "loss of context" issue are under consideration; however, the potential impact of their application on system performance is significant and the problem becoming more pronounced as a function of the number of data sources being aggregated (Oliver, 2014).

Removal of personal data from data stores becomes further complicated due to standard data retention, backup, and archive practices (Burn-Murdoch, 2013). Burn-Murdoch noted that at any one time there can be eight or more copies of collected data in different data stores across different locations. The technology challenge is how to locate and anonymize this data when the owner initiates the request and how to demonstrate compliance with government regulations. Current data archive technology introduces additional complexity since backed up data could be stored on tape in a secure, offsite location and therefore not directly accessible.

Security

Securing Big Data, specifically PII, is a leading concern and focus for security experts responsible for protecting a company's reputation and customer's data as well as governments protecting their citizens and their sovereignty. Prior to the advent of Big Data, security standards and policies focused on securing information from breaches or leakage stored within a company data centers. The exposure of a security breach have been well documented and sensationalized in the media, the recent Target data breach affected tens of millions of customers and will have a significant impact on Target's profits in the near term and likely in the longer term. Security research is increasing in areas such as intrusion detection in high speed networks, anomaly detection, and cryptography in response to the increasing number of cyber-attacks. Experts do not anticipate a increase, in the number of security breaches and attempts and Big Data storage will increasing become the targets of such attacks in part due to the large amount of data stored.

Organizations such as NIST in the US, the European Union, the Cloud Security Alliance, etc. have ongoing efforts to identify, prioritize, and address Big Data security issues. These organizations along with others in the private and public sectors are focusing on accessing the risk and approaches to securing Big Data collection, processing, and analysis. These efforts will in turn result in a multitude of laws, policies, standards, and recommendations. The impact on people attempting to understand the understand security issues and how to assess risk have yet to be studied. In a compilation of research on the psychology of security Schneier identifies a number of factors that interact when a person assess risk (Schneier, 2008). Schneier identifies perception as playing a significant role in human security decision making. Factors such as media coverage, familiarity, the recency effect, complexity of the risk, and the ability to make trade-offs in risk avoidance together impact how a person assess their exposure to risk and how they will respond. Combining all of these factors together at a single decision point during which a person must decide whether to share their personal information or not presents a difficult design problem for a company attempting to properly educate a person prior to engaging them in a service relationship.

Data Sovereignty

Data sovereignty laws attempt to protect the data of governments and their citizens from unlawful or illegal use. Typically these laws govern the transfer of data outside of a country's border in order to ensure the data is handled in accordance with the security and privacy laws of the host country. In addition there are laws that govern the availability and access to information stored within systems in a country (e.g., U.S. Patriot Act). These types of regulations place restrictions on the movement of certain types of data across borders, making it seemingly more difficult for a business to adopt cloud applications with datacenters located outside their home country.

Data sovereignty issues have largely been government and in-country focused. These laws are intended to protect citizens, business, and governments of that country and do not extend to citizens when they travel to other countries with conflicting or less stringent privacy and security laws. The violation of a person's home country privacy while in a different country has yet to be contested in the court systems. Can a violated citizen seek judicial judgments in their home country for the transgressions that occurred while not in their home country?

Data Collection

Securing data at the point of collection is a difficult problem to solve. Devices such as cell phones, kiosks, POS terminals, laptops, tablets, and smart watches expose people to one set of possible risk while images from security cameras, Facebook posts, tweets, etc. preset a different type of exposure. Data can be collected clandestinely, without user's knowledge or consent through a compromised sensor or network device. The value and utility of even the most mundane information is not well companies will err on the side of collecting it all since storage is cheap and it is a lot easier to throw away captured data than it is trying to capture it after an event has happened. How can we inform and advise people on securing their personally information when faced with this multitude of breach points? Would an informed person destroy their smartphone, lock themselves in their houses, and turn off their WI-FI. Establishing a balance between an informed user and the risk-rewards of utilizing service is an area where further research is needed.

