Incorporating Ethics in Big Data and Analytics Curriculum: A New Degree and New Opportunity

Gwen White
whiteg@xavier.edu

Thilini Ariyachandra
ariyachandra@xavier.edu

Management and Information Systems
Xavier University
Cincinnati, Ohio, 45207, USA

Abstract

Big data, the latest technology innovation promising to enhance the quality of decision making in the economy, presents a new inflection point for business to compete and transform. As more data is integrated, organizations can gain an intimate view of transactions and individuals. This can cause companies to be tempted to push ethical boundaries in search of higher revenue. Meanwhile, business still struggle with a shortage of business intelligence, big data and analytics talent. As academic institutions begin to produce students with the required expertise to deal with big data, it is essential that students get an opportunity to reflect on the impact of ethics on big data as they are exposed to big data curriculum.

Key words: big data analytics, ethical theories, big data analytics curriculum, ethical reflection, big data framework

1. IMPORTANCE OF BIG DATA ANALYTICS

The utilization of data has become part of the essential input needed to run every sector of the global economy along with human capital and other fixed assets (McGuire, Manyika, & Chui, 2012). Big data, the use of torrents of multi-structured data to discern patterns that enhance decision making, has become the latest information technology based innovative trend to be adopted by business to achieve sustainable competitive advantages. Not surprisingly, over 73 percent of organizations have already invested or plan to invest in big data by the end of 2016, while estimates suggest the use of big data in healthcare could save as much as 300 billion dollars for the sector each year (Marr, 2015). In the retail sector, McKinsey Research (Court, 2015) suggests retailers who leverage the full power of big data could increase their operating margins by as much as 60 percent.

In order to seize these benefits, companies are implementing big data infrastructure and analytics as well as attempting to ramp up their existing talent and expertise in this field. Unfortunately, the demand for skills has not translated to employees with big data skills at the pace needed by organizations. This has created a major skills gap (Terdoslavich, 2015; Chiang, Goes, & Stohr, 2012) and pressure is mounting on academia to produce more graduates that meet the demands of this field (Davenport, 2014). Academic institutions and vendors have stepped up to the plate to help ease the shortage.

A global survey of academic institutions in 2014 revealed that increasingly, a wide variety of universities currently offer graduate programs, undergraduate programs, certificates, concentrations and courses on business intelligence, big data and analytics (Wixom, Ariyachandra, Douglas, Goul, & Gupta, 2014). The big data analytics programming offered can take on many forms. Some are more technical than others and others are focused on specific industries like healthcare or focused on customer oriented analytics.

The big data framework submitted by Phillips-Wren, Iyer, Kulkarni and Ariyachandra (2015) is broken down into a process view which includes data sources, data preparation, data storage, analysis and data access/usage (Figure 1). First, sources include the collection of data from social media, video, text, multimedia, external and internal data. Second, data preparation is the cleaning of the data that was collected and prepared for use. Third, the data is stored either in a data warehouse or even a Hadoop cluster. Fourth, the data is analyzed and finally it is analyzed by end users for their required reporting. The big data framework outlines the skills needed for curriculum that emphasizes data analytics (Phillips-Wren, Iyer, Kulkarni, & Ariyachandra, 2015).

Most of the main subject areas suggested in the framework are captured by academic institutions that have a big data related curriculum. The survey on academic institutions in 2014 (Wixom et al., 2015) further indicates that the subjects and concepts that should be covered from a practitioner perspective include both technical and soft skills (Figure 2). As the figure indicates, the technical skill most in demand was SQL and querying skills. Overall, the skill that was identified as most important was communication skills, a soft skill. When considering the content and skills that students currently develop through education, ethics is not identified as a topic of importance.

Ethics in Existing Curriculum
At Xavier University, ethics is incorporated as a part of the Management and Information Science course normally incorporating a two-week timeframe. For example, INFO 220 Management of Information Technology has a two-week ethics component incorporated into the course. Students are required to participate in a group project where they are assigned a case study. The students have to review the case and create a solution along with a video that explains the problem and solution through role play. The video solutions are exhibited to faculty and other students to encourage learning of ethics and various solutions to various ethical dilemmas. This is one way to include ethics as a part of a curriculum, but it seems to be the exception.

Ethics as a subject has a perception that it should be "caught" and not "taught" (Mahajan, Aruldhas, Sharma, Badyal, & Singh, 2016, p. 2). Most students pick up ethics through the actions of their professors and not from actual instruction. In a medical school environment, the infusion of ethics in a curriculum was beneficial because it forced students to understand they should do the right thing and not just think about it (Mahajan et al., 2016).

