ICTN 6823

Dr. Lunsford

July 24, 2014

Michael Davis

Ethical Questions in Big Data

The growing world of big data presents many interrelated ethical concerns in a business environment. These concerns touch a broad spectrum of ethical concerns including the legal, moral, financial, and social impacts of ethical planning and decision making in emerging situations of data analytics and privacy. As the data continues to grow exponentially in size, the moral concerns will grow as well.

With this speed of change in an area that can have so much affect on people's lives, it is critical that moral decisions be made now and planned for well now to guide future choices later. Information Security experts are on the frontlines of this changing landscape.

This research will focus on an initial look at the goals of an ethical approach in the first place, and then look at several business aspects of ethical decisions relative to big data analysis. These include the analytical process development itself, ethical research principles, and several proposed models of ethical decision-making.

When discussing any important point with a lot of detail with others, it we important to clarify the operating assumptions behind the discussion. This point is especially true when it comes to sets of moral principles that can differ from situation to situation. It is this transparency in the initial discussion that sets up the chances of support of the discussion.

The initial point to clearly understand is the meaning of what business ethics is all about, and really the relevant interpretation of ethics as it applies to business data situations.

Ethics cannot be discussed in a vacuum. When looking at an ethical question, it is important to clearly understand the ethical framework under which the question is poised. It is not enough to announce the imperative "Do Good!" and then carry on, satisfied that your ethical requirements are met. It is important to clearly decide on the context of the question and the associated assumptions of that context before drawing any conclusions about the question.

This is not focused on the idea of relativistic ethics that change with the relative situation. This is more appropriately regarded as a clarification from all involved parties of any assumptions that are being made. Not everyone will see things the same way and this removes an initial problem area.

A standard way of looking at this situation is through the use of imperatives. These are points or situations that are considered absolutely necessary to include. Of course, sometimes two imperatives will be at cross-purposes, so that situation requires still deeper examination of the root factors of the question.

Moral imperatives should be the primary driver for a system of business ethics. These should be listed in a company's vision statement and be guiding principles of action for the company. However, these should also extend to the nuts and bolts of daily action within the company.

Security imperatives occur when situations of public safety temporarily overrule in priority protections of personal privacy. It is essential to note that these are not supposed to be long-term approaches to privacy priorities. It is only the imminent and urgent need that allows this to occur.

There are several direct impact areas for an InfoSec professional, involving both proactive and reactive focus. From the proactive perspective, it falls to this area to design and implement standards and technologies to provide maximal safety and convenience. Or was that convenience and safety?

Big Data is big, and it's growing fast. In the last ten years, the total amount of stored data has grown by five times in size. In the future it is expected to grow even more quickly. "The Big Data technology and services market represents a fast-growing multibillion-dollar worldwide opportunity and is expanding rapidly. A recent forecast shows how that the Big Data technology and services market will

grow at a 27% compound annual growth rate (CAGR) to $32.4 billion through 2017 - or at about six times the growth rate of the overall information and communication technology (ICT) market." (1)

As the total data size increases, information security really becomes the keystone piece of framing many of the ethical considerations concerning emerging technology. In truth, security is the control enforcement behind access to the technology, and the having the control requires the related ethical choices.

Increasing data security responsibilities require proactive examination of the technologies used to protect data in storage and in transmission. It is a significant part of InfoSec management responsibilities to ensure that proper safeguards are in place and then to review those assurances on a continual basis. More to the point, his type of privacy assurance needs to be designed into a closed loop system to eliminate or minimize the chances of big data exposure through company negligence.

It is not enough anymore to simply say that senior management was not aware of the situation. In this day and age, it is a big responsibility of senior management to be aware. The CEOs under fire may well be telling the truth about their ignorance of the particular malfeasance under the microscope, but still many heads have rolled.

This reactive focus for InfoSec is the primary sphere in which the all-important questions of ranking imperatives are debated and it is not enough to merely react. Determining what is the right thing to do to respond to an ethical question that has grown from emergent technological advancement requires proactive coordination and is not an easy task.