Data Retention

Data retention, the period that collected data should be stored, and determining when it can or should be deleted is

another complex Big Data issues. Data retention is intricately inter-linked to privacy, security, ownership, and use policies. When a consumer gives a company permission to capture and use data is there or should there be a shared agreement of the period during which that data can be analyzed and put to use. From a Big Data perspective the goal would be to never delete the data since it would be more important to have the data available to use then to have deleted it and not have it available when it might be valuable. A person on the other hand would seemingly want to have more control over how the data was used and when; along with the ability to delete it when they feel it is no longer of value to them.

Data Usage

When a person shares their personal data there is usually an expectation of how it will be used by the person or entity they are sharing it with based on the context it is provided. When a user provides their email address or name and phone number in order to register to win a free iPad, that user expects to receive some email advertisement or phone solicitation from the drawing sponsor trying to sell them something. This is an expected or assumed use by the person which could be verified by reading the fine print on the registration website. The extent to which collected information can be used for other "purposes" and the question of whether a person must give specific consent for each type of use are issues that have not yet been addressed by policy makers. The loss of usage context, like the loss of security context during data aggregation must be addressed as standards and policies evolve.

Data Ownership

"Who owns the data?" is a question whose answer impacts how the other Big Data issues are addressed. This question has gone largely unaddressed outside of personally identifiable information laws, policies, and standards. Clearly a person owns their own PII and can lend it out to businesses and governments for use. In such a situation the information is shared for a specific use with an expectation that it can and will only be used for a prescribed purpose. There is an underlying expectation that the information will either being deleted or anonymized when that purpose is served. People are most familiar with demographic and statistical data when it is aggregated with other data and presented in charges and statistics describing large groups of people. The value of a single data point in a aggregated set of data is relatively small and insignificant and is lost without the context of the other data in the set. The value of non-PII data that is continuously being collected and captured tied to a particular person is not as obvious. As more and more companies amass large data sets of business and people habits and activities, the fuel needed to run the Big Data analytics engines, which other companies will pay them for, there is clearly value to an individual's data and the right to use it. How much would a pharmaceutical company being willing to pay a person who maintains and extensive electronic medical history, including exercise regimen, and eating habits and has taken that companies prescription drug for a decade or two? Who has ownership rights to the data? Ownership of medical history is straightforward having been consistently identified as personal information. Ownership of exercise and eating habits becomes complicated if a person uses and equipment suppliers website or a free for use application to track and store that information. The ability for a person to sell the non-medical information will depend on the usage agreement they accepted when signing up to use the provided service. Creating standard agreements for these situations would enhance Big Data's access to this information and address privacy concerns for the people agreeing to them.

Data Aggregation – Big Data Security Risk

As noted above aggregating data from different data sets has the problem of losing security context. A user who provides data in one context, say to apply for a home equity loan, may not have wanted to allow that data to be consolidated with data they provided when they registered for a sweepstakes drawing to support a marketing campaign for houses for sale in Aruba.

Compliance and Assurance

Addressing Big Data security policies and standards results in another Big Data problem - oversight and regulatory compliance requires administering and reporting on data security activities. Capturing and storing information whenever protected data is access produces new opportunities for Big Data analysis. Glick (2013) postulated that in order for a business to demonstrate compliance with the laws that will eventually be enacted to support privacy and security in the Big Data space each data element that is governed by these laws will need to be tracked. Businesses will need to demonstrate that PII data is secure and has not been accessed illegally, that the data has been

anonymized per owner request, and that it has been properly and completely deleted when required. This produces significant overhead to companies managing Big Data and on the companies developing Big Data tools. Glick posited that technology to address the compliance assurance problem should not be expected any time soon and when developed will not be able to address existing data stores seamlessly.

SERVICES SCIENCE AND BIG DATA

Service science, service innovation, and service design disciples have the opportunity to shape and influence how Big Data addresses these issues through establishing interface between service businesses and the people that use them. The identification and definition of service centered interactions between users and business which data collection, usage, ownership, privacy, and security in the context of service delivery and co-creation of value (Vargo, 2008) would provide a context in which solutions can be developed. Determining the impact of Big Data on the "service-for-service exchange" and service value creation where the value equilibrium between service provider and service consumer may be disrupted by the change in service relationship. Understanding the roles and motivations of the participants in the service relationship and the development of value-in-context models is critical to addressing Big Data application to services.

