All the data we can get:

A contextual study of learning analytics and student privacy

By

Kyle M. L. Jones

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Library and Information Studies)

at the

UNIVERSITY OF WISCONSIN-MADISON

2015

Date of final oral examination: 12/11/2015

The dissertation is approved by the following members of the Final Oral Committee:

Greg Downey, Professor, Journalism and Mass Communication/Library and Information Studies
Kristin Eschenfelder, Professor, Library and Information Studies
Richard Halverson, Professor, Education Leadership and Policy Analysis
Michael Zimmer, Associate Professor, Information Studies (University of Wisconsin-Milwaukee)
Alan Rubel, Assistant Professor, Library and Information Studies

Abstract

In this dissertation, I explore student privacy perspectives of higher education actors building capacity for learning analytics. Learning analytics is a socio-technical practice that analyzes data captured from information systems and networks students interact with to inform students' educational experience, among other ends. Student privacy concerns are inherently linked to learning analytics due to the comprehensive and sensitive nature of the data and information the technology analyzes. The extant literature indicates that very little is known about what student privacy issues institutions are encountering when they adopt learning analytics; equally little is known about how to resolve privacy problems in practice.

The following questions motivated the study: 1) How do institutional actors perceive student privacy issues related to learning analytics technologies, 2) how do they resolve the privacy issues when they emerge, and 3) what contextual factors influence student privacy practices? To answer these questions, I employed an interpretive case study design of two unique public, higher education institutions. Interviews with institutional actors served as the primary data source; I analyzed the data using constructivist grounded theory methods.

Findings revealed that powerful actors wish to gather as much students data as possible to develop advanced analytic insights. Current institutional policies and federal privacy law provide colleges and universities freedom to use student data and information with few limitations for learning analytics; also, institutional policy is not advanced enough to handle the emerging privacy problems. Actors revealed that they valued transparency with students about learning analytics and were worried about the negative effects of predictive analytics on students. Yet, the institutions had not created any systematic way to be transparent or reduce harms. Finally, there was notable conflict with regard to whether or not students should be able to control personally identifiable information and data for learning analytics.

In the discussion, I use the framework of contextual integrity and conclude that student privacy at the two case sites was under threat, yet in unique ways. I assert that students need greater control over their information by building identity layers into technological systems that can respect privacy preferences.

Acknowledgments

Writing a dissertation is not a solitary act. Scholarly writers, like myself, are surrounded at all times by a widely diverse community of friends, professional colleagues, intellectual peers, and advisors who all work in their own way to keep us moving forward. Without these individuals, any progress I have achieved in the past years would never have materialized. They deserve my thanks, and I hope by putting my gratitude in these pages they understand how much I have appreciated their presence in what has been the most intellectually challenging and personally testing period of my life.

The School of Library and Information Studies (the iSchool) at the University of Wisconsin-Madison, the Robert F. and Jean E. Holtz Center for Science and Technology Studies, and Beta Phi Mu's Eugene Garfield Dissertation Fellowship deserve recognition for funding my research.

To my dissertation committee–Kristin Eschenfelder, Alan Rubel, Greg Downey, Richard Halverson, and Michael Zimmer–I thank you for your advice *and* criticism. I appreciated the long, hard conversations as much as the short, enjoyable ones. Each of you has left your mark on me in your own way, and I recognize that my future success is linked to you and the guidance you provided me while at Madison.

I would like to especially thank a few outstanding faculty members who have impacted me throughout my academic career. My thanks go out to Mary Kay Mulvaney, Ted Lerud, Dianne Chambers, Bridget O'Rourke, and Janice Fodor at Elmhurst College; Kate Marek at Dominican University; and Allison Kaplan, Michele Besant, and, once again, Kristin Eschenfelder and Alan Rubel at the University of Wisconsin-Madison. I would like to especially recognize Christena Nippert-Eng for providing perfectly timed support and encouragement.

Friends are an especially important part of a dissertator's support crew, and so I would like to recognize some wonderful people who gave pep talks and shared some great times with me over the past years. My thanks go out to Mei Zhang, Rachel Williams, and Tammy Mays, who were wonderful doctoral student colleagues and confidants. Chad Shorter and John Thomson, in addition to being outstanding educational technologists, were great sounding boards and true supporters of my work (in addition to just being great guys). I am appreciative of the mentorship Jacob Hill gave me in the early years of my professional career in librarianship, but I am even more thankful for his continued support and friendship. Finally, I have to thank my friends at the Sow's Ear Cafe in Verona, Wisconsin who served me copious amounts of legal addictive stimulants (coffee, to be clear) and genuinely took an interest in my well-being: Liz, Ben, Will, James, Jen, Steve, and Al.

My family has been especially understanding while I pursued my doctoral degree. Dave and Lynne Lauer, my in-laws, have been completely supportive of my academic ambitions and modeled for me what a student-centered faculty member should value. My parents, Alan and Kris, have inspired me with their resilience in the face of adversity, and I cannot thank them enough for all the ways they have helped me reach this point in my life. My brother, Bryan, deserves special recognition for always running faster, scoring better grades, and, now, flying much, much higher as a highly successful pilot; I have used his success and ambition as a yardstick to motivate me to push my own limits–but I will leave the sky to him, however.

I have left two people absent from this list thus far, and that it is because they deserve my final thanks. My wife, Elizabeth (Liz) Lauer-Jones, deserves credit for her unwavering support and love in challenging times. And, finally, Michael Stephens–my mentor while at Dominican University and now one of my best friends–merits my heartfelt appreciation for inspiring me to enter into the intellectually engaging world of library and information studies, but I thank him most of all for his kindness. I dedicate this work to Liz and Michael.

Table of Contents

Chapter 1. An Introduction to the Study

1.1. Introduction

Organizations, institutions, and businesses now understand the potential value of Big Data.[1] They have begun to analyze "data at extreme scale" (Hopkins & Evelson, 2011, p. 4) to gain insights into those who use their services and products. To mine troves of data, stakeholders hire individuals with analytic skills and build or buy advanced technologies for their organizations in order to turn nascent data into actionable information. Lewis's (2003) Moneyball–made even more popular by the movie of the same name–introduced the world to the use of Big Data in United States professional baseball; the big-box store Target used its large data stores and analytic insights to market prenatal vitamins to pregnant women (Duhigg, 2012a); and, in both 2008 and 2012, Nate Silver used Big Data techniques to successfully predict the United States presidential elections (Taylor, 2012).

While analyzing Big Data can create new insights, related practices have inherent privacy issues. The technologies often use personally identifiable information, and since the analysis can be used to influence an individual's behavior, analytic technologies may intrude into various spheres of private life. Moreover, Big Data as a socio-technical movement is advancing at a rapid rate and it is unclear if current practices, policies, and law are up to the task of protecting individuals from privacy harms.

Higher education has initiated its own Big Data program using technology called "learning analytics," which advocates argue will improve student learning outcomes and make institutional programs more efficient and effective (Fain, 2012). Like other Big Data initiatives,

[1] Like boyd and Crawford (2014), I capitalize "Big Data" throughout to emphasize its status as a unique phenomenon, and to differentiate between the phenomenon and large sets of data.

learning analytics is inherently tied to questions of privacy. Personally identifiable data and information about students is at the core of all learning analytics projects, so there are outstanding questions about student privacy as colleges and universities build capacity for learning analytics on their respective campuses. I take up questions of student privacy and learning analytics in this study.

In the following sub-sections of this chapter I provide an overview the study. Specifically, I introduce the study's aims; layout my research problem and related questions; briefly describe the theory and methods that inform its design; and discuss the structure of the study as a whole by summarizing each chapter.

1.2. Aims of the Study

Regardless of the rapid pace of development around learning analytics, institutions are struggling to answer "big questions" (Willis, Campbell, & Pistilli, 2013, para. 1) regarding data use, data governance, and privacy problems that emerge from using sensitive types and comprehensive amounts of student data. When institutional stakeholders were asked in a survey if learning analytics would affect student privacy, most agreed that it would; yet, the same stakeholders were not clear about how, specifically, privacy would be impacted (Draschler & Greller, 2012, p. 126). Part of the problem is that stakeholders do not know how to communicate privacy issues, and some fear that conversations about privacy will stop progress with learning analytics, thus they may choose to avoid them all together. At the annual Learning Analytics Summer Institute (LASI), George Siemens (Society for Learning Analytics Research, 2014a) said that ethical conversations have the effect of dropping one's pants: "If you go to a party and you drop your pants, everybody walks away from you," remarked Siemens, "It kills conversation once you start talking analytics and ethics." If little is known about how stakeholders perceive privacy problems and address them while building capacity for learning analytics, it is plausible that the privacy issues will "continue to plague" (Kruse & Pongsajapan, 2012, p. 2-3) institutions.

Of the literature that does exist regarding building capacity for learning analytics, researchers commonly discuss issues related to creating technological and organizational capacity to support data practices (Norris & Baer, 2013), promoting analytical skills and aptitudes among staff (Baer & Campbell, 2012), and managing costs associated with adopting or building learning analytics technologies (Bichsel, 2012). But to date, only a select group of authors have written about institutional perceptions of capacity issues (see Arnold, Lynch, Huston, Wong, Jorn, & Olsen, 2014; Bichsel, 2012; Draschler & Greller, 2012; Macfadyen, Dawson, Pardo, & Gašević, 2014; Scheffel, Drachsler, Stoyanov, & Specht, 2014; Stiles, 2012). Researchers recognize that for institutions to work out the privacy issues inherent to learning analytics, they must take a multi-faceted and contextual approach (Macfadyen et al., 2014). Not all privacy problems are purely technical in nature, neither do they exist as only social issues. And while colleges and universities may share similar concerns (e.g., how to manage access to geolocation data), each will address problems differently based on the norms, values, goals, and socio-technical infrastructures unique to each institution.

Even though higher education institutions are equipped with knowledgeable and skillful individuals, they tend to lack the capacity to work through the ethical issues (Manyika et al., 2011). On the ground, institutions address social, technical, and policy related problems as they encounter them, which often leads to instability and resistance (Macfadyen & Dawson, 2012) in the wake of poor planning and leadership vacuums (Norris & Baer, 2013). As a result, it is not uncommon for stakeholder perspectives on issues to be highly varied, which can lead to different and sometimes incompatible practices. More research is needed into contextual issues at the institutional level, especially related to student privacy, in order to understand how colleges and universities perceive and address these problems in practice. This study addresses that gap in the literature.

1.3. The Research Problem and Related Questions

The research *problem* guides my research *questions.* The research problem establishes boundaries within which to seek answers to the research questions. By setting boundaries, the research problem delineates topical restrictions. Without these, it could be possible to lose sight of the main goal of the study by chasing ideas and concepts outside the study's scope (Myers, 2009). For this study, I developed "general, flexible, and open" (Ng & Hase, 2008, p. 158) research questions to allow my participants to tell their story and provide me freedom to explore the research problem (Charmaz, 2014; Corbin & Strauss, 2008). My research problem is as follows:

- Learning analytics is an emergent socio-technical practice in higher education, and we are unaware of how institutional actors in higher education institutions perceive related student privacy issues and address them while they build capacity for the technology.

This research problem details that the scope of the project is limited to:

1. Learning analytics technologies,
2. higher education institutions,
3. institutional actors,
4. student privacy,

5. perceptions of student privacy issues,
6. and how perceptions inform practices while building capacity for learning analytics technologies.

My research questions follow:

- How do institutional actors who are building capacity for learning analytics perceive related student privacy issues?
- How do those perceptions influence their practice with the technology?
- How do actors resolve privacy problems as they encounter them?
- At each institution, what contextual characteristics influence student privacy practices related to learning analytics?

1.4. The Design of the Study

This study was an interpretive case study of two unique, public higher education institutions. Both institutions were building capacity for learning analytics technologies and had active learning analytics projects. Comparing two unique cases provided me the opportunity to develop a wider understanding of student privacy perspectives, especially in unique contexts.

Contextual characteristics were a key concern in this study given my social informatics orientation and theoretical perspective with respect to privacy. A social informatics approach to privacy issues helps researchers and readers alike see privacy as a social value that is impacted by the the interrelation of social and technical elements in specific social contexts (Kling, Rosenbaum, & Sawyer, 2005; Waldo, Lin, & Millett, 2007). Where privacy theory is concerned, I adopted Helen Nissenbaum's (2004, 2010) theory of contextual integrity in this study, as a contextual approach to privacy issues helped me analyze and assess specific student privacy problems across my institutional contexts.

Interviews served as the primary data source for this project. In total, I conducted 20 interviews with 19 participants from the two institutions. The participants represented a variety of institutional actors, such as data scientists, instructional technologists, representatives from campus legal counsels, registrars, information officers, system administrators, instructors, and student advisors. I digitally recorded nearly every interview, which resulted in a large textual dataset. I also reviewed publicly available multimedia and documentation about my case sites' respective learning analytics projects and the technologies they employed.

To make sense of the data, I used constructivist grounded theory methods, which are well-matched to interpretive case study designs (Charmaz, 2014). Constructivist grounded theory approaches seek to understand how participants construct meaning and enact action in particular contexts; as such, this variant of grounded theory supported my inquiry into how institutional actors perceived student privacy problems related to learning analytics and what they did to resolve those problems.