Educators in a GIS major were surveyed to determine if they were taught ethics and privacy in undergrad and graduate courses. The potential of the unethical use of data was a possibility in the GIS industry but was it truly introduced in the courses? First, GIS educators in the United States were surveyed (N=232, n=150) and 115 responded. Many educators were not exposed to ethics and privacy in their degree programs.

Instructor Exposure to Ethics in Degree 1994 - 2006			
None	Limited	Moderate	Significant
25.2	49.6	14.7	10.3

Table 1. Adapted from Scull, Burnett, Dolfi, Goldfarb, & Baum, 2016

In schools the topic is not significantly covered at an introductory or graduate level. Students were surveyed and asked if they were exposed to ethics or privacy during the undergraduate or graduate major.

Topic of Ethics and Privacy in GIS					
	None	1	2-3	3-5	5+
Intro	37.1	41.9	19	1	1
Adv	40.2	27.8	24.7	1	6.2

Table 2. Adapted from Scull et al., 2016

Instructors concluded the difficulty to include ethics and privacy was due to the analytical nature of the program (Scull et al., 2016). Much of the time in the program is spent on mathematics and technology. There was a desire to include more ethics in a course, but finding where to incorporate the topic was difficult. Students desired more on ethics.

There is some disagreement that ethics should be included in the curriculum however it has been proven that ethics education can have a positive impact on their careers after graduation (Dahm, 2015). A pre and post-test were conducted to determine if their perception of ethics changed during their time in school. The exposure to ethics changed their way of thinking. It changed from just logic and included feelings in their decision making processes (Dahm, 2015). The students that were exposed to ethics developed skills like moral reasoning and ethical dilemma resolution. They would be better prepared for their careers and ethical situations in the future.

This paper attempts to address this gap by identifying the importance of big data, the current position of ethics in the curriculum, ethical theories that apply to big data and present an example of inclusion of ethics in a big data analytics curriculum.

2. ETHICS IN DATA DRIVEN SOCIETY

Society has changed in many ways and that includes the increase of data in everyday decision making. The number of ethical violations has increased since 2001 and is expected to grow each year (Tene & Polonetsky, 2013). There is a need for ethical rules that govern big data that are tailored specifically for the big data and data analytics. Ethical principles in industries like healthcare, law, real estate, and banking can be used to develop ethics for big data and data analytics. At this time the industry is borrowing these principles and making modifications but the documentation of the actual ethical principles is lacking.

The amount of big data is overwhelming, and this makes it next to impossible to protect and ensure its clarity. There are so many outlets for data that it is touched by so many, it can be easily corrupted (Tene & Polonetsky, 2013). Businesses make decisions in an attempt to understand their customers, suppliers, and competitors. There are many participants in this process and without ethical governance, there will be problems.

In the race to get results, sometimes the big data analyst might cut corners. What is there to stop them? This is one of the most important arguments for ethical guidelines for big data. It is the responsibility of academia to shape students that have a moral ethical conscious when analyzing data, beyond merely teaching them the skills necessary to meet industry needs. It can start with the incorporation of ethical themes in the big data curriculum

3. ETHICAL THEMES

Overall there are four themes that intersect these industries including privacy, security, ownership and evidence based decision making.

Ethical themes such as privacy, security, ownership and evidence based decision making are present to ensure that data is properly handled (Table 3). Privacy is the ability to keep data secret from those who should not have access. Big data collection will continue with little regard to privacy fears (Davis, 2016). People who allow the use of their personal data expect a basic level of protection (their data will not be shared with unauthorized agencies or individuals or stolen) (Tene & Polonetsky, 2013). There are federal and state guidelines that protect patient privacy, but in relation to big data, it is up to the company to withhold personal data (Tene & Polonetsky, 2013) Security is the protection of data from unauthorized use. More and more data is vulnerable to hacks and existing security might not be sufficient to protect data. It is important to ensure data protection because it affects other categories of ethics like privacy and ownership. Ownership is the ethical theme which evaluates who has control over the data after it has been gathered. Collected data comes from a variety of sources, and many think they still have control of the data after it is collected (Tene & Polonetsky, 2013). People believe informed consent protects their information – not necessarily true (Penn State Health and Human Development: School of Nursing, 2000) Evidence based decision making looks at the methods of decision making and its outcomes (Martin, 2015). Decisions are made based on information received and the consequences might not be favorable or biased towards the intended recipient (Zwitter, 2014). The interesting part is these decisions are something not intentional and not governed by ethical rules. These four themes combined with the ethical theories make up the foundation for creating a curriculum that incorporates ethics.