That the issues specific to emergent technologies and cannot be overstated in this discussion. In older traditions, people generally considered ethical methodologies based on generally accepted practices of the time. In data ethics, many times these situations have not happened before so there is no established practice to follow. Another situation relative to this might be that the situations

have happened, but in such a remarkably different format that there has not been enough time to allow ethics to develop.

Personal privacy and its limits are under constant debate in this age of cameras everywhere and everything one writes is now almost certainly archived somewhere. The old school threat is now true; if you make a big mistake, it really will go down on your permanent record. It is not very comforting when a major part of an individual's expectation of privacy is that he or she is just a very small individual in the middle of billions of other individuals, all of whom are being spied on too. These indelible reflections on people's actions can affect their lives in significant ways and the herd approach is not very comforting.

Building on the basic understanding of ethics relative to this type of discussion, emergent ethical considerations have some special characteristics. With the growth of big data, situations arise that require the fast and correct application of established principles to new situations that they were not originally designed to cover. It is this reactive application of social and legal principles that make the examination of ethical questions in big data so important.

With the growth of new ethical paradigms supporting the current data structure systems, there are several points to consider. Both moral and security imperatives play a role in societal control of privacy issues. As in most things, a good balance between the needs of each is the best approach.

To initially consider the privacy versus security debate, it must be accepted that even though no one likes it that way, safety does trump privacy. We agree by being part of a community that we have to live by the rules that protect and strengthen that community.

Everyone is entitled to his or her own personal privacy except in case of strong societal social need to intervene. At least in the US and some other countries, personal privacy is a valued right. The exact placement of the line in a question of fairness is an oft hotly debated question.

After considering the community based safety issues, the next underlying point is that an ethical decision should try to be fair to all concerned parties, even those that don't yet know they are concerned. The moral requirement for fairness is

a primary concern in almost all ethical thought systems as a universal concept, so it is a strong contender for a base component for responsive ethics.

Appropriate transparency is also vital to the proper moral management of the considerable amount of sensitive personal data. This auditing component is an important part of a flexible social structure. In times of emergency, a person or group needs to be empowered to make a choice and give a timely response in a relevant period. However, after the emergency passes, a period of review by a group of other knowledgeable people in the field should take place.

This type of respond and review structure allows both a flexible and stable approach over the long run. That is why it is used a lot in our government, such as the Presidents ability to respond quickly if attacked, but it is up to Congress to evaluate the situation and make decisions on a continuing basis.

Security imperatives do have the right to overrule moral imperatives, but only in a limited sense and only temporarily to stabilize a situation. It is also important that when this type of scenario is active, that it is clearly focused on an emergency and that the goal is to return to standard procedures as quickly and as safely as possible.

Several real-world examples should effectively highlight many of the potentials and pitfalls of business ethics and the choices that must be made to effectively respond to the challenges of big data.

A good initial real-world example of the ethical issues with big data was the fairly recent NSA leak and related commentary by former contractor Edward Snowden.

His actions in releasing confidential information publically sparked a firestorm on both sides of the privacy security debate. The US government's intelligence personnel were livid and claimed that this data release had harmed national security interests. By contrast, those in favor of a larger focus on personal privacy were outraged at the wiretapping and irrelevant personal data collection.

This debate continues to actively engage. In a recent post, Mr. Snowden pointed out that analysts regularly shared compromising data and pictures that were discovered during the processing of official information. This extraneous

personal information was in no way included in the required area of observation; it was just collateral damage, so to speak.