1.5. The Structure of the Study

I have written this study in nine separate chapters. Following this introductory chapter, the study begins with historical touchstones that highlight the ways by which higher education institutions used student information, starting in the 17th century and ending at the present. I pay special attention in this chapter to the overlap of institutional values, technologies, and student privacy rights. I thematically address *in loco parentis* (i.e., in the place of the parent) justifications for accessing and using student information; the rise of comprehensive student records and technologies that supported their creation; then, I turn to a discussion about data-based information systems in higher education; and, finally, I cover the growth of the information technology ecosystem on campuses and some student privacy problems ubiquitous information systems create.

I turn to a discussion of learning analytics after finishing the historical chapter. I address what distinguishes learning analytics from other prominent terms, such as "business intelligence" and "academic analytics." In doing so I highlight the contested nature of the definition of learning analytics. Next, I situate learning analytics as a Big Data practice, and I point out the ways Big Data-style analytic practices can be put to use in higher education to aggregate and analyze student data and information. At the end of the chapter I highlight how campuses have already begun applying learning analytics practices, discuss the ends to which they have applied learning analytics, and point out the relationship between learning analytics and the culture of assessment in higher education.

As mentioned, Big Data has inherent privacy issues, and I address a number of those problems in the chapter that follows the discussion of learning analytics. Specifically, I address privacy issues related to informed consent, surveillance, and a lack of transparency among those who adopt Big Data as means to analytic ends. I also consider the ways by which those who control Big Data technologies and practices hold power over their data subjects, and I address particular issues related to analyzing human behavior in Big Data projects. I conclude the chapter by examining the paradox that surrounds deidentifying Big Data as a privacy protection, as researchers have proven that deidentification practices are problematic since databases can be easily linked and, therefore, make it possible to reidentify individuals.

The discussion on Big Data's privacy problems sets up a more targeted conversation of learning analytics and its particular privacy issues in the next chapter. Here, I address a lack of information policy to protect student privacy. I also consider how a lack of transparency regarding data-driven practices in higher education is problematic for privacy, and I take stock of extant conversations about informed consent and opt-out practices to protect student privacy. I follow these conversations with an analysis of unsettled debates regarding student information controls and whether students should have a so-called right to be forgotten. There are also outstanding questions in the literature about data governance and stewardship, which I take up before turning to a discussion regarding concerns over third-party access to private student information in light of new data-driven practices. I end the chapter by detailing ethical concerns about whether or not institutions are obligated to act on the student information they acquire, address how learning analytics implicates student autonomy concerns, and discuss existing ethical codes to guide learning analytics practices.

In the following chapter I setup the study's empirical methods and move away from the extant literature about student privacy and learning analytics. I lay out my research motivations and justify the theoretical perspective that guides my study of privacy issues, which is the framework of contextual integrity. I follow these conversations with a description of my methodological perspective and how I employ constructivist grounded theory. The final parts of this chapter address my data sources, how I collected and analyzed data, and the evaluative techniques I used to increase the study's rigor.

Next, I move to a chapter that discusses the thematic findings I discovered in the data. In this chapter I provide a meta description of all the relevant data that informed the findings and discuss relevant contextual factors. I then move to a discussion about privacy flags, which represented to my participants situations when student privacy was especially at risk. Following this discussion I detail themes discussing data governance, institutional policy, and interpretations of federal student privacy law. Another theme considers potential negative effects of aggregating large amounts of sensitive student data and information. The chapter ends with a conversation about "self-fulfilling prophecies," or situations where predictive analytics may harm students, and a discussion about participant perspectives with regard to whether or not students should have control and ownership over their information.

After the findings, I take up theoretical and practical considerations in the following chapter. Specifically, I use a decision heuristic developed in my theoretical framework to determine if learning analytics does, in fact, present unique challenges for student privacy. I then build an argument for why the framework of contextual integrity needs to consider micro-level privacy concerns as much as meso- and macro-level ones. Finally, I argue that students should ultimately be given more control over their data and information to protect their privacy; to this end, I propose that higher education institutions and technology vendors should build data dashboards and identity layers.

The study concludes with a final chapter that summarizes the study and details its contributions. Here, I also consider implications for practitioners and information policy, as well as provide recommendations for future research on student privacy and learning analytics; I also address the study's limitations.

Chapter 2. Historical Uses of Student Data and Information in Higher Education

2.1. Introduction

In this section, I address the overlap of institutional values, technologies, and student privacy rights throughout the history of higher education in order to examine the role of student information and data in college and university practices. I begin by examining the *in loco parentis* (i.e., in the place of the parent) justification higher education institutions employed to obtain and use student information. Next, I turn to the creation of cumulative records about students, especially in relationship to the Hollerith punched-card system. After this, I discuss data bank technologies and concerted legal efforts to protect student privacy; this conversation informs the subsequent section on the rise of data-based systems in higher education, which use student data to respond to stakeholders' demands for accountability. Finally, I discuss the growing information technology ecosystem on higher education campuses and the contested place of student privacy with respect to data-driven systems.

2.2. In Loco Parentis

For over 300 years, American institutions of higher education pursued religious, social, and academic policies based on the belief that they acted in loco parentis. During this time, in loco parentis was a method of "inculcating within [students] discipline, morals, and character" (Cohen & Kisker, 2010, p. 27), which enabled institutional actors to regulate the lives of their students and to punish them (sometimes severely) for their transgressions as a father would his son. To accomplish these aims, higher education institutions sought and used private information about students to control, discipline, and punish them.

One of the avenues through which such control initially emerged was religion. In the early days of the American colonies, religion and religious values were central to communities and the colleges therein. When children went to college, the parents–or more accurately put, the *father*–delegated his parental rights to the institution. The delegation of responsibilities worked as such:

> A father paid a schoolmaster to educate his child (mostly male children, who were pre-fathers) and the schoolmaster agreed to educate the child. To make this arrangement work, the father would have to give the schoolmaster much of the nearly limitless power over the child. (Lake, 2013, p. 20)

The expectation of the father, and one of the goals of the institution, was that the institution would imbue the student with "mental discipline," a potent but balanced mixture of "psychological, theological, and moral convictions" (Veysey, 1965, p. 22).

Colleges developed strict campus laws to govern academic, religious, and social behavior to fulfill their *in loco parentis* responsibility. In short, they told students where they could go, when, and with whom. For example, Yale College's rules stated:

> No Student of this College Shall attend upon any Religious Meetings either Public or Private on the Sabbath or any other Day but Such as are appointed by Public Authority or Approved by the President… If Any Student shall Prophane the Sabbath by unnecessary Business… or making any Indecent Noise or Disorder… He Shall be punished… No Student Shall walk abroad, or be absent from his Chamber, except Half an hour after Breakfast, and an hour and an half after Dinner. (Dexter, 1896, pp. 4–5 as cited in Cohen & Kisker, 2010, p. 28)

Student handbooks, manuals, and pamphlets dictated to students the rules of the institution in painstaking detail, going so far as to force loyalty oaths on students and subject them to "treason codes" (Veysey, 1965, p. 33). The faculty enforced the laws with physical punishment in the manner of "authoritarian regimes" (Veysey, 1965, p. 35).

It is one thing to set a behavioral policy, but another entirely to enforce it. To do so, colleges spied on their students and developed extensive record-keeping practices. To create a "controlled environment" (Veysey, 1965, p. 35) and discipline students to behave in particular ways, administrators used a "spy system" (McCosh, 1878, p. 434) on college campuses to further inform faculty and administrators about students' private lives. For example, Noah Porter, president of Yale College, instructed his faculty to secretly watch students by peering through windows and listening at keyholes (Veysey, 1965). Porter (1870) grounded his student surveillance in theological orthodoxy, claiming that "God spies" on students at all times in order to judge them, and so should his academic ministers–the faculty (Veysey, 1965, p. 35).

Archival documents reveal the breadth and depth of information gathering practices by institutional actors. For instance, Harvard University's archives included "detailed records of student board, room, and other expenses"; they also disclosed what students ate, how much beer they drank, and if they were fined for breaching college laws (Holden, 1976, p. 461). As the years past at Harvard College, student records became even more intrusive and inclusive. The archives show that institutional actors even kept track of the cleanliness of a student's room.

Surveillance and record keeping supported an institution's policy of acting in place of a student's parents; as parents would do, institutions used the information they had on student transgressions to exact punishment. Many historians of higher education note that the raucous nature of adolescents, sprung free from home and densely situated in residence halls, made for a bad mix. In response, institutions developed "an obsession with order" and enacted "disciplinary measures… considerably more stringent than parents would have prescribed" (Cohen & Kisker, 2010, pp. 75–76). At one end of the spectrum, institutions simply expelled an insubordinate student for behavioral offenses, but on the other end *entire* classes of students were expelled, as was the case with Dickinson College's junior class in 1851 (Cohen & Kisker, 2010); and at the extreme, *in loco parentis* enabled institutions to enact corporeal punishment. Lake (2013) noted that the delegation of power from parent to institution "was not to coddle, protect, or nurture," but to "restrain [and] correct" (p. 20) the student–even by force. Consider the treatment of a student at Harvard College (now, Harvard University) in 1674: for blasphemy, President Leonard Hoar read aloud the student's infraction and penalty to the community in the college library, said a prayer for the student, and then publicly whipped the boy before concluding with another prayer (Thwing, 1900).

Until the middle of the eighteenth century, *in loco parentis* was simply a social norm, an accepted way for institutions to administer education to students, but during this time it also became law. In English family law, which Lake (2013) notes was adopted to a great degree in America, renowned legal commentator Sir William Blackstone made the first legal reference to *in loco parentis* around 1750. Blackstone affirmed that the parental role institutions filled was "virtually immune from legal scrutiny and liability" (Lake, 2013, p. 20), much like a father, as a master of his own family, was free to run his household as he pleased. Similar to his legal colleague in England, American James Kent reasserted in 1820 the validity of *in loco parents* in a famous treatise, which American courts began to cite in cases concerning student discipline (Lake, 2013). Under *in loco parentis*, "constitutional rights stopped at campus gates" (Lee, 2011, p.

67), and students were afforded few if any protections against invasive information gathering practices and the disciplinary measures built upon them.

Three landmark cases in particular used *in loco parentis* in order to arrive at a decision: *Pratt v. Wheaton College* in 1866, *Gott v. Berea College* in 1913, and *Stetson University v. Hunt* in 1916. At Wheaton College, E. Hartley Pratt was suspended for joining a secret society. The court ruled as follows:

> A discretionary power has been given [to college authorities] to regulate the discipline of their college in such a manner as they deem proper, and so long as their rules violate neither divine nor human law, we have no more authority to interfere than we have to control the domestic discipline of a father in his family. (Pratt v. Wheaton College, 1866, p. 90)

Drawing in part on *Pratt*, the outcome of *Gott* determined that institutions could prohibit student travel off campus. For years, Berea College had listed in the student manual under the sub-section "Forbidden Places," locations of off-campus establishments students were not allowed to enter, one of which was Gott's restaurant. The court found that where the "physical and moral welfare, and mental training of campus pupils" (*Gott v. Berea College*, 1913, p. 379) was concerned, higher education institutions could make "any rule or regulation for the government, or betterment of their pupils that a parent could for the same purpose." With *Hunt* (1916), the issue at stake was whether or not Helen Hunt was "maliciously, wantonly, and without cause in bad faith expelled" (p. 513) from her institution, John B. Stetson University, after being accused of "hazing the normals, ringing cow bells and parading in the halls of the dormitory at forbidden hours" (p. 514). In the same vein as *Pratt* and *Gott*, the court found that the university acted fully within its rights and responsibilities to punish Hunt for her failure to abide by Stetson University's rules, which were supported by *in loco parentis*.

In loco parentis is often wrongly invoked as an institutional duty of care for students, when in fact *in loco parentis* was "a power, not a responsibility" (Lake, 2013, p. 29). For example, some students lost eyes in egregious laboratory accidents and received horrible medical care at student care centers; in both cases, the court found that universities had no obligation to look after the health and safety of students, and the students received no legal or financial recourse (see *Davie v. Board of Regents, University of California*, 1924; *Hamburger v. Cornell University*, 1925; *Parks v. Northwestern University*, 1905). Consequently, colleges and universities had no genuine responsibility to protect, in a parental way, their student bodies; yet, they had enormous freedom to direct and control their behaviors as they saw fit. However, these institutional liberties would not last due to student student discord and changes to legal reasoning.

Scholars agree that the reign of *in loco parentis* as a social norm and legal policy ended around the 1960s due to two major factors: the first factor was social unrest by the students themselves, and the second was a shift in legal thought regarding what universities were responsible for in terms of caring for students. As the 1800s waned, students grew frustrated and disillusioned with the "sword" (Lake, 2013, p. 24) of control *in loco parentis* provided to colleges and universities. Disciplinary measures and inquisition-like interrogations pushed students–especially at Wesleyan University–to action, where they called for their institution to treat them as "reasoning and self-responsible beings" (Peterson, 1964, p. 115). Student newspapers, like the *Williams Argo*, argued for an end to the *in loco parentis* regime, stating that "few parents would attempt any such government of 20 year olds as do colleges of their students" (as cited in Peterson, 1964, p. 146). Other students were less peaceful with their protests and took to rioting against their institutions, and some even organized the ouster of their institution's president (Lipset, 1993).