4. ETHICS IN BIG DATA CURRICULUM

The inclusion of ethics in existing big data curriculum in academic institutions is meager. Key publications in the IS education space also suggests the same. A study that reviewed the proceedings of Information Systems and Computing Education Conference by Barnes,

Clark, Athey and Plotniki (2015) uncovered that big data analytics and/or business intelligence are among the skills most in demand by industry in 2015. In examining the major trends from 1982 to 2014, the study does not suggest that skills in ethics as a distinct topic that has grown in relevance in curricula. In order to get a better assessment of how ethics related to big data analytics education has been discussed in information systems education, the archives of the journal of information systems education (JISE) were examined.

The review of articles in JISE identified several papers that discussed the exponential growth of data in the global economy and its implications on analytics curriculum and education. For instance, Borkovich and Noah (2014), suggests that the exploration of big data is rooted and involves the application of communication theory. Liu and Murphy (2014) examined the growth in documents as related to the growth in big data and discussed its implications on the curriculum. Mamonov, Misra and Jain (2015) revealed topics that could be added to analytics curriculum using a competency based curriculum perspective. Another article advocated for the use of data science courses as an alternative means for students to meet the proficiency in scientific literacy requirement (Ceccucci, Tamarkin & Jones 2015).

Schwieger and Ludwig (2016) discussion about protecting privacy in big data, is perhaps one of the very few papers that present the importance of ethics to big data analytics curriculum. The paper describes how to incorporate content on privacy into the IS2010 model core curriculum. Another article by Molluzzo and Lawler (2015) propose how to develop a concentration in analytics for information systems majors. Frydenberg (2015) describes the manner in which to include big data topics in introductory information systems course and touches on discussing privacy issues when dealing with big data to build greater awareness of data privacy concerns.

As the past publications suggest, there are a few papers that specifically look at the privacy aspects of big data ethics. Molluzzo and Lawler (2015) identify the importance of ethics in big data (Molluzzo & Lawler, 2015) when creating a concentration in analytics. Most other papers focus on specific vendor software or use of case studies to teach business intelligence (BI). However, ethics should be front and center when teaching big data curriculum to the next generation of big data analysts.

There is current curriculum on ethics, but mainly it focuses on businesses, counseling healthcare and legal (Dahm, 2015). Ethics should be more than a glancing blow. Business executives are placing greater value on ethics due to some major violations in the past ten years (Reyes, Kim, & Weaver, 2016). Studies have proven that ethics is beneficial to students, especially when it is configured to a student's learning level (Dahm, 2015). There is a need to increase the awareness of ethics in big data analytics. This awareness can be presented in the form of a single course or an entire curriculum for students. When students graduate they have to understand how ethics impacts the decisions they will make.

The delivery of ethics education varies depending on student readiness. There are three traditional methods of ethics education: clarification approach, ethics instruction and critical thinking value outcomes (Dahm, 2015). The first approach makes students confront their own values and inconsistencies. The second approach is the traditional model using existing ethical theories and decision making as a part of the curriculum. The third approach is the use of critical thinking to allow students to fully commit and develop defense strategies of their values.

Students have defined ethics as "the ability and choice to do the right thing by a set of standards that have been established" (Dahm, 2015, p. 34) or to know how to deal with situations in the most positive way and try and benefit both sides in the workplace in everyday life – being able to identify both positive and negative outcomes that arise, and knowing how to handle them in the best possible way" (Dahm, 2015, p. 34). These quotes show that student can benefit from an ethics education, regardless of its delivery. These quotes also show that an ethics education will remain with the student after graduation.

Ethics should be considered an important part of a big data analytics curriculum. It can be difficult to incorporate ethics into an existing set of courses due to the technical nature of the course (Dahm, 2015). At first, instructors can make small modifications to include ethics. Over time ethics can become a significant part of the course without being intrusive. Students will become acclimated to the inclusion of ethics as a part of their coursework in big data analytics.

5. ETHICS AT XAVIER (THE FUTURE)

The learning goal throughout the Williams College of Business is:

"MBAs are able to foster an ethical climate in their roles and responsibilities in business and society" (Xavier University, 2016)

This includes a corresponding objective:

"MBA students will recognize ethical issues and demonstrate the skills necessary to analyze information and make informed, ethical decisions in complex, conflicting or ambiguous environments or situations" (Xavier University, 2016).