This example perfectly illustrates how hard it can be to control the human element. After the Edward Snowden revelations about government surveillance capabilities, “We are at the beginning of a big national conversation on government surveillance and the role private companies play in that surveillance, this is a prime moment for scholars to intervene in these discussions and try to work our way through these competing values of privacy and security, and hopefully come up with a system that can both protect our bodies from terrorist attacks and protect the integrity of our minds – our ability to control what the government and other people know about us, our thoughts and desires, and our ability to try on unpopular ideas.” (2)

To try to respond well to government financing transparency issues, the current administration has created ethics.data.gov. to allow clear and open finance tracking of all publicly funded civic leaders. “The creation of Ethics.Data.gov is part of President Obama's commitment to promoting ethics, transparency, and account-ability across government.” (3)

Another field of application of data ethics is the various types of research that is being done. Review boards at the Federal level provide a good check and balance system that could be enforced as well at the private level. Planning ethical considerations into research and data use proactively is very important. “Researchers whose work is funded by the federal government, or who work for organizations that receive federal dollars, must submit their research proposals to review by a board of members from the community who consider the ethical implications of the research. These boards are often known as institutional review boards or IRBs.” (4)

Constant vigilance is the watchword in this business environment. There are many positives. Business actions follow where the intelligence senior management receives guides their choices. Big data analytics have provided excellent quality information to businesses. This has provided many benefits to consumers as the

businesses can more efficiently control costs and variances if they have good information on what their customers want.

The problem arises when companies adopt an approach where it is assumed that if some analytic information is good, then more must be better.

This approach can quickly cause issues, as people often feel their privacy has been invaded. “Tracking customer preferences and purchases can reveal all kinds of private information, like illnesses, financial problems, even pregnancy, the latter of which created an embarrassing incident for Target in 2012, when it sent coupons to a teenager for baby products by analyzing her shopping patterns. The girl’s father complained, only to find out that his daughter was expecting a baby –something he discovered only because of Target’s custom advertising technology.” (5) The father later apologized to Target, but it was and is an embarrassing situation for them to be exposed for doing such targeted profiling. In interviews about the situation, many people noted that it made them feel uncomfortable.

This example is a good one as it illustrates that mishandling or not correctly interpreting data can have significant consequences. In this case, Target did a good technical job of analyzing the users purchases even to the point where they could send tailored coupons for specific pregnancy stages. However, they seem to have been too preoccupied with their on point analysis to check to discover that the customer was a high school girl and presumably would have wanted her purchases kept private as opposed to the coupons.

These days, social media must be considered as a prime example of the active use of big data in targeting different customers with information used from their existing profile. This is a significant area of focus in the discussion of emergent ethics. With the growth of this medium and the daily impact it has, ethical decisions and methodologies have to be made that keep everyone’s best interests at heart.

As a directed example Twitter data is used on a considerable basis to track very general population trends, far from the individual level. “The strength of tweets as a data source is in the volume; collection through the garden hose API brings in approximately 60,000–100,000 tweets per day. However because tweets are short and often lack context, it is difficult for computers to determine tweet

content automatically. For this reason, researchers primarily use tweet data to conduct population-level research concerned with trends and patterns." (6)

Data collection methods must reflect ethical considerations especially in this era of social media. Facebook has been taken to task over this under allegations of unclear privacy policies and controls for users. Each additional control in this area really limits one of Facebook's primary revenue sources.

"The opportunities that big data offer to impact social, cultural, political change in our lives are promising. There are lots of people who are very excited about that. The challenge on the flipside of that is the risk of unintended consequences." (7) Planning for those unintended consequences is key, but that can be difficult to accomplish as the specific goal is pretty amorphous.

It can be seen as somewhat of a circular progression in that the human element is difficult to quantify, unless you gather a lot of specific data. Doing that does make the information gleaned more targeted, but it also requires a much deeper invasion into the data structure.

The capability the computer gives of being able to assemble these seemingly innocent and insignificant facts into a comprehensive personal profile and to make it widely available gives that information a different significance. Even though limited groups of people may have legitimate reasons to have access to some of those facts for specific purposes, when the facts are all put together into a dossier they become much more personal and invasive. They thus present many of the dangers of other invasions of privacy. The information can be used for purposes other than those for which it was intended. (8)

Big data and analytics, which "marries large data sets, statistical techniques and predictive modeling [to mine] institutional data to produce 'actionable intelligence,'"[1] present big questions to those of us in higher education. Systems now exist that link students' demographic data with data on past and current educational performance, engagement with online course materials, or in-class participation levels to predict, with startling accuracy, specific outcomes such as final grades within a course. When applied to collegiate administration, this raises several

ethical questions, including; What, specifically, is the role of big data in education? How can big data enrich the student experience? Is it possible to use big data to increase retention? To what extent can big data contribute to successful outcomes?