For the first time, *in loco parentis* faced a stout challenger in *Dixon v. Alabama State Board of Education* (1961) as the Civil Rights movement gained momentum. St. John Dixon and five of his student peers, all of whom were black, entered into a publicly-owned restaurant and were refused service; the restaurant subsequently closed for the day, but the students refused to leave. President Trenholm of Alabama State College expelled the students for civil disobedience without a hearing. The district court found in favor of the defendants initially, but in the court of appeals the circuit judge ruled in favor of the plaintiffs, arguing that the students were denied a right to due process of a hearing and that they were expelled improperly. The judge wrote that such a right was "fundamental to the conduct" of society and a "constitutional principle" (pp. 157–158). With *Dixon, in loco parentis* lost its illegal immunity. Students could now challenge disciplinary measures of their institutions and seek damages.

Towards the middle of the twentieth century, *in loco parentis* and most of the practices that supported it no longer existed, at least under that particular banner. The push for mental discipline rooted in theological orthodoxy faded away and the values of Puritan America lost their iron hold on higher education. Colleges and universities, and especially the incoming presidents who took over for their embroiled predecessors, ended the "automatic assumption" (Peterson, 1964, p. 147) that they should and could act in the place of parents. With the demise of *in loco parentis*, institutions turned from oppressive autocracies to liberty-promoting democracies, which afforded students more say in their education, freedom from institutional rules, and defensible legal rights (Petersen, 1964). Since higher education institutions no longer had the legal immunity to control and severely discipline their students, most of the vestiges of the old spy system faded away, except for comprehensive student records.

2.3. A Cumulative Record

The late nineteenth century into the middle of the twentieth century marked a formative era in history of American higher education, one that was "fundamentally [different] in character and purpose" (Gruber, 2007, p. 261) than any other time before it and which "profoundly altered" (Goldin & Katz, 1999, p. 37) the shape of colleges and universities for the foreseeable future. Scholars trace three distinct changes that affected colleges and universities, all of which began in the late nineteenth century. First, colleges and universities experienced a series of "technological shocks" (Goldin & Katz, 1999, p. 40) in the form of a splintering and specialization of knowledge work inside the academy. Second, student populations boomed and institutions found themselves floundering under new administrative requirements and pressures. And third, institutions created new managerial roles and administrative practices. In response to these significant changes in higher education, institutions found new value in bolstering student records of the past with new, usable information to help inform the management of the university and its students. I address these three changes in turn below to frame and analyze novel uses of student information and new technologies in this era.

Higher education institutions experienced a disciplinary splintering effect in the late 1800s. Some faculty "subdivided and specialized" (Goldin & Katz, 1999, p. 38) their work from their main discipline, others simply broke off into completely new areas. A primary reason for this change was due to faculty adopting the scientific method and its related practices, which included new tools and techniques associated with laboratories. Many faculty split with traditional scholarly agendas and disciplines to pursue a scientifically-driven area of inquiry, leaving the vestiges of the Enlightenment-era college to some of their peers. For example, economics, political science, and sociology all separated themselves from the "hodgepodge curriculum in moral philosophy" (Gruber, 2007, p. 262). Splitting a discipline was not always amicable; it was often a highly-charged debate centering on concerns over what should be considered scientific research. Some disciplines dug in their heals with others formed new alliances.

With the splitting of disciplines and a growing emphasis on research emerging, the structure of the American college and university irreversibly changed. "The old-time professor," wrote Veysey (1965), "who was a jack-of-all-disciplines rapidly disappeared from all but the bypassed small colleges" (p. 142); the specialized scientist took the empty place at the lectern. Many institutions adapted internally to the changes within their academic programs; others either created new departments or bolstered existing departments in response to social and governmental pressures to develop scientists–especially physical and chemical scientists–for the workforce and special federal projects (Goldin & Katz, 1999). In some cases, brand new institutions were established with research needs in mind, including Johns Hopkins University in 1876, Clark University in 1889, Stanford University in 1891, and the University of Chicago in 1892 (Atkinson & Blanpied, 2008).

Student enrollment increased along with the number of institutions. In their analysis of enrollment figures across two-year schools, collegiate and graduate institutions, and professional programs, Goldin and Katz (1999) found a fivefold increase in the student population from 1890 to 1940. Around the same time, a little over 200 new college and universities emerged. While some may believe that the first Morrill Act of 1862 contributed significantly to the increase due to new land-grant institutions, the data shows that a vast majority of the new colleges and universities were actually private. In fact, it was not until the second Morrill Act of 1890 that *public* higher education saw a notable increase in enrollment (Williams, 2007). All the same, the the growth of American higher education reflected the new demands of an enlarging student population. In part, immigration and an increase in population throughout the United States impacted enrollment figures due to growing numbers of potential higher education students, and institutions were increasingly admitting women along with men, as well. But a plausible answer also exists at the secondary school level. High schools expanded their curriculum during this time and began to bridge the divide secondary and post-secondary levels of education (Cohen & Kisker, 2010). Additionally, Goldin and Katz (1999) argue that more students graduated from high school than ever before. And even though the cost of pursuing a degree was high, the return on the investment was substantial: it made financial sense for students to continue their education past the secondary level.

Combined with the ever-increasing number of students and faculty and the decentralized nature of higher education, institutions faced a managerial nightmare. Colleges and universities used to be managed by one chief administrator–the campus president. Given the significant change in scale and complexity on campus, the president could no longer oversee "every detail of campus management" (Rourke & Brooks, 1966, p. 4). As managerial problems compounded, external and internal pressures mounted for bureaucratic efforts. Society would not abide a complex institution, such as higher education, that could not effectively manage its own house.

As Progressive Era thinking grew in prominence and scientific management put down roots in factories and businesses, colleges and universities heard loud calls to reform their administrative structures. Morris Cooke, a researcher for The Carnegie Foundation for the

Advancement of Teaching and a “disciple” (Rourke & Brooks, 1966, p. 6) of Fredrick W. Taylor, stated that “only good can come to an organization–whether it be commercial, educational, or religious–when a *friendly* hand turns the light of public scrutiny upon its methods, resources and aims” (Cooke, 1910, p. v, emphasis added). Yet, Cooke was no friend to higher education, and he was mistaken to push scientific management practices borne from hard Taylorism on colleges and universities. Only a form of soft Taylorism could work on campus that took into consideration the particular needs related to doing academic work (Newfield, 2007).

Soft Taylorism manifested itself in the dual governance system, which emerged during this time. Dual governance allowed the faculty and administrators to manage and direct an institution, but the faculty generally did not want to be managed nor participate in management practices, as they saw such efforts as an attack on their autonomy and individualism (Veysey, 1965). As a result, “the care of the organization” (Newfield, 2007, p. 357) went to boards, presidents, provosts, and deans who took the managerial reins and responsibility for the extensive business affairs of the institution, forming the other half–the more powerful half–of the dual governance system. Faculty primarily kept to their classrooms, labs, and research, only flirting with institutional issues when they felt their participation in “unwieldy” (Newfield, 2007, p. 356) academic senates *might* effect change. On the whole, college and university administrations felt the brunt of the rolling changes in higher education, but they had the freedom to participate in Progressive reform movements and to craft much of the future of higher education in the absence of an engaged faculty (Thelin, 2007).

With new administrators came new administrative practices. Notably, clerical work rose in prominence to meet new demands and calls for efficient work. Veysey (1965) remarks that typewriters simply “appeared” (p. 306), flooding the campus with new correspondence and documents. At Columbia University in New York, newly inaugurated president Nicholas Murray Butler brought with him a small army of support staff, including "three secretaries, five stenographers, and two office boys" (Veysey, 1965, p. 307). Running a larger, fractured university took great efforts to communicate to the masses and ensure information was documented and flowed as needed.

Where student information was concerned, it, too, became enmeshed in bureaucratic work. As some institutional actors argued, "[student] record-keeping and adequate reporting of factual data are essential to good administration" (Russell & Reeves, 1936, p. 245 as cited in Miller, 1938). Peterson (1944) went as far as suggesting that the "seriousness" of an institution was represented by "the kind of [student] records it keeps" (p. 191). Administrators saw and exploited a new opportunity to use their clerical staff and its technologies and techniques to manage an ever-growing student population and expanding campus. Veysey (1965) wrote:

> The danger was no longer so much one of riots or other forms of open rebellion as it was one of drift, laxity, and illegitimate pursuit of personal or factional advantage. Techniques of control shifted from the sermon and the direct threat of punishment toward the more appropriate devices of conference, memorandum and filing system. (p. 315)

Individual behavior could be controlled and academic growth could be monitored using a technological armory of paper, pen, typewriter, filing cabinet, and–as time went on–Hollerith punched cards. There was no longer a need for faculty spies and paternal discipline as new technology-enhanced information practices emerged.

In 1927, the American Council on Education (ACE) released its report in the *Research Bulletin of the National Education Association* advising higher education institutions to adopt what they called a “cumulative personnel record folder.” Tansil (1941) argues that the cumulative record system significantly affected how institutions gathered student information and recorded it for use, but the “greatest contribution” (p. 170) of the ACE report was that it helped develop “a consciousness” in higher education regarding the need for comprehensive student information. Due to this “consciousness,” and in part due to administrative needs, institution began to seek new value within student records and the information held within. To facilitate record-keeping and use practices in aid of student development, new administrative offices, positions, and roles emerged, such as the “Admissions and Personnel Office,” the creation of a Dean of Students, and new faculty responsibilities to work as dedicated student counselors (Kirkpatrick, 1941, p. 317).

At this time, student information was strewn throughout the university in the offices of individual faculty and others, limiting its accessibility. Creating cumulative records provided the opportunity for institutions to audit what information was available and aggregate it in a central, accessible file. At Bethany College, Kirkpatrick (1941) touted the institution’s success in centralizing student records in new offices, which were served by a number of clerical and stenographical staff. When the office was called by someone on campus, it had the ability to “provide accurate information regarding a student [including a] comprehensive and cumulative summary covering all phases of his life and activity on the campus, and a rather complete history of his educational career and interests” (Kirkpatrick, 1941, p. 316–317). Similarly, the “Domesday Book,” University of Toledo’s comprehensive three-volume set of records about all of its 1,700-plus student population, was created in part with this purpose in mind (McClure, 1936). Its name was derived from the actual *Domesday Book* developed in the time of William the Conqueror, which was a detailed record of land ownership within the Norman leader’s vast kingdom. Each volume of the university’s book was tabulated using machines (presumably, Hollerith punched-card systems) and replicated using mimeograph technology; it included past and current academic information about each student (including ranking) and was painstakingly indexed for easier use by anyone on campus who sought it.

Some colleges and universities made great efforts to assess the quality and comprehensiveness of the cumulative records they were creating, using the opportunity to match the record with the objectives of the institution (Tansil, 1941). If the record was in some way deficient, or if the record failed to meet the needs of institutional actors who used it, institutions made adjustments and bolstered it with new information. At the State Teachers College of Towson, Maryland, the cumulative record was reviewed three different times over six years in order to optimize the record for individualized instruction and to support a new curriculum. The final record consisted of 59 separate items about a student in order to "give as complete a picture of the individual as possible" (Tansil, 1941, p. 185). The final record included financial and academic standing, employment status, and familial history, among other details. Cumulative records at other institutions also included what today seems like fatuous information, such as the number of "sound," "special," and "normal" teeth a student had when she matriculated (Miller, 1938, p. 225). If the information was valuable to the institution, it was included in the cumulative record.

Records served a number of purposes, and institutions often developed them with third parties in mind. On campus, institutional actors could review student records, write comments, and assess student abilities using the social, academic, financial, and familial information held within (Tansil, 1941). Off campus, potential employers could review a student's comprehensive record. To this end, the University of Ohio's job placement department made "photostatic" (Peterson, 1944, p. 194) copies of student records for potential employers to have and use in order to understand "the complete story of the student" (p. 192).

Institutions continued to develop their student information practices, and in doing so discovered new ways to glean new value by turning to emerging technology. Cumulative records at their most basic level a comprehensive representation of an individual student which was valuable in and of itself. However, once the records were completed and aggregated, institutions began to realize that they had a significant store of information that could potentially reveal new insights into successes and failures in the administration of the institution and the education of the student body. But file drawers of "static and inflexible" (Constance, 1935, p. 59) student records were of "slight direct value" in an "age of inquiry and research." Even though the information was aggregated, it was inefficient to analyze and use.

Administrators widely adopted Herman Hollerith's punched-card system to work through masses of student information, especially to prepare grade reports, draft tuition bills, and register students in courses, among other clerical and statistical jobs (see Fichtenbaum & Shipp, 1947). Simply described, punched cards store information "coded as holes" (Ceruzzi, 2012, p. 6). Once punched cards were completed, other parts of the system performed additional tasks with the cards. A "verifier" checked for errors; an "alphabetic printing punch" interpreted punches and printed limited characters; a "gang punch" punched multiple sets of cards at once; a "reproducing punch" duplicated already punched cards; a "tabulator," the precursor to the electric calculator, did mathematic calculations; and a "sorter" organized cards into different slots based on preset sorting schemes (for detailed descriptions of each, see Arkin, 1935, pp. 9–20). Institutions invested heavily in specific machines, purchasing a part of Hollerith's system to custom fit their needs, hired specially trained clerical staff, and rearranged entire offices to promote efficient work practices reliant on Hollerith machines (for an example, see the figure in Cobb & Bray, 1935, p. 105). Punched-card, record-creating practices no longer required armies of temporary clerical workers; all could be done with "great convenience" (Arnsdorf, 1935, p. 30) and "freedom," all while increasing "operating efficiency" and reducing costs.