Xavier University is an ethics focused institution and takes great pride to ensure students are surrounded with opportunities to learn about ethics and how it will become a part of their career after graduation. Emphasis on ethics can be found throughout the university in a variety of majors, especially in the business college.

At Xavier University the Master of Science in Customer Analytics (Table 4) is a new degree designed to help fill a projected shortage of data analytics professionals (Xavier University, 2016). This degree focuses on big data analytics and began in the Fall of 2015. It has been proposed within the curriculum a minimum of 2 credit hours of ethics instructions (30 contact hours). This will include the ethical themes of privacy, security, ownership, and evidence based decision making. Students will be exposed to ethics through lecture, discussion, assignments or outside classroom activities including visiting lecturers and site visits.

Most instructors will include one lecture related to the ethical theories mentioned earlier. Students will be given assignments or participate in discussions. Other instructors will encourage students to attend the variety ethics lectures held on campus throughout the year. Assignments and discussions will be incorporated into the course. Examples of emphasized ethical themes covered in each course are listed in Table 5.

6. CONCLUSIONS

Ethics should be at the forefront of big data education. As sensors, machine to machine communication, social media and online transactions continue to grow, businesses can create data mosaics that normally would not have been possible. The integration and analysis of big data are creating insights that now trend the boundaries of ethical behavior. As the demand for new employees who are able to conduct self-service BI and analytics and advanced data science increases, it falls upon universities to make sure that future analysts will consider ethical implications when dealing with big data analytics. When presenting curriculum on big data, it can be helpful to stop at various teaching moments to reflect on ethical questions that students should consider. By tapping into ethical frameworks used in business, students can be better equipped to effectively navigate the data driven decision making world.

7. REFERENCES

Barnes, J., Clark, J., Athey, S., & Plotnicki, J. (2015). A topic analysis of ISECON Conference Proceedings from 1982 through 2014. In *Proceedings of the EDSIG Conference* (p. n3441).

Ceccucci, W., Tamarkin, D., & Jones, K. (2015). The effectiveness of data science as a means to achieve proficiency in scientific literacy. *Information Systems Education Journal*, *13*(4), 64.

Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems (TMIS)*, *3*(3), 12.

CPCM, D. J., & Noah, P. (2014). Big data in the information age: exploring the intellectual foundation of communication theory. *Information Systems Education Journal*, *12*(1), 15.

Court, D. (2015, January). Getting big impact from big data. Mc*Kinsey Quarterly,* retrieved on July 23, 2016 at http://www.mckinsey.com/insights/business _technology/getting_big_impact_from_big_d ata?cid=other-eml-alt-mkq-mck-oth-1501¿.

Dahm, M. J. (2015, September). A dilemma-based approach to teaching ethics; Life lessons for family and consumer sciences college seniors. *Journal of Family and Consumer Sciences Education, 32*(1), 29-41.

Davenport, Thomas. *Big data at work: dispelling the myths, uncovering the opportunities*. Harvard Business Review Press, 2014.

Davis, J. (2016, January 4). Big data predictions for 2016. Retrieved September 3, 2016, from Information Week: http://www.informationweek.com/big-data/big-data-analytics/big-data-predictions-for-2016/d/d-id/132367

Frydenberg, M. (2015). Introducing big data Concepts in an introductory technology Course. *Information Systems Education Journal*, *13*(5), 12-25.

Liu, D., and Murphy, B. (2014). The document explosion in the world of big data – curriculum. *Information Systems Education Journal*, *12*(6), 83-95.

Mamonov, S., Misra, R., & Jain, R. (2015). Business analytics in practice and in education: A competency-based perspective. *Information Systems Education Journal*, *13*(1), 4.

Martin, K. E. (2015, June). Ethical issues in the big data industry. *MIS Quarterly, 14*(2), 67-85.

McGuire, T., Manyika, J., & Chui, M. (2012). Why big data is the new competitive advantage. *Ivey Business Journal*, *76*(4), 1-4.

Lawler, J., & Molluzzo, J. C. (2015). A proposed concentration curriculum design for big data analytics for information systems students. *Information Systems Education Journal*, *13*(1), 45.

Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015, August). Business analytics in the context of big data: A roadmap for research. *Communications of the Association for Information Systems, 34*, 448472.