Human in the Middle

There are a number of organizations that are participating in establish the direction of Big Data. Figure 3 represents these organizations and how the human is in the middle of the Big Data model of the future. Typically the human is viewed as a data source, a target for products and services, or an entity that needs to be protected. The data source is capturing how the human operates in the world which includes both the physical and the virtual.

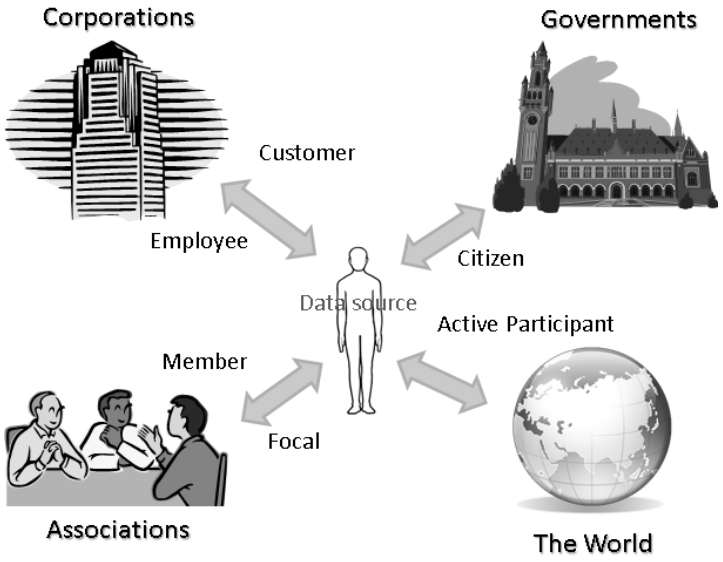

Figure 3: Human in the Middle

Government involvement in the Big Data journey is motivated by the need ensure the protection of their citizens and to protect and defend the country. Laws and policies have been and will continue to be developed that address project of citizen privacy and data sovereignty. One of key provisions of the US PATRIOT Act (USA Congress, 2001) the right to perform "searches of business records" has resulted in difficulties to many global businesses who have data center storage within the United States since this provision allows for the US government to ignore or violate the privacy of citizens of other nations. Governments are also considered one of the biggest potential sources of data that can be used to seed Big Data projects since they potential hold data sources for information for all of the citizens of a country (e.g., Welfare, Taxation, Healthcare, etc.). Since governments have been selling this data for decades it also makes sense that they want to protect this revenue source. Lastly, governments have the opportunity to put the data they already have to the benefit of their citizenry through the proactive delivery of services based on

analysis of the data they already have available to them.

Corporations play multiple roles in Big Data. While can be sources of data, like humans, they will most likely be the biggest consumers of it as well. In addition, corporations need to protect themselves and their employees from unauthorized or illegal use of their data. Consequently corporation's involvement in Big Data policy and standards spans across all areas of Big Data and often times they are at odds with themselves.

Associations are similar to corporations in as a group they are involved in the creation of Big Data policies and standards and are also consumers of them, however typically the same association is not involved in both. Standards bodies focus on developing standards that address security and privacy issues while the collections of like-minded people (e.g., Cancer Society, National Audubon Society, etc.) seek to use those standards and policies to protect their members.

Humans are the basic elements of all of these organizations and are consequently differentially affected through the roles play in each.

BIG DATA IN HUMAN TERMS

In order for service systems to properly adapt to the introduction of Big Data and Big Data analytics as an extension to co-creation of value system the following human impacting issues need to be addressed:

Privacy – people engaging in services requiring the sharing of personal information will need to clearly understand what information is being collected and how it will be safeguarded. The service must provide the ability to opt-in to the use of this data for purposes other than those agreed to when the service relationship was initiated. In addition, the ability to anonymize and delete personal data once it is no longer required for the service must be provided.

Service providers will need to identify all conditions under which a person's privacy could be violated such as compliance with laws that permit access to private information in support of criminal investigations and national security concerns.