Hollerith's machines enabled colleges and universities to streamline their record-keeping practices and treat student records as analyzable slices of data. Institutions either copied in whole or targeted in part segments of the information within cumulative records to be replicated in punched-card form depending on what they valued. Using punched-card records, institutions ran specialized reports about their student body, including class rankings and aggregate grading statistics. For examples, colleges and universities developed enrollment reports using personal information to describe incoming and current classes of students to uncover, for example, the gender ratio of the sophomore class or the state residence breakdown of the entire student body. And, as was done at the University of Texas, institutions correlated academic performance with student information to better understand learning outcomes, foreshadowing what has become the norm with learning analytic technologies (Fichtenbaum, 1935, p. 47).

2.4. The Rise of Data Banks

The post-World War II era brought about significant changes in information technology throughout society and within higher education. The punched-card system, which was arguably the "forerunner" (U.S. Department of Health, Education, & Welfare, 1973, p. 197) to the modern-day computer, began a new phase in information storage, retrieval, and analysis, one that can be characterized as monumental. But regardless of the efficiencies the punched-card system helped universities gain, it had its problems. The act of manually creating thousands of punched cards and then tabulating each one mechanically took significant time. Moreover, the process required centralized storage of the cards and specialized technical understanding to make good use of them. Major evolutions in information technology addressed both of these deficiencies in record keeping with the development of electronic data processing and data banks, which once again transformed how information was recorded about students and how institutions used it.

Drawing on their research of 55 case sites of governmental, commercial, and non-profit organizations, Westin and Baker (1972) found that electronic data processing in data aggregation and use practices were "qualitatively different from anything which existed in the manual era" (p. 231) of record keeping. In higher education, electronic data processing advancements proved to be a boon to administrative offices. Rourke and Brooks (1966) found that computers with data storage technology had permeated work in offices of student affairs, institutional financial planning and management, and the physical plant. And once a "sufficient backlog" (Rourke & Brooks, 1966, p. 28) of data was acquired, these offices, and especially those of institutional researchers, began using the data along with statistical measurements to develop policy and model future visions of their university.

With electronic data technology opening new doors to information processing opportunities once unfathomable, university administration envisioned the "development of a total administrative information system" (Rourke & Brooks, 1966, p. 33). It was thought that such a network of computers with advanced data processing capabilities would enable the president and her fellow administrators the ability to monitor their institution "in the same way in which a modern general… can observe the progress of a military action" (Rourke & Brooks, 1966, p. 34). Regardless of their grand illusions, this remained a vision in part due to the slower than expected pace of technological development and certain cultural factors, including concerns about the mechanization and dehumanization of the university and potential threats to departmental and faculty autonomy. For example, at one university, departments had remained in control of their punched cards. When the administration began converting them to magnetic tape for future data processing, faculty and staff were disinclined to handover their records due to concerns about the increased scrutiny that data aggregation and analysis could bring to their respective departments (Rourke & Brooks, 1966).

Although higher education institutions were on the cutting edge of research into advanced computing technologies and had visions for information technology, the growth of computer applications for administrative and record-keeping purposes had mostly "grown slowly" (Westin & Baker, 1972, p. 169), except, that is, where student records were concerned. In the mid 1950s to late 1960s, colleges and universities began creating data banks of student records, which included student information from registrar and admissions offices, disciplinary details from deans' offices, and other "special facilities" information from health and wellness departments (Westin & Baker, 1972, p. 169). In part, data practices increased because outside stakeholders, like legislative bodies and state and federal agencies, "pressed institutions to improve their data systems" (Balderston, 1974, p. 51). In response, colleges and universities began to demand hard data to prove the efficacy of institutional funding requests and better understand their student bodies (Bagley, 1967).

Consider a few examples of data-based student information practices. At the Massachusetts Institute of Technology, its disparate offices began combining their data on individual students in an ad hoc fashion into a centralized data bank in the late 1960s. In the early 1970s, the institution codified data centralization efforts, creating a new office of

Administrative Information Systems to streamline aggregation efforts and improve data accessibility. New data-based records successfully captured details of a student's life from application for admission, to matriculation, and beyond graduation as an alumni (Westin & Baker, 1972). The University of Toronto's Faculty of Medicine believed that electronic data processing would "relieve the administrative workload" (Clancy, Hoke, & Mullan, 1975, p. 84) that came with serving a growing student body and sprawling medical campus. Its student record program subsumed demographic, financial, course scheduling, academic, and electronic examination data created from *within* and *outside* the institution by including demographic data provided by a central admissions service, which at the time served all Ontario medical schools.

Student data banks may have increased administrative information practices, but they also presented previously unknown issues. As more offices contributed student and other information to data banks, and as computer technology increased in ubiquity, institutions were beset with overwhelming amounts of data. Technically, colleges and universities had to confront ever-changing needs for information storage and retrieval, as well as security. Socially, institutions, for the first time, had to consider the legal and moral implications of storing comprehensive records about students that were far different in character and scope than ever before (Bagley, 1967). In response, higher education institutions were forced to draft information policy that could account for growing digital records about students and govern their access, use, and disclosure within and outside of campus borders (Fincher, 1977).

Of primary concern was the potential data banks posed to deteriorate a student's information privacy. In the past, higher education was not under a "social and legal dictum," as Virunurm and Gaunt (1977, p. 56) wrote, to assure students that their information would be protected and their privacy upheld. Neither norms nor laws existed or were sufficiently influential to spur institutions to protect sensitive information about students and respect their privacy.

Data banks increased worries about privacy at a societal level during the 1960s and 1970s. The core of such concerns centered on large-scale government surveillance and the opportunities data banks created for government actors to control individuals (Regan, 1995). As it was with higher education institutions, government agencies began aggregating data from separate databases in order to improve operational efficiencies. While the technical capacity (i.e., the ability to store, access, search, and retrieve personal information) of such expansive national data banks increased concerns, it was the potential misuse of the technology by the government that propelled privacy conversations to the fore. At the center of the debate was the Social Science Research Council's proposed Federal Data Center in 1965 and the subsequent proposal for a National Data Center in 1968; both programs sought to aggregate government data for research purposes. Even though both proposals were denied, they motivated privacy discussions in academic, social, and congressional circles.

During this time, congressional committees and government agencies sprung into action to consider the many facets of privacy and their connection to emerging computer technologies. Consider the following timeline of events: in 1964, a Special Subcommittee on Invasion of Privacy was established in the House; special hearings were convened in both chambers in the late 1960s; the Senate Judiciary Committee's Subcommittee on Constitutional Rights began a four-year study of government data banks in 1970; and congressional fervor over privacy only increased after the Watergate revelations in the early 1970s. The four-year study of data banks concluded in 1973, and the Department of Health, Education, and Welfare (HEW) subsequently published its *Records, Computers, and the Rights of Citizens* report. The report developed the Code of

Fair Information Practices (CoFIP), which framed many information privacy laws and continues to this day to influence privacy rules and regulations in "nearly every U.S. industry" (Ramirez, 2009, p. 8). Notably, the Fair Credit Reporting Act of 1970, the Family Educational Rights and Privacy of Act (FERPA) of 1974, the Privacy Act of 1974, and the Right to Financial Privacy Act of 1978 were all passed in order to protect individual privacy against the potentially negative consequences of electronic and other records held by organizations, institutions, and government agencies. Of all the privacy legislation passed in the post-World War II era, FERPA was, and continues to be, the most formative law to address student privacy in the context of higher education.

Before FERPA, student privacy was indistinguishable from individual privacy protections provided under an amalgamation of constitutional rights, including the Fourth Amendment, the due process clauses of the Fifth and Fourteenth Amendments, and the tort right to privacy (Caruso, 1971; Stevens, 1980). Some state statutes provided students protections regarding their records, but they varied from state to state (Thomas, 1978). As some scholars have argued, student records protections were also governed under explicit or implied contracts between the student and the institution, since the courts historically used contract law to determine student rights after *in loco parentis* ceased to be (The confidentiality of university student records, 1976; Thomas, 1978). While privacy protections existed, if dispersed throughout constitutional, common, and state law, few cases specifically addressed student privacy, especially information privacy (Caruso, 1971; The confidentiality of university student records, 1976). When case law did address student records and information privacy, courts often "disposed of the issue without articulating a legal basis" (The confidentiality of university student records, 1976, p. 479).[2] FERPA essentially filled a number of gaps in the legal fabric and provided much clearer direction to students, parents and guardians, and educational institutions regarding student privacy rights and institutional record-keeping practices.

At the core of FERPA rests a motivating principle: Students should have the right to access and maintain some control over their educational record. The framers of FERPA, senators James L. Buckley (R-NY) and Claiborne de Borda Pell (D-RI), feared that a student's record may serve as a damning dossier that could haunt her in her personal, academic, and professional future. If libelous or intrusive, the record may live as a burdensome file that could ultimately limit the student's control over her own life.

As mentioned previously, FERPA was built on the CoFIP developed by the HEW in 1973, which included the following principles:

1. there must be no personal data record-keeping systems whose very existence is secret;
2. there must be a way for a person to find out what information about the person is in a record and how it is used;
3. there must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person's consent;
4. there must be a way for a person to correct or amend a record of identifiable information about the person;

[2] Van Allen v. McCleary (1961) is a notable exception in the public education context. Here, Van Allen petitioned for access to the public school records of his son. The court found that the state constitution of New York, state statutes, and state agency rules and regulations did not inform the decision; therefore, the court used common law to rule in the father's favor, making analogous comparisons with a patient's interest in her health records, a stockholder's interest in corporate records, and a client's interest in her attorney's file.

5. and any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.

All five principles are found embedded in FERPA in some form. For example, if any type of file, form, document, or similar recording relates to the student, §99.10 states that she has "the right to inspect and review" those records, which attempts to shine light on secret record-keeping systems and practices about the student, enables her to audit her record, and in effect, conforms to CoFIP numbers one and two. Additionally, §99.30 satisfies CoFIP number three by limiting disclosure without consent; section §99.21 aligns with CoFIP number four by enabling students to request a hearing to "challenge the content" of their records; and CoFIP number five is strewn throughout multiple sections, including §99.32 which requires institutions to maintain the integrity of the record by keeping track of who has "requested or obtained access" to an individual student's record.

What is pertinent about FERPA is that it was drafted in response to a growing societal concern about organizations and institutions, like higher education, that were building technological capacity for and increasing their reliance on personally identifiable digital records. New technologies enabled colleges and universities to record, store, aggregate, search, and analyze the store of student records they had been growing over the years. FERPA acted as a counterbalance to the unlimited power universities had to create student profiles without affording students privacy rights. This counterbalancing role, however, should not be misconstrued as an effort to limit the scope of information gathering practices: it placed almost no limits on the types of data and information that could be included in an educational record and how those records could be used within the institution. Colleges and universities could still create digital cumulative student records, but now there were some limits regarding record disclosures.

2.5. Ubiquitous Systems

As database technology evolved, and as terminal computers changed into the personal computers we know today, processing power and storage capacity transformed stores of digital data and what could be done with it. After the 1970s, computers, such as the Xerox Alto, Apple-1, The Commodore, and many others, quickly came into existence and were just as quickly overshadowed by subsequent models; software, as well, experienced a short lifespan. While modern computing technology does not experience the same turnover as it once did, it still evolves at a fast rate. However, where student records are concerned, little has changed primarily because they continue to be stored in and accessed using databases. What has changed in modern times, however, is higher education’s reliance on information technology and the speed at which student data is created by faculty, staff, and students as they interact with a university's information infrastructure.

Strewn throughout university campuses are miles of fiber-optic cables, the so-called “pipes” upon which all networked information technology relies. Gone are the days where a few terminal computers were available to a select group of campus actors. Today, networked technologies are common throughout all institutional departments and offices, and the information on which university employees rely is often born digital. File cabinets still exist, but they do so more as a relic or as a part of an analog information practice soon to be digitally transformed. Instead, digital records are created in unique systems, and institutional actors can access, modify, duplicate, and send the records across the campus and the world in real time.

An increase in networked computers and a growing dependence on data-based technologies brought about a need for specialized systems that could handle new information practices. Gorr and Hossler (2006) noted that as higher education institutions continued to grow, both in terms of their physical footprint and population of students, they turned to enterprise resource planning systems to manage core administrative functions, and, over time, these systems have "become a dominant concern" (Hossler, 2006, p. 76) in higher education. Such systems were not cheap, however. In 2006, some institutions spent up to $500 million (Hossler and Gorr, 2006); furthermore, the total cost of ownership of systems often have forgotten or hidden costs institutions do not plan for, such as system upgrade requirements and the need to invest in new staff to support complicated system installations (Babey, 2006; Hossler, 2006). Nonetheless, offices of campus information technology continue to spend a significant amount of their budgets on technical and personnel resources to implement and support new systems.

Included with the crop of newly adopted enterprise systems was the student information system, a new technology to create, store, and analyze records about students. Arguably, no system is more important than a campus's student information system, as it documents the academic life of all individually-identifiable students on campus. By recording when the student applies for admission, matriculates, and graduates, the system captures the beginning and endpoints of the student's relationship with the institution and relevant academic data, such as grades, course status (e.g., enrolled, passed, failed), and the student's stated program major and minor, among other details. But such systems also document a significant amount of personal, familial, and financial information with such exactness that the student records they store rival the cumulative records of the past. This is due in part to the breadth of information applications for admission require of students, as well as other information federal and institutional applications require for students to apply for financial aid.