More specifically, we must ask what it means to "know" with predictive analytics. Furthermore, once an administration "knows" something about student performance, what ethical obligations follow? " (9)

One option to follow is the Big Data ethics model; "development of a "Big Data Ethics," a set of four related principles that should govern data flows in our information society, and inform the establishment of big data norms. First, we must recognize "privacy" as an inevitable system of information rules rather than merely secrecy. Second, we must recognize that shared private information can remain "confidential." Third, we must recognize that big data requires transparency. Fourth, we must recognize that big data can compromise identity. (10)

In summary, big data presents a lot of positives as humanity seeks to understand the patterns in the universe around us and use that understanding for the betterment of humanity as a whole. However, it is vital that constant vigilance is required. It must be recognized that there is danger present in this situation as well if proactive methodologies are not taken to ensure proper ethical consideration of the potential invasiveness of big data.

References

1. Martin, E. (2014, March 27). The Ethics of Big Data. *Forbes*.

2. Canon, H. B. (2014, April 1). U.Va. Holds First National Conference to Tackle 'Big Data' Ethics, Law and Policy. . Retrieved July 24, 2014, from https://news.virginia.edu/content/uva-holds-first-national-conference-tackle-big-data-ethics-law-and-policy

3. McFarland, M. (2012, June 1). Ethical Implications of Data Aggregation. *Ethical Implications of Data Aggregation*. Retrieved July 24, 2014, from http://www.scu.edu/ethics/practicing/focusareas/technology/internet/privacy/data-aggregation.html *

4. Richards, N., & King, J. (2014, April 1). Big Data Ethics. *by Neil M. Richards, Jonathan H. King*. Retrieved July 24, 2014, from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174

5. Event Summary: The Social, Cultural, & Ethical Dimensions of "Big Data". (2014, March 17). . Retrieved July 24, 2014, from http://www.datasociety.net/pubs/2014-0317/BigDataConferenceSummary.pdf

6. Riglian, A. (2012, November 1). 'Big data' collection efforts spark an information ethics debate. *'Big data' collection efforts spark an information ethics debate*. Retrieved July 24, 2014, from http://searchcloudapplications.techtarget.com/feature/Big-data-collection-efforts-spark-an-information-ethics-debate

7. Ethics | Data.gov Communities. (2014, January 1). *Data.Gov*. Retrieved July 24, 2014, from https://explore.data.gov/ethics

8. Azzara, M. (2014, April 1). Big Data Ethics: Transparency, Privacy, And Identity–Oh My!. *Big Data Ethics: Transparency, Privacy, And Identity–Oh My!*. Retrieved July 24, 2014, from http://www.cmo.com/articles/2014/4/3/big_data_ethics_tran.html

9. Rivers, C., & Lewis, B. (n.d.). Ethical research standards in a world of big data . *Ethical research standards in a world of big data*. Retrieved July 24, 2014, from http://f1000research.com/articles/3-38/v1 *

10. Stromer-Galley, J. (2014, July 1). Facebook Users or Lab Rats: Ethical Research in the Age of Big Data. *Information Space Facebook Users or Lab Rats Ethical Research in the Age of Big Data Comments*. Retrieved July 24, 2014, from http://infospace.ischool.syr.edu/2014/07/02/facebook-users-or-lab-rats-ethical-research-in-the-age-of-big-data

11. Ethics, Big Data, and Analytics: A Model for Application (EDUCAUSE Review) | EDUCAUSE.edu. (2013, May 1). *Ethics, Big Data, and Analytics: A Model for Application (EDUCAUSE Review) | EDUCAUSE.edu*. Retrieved July 24, 2014, from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application

12. Big Data and Analytics. (n.d.). . Retrieved July 24, 2014, from http://www.idc.com/prodserv/FourPillars/bigData/index.jsp