Security – service descriptions should identify the security measures that a service provider utilizes to protect collected data including the security policies and standards that are employed and the mechanisms used to ensure compliance. In addition the description should include the identification of risks to the user and what remedies that will be employed should a data security breach occur.

Ownership – an explanation of data ownership and right-to-use policies should be included in the service descriptions along with the identification of the length of time that data will be used, what will happen to the data at the end of this time period and at the end of the service relationship (e.g., deleted, anonymized).

Data Collection – identification of the information that will be collected during the use of the services and how it will be stored must be provided to the service user. This should include how the information will be collected and how the data will be used. The ability to opt-out of certain information from being collected should also be provided.

Notification - services should identify the method the user will be notified should a data security breach occur including breaches where their data was and was not affected. In addition, the method for communicating changes to the service privacy, security, and data usage policies should be identified.

Data Integration – a description of how collected data for the service will or may be aggregated into other data sets and how the user's data will be anonymized when aggregated should be included within the service description. In addition users must be given the opportunity to opt out of the having their data aggregated.

CONCLUSIONS

The purpose of this paper was to identify issues within Big Data and Big Data analytics that impact the people who own, produce, and are targets of this new wave of innovation. An exploration of the impact on big data was briefly explored and as well as areas of the research that should be undertaken to both adopt Big Data capabilities within service science. Service design and innovation have an opportunity to help shape the standards and policies being developed by governments and standards body by identifying and championing the human perspective in these efforts. There is a need for an initiative in service design and service innovation to address Big Data issues, particularly those dealing with best practices that will enable the transition from small data to Big Data.

REFERENCES

Burn-Murdoch, J. (2013), *Data security and privacy: can we have both?* The Guardian Website: http://www.theguardian.com/news/datablog/2013/jul/31/data-security-privacy-can-we-have-both

Garfield, B. (2013), *The Declassification Engine.* On the Media Website: http://www.onthemedia.org/story/declassification-engine/

Glick, B. (2013), *Information security is a big data issue.* ComputerWeekly.com Website: http://www.computerweekly.com/feature/Information-security-is-a-big-data-issue

McCallister, E, Grance, T, Scarfone, K. *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII).* National Institute of Standards and Technology Special Publication 800-122.

Minard, C.J, (1869), Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813. Regnier et Dourdet.

Manyika J., Chui M., Brown, B., Bughin, J., Dobbs R., Roxburgh C. , Hung Byers, A. (2011), *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute Website: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Oliver, A. (2104) *Trust me: Big data is a huge security risk.* InfoWorld Website: http://www.infoworld.com/d/application-development/trust-me-big-data-huge-security-risk-236684

Radicati, S. (2013), *Email Statistics Report, 2013-2017.* The Radicoti Group, Inc. Website: http://www.radicati.com/wp/wp-content/uploads/2013/04/Email-Statistics-Report-2013-2017-Executive-Summary.pdf

Schneier, B. (2008), *The Psychology of Security*, AFRICACRYPT 2008, LNCS 5023, Springer-Verlag, pp. 50-79.

Steel, E. (2014), *Google reinforces online fraud squad with Spider.io acquisition*, Financial Times Website: http://www.ft.com/cms/s/0/352c7d8e-9acc-11e3-946b-00144feab7de.html#axzz2tyuWh4aP.

Sweeney, L, Abu, A, Winn, J. (2013), *Identifying Participants in the Personal Genome Project by Name.* Harvard University. Data Privacy Lab. White Paper 1021-1.

USA Congress (2001), *Uniting and Strengthening America By Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA Patriot Act) Act of 2001*, Public Law 107–56—OCT. 26, 2001. Legislative History—H.R. 3162:

Vargo, S., Maglio, P., Akaka, M. (2008). *On value and value co-creation: A service systems and service logic perspective.* European Management Journal vol. 26, pp. 145–152

Von Worley, S. (2009), Where The Buffalo Roamed: How Far Can You Get From McDonald's? Datapointed.net WebSite: http://www.datapointed.net/2009/09/distance-to-nearest-mcdonalds/

Wood, P. (2014), *How to tackle big data from a security point of view.* ComputerWeekly.com Website: http://www.computerweekly.com/feature/How-to-tackle-big-data-from-a-security-point-of-view