For institutional needs, student information systems provide a trove of data to administrators to manage the day-to-day activities of the campus. By knowing the exact number of students registered for a given course, institutions can engage in detailed space planning by analyzing stresses on instructional technology resources, lecture halls, entire buildings, and heavily trafficked public spaces around campus. Student information systems also keep track of and control access to course registrations from semester to semester. And since student information systems store demographic data about the entire student body, institutional research offices often mine digital student records to create reports regarding student retention, especially among historically disadvantaged or at-risk students. Access to student data by institutional researchers and other administrative offices has become especially important, as it enables them to report on the success of the institution to external stakeholders, such as accrediting bodies like the Higher Learning Commission. In essence, the student information system serves as a "front door" (Halverson & Shapiro, 2013) to student data for institutional actors and authorized third parties.

What an institution keeps on record about a student primarily depends on the ends to which the record will be put, but also on the technical affordances of its student information system. For example, student information systems may not be able to retrieve data from other information technology systems and vice versa. Colleges and universities often purchase systems from vendors who design the technology with a generic institution in mind: they are typically not custom-fit to the specifications of an individual institution. This has been especially problematic since systems were not designed with a concern for data interoperability. In more recent times, data standards related to student data have come to fruition, enabling crosswalks from one data schema to another. The Common Education Data Standards (n.d.), for instance, aims to develop a data library to enable student records to flow between systems from one educational institution to the next throughout the entirety of a student's learning history. In lieu of set data standards and designed interoperability between systems, institutions have turned to data warehousing technologies in order to aggregate student data from disparate systems on campus.

Institutional data warehouses allow colleges and universities the ability to bridge the nodes in the vast network of campus information technology. As offices and departments either built or adopted various information systems to fit their needs, data became stuck in technological silos, such as databases, spreadsheets, or other electronic files. Student information systems continue to hold the majority of information about students, but data warehouses allow institutions to "desilo" data and "allow multiple users to extract meaningful, consistent, and accurate data" (Ingham, 2000, p. 132) for various purposes. Students interact with any number of student offices, including that of advising, and provide information about themselves in order to receive potential benefits–like grants and scholarships–and actual benefits, such as campus employment, access to technology, and the use of bus services. This sort of information, when aggregated in a data warehouse, adds to a rich, revealing profile of a student's personal and academic experience while on campus.

Data warehouses empower colleges and universities to extract useful information about students and the institution for important reporting purposes. And quality data and reporting is important, especially given the role of accountability measures and processes in American higher education. While a trust-based contract once bound the relationship between higher education institutions and outside stakeholders, including government bodies, tax-payers, and those footing the bill for tuition (primarily parents), since the 1980s the same stakeholders have turned towards data transparency and accountability to force institutions to use measures of efficiency and effectiveness to prove that their educational programs are of a certain quality and deserving of the tuition and fees they charge (Huisman & Currie, 2004; Trow, 1996). Pressures, especially from government, for accountability were spurred by the critical Spellings Commission Report (U.S. Department of Education, 2006). Among other facets of higher education that the Spellings Commission Report found deficient, it highlighted the inability of stakeholders to hold institutions accountable due to "limited and inadequate" (p. 4) data systems and reporting among higher education institutions. Data warehouses are not *the* solution, the report found, but they may be part of a solution that enables institutions to use data about students more effectively to respond to institutional needs and stakeholder pressures.

Stakeholders often cite low student retention measures as a key concern in their accountable arguments. Defined, student retention is the ability of a university to shepherd a student successfully to graduation; it is also central to an institution's ability to maximize resources–if a student drops out, the resources the university invests for that student are for naught. At a more abstract level, higher education has a responsibility to society to retain, educate, and successfully prepare students for a fulfilling life, personally and professionally. When colleges and universities fail to retain students, they are culpable and their relationship with society is weakened. Students who fail to graduate are often burdened with significant financial responsibility, the institution's reputation may be at stake, and society may lose the skill and knowledge that those students once had to offer (Crossing, Heagney, & Thomas, 2009; Yorke & Longden, 2004).

Due to the importance of student retention, academics, institutional policymakers, and stakeholders have put significant effort into research in this area to identify which variables positively correlate with higher retention rates. Peltier, Laden, and Matranga (1999) scoured the literature and described the breadth of variables often used in retention studies, which include: student involvement, demographic characteristics (e.g., ethnicity, age, and gender), residential status and location, and affiliation with sororities and fraternities. Some of the data Peltier et al. (1999) mention is commonly available in student information systems, while others, like a student's sorority affiliation may only be available in a specialized database maintained by an office for Greek life. Still, this data only provides a snapshot of a student's life. "Data famines" (Buglear, 2009, p. 383), or the inability for institutions to capture more data about a student, create "an insufficient basis" to explore variables related to retention, reduce the ability for institutions to act in a timely manner, and hinder the development of institutional retention policies. Data warehouses fill this crucial gap by providing institutions with student data that could reveal explanatory information in retention studies.

The federal government has not been patient with colleges and universities to demonstrate institutional efficacy with student data. Each year, over 7,500 state, private, and for-profit higher education institutions must submit a vast array of data about student enrollment, program completion, graduation rates, and institutional characteristics in order to participate in federal student aid programs (About IPEDS, n.d.), as required under Title IV of the amended Higher Education Act of 1965. In aggregated, anonymized form, higher education institutions submit data for inclusion in the Integrated Postsecondary Education Data System (IPEDS). Since 2003, IPEDS has specifically tracked first-year retention rates, and graduation rates have been tracked since 1990 as a requirement of the Student Right-to-Know and Campus Security

Act. While the IPEDS has provided "large quantities of big picture information" about colleges, universities, and to a limited extent, student achievement, some argue that the system "was not designed to ask many of the questions that students, families, institutions, and policymakers" (McAnn & Laitinen, 2014, p. 4) ask about the success of individual institutions. In response, the Department of Education has pushed for a federal unit record system (FURS) since 2005 to replace the IPEDS summary data system to comprehensively track identifiable students throughout their educational career, from pre-kindergarten through postsecondary levels, and there are future ambitions to aggregate workforce data, as well.

Policymakers and higher education administrators have sought an "integrated, inclusive, longitudinal student-level data system" like the "Holy Grail" (Hearn, McKlendon, & Mokher, 2008, p. 665), and nothing comes as close to what they seek like a FURS. The FURS idea was introduced in 2005 by the National Center for Education Statistic (NCES) in a feasibility study in response to "growing interest" (Cunningham & Milam, 2005, p. iii) from postsecondary stakeholders in "more accurate measures" and due to "congressional desire to hold postsecondary institutions accountable for student outcomes." The National Commission on Accountability in Higher Education, one of the powerful postsecondary stakeholder groups seeking change, argued that the IPEDS system for student data analysis was "outmoded and inaccurate," and a new federally-based system could push forward the state of student data analysis (Fischer, K., 2005, para. 6). The Spellings Commission directly referred to the FURS recommendation within its report, stating that "the commission supports the development of a *privacy-protected* higher education information system that collects, analyzes and uses student-level data as a vital tool for accountability, policy-making, and consumer choice" (U.S. Department of Education, 2006, p. 21, emphasis in original). But the degree to which such a system can protect student privacy is limited by the fact that it requires personally identifiable information. As planned, a FURS record would encompass a number of different demographic and academic data points about a student, and include the student's name, Social Security Number, address, and date of birth. In essence, a FURS student record would serve as a long-lasting digital dossier for *all* students within American institutions, if the student or the institution received federal aid. And even if students did not receive *direct* federal aid themselves, Cunningham and Milam (2005) argue that they received *indirect* federal support, and therefore, their student data should be included in a FURS irrespective of the student's wishes.

A FURS immediately raised privacy concerns. While the report directly recognized that privacy, confidentiality, and data security issues existed, it did not give them much attention, citing that the NCES had always protected identifiable data using high-level security measures. Additionally, the NCES was, and continues to be, subject to "stringent requirements" (McCann & Laitinen, 2014, p. 10) to uphold student privacy by a collage of federal laws, and large financial penalties–up to $250,000–or a felony conviction await those who improperly disclose identifiable student data (Confidentiality laws, n.d.). Furthermore, the report pointed out that the use of a Social Security Number, while concerning, was already a common practice for applications for Federal Student Aid.

These assurances did little to assuage the concerns of critics of a FURS, who often invoked chilling visions from George Orwell's *1984*. Katherine Haley Will, former president of Gettysburg College and a vocal critic of a FURS, cast it as an "Orwellian scheme" (2006, para. 9) under which "information could all too easily be shared with other government agencies or even with the private sector" (2005, para. 5). Similarly, Sarah Flanagan of the National Association of Independent Colleges and Universities (NAICU) felt that an immense government database full of student records had “a Big Brother aspect to it” (“Public-private split,” 2004, p. 2). Legislative representatives, like Representative John Boehner (R-Ohio) (2005), dramatically argued that a “monster database” (para. 10) of student records was what Big Brother dreamed about at night. Officials with the Free Congress Foundation and the Eagle Forum were particularly concerned with the possibility that the student database would experience breadth, depth, and “creep” (Burd, 2005, para. 27) over time.

Public opinion also opposed the use of federal unit records about students. Shortly following the release of the Spelling Commission Report, the NAICU conducted a three-question poll to garner the opinion of the American public. It found that 45 percent of respondents strongly opposed the proposed FURS, and 60 percent opposed requiring colleges and universities to report individual student information to the federal government; many of the respondents expressed concern over data safeguards, as well (Powers, 2006). Since the poll was sponsored by the leading detractor to a FURS plan, its validity was naturally called into question by opposing parties. Nonetheless, the poll galvanized legislative actors in the name of student privacy.

There were two legislative responses to the FURS recommendation: 1) increase access to information about higher education institutions at the aggregate level, or 2) ban any kind of federal unit record. With regard to the first approach, representatives John Boehner (R-Ohio) and Howard P. McKeon (R-California) introduced the College Access and Opportunity Act in 2005 (2006), which aimed to increase access to the amount of data provided to IPEDS and to create a college affordability index to inform students and parents about the cost of individual institutions in comparison with their institutional peers. With regard to the second approach, legislation outright banned federal student databases, which is what the 2006 version of the College Access and Opportunity Act did in §109; it reads:

> [N]othing in this Act shall be construed to authorize the design, development, creation, implementation, or maintenance of a nationwide database of personally identifiable information on individuals receiving assistance, attending institutions receiving assistance, or otherwise involved in any studies or other collections of data under this Act, including a student unit record system, an education bar code system, or any other system that tracks individual students over time.

Neither the 2005 or 2006 act became law, but they set the framework for future legislation. In August of 2008, the Higher Education Opportunity Act, which did become law, outright banned a FURS using much of the language introduced by Representative Boehner in 2006.

While the Higher Education Opportunity Act prohibited a federal system, it gave states the option to create their own state unit record systems (SURSs), a legislative move which has been enthusiastically supported by presidents George W. Bush and Barack Obama. In 2005, the Department of Education awarded over $50 million to 14 states to develop such systems. In 2009, the government provided an additional $250 million from economic stimulus funds, doubling the number of states using SURSs up to 31 states (Basken, 2010). And since 2009, according to McCann and Laitenen (2014), the federal government has invested over $500 million in state-based student data systems, in part to work around the FURS ban.

Recent legislation has sought alternatives to the FURS ban introduced in 2008. Senator Ron Wyden (D-Oregon) drafted the Student Right to Know Before You Go Act of 2012 (S. 2098, 2012) in order to force institutions to participate in SURSs and allow data to be collected in a third-party system, in much the same way that the National Student Clearinghouse is used to aggregate and analyze data required by the federal government to meet financial aid requirements; the same bill was introduced by Representative Duncan Hunter (R-California) in the House (H.R. 4061, 2012). "How can it be that we lack the data," Wyden and fellow Senator Marco Rubio (R-Florida) (2012, paras. 3–4) opined in *USA Today* about the lack of usable information about higher education institutions, "[T]his is a little bit shocking considering we live in a data driven [*sic*] world." In a separate piece, Wyden and Rubio (2014, para. 2) asserted that current educational policy was ill-informed due to data deficiencies:

> Any policy debate should start with a clear picture of how the dollars are being spent and whether that money is achieving the desired outcomes. Unfortunately, a lack of accurate data makes it impossible to answer many of the most basic questions for students, families and policy makers who are investing significant time and money in higher education.

Both versions of the Student Right to Know Before You Go Act of 2012 stalled in their respective chambers of the 112th Congress. However, both bills were reintroduced in the 113th Congress beginning in 2013 with a notable change: they required the Department of Education to create a FURS, putting to rest their original SURS requirement and reinvigorating the conversation about a federal system (H.R. 1937, 2013; S. 915, 2013). Neither bill has yet to pass in either chamber.