Penn State Health and Human Development: School of Nursing. (2000). *Theoretical approaches to health care ethics*. Retrieved July 22, 2016, from Penn State Health and Human Development: School of Nursing: http://www.personal.psu.edu/faculty/d/x/dxm12/n458/theoretical_approaches.htm

Reyes, G. d., Kim, T. W., & Weaver, G. (2016, June 15). Teaching ethics in business schools: A conversation on disciplinary differences, academic provincialism and the case for integrated pedagogy. *Academy of Management Learning and Education, 2*.

Schwieger, D., & Ludwig, C. (2016) Protecting privacy in big data: A layered approach for curriculum integration. *Information Systems Education Journal*, *14*(3), 45-58.

Solas, J., & Strike, K. (2009). *The ethics of teaching.* New York: NY: Teachers College Press.

Tene, O., & Polonetsky, J. (2013, April). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*(5).

Terdoslavich, W. (2015, November). Bridging the big data skills gap with online training. *Information Week*, Retrieved on July 15, 2016.

Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., ... & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems*, *34*(1), 1.

Zwitter, A. (2014). Big data ethics. *Big Data and Society* (July - December), 1-6.
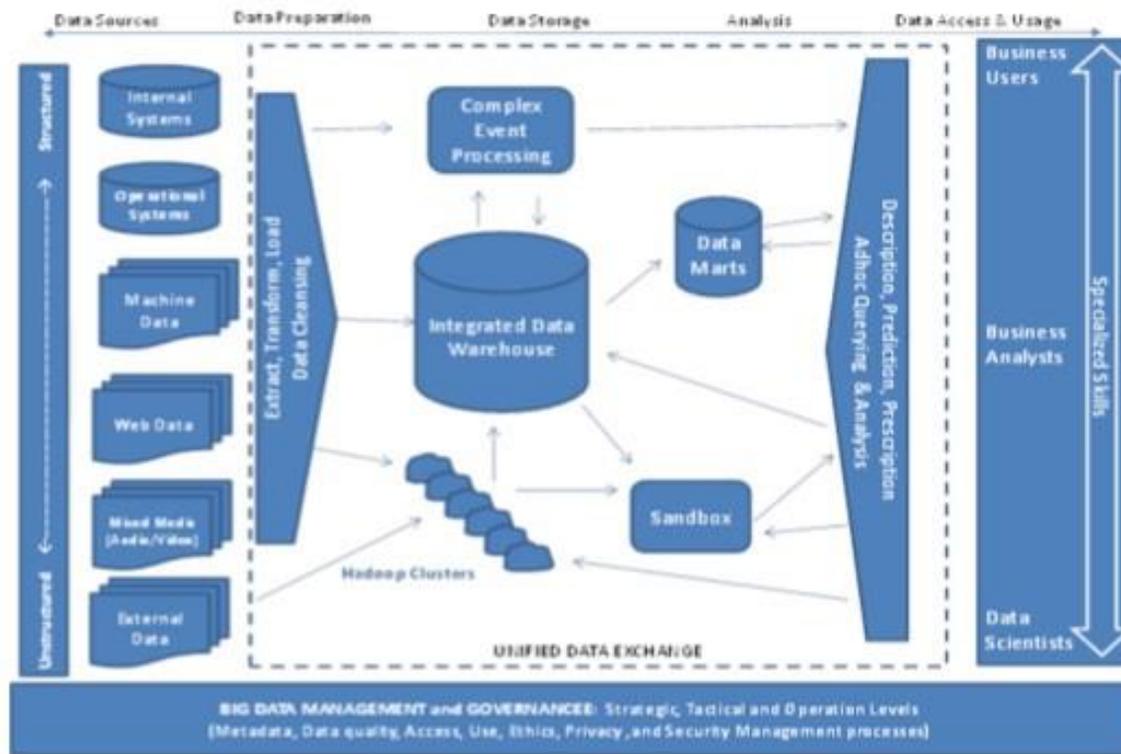
APPENDIX