Chapter 3. A New Era: Learning Analytics as a Big Data Practice

3.1. Introduction

Metaphors can be powerful, and they have been applied to Big Data with great effect. Often, Big Data is referred to as the "new oil" (see Lane, 2014; Mayer-Schönberger & Cukier, 2013; Moss, 2014; Rotella, 2012) or a "gold mine" (see Asay, 2013; Peters, 2012; Steinberg, 2013). Held within these illustrative comparisons is the idea that Big Data is a resource, not unlike the most coveted of resources in our society, that holds *potential* political, financial, and social value for those who are able to control, refine, and maximize it. With significant returns a possibility, it is easy to understand why much of society is interested in Big Data's allure, including higher education. Colleges and universities are now beginning to capitalize on the promise of Big Data by adopting new information practices and technologies–often under the umbrella term of "learning analytics"–to analyze student data and information. This section begins by describing the emergence of learning analytics with a special emphasis on issues regarding its definition. Next, I address some motivating factors that are helping to build a foundation for and drive learning analytics' development. Following this, I situate learning analytics as a Big Data practice and illustrate through examples how institutions are using learning analytics or plan to in the future. Finally, I consider the growing ties between learning analytics practices and the culture of assessment in higher education.

3.2. Analytics in the Academy

Analytics is an amorphous term. Since the early 1970s, its label, definition, and the applications to which it has been put have all changed. According to Davenport (2014, p.10), the terms “decision support,” “executive support,” “online analytical processing,” “business intelligence,” and “big data” are all types of analytics that encompass more or less the same thing: using data to make sense of a particular area of the world. The basis for each term is dependent on the social milieu of the time, the technological affordances of the day’s systems, and the politics and values driving data analysis. As such, the definition of analytics and applications thereof are context dependent. Higher education is one of those specific contexts, and it is increasingly adopting analytic technologies to mine student data and gain institutional acuity.

Two distinct forms of analytics have emerged in colleges and universities, the first of which being academic analytics and the second being learning analytics. The degree of separation between the two terms, at first thought, seems minimal: “academic” and “learning” are etymological relatives. Yet, there are important differences in each term deserving of individual attention.

Academic analytics is the precursor to learning analytics. In their first-of-a-kind report on academic analytics, Goldstein and Katz (2005) explained that “the challenge [to colleges and universities] is no longer the lack of access to timely information” (p. 11), it is the ability to make actionable decisions based on available information. Regardless of the multimillion dollar investments higher education institutions made in information technology from the 1970s forward, they have done little to capture, analyze, understand, and make use of the various stores of data to which they now have access. Academic analytics, argued Goldstein and Katz (2005), could change that by optimizing and creating new revenue streams that recoup past financial investments in information technology.

The roots of academic analytics are firmly planted in the business intelligence field. Like business intelligence, academic analytics is the use of various technological systems and applications to analyze accessible institutional data in support of decision making. In fact, there is very little difference between the two terms; however, to Goldstein and Katz (2005), "[business intelligence] rang hollow to [their] delicately trained academic ears" (p. 6), so they rebranded the term as they saw fit. Their intent in doing so was to realign business intelligence to the particular needs of the academic environment, including the complex interplay between financial, operational, and academic needs and interests.

Proponents of academic analytics argue that information derived from analytic practices can be used to defend against mounting accountability pressures. There are a number of hurdles for and weaknesses in American higher education, according to Campbell, DeBlois, and Oblinger (2007). One is that emerging countries like India and China, whose growing economies rely on an educated population, are investing heavily in education in order to increase the proportion of their respective populations with postsecondary degrees, whereas the United States has slipped in its ranking among industrialized nations with postsecondary degrees. Another issue is weak educational attainment among growing minority populations. Where the American economy is concerned, growth sectors are requiring postsecondary education and specialized skills; if gaps continue to grow between sectors requiring degrees and lower postsecondary educational advancement, Campbell et al. (2007, p. 42) cite research that puts United States per capita income at a 2 percent loss over 20 years. The downstream effect would be a less-educated workforce and a weaker economy. The salve for these systemic problems, argue Campbell et al. (2007), is academic analytics, writing:

> Thanks to enterprise-wide systems that generate massive amounts of data, data warehouses that aggregate disparate types of data, and processing power that sifts, sorts, and surfaces patterns, academic analytics is emerging as a new tool that can address what seem like intractable challenges. (p. 42)

Whether academic analytics has or will resolve such systemic issues is still an open question.

Some institutions have initiated programs to capitalize on the purported gains of academic analytics. Baylor University, the University of Alabama, Sinclair Community College, and Northern Arizona University have all implemented academic analytics initiatives to address student management issues surrounding enrollment and retention, using predictive modeling in the programs to identify and quickly address emerging problems (Norris, Baer, Leonard, Pugliese, & Lefrere, 2008). Arizona State University's systemic analytical programs, which are arguably more mature than most other institutions, found that academic analytics helped the institution make more fully informed decisions, accomplish strategic objectives, gain a competitive advantage over its peer institutions, and improve enrollment and retention rates (Goldstein & Katz, 2005, p. 14).

Academic analytics predates learning analytics by about five years, when it became a prominent topic among educational technology professionals in 2005, but after that learning analytics slipped into common parlance and has now replaced academic analytics entirely. Often a harbinger of educational technology terminology shifts and trends in higher education, EDUCAUSE's annual *Horizon Report* identified learning analytics language and recognized the nascent value in the technology as early as 2010. In that year's report, the authors identified the use of advanced computational methods and data visualization techniques as a potentially influential strategy for understanding "the most complex of social processes" (Johnson, Levine, Smith, & Stone, 2010, p. 30): student learning. By aggregating, streamlining, and making accessible large stores of data, institutions could "make visible" (Bienkowski, Feng, & Means, 2012, p. ix) student data once "unseen, unnoticed, and therefore unactionable," enabling institutional actors to interrogate relationships and patterns related to learning. The 2010 report used neither the "academic analytics" or "learning analytics" terminology, but we can see a shift in focus emerging. Academic analytics focused on the institution as a data subject for analysis, but this pivot in the report puts attention on students and their learning behaviors as the new target of data analytics.

In 2011, the *Horizon Report* named "learning analytics" as the key type of analytics to be adopted on college campuses by 2015 (Johnson, Smith, Willis, Levine, & Haywood, 2011). The authors of the report defined learning analytics as "the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues" (p. 28). The goal of learning analytics, they wrote, is to "tailor educational opportunities to each student's level of need and ability" (Johnson et al., 2011, p. 28). Yet, the report makes it clear: learning analytics is *not* just about learners, it is also about the learning context and can be used to "assess curricula, programs, and institutions" (p. 28). In that respect, learning analytics is clearly an offshoot from academic analytics. In the most recent *Horizon Report* (Johnson, Adams Becker, Estrada, & Freeman, 2015), learning analytics permeated much of the language regarding data-driven educational technology, so much so that EDUCAUSE now treats it as a common, accepted term.

Although the *Horizon Report* clearly defined and now freely uses "learning analytics," the definition continues to be contested by researchers and practitioners alike. In hopes of gaining definitional clarity, Van Barneveld, Arnold, and Campbell (2012) set out to synthesize the literature by describing characteristics of learning analytics. They arrived at seven varied "conceptual and functional" terms, which include: analytics, business analytics, academic analytics, learning analytics (academia), learning analytics (industry), predictive analytics, and action analytics. They attached each term to a specific level or levels of focus, such as whether or not the analytics concerned the entire institution, departments, instructors, or students. The term and its level of focus, Van Barneveld et al. (2012, p. 3) stated, depended on the specific "goals and objectives" to which analytics was put. Based on their fusion of the various definitions of analytics in higher education, they conceived the following unified definitions:

> - Analytics: An overarching concept that is defined as data-driven decision making.
> - Academic analytics: A process for higher education institutions with the data necessary to support operational and financial decision making.
> - Learning analytics: The use of analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals.
> - Predictive analytics: An area of statistical analysis that deals with extracting information using various technologies to uncover relationships and patterns with large volumes of data that can be used to predict behavior and events. (Van Barneveld et al., 2012, p. 8)

It is important to consider the change in terminology from academic to learning analytics. There are three possible reasons for this change. The first is simply the fact that analytics in the academy was a novel practice. The technology and science behind analytical work using large datasets in the context of higher education was new, and those involved were trying to accurately name this emerging type of work. Over time, the terminology naturally shifted from academic to learning analytics.

A second answer that explains the shift is found in the above definitions. In the early history of the literature, academic analytics is focused heavily on improving institutional decision making, efficiencies, and outcomes. When learning analytics rose in prominence, the focus on the institution decreased and turned instead to the learner; learning outcomes, not institutional efficiency, formed the target of analytic practices. More simply put, the goals of analytic practices changed and that warranted a new, more accurate term.

Another plausible answer is that the evolution in the terminology was political. Advocates of business intelligence practices in higher education, on which academic analytics was founded, recognized that data-driven decision making would grate on some in the academy who see their work as morally superior to concerns about efficiency, effectiveness, and revenue streams. In an attempt to refocus the conversation away from core business intelligence principles, EDUCAUSE and others recast the term and reset the conversation. As a result, learning analytics emerged with a new focus on the learner, or so it seemed.

Regardless of the shift in terminology, academic analytics goals are embedded in emerging learning analytics initiatives and conversations, which has prompted stout learning analytics advocates to remind researchers and practitioners that learning analytics should be refocused on learning (see Gašević, Dawson, & Siemens, 2015). This may be a useful call to action for learning analytics researchers, but present-day learning analytics conversations continue to include concerns about institutional efficiencies and effectiveness alongside behavioral learning patterns and improved learning outcomes.

3.3. Building a Foundation for Learning Analytics

The *Horizon Report* and others continue to note that learning analytics is rising both in prominence and adoption rates in higher education. And although Gartner's *Hype Cycle for Education* report remarks that mainstream adoption of learning analytics will experience a "bumpy road" (Lowendahl, 2013, p. 32) in the coming years, the research organization maintains that once the foundation is laid, it has the potential to transform education. The following areas address how a foundation for learning analytics is being built and by whom.

The community of learning analytics researchers is growing significantly due to researcher interest in the area and, subsequently, increased scholarly output. Notably, the Society for Learning Analytics Researchers (SoLAR) developed in late 2011, which continues to bring together interested researchers and practitioners alike to share research findings, build a collaborative network, and create opportunities for community development with face-to-face and virtual events, conferences, and workshops. SoLAR also created the open-access, peer-reviewed *Journal of Learning Analytics* in early 2013 and published its first issue in 2014. Other publications, such as yearly proceedings from the Learning Analytics and Knowledge (LAK) conference, have become established places for the field to publish its research, but citation analyses and assessment of learning analytics research also shows that the field has become quite diverse in its approach with regard to the venues in which it publishes its work (Dawson, Gašević, Siemens, & Joksimovic, 2014).

Institutional collaboration has in the past and continues at the present to build up institutional capacities for learning analytics–especially in relationship to technological infrastructure. The following four partnerships deserve special attention due to their prominence: IMS Global's Caliper standards, the recently announced Unizin Consortium, the Apereo Foundation, and SoLAR's Open Learning Analytics (OLA) initiative.

IMS Global is developing a standards framework, called Caliper, to improve learning analytics platforms. IMS Global's mission is to "advance technology that can affordably scale and improve educational participation and attainment" ("About IMS Global," 2014), and by developing a "learning measurement framework," it expects that new standards will help facilitate the capture and analysis of wide varieties of learner data for analytics purposes (IMS Global, 2013). Caliper standards work in coordination with other IMS Global standards, which are set by member institutions. Unlike some of the collaborations mentioned below, the pay-for membership of IMS Global consists heavily of for-profit companies (e.g., Educational Testing Service, ACT, Blackboard, McGraw Hill) with a mix of non-profit educational institutions (e.g., University of Michigan and Pennsylvania State University), and for-profit educational companies (e.g., University of Phoenix and the American Public University System).

In contrast to IMS Global, Unizin consists only of American, non-profit higher education institutions. The consortium's aim is to maintain "control of the content, data, relationships, and reputations" (Unizin, 2014a, para. 1) the member institutions create in order to "bias things in the direction of *open* standards, interoperability, and scale" (emphasis added). Some driving goals of Unizin are to create a "digital learning ecosystem" (Unizin, 2014b, para. 5) of sharable digital learning resources, to improve interoperability between various learning systems, and develop cross-institutional learning analytics technology, policies, and research (Burns, Hilton, & Patterson, 2014). Ironically, the Unizin blog post is also quick to note that the standards and infrastructure it develops will only be available to member institutions. So, it is unclear to what degree the consortium's work in the area of learning analytics will actually be open to the benefit of higher education and educational technology communities, if at all.

The Apereo Foundation supports an international membership of higher education institutions along with a minority of technology companies. Its aim is to incubate, develop, and share educational technology resources and solutions for all fee-paying members. To date, their Learning Analytics Initiative has brought together already successful projects from Marist College and the University of Amsterdam to "accelerate the operationalization of Learning Analytics software and frameworks, support the validation of analytics pilots… and work together… to avoid duplication of efforts" ("Learning analytics initiative," n.d., para. 1; Zeckoski, 2014).

Finally, SoLAR's OLA initiative aims to provide leadership, strategic vision, and advance technology in the area. SoLAR has in the past and will in the future host strategic meetings–"Summits," as they call them–and facilitate collaboration at conferences in order to hone the work done in the OLA area. For example, a major outcome of the 2014 Summit was the selection of four major domains of focus for OLA: open research, institutional strategy and policy issues, a focus on learning sciences and design, and the development of open standards and software (Society for Learning Analytics Research, 2014b). By developing a technological platform for learning analytics, SoLAR aims to build an "analytics engine" and user interface dashboards as part of a broader, modularized learning analytics "toolkit" to support research into and application of learning analytics technology (Siemens et al., 2011).

In addition to institutional partnerships, the United States federal government has begun a concerted effort to gain access to and provide analytics services around student data. To bring forward "a new standard of openness," President Obama (Transparency and Open Government, 2009) wrote a memorandum emphasizing that government should be transparent, participatory, and collaborative (p. 4586). A number of governmental departments created projects in response to create openly accessible data sets as an act of "smart disclosure" (Sunstein, 2011). Smart disclosure allows the public to access and use standardized, machine-readable open data for analysis as means for individuals to make “informed decisions” (Sunstein, 2011, p. 2). As part of the “smart disclosure” memorandum, agencies were empowered to require or encourage entities (e.g., companies, organizations, departments) they oversaw to make information available directly to individuals or provide it to the overseeing agency for proper disclosure: MyData projects are a direct result of this initiative.

The administration pressed into service its MyData programs in order to provide individuals secure access to their data; it also aimed to stimulate private-sector innovation and services around that data (The Presidential Innovation Fellows, n.d.). While this initiative has sprouted data projects in the Department of Energy and the Department of Veterans Affairs, work has just begun in the Department of Education. The aim of an education-focused MyData project is to provide “learners of all ages” downloadable copies of their academic transcripts, records of their online learning activity, and financial aid information (My Data, n.d.; The Presidential Innovation Fellows, n.d.). Not only will the government provide the data, especially from Federal Student Aid and the Education Data Community at data.gov, but strategic public and private partnerships will also serve as data sources. Most notably collaborations with Microsoft and Pearson aim to increase data access and analytic tools for the MyData program (Chopra & Smith, 2012; Office of Science and Technology Policy, 2012).

Unique partnerships with non-profits are bolstering the education data access efforts the White House aspires to accomplish. The Next Generation Learning Challenges (NGLC) is a collaborative research initiative that funds projects that aim to “dramatically improve college readiness and college completion” through innovative uses of technology (About, n.d., para. 1). Led by EDUCAUSE and funded primarily through the Bill and Melinda Gates Foundation, the

NGLC has awarded over $37 million in grants, and of that total higher education grantees received nearly $24 million (Grantees, n.d.). Of the four specific challenges set out by the NGLC, the fourth challenge charges researchers to help "institutions, instructors, and students benefit from learning analytics" (Gammon, 2010, para. 4). In the first wave of NGLC's funding, six learning analytics projects were funded to a total of nearly $3 million (Wave I, n.d.).

The White House, EDUCAUSE, and other proponents of learning analytics all understand that large-scale analysis of student data requires a sound and robust digital environment. Without it, the data is inaccessible (or unavailable) and the analysis cannot be done. And until recently, such an environment was still immature. However, the 2014 *Horizon Report* acknowledged that the infrastructure supporting learning analytics and practices, which drive student data creation, are aiding the maturation of a stout, networked ecology of technological artifacts, data, and systems (Johnson, Adams Becker, Estrada, & Freeman, 2014). The report also posits that students spend a significant amount of time on the Internet, that the devices they use create a ubiquitous connection, and the social media that students use help them to establish flourishing social networks for personal, professional, and academic purposes. As more students engage in fully or partial online learning, an environment that requires and is supported by a vast technological network, more data may become available for analysis by learning analytics systems. Next, I turn to a description of this networked ecology in higher education in order to lay out actual and potential data sources for learning analytics.

3.4. (Big) Data for Learning Analytics

To many, learning analytics signals the arrival of Big Data in higher education. Big Data is a "cultural, technological, and scholarly phenomenon," according to boyd and Crawford

(2012, p. 663), and it has set off new conversations about the role of data in dissimilar areas of society. From the role Big Data can play in keeping Americans safe from terrorism to its usefulness in personalizing marketing, Big Data is quickly becoming a normative practice.

Defining Big Data, however, is difficult. The difference in technology that underpins it, the disparity in goals that drive it, and the innumerable variations of what is and is not considered Big Data does not allow for a "rigorous definition" (Mayer-Schönberger & Cukier, 2013, p. 6). In order to capture the *essence* of Big Data, however, Mayer-Schönberger and Cukier (2013) propose the following approach:

> 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, organizations, the relationship between citizens and governments, and more. (p. 6)

Inherent to Big Data is the notion that there is something different about modern datasets. And for us to consider the role of Big Data in higher education, we first must establish what makes the data part of Big Data unique.

Commonly, pundits explain that the data in Big Data can be distilled to three defining characteristics, which Doug Laney (2001) termed the three Vs: *volume, velocity, and variety*. Laney (2001), in his observation on fast-arriving changes in data, explained that the traditional features of data and datasets were outmoded. Data had transformed in size, the speed at which it was created had accelerated, and its once rigid column and row structure had dissolved with new formats and incongruous data structures. The volume of Big Data is the most striking characteristic. Since the beginning of civilization to 2003, humans created five exabytes of data; in 2013, humans produced the same amount of data *every two days* (Miller & Chapin, 2013 as cited in Lane & Finsel, 2014). The growth of social media applications and the high adoption rate among users has, arguably, led to the greatest increase in data. Consider the following social media usage statistics aggregated by Connor in 2012:

- Twitter users tweet 340 million times a day;
- Facebook users post more than 684,000 bits of content a day;
- YouTube users upload 72 hours (259,200 seconds) of new video a minute;
- Online consumers spend $272,000 shopping a day;
- Google receives over 2 million search queries a minute;
- Apple receives around 47,000 app downloads a minute;
- Companies on Facebook receive more than 34,000 'likes' a minute;
- Tumblr blog owners publish 27,000 new posts a minute;
- Instagram photographers share 3,600 new photos a minute;
- Flickr photographers upload 3,125 new photos a minute;
- Foursquare users check-in over 2,000 times each minute;
- Individuals and organizations create 571 new websites a minute; and,
- WordPress bloggers publish near 350 new blog posts a minute.

Each tweet, post, video, photo, and like represents new data. And since these statistics are focused solely on social media usage, they fail to account for the data created by sensors embedded into smartphones and buildings, data created in workplace systems, and data created by emerging "smart" infrastructures–like power grids–that can automatically monitor system loads, adjust for peak usage times, and predict future failures. Even with the rapid pace at which users contribute to social media websites, Davenport (2014) argues that the largest contributor to the volume of data, at present and in the future, will be due to sensors. Remarkably, as of 2008, Internet-connected sensors far outnumbered the human race (Evans, 2011). Combined, "increasingly networked" (p. 11) sensors and humans, with their Fitbit straps, Nike+ shoes, Apple Watches, and Nest-enabled homes will continue to add to the deluge of data.

The increase in data volume is, in part, due to the rise of participatory culture and the willingness of individuals to create, share, and communicate about web-based content, but it is also due to a new "term of art" (Mayer-Schönberger & Cukier, 2013, p. 113): data exhaust. Data exhaust is the creation of data as part of an individual's interaction with sensors, information technology systems, applications, and Internet websites. Based on human actions and timestamps, new data exhaust is created that leaves a historical track of behaviors. In most cases, the data exhaust is more metadata than actual data, meaning that it has less to do with content (e.g., videos, blogs posts) and more to do with describing the content (e.g., by associating descriptive data, timestamps, and click paths with the content). Users unknowingly create this data, and anecdotal evidence usually points out that they rarely care about it, but data scientists often mine it to extract new insights into human behavior. And by doing so, system developers gain a better sense of how an individual user–or a profile of a certain type of user–navigates within and interacts with digital and, increasingly, physical spaces.

Big Data conversations about data velocity and variety are closely linked to issues of data volume. There is simply more opportunity for individuals to create data in growing formats, given the social aspects of today's Internet and the ubiquity of Internet-connected devices. Furthermore, all of the data and data exhaust from user-system interactions with devices, applications, and sensors inherently increases and complicates data structures. Traditional data, or data that fits nicely within a SQL database in organized rows and columns, still exists, and many computer applications still format their databases in this arrangement. But new technologies, such as Tableau and NoSQL, hold promise for individuals to explore the growing warehouses of unstructured data unfettered by past requirements to delimit the scope of datasets and clean data for relevant purposes. In the past, statisticians, researchers, and data scientists "winnowed" (Mayer-Schönberger & Cukier, 2013, p. 20) sets of data, paring them down to within the established scope of the project while cleaning them of unnecessary data. With Big Data, all data is within scope and relevant until proven otherwise.

In the context of higher education, Big Data practices, such as learning analytics, target the increasing amounts and sources of student data on campus. Consider a day in the life of a typical student as an illustrative way to understand the extent to which an institution captures data about and created by a student. For this scenario, let us follow "Jeremy." Jeremy rises in the morning to the alarm on his iPhone, which he picks up and immediately checks two applications for overnight updates: Facebook and his campus e-mail. After prepping for the day, he returns to his laptop, logs in to the central sign-on system with his campus-provided credentials, and connects to the residence hall's WiFi network in order to get ready for his upcoming courses. He syncs his campus calendar to his laptop and iPhone, making sure that he will not miss important upcoming meetings with his advisor and the financial aid office. Next, he logs into the learning management system to get caught up on some discussion forum posts for his online class, download required readings to review later, and check his grades. As he leaves his dorm room, he touches his ID card against a black box near the door's frame, which reads the RFID chip in his ID and locks the door.

Before heading into the lecture hall for Biology 101, he stops in at the campus market, purchases a doughnut and a coffee, and pays by swiping his ID card. Jeremy rushes to lecture, arriving a little bit late, but he is not worried about being marked absent for class: the sensor in the doorway automatically recognizes his RFID-enabled ID and records his presence in the attendance system (albeit five minutes after class began). After class, he strolls into the library, checks out some physical and electronic books for his upcoming educational theory essay, and does some related searching in library databases.

Next on his agenda is the meeting with his advisor. Together, they review his course schedule for the rest of the semester, discuss some study strategies, and review his academic plan for the next three years; he is on track to graduate on time–at the moment. The "activity stream" in the eAdvising system keeps track of campus activities Jeremy has participated in (which were automatically recorded when he registered online or swiped his ID card at the event), along with communications from his professors to his advisor. They use this information to discuss his involvement on campus and available academic resources that may help him improve his grades.

The meeting with financial aid follows. About a week ago, Jeremy filled out an online survey about his personal and professional interests, and he allowed the system to "grab" his academic information from the university's student information system. In the meeting, Jeremy and a staff member discuss what loans and grants match his needs and interests by reviewing the system's results: a personalized listing of financial aid resources custom matched to his personal profile. The staff member sends a message through the eAdvising system to Jeremy's advisor, noting what scholarship he decided to apply for and its due date.

Jeremy's day ends with a study session at the campus cafe with a few friends from Calculus 210. Using their laptops and tablets connected to a campus WiFi hotspot, they work collaboratively using Google Documents, socialize on Facebook, and discuss the lecture from the last class period. As the night wraps up, Jeremy creates an item in his web-based campus calendar to schedule the next date and time of their study group, sending electronic invitations about the meeting to the rest of the group; all but one of the students accept the meeting invitation.

At each step of the day's progression, Jeremy's movements are captured by information technology systems. From the moment he wakes to the second before he sleeps (and sometimes when sleeps with wearable technology), a system, app, and device records and stores data about Jeremy. Stitched together, all of this data creates a timeline of his digital activities and his physical movements on campus. It also creates a network of communications and connections, linking Jeremy to other individuals with whom he spoke or associated throughout the day.

Much of the data is created purely as a part of digital exhaust. System interactions, card swipes, single sign-on systems, and WiFi hotspots create identifiable data about Jeremy. Consider, for example, the WiFi hotspots Jeremy connects with. The purpose of such systems is *not* to track where Jeremy is located on campus and when, but if the university requires him to register his device's MAC address to confirm his identity, it would be relatively easy to establish his geolocation and map his movements from hotspot to hotspot. To comply with the Higher Education Opportunity Act of 2008, many universities monitor network activities for illegal activity, such as peer-to-peer downloading of copyrighted works, and require MAC and IP address registrations to identify students who are in violation of the Digital Millennium Copyright Act. The technological infrastructure already exists to capture not only geolocation data, but also to monitor network usage by specific students, which is data learning analytics proponents may find particularly useful.

Other data and information students create about themselves is purposeful and intended to meet a particular need. Beginning with the admissions process, students disclose personally significant data and information related to their socioeconomic status, demographic profile, personal and familial academic histories, hobbies and interests, as well as their professional ambitions. Student support offices and academic departments often survey students as they progress through their educational career in order to update and augment this information. As students participate in online courses (and increasingly in face-to-face courses, as well), post assignments in online discussion forums, submit assignments to the LMS, and build ePortfolios about their academic success and professional goals, they furnish even more analyzable content.

When data exhaust and the information students purposely provide to their institutions remains siloed in disconnected databases, its potential insights and its untapped value remains out of reach, which is the *exact* problem Big Data attempts to resolve. More than ever before, campus information systems are becoming interconnected, data warehouses are growing in capacity and capability, and the walls between datasets are coming down. And by breaking down the technical barriers between sets of data, higher education institutions are beginning to go on "fishing expeditions" (Mayer-Schönberger & Cukier, 2013, p. 29), trawling for correlations, trends, and statistically-driven stories borne of data to help understand student learning and the complex nature of higher education.