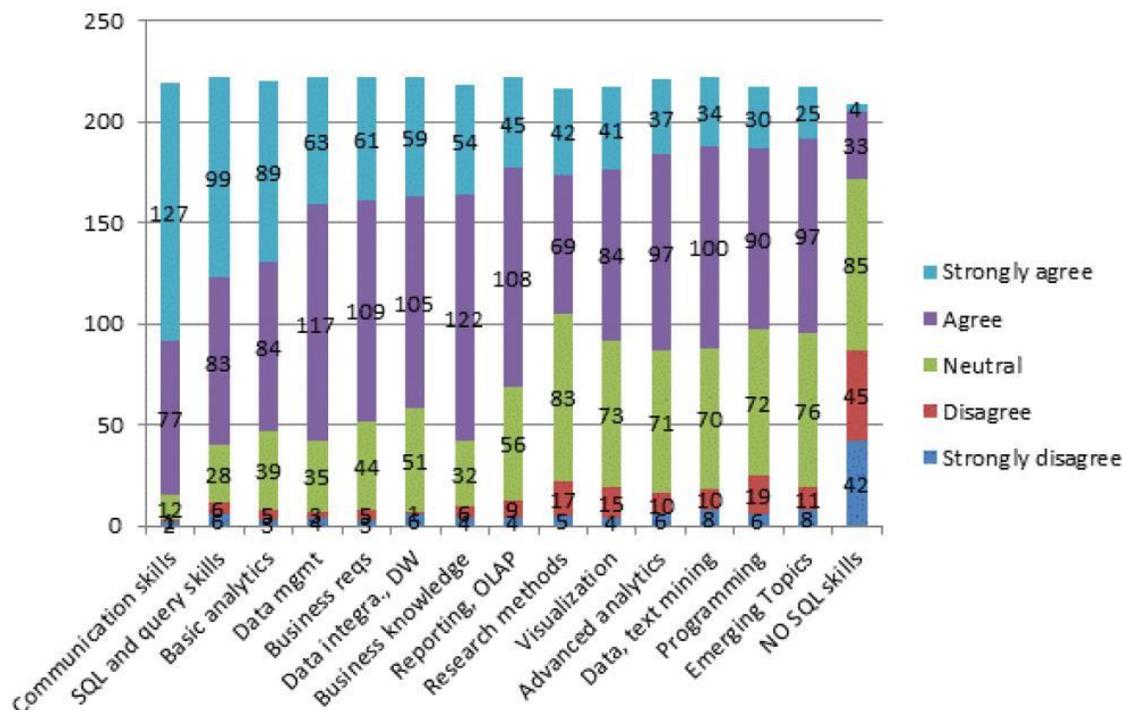Figure 1. Big Data Analytics Framework (Adopted from Wren, Kulkarni, Iyer and Ariyachandra, 2015)

Figure 2. The Content Students Need for Big Data Analytics Roles (Adopted from Wixom et al 2015)

Theme	Ethical Challenge	Example
Privacy	Sharing of personal information without permission – de-identifying information	Data used to determine Ebola outbreak in 2015 Facebook study in 2012 to test user's emotions without their consent
Security	Protection of data from outside threats	Hospital data ransomed in 2016 due to lax security
Ownership	The rightful ownership of the data used for analytics	Research in illegal behaviors where the courts want the data to build a case against a person
Evidence Based Decision Making	The use of data to make decisions about a population based solely on quantitative information	States make decisions about welfare guidelines using income as the sole factor

Table 3. Ethical Themes (Adopted from Big Data and Ethics: Examining The Grey Areas of Big Data Analytics, 2016)

Master of Science in Customer Analytics
• BUAD501: Leadership & Communication • BUAD 605: Analytics Practicum • MKTG 550: Marketing Strategy • MKTG 602: Marketing Research • MKTG 605: Applied Multivariate Analysis • MKTG 640: Consumer-centric Category Management • MKTG 664: Consumer Behavior Theory • INFO 665: Spreadsheet Applications Decision Making • INFO 674: Database Management • INFO 680: Intro to Data Mining for Managers

Table 4. Master of Science in Customer Analytics (Xavier University, 2016)

Ethical Themes in Master of Science in Customer Analytics				
Ethics Theme	Privacy	Security	Ownership	Evidence Based Decision Making
Course	Theme			
BUAD501: Leadership & Communication	X			
BUAD 605: Analytics Practicum	X	X	X	X
MKTG 550: Marketing Strategy	X		X	X
MKTG 602: Marketing Research	X			X
MKTG 605: Applied Multivariate Analysis				X
MKTG 640: Consumer-centric Category Management	X		X	X
MKTG 664: Consumer Behavior Theory	X	X	X	
INFO 665: Spreadsheet Applications Decision Making	X	X	X	X
INFO 674: Database Management	X	X	X	X
INFO 680: Intro to Data Mining for Managers				X

Table 5. Ethical Themes and Courses in the Master of Science in Customer Analytics (Xavier University, 2016)