The technological culture of higher education and the vast infrastructure that supports it provides some of the most sought after conditions needed to capitalize on Big Data. In their students, universities have a captive audience. At residential universities, students live, work, play, and learn using the networked and data-driven infrastructure of the institution. At online universities, completely digital learning environments maximize the extent to which institutions can observe and analyze learning behaviors along with any other stores of information about their students. In contrast, commercial organizations, like the superstore Target, analyze significant amounts of data based on the purchasing behaviors of customers, but they will never be able to cast a data net as wide as higher education's without purchasing expensive third-party datasets from information brokers like Experian or by somehow making their stores a much larger part of their customers' lives.

3.5. Learning Analytics in Action

It is informative to consider the variety of student data available and postulate about potential analytical uses thereof, but examining learning analytics projects in action helps us to understand the ways in which new uses of student data and information are impacting higher education institutions. The following descriptions detail common student data-based analytics and reveal the variety of ends to which institutions put learning analytics.

Early iterations of learning analytics technology relied heavily on data created by students as they participated in online courses in learning management systems (LMSs). Sixty-three percent of online courses use LMSs to serve the online learning needs of students and instructional goals of their teachers (Green, 2013) and nearly 34 percent of students take at least one online course (Allen & Seaman, 2014), demonstrating that LMSs serve as rich systems for data creation and capture.

Many institutions use some form of basic learning analytics technology as part of their LMS platforms. The problem with LMS datasets is that they are relatively descriptive and limited. However, initial research into learning analytics found that when the data is augmented with other data sources, it leads to richer insights into learning behaviors (see Dawson, McWilliam, & Tan, 2008; Macfadyen & Dawson, 2012; Mazza & Dimitrova, 2007). The difficulty is, however, that research is still nascent in this area, and few if any best practices exist extolling sources and types of data to effectively augment LMS data for learning analytics. In response, some have advocated for a "smorgasbord" (Diaz & Brown, 2012, p. 13) approach to blend various types of data, thereby distilling the most insightful combinations.

Purdue University is, perhaps, one of the most widely known institutions to employ learning analytics and blend data sources, due in part to their early development of the technology in 2007. Their homegrown solution–Course Signals–measures a student's performance (i.e., the grades she has earned in the course) and effort (i.e., her interaction in the course as compared with her peers), and using her past academic history and other personal information (e.g., residency status, age, and credits attempted), the system predicts if she is at-risk of being unsuccessful in the course; the student and her instructor are made aware of the student's risk level by way of a red, yellow, green visualization within the system (Arnold & Pistilli, 2012). One of the motivating goals for Course Signals was to arm instructors with actionable information to inform interventions early on in a student's experience with a course; should she become at-risk, instructors could communicate the academic issues and provide resources expeditiously (Arnold, 2010; Arnold & Pistilli, 2012).

Other institutions have aggregated various sources of data for their learning analytics projects, as well. The University of Phoenix, a for-profit institution, has tapped into the data produced by its online student body–the largest in North America–in order to to identify students in danger of failing an individual course (Barber & Sharkey, 2012; Brown, 2011). Similarly, Rio Salado College, an online community college, created RioPACE to identify students at-risk of earning "C" grades or lower (Grush, 2011; Smith, 2012). The University of Alabama, a public institution, aimed to improve freshmen to sophomore year retention rates using data and analytics to assess which students would need assistance as they progressed from year-to-year (Campbell, DeBlois, & Oblinger, 2007).

Some institutions have implemented eAdvising, a form of learning analytics, to counsel students about their course selection and keep them on track to their degree. Austin Peay State University developed their own analytics system–Degree Compass–a course recommender that uses predictive analytics based on grades and enrollment data to "measure how well each course might help the student progress through their program" (Austin Peay State University, n.d., para. 2) and to help students make time-sensitive choices using "decision support system[s]" (Kularbphettong & Tongsiri, 2014, p. 21). The system takes into consideration what students select as a major, their achievement in past courses, and compares them with their peers' success in courses (Young, 2011); the algorithm for Degree Compass then prioritizes course recommendations based on courses necessary for the student to graduate, courses core to the university's curriculum, and courses in which the student is expected to be academically successful (Denley, 2013).

Arizona State University uses eAdvising analytics in a similar vein as Austin Peay State University, but at Arizona the penalty for making less-than-stellar academic progress or not enrolling in a system-recommended course is higher. "If they fail to sign up for a key course or do well enough," writes Parry (2012, para. 16), "the computer cracks a whip, marking them 'off-track.'" As a result of "wandering" off their academic path, students may be required to change majors, or at least meet with a real-life advisor. In addition to their eAdvising analytics program, Arizona State University has formed strong alliances with the adaptive learning software company Knewton and the educational content giant Pearson to create personalized online learning experiences driven by extremely large amounts of data created by student users (not just Arizona's students) of both companies' technologies (Kolowich, 2013).

Campus Labs' Beacon product, another eAdvising platform, aggregates data from specially developed enrollment surveys, entrance exams, and student activity data gathered from ID card swipes used at campus events. Using triggers setup by advisors, the system sets in motion automatic interventions. If, for example, a student reported on a survey question that they felt socially disengaged, campus housing and that student's resident advisor would be informed (Campus Labs, 2014a). Like other eAdvising systems, Beacon has predictive capabilities. Instructors, advisers, and other institutional actors can see a student's predicted retention rate and academic success probability (Campus Labs, 2014b).

Primarily in the United Kingdom, work has been done to integrate library data about students into types of learning analytics programs. The Library Impact Data Project (LIDP), for example, found a statistically significant positive correlation between individual use of the library and higher levels of degree attainment (Stone & Ramsden, 2013). The study's quantitative outcomes revealed that the more an individual student borrowed books and accessed electronic resources, the greater the student scored on five different levels of degree attainment. The LIDP and others like it continue to aggregate student data for evaluation and reporting purposes. Obviously, this work is currently limited, as–unlike other learning analytics projects–it does not report the data back to the specific student about whom the data was gathered. But the LIDP project shows great promise. It has future goals to begin to further drill down in the data to examine correlations at the school and course levels and collaborate directly with instructors to "direct student support and education" (Stone & Ramsden, 2013, p. 555). Graham Stone of LIDP is beginning work with the Library Analytics and Metrics Project (LAMP) (2014) to develop an analytics platform specifically for academic library needs. The goals of LAMP closely parallel the overarching goals of many learning analytics projects by focusing on student success and improving library efficiency while decreasing costs (Showers, 2013).

In the United States, the University of Minnesota (n.d.) has made a notable push towards participating in learning analytics discussions and projects with their Library Data and Student Success (LDSS) project. The LDSS, much like the work in the United Kingdom, uses student data from the demographic (e.g., gender, ethnicity, etc.), academic (e.g., semester and cumulative grade point average (GPA)), and library (e.g., circulation and library usage statistics) domains to correlate GPA and student retention metrics (Nackerud, Fransen, Peterson, Mastel, Soria, & Peterson, 2012).

Work in the United States tying personally identifiable student data from the library to learning analytics applications and practices, however, has its share of hurdles. Professional ethics commitments developed by the American Library Association (2008) guide library professionals away from using identifiable data about library and collection usage to protect user privacy. Salo (2014) notes that even if library professionals wanted to use personally identifiable data to improve services, there exist no "best-practice documents, charts and checklists, [or] sample policies" (para. 6) available to guide professionals through the serious ethical questions.

As previously demonstrated in the day of the life of Jeremy, data gathered from RFID chips and magnetic strips embedded in student IDs are potentially valuable for learning analytics. At Northern Arizona University (NAU), students use their IDs to gain entry to buildings and their dorm rooms, to purchase meals from campus unions and restaurants, and they often use their ID as a debit card at campus stores, such as the bookstore (O'Connor, 2010). Using federal stimulus funds, NAU sought to advance its use of the RFID-enabled IDs by streamlining student attendance tracking by automatically reading RFID chips as students enter classrooms and lecture halls (Brazy, 2010). Using a similar setup at the University of Mississippi, its attendance tracking system automatically notifies a freshmen student's resident advisor if she has been absent 3 or more times from class; after the fourth absence, the student is marked "at risk" and the instructor and the student's advisor are also alerted (Gates, 2014).

Data associated with student IDs is not only a powerful indicator of student movements on campus, it also holds the potential to reveal more personal (and unseen) details about a student's life. Matthew Pittinsky (formerly of the LMS giant Blackboard) argues that geolocation information gathered from student IDs can "model, at a high level, the social network of the college" (Parry, 2012, para. 57), which could enable the institution to identify and intervene with students who are not socially integrated into the campus community. Advocates of student ID-based data captures also point out that card swipes can reveal dietary patterns and caloric intake based on food purchases (Ash, 2010). When this data is correlated with other datasets, like a student's time at campus recreation centers, it may reveal insights into the student's quality of health.[3] Researchers hypothesize that when RFID and card swipe data is associated with learning outcomes and processes, social behaviors, and environments, the combination may help educators better understand correlations between learning contexts and academic development (Adorni, Coccoli, & Torre, 2012).

Data created or associated with students outside of campus domains on the social web is also of interest to learning analytics advocates. Phil Long and George Siemens (2011) forecast that learning analytics will include information from social media profiles as an added source of data to compare with the wealth of information in the LMS, due in part to the significant "digital footprint" (p. 32) students leave in these types of environments. Samford University uses an

[3] Researchers also argue that mining data that reveals student behaviors and health information may be able to predict which students are vulnerable to suicide (Mandge, 2013).

application to cull data from their "Class of 2017" Facebook group by analyzing "social and behavioral data patterns" (Hoover, 2012, para. 2) to examine and predict who may enroll in the institution. Predictions enable its admissions department to strengthen connections with students "on the fence" and to maintain relationships with students who are likely to matriculate. Furthermore, Facebook-based data analytics empower institutions to target specific types of student demographics by analyzing the profiles of interested students in order to carefully craft a a diverse incoming student body. Samford University's analysis of social media data is not used to directly improve student learning, but it does inform the institution's understanding of the evolution of its student population. Social media data gleaned from *currently* enrolled students may communicate to instructors and advisers other metrics related to campus engagement and student needs.

3.6. The Narrow and Wide Focus

Colleges and universities pursue learning analytics as a means to various ends depending on an institution's needs. But in general, Duval and Koskinen (2014) argue that the focus of learning analytics projects are either "narrow" or "wide" in focus (p. 2). A narrow focus, they write, emphasizes institutional goals, such as advancing admissions practices and raising retention percentages. In contrast, a wide focus highlights student learning goals, like personalizing learning and helping students self-regulate their academic progress. Below, I address the distinguishing elements of the narrow and the wide focuses.

Higher education institutions are defined in part by their student bodies, and cultivating a student body that represents the quality of an institution and its values begins with the admissions process. Courting and choosing specific students for admission is not simple. It is often driven by a university's ever-evolving needs to populate specific academic programs and admit diverse students, not to mention compete with other universities for the best and brightest of the incoming class. While anecdotal evidence and professional hunches once informed admissions decisions, those who consider students for admission have ultimately "concluded that the more institutions know about… [their] students' interests, background, abilities, prior performance, and aspirations" (Goff & Shaffer, 2014, p. 94), the better it can inform admissions processes, who they admit, and how they lead students to enroll.

In order to optimize enrollment, admissions offices have turned to analytics. Statistically-driven analyses and in-depth market research are what help institutions identify, pursue, and enroll students. First, admissions departments segment a student into atomistic bits of data–data borne from her admissions application, high school transcripts, and entrance exams (from which the ACT has extracted 265-plus data fields about her). Next, they compare this personal data with enrollment patterns and performance trends of their own institution and that of their peers. These types of analytic practices enable a university to model an incoming class and to predict the likelihood that a student will enroll, which creates rich, actionable information for the university in a number of areas, including space planning and course scheduling for the upcoming year.

Amazon, Netflix, and dating sites, such as eHarmony, also serve as unexpected models for admissions departments. In these companies, admissions staff see the potential benefit of using data to match services, resources, and information to student needs in timely, efficient, and customized ways. So, understanding who is applying for admission is only one half of doing data-driven enrollment, while the other half is utilizing all the available data in order to "know 'everything' about a student's needs and abilities" (Goff & Shaffer, 2014, p. 108, 110) and respond in "hyper-personalized" ways.

In action, personalized enrollment practices make it possible for an institution to use unique information about individual students for various ends. For example, data could reveal a student's learning style or particular learning needs–before she even steps foot on campus. With this data at hand, an admissions staff member could match specific resources to the student's needs, and in doing so demonstrate that the institution can provide an optimal learning environment.

The personal interest profiles students complete as a part of the application for admission may serve as another source of rich information. These profiles detail a student's particular hobbies, interests, and academic goals. Admissions staff members could use the information within the student's profile to match her with leaders from specific clubs, in order to integrate the student into the life of the university more quickly.

Data derived from personal interest profiles also drive predictions about a student's potential for academic success. If a student indicates that she would like to major in pre-law, data-driven enrollment systems can analyze data about her and other students *like* her who are pursuing pre-law. Comparing profiles (and the data they contain) about her and students similar to her enables an institution to predict the likelihood of her success on this path.

Enrolling quality students is valuable in and of itself, but one of the primary goals of admissions offices is to admit students they can retain, and who will remain committed to the institution through to graduation. Data-driven admissions and enrollment practices have the potential to better match a student to an institution (and vice versa), and in doing so, positively impact student retention. Research indicates that when the culture of the institution (e.g., its "goals, values, and attitudes" (Goff & Shaffer, 2014, p. 99)) closely align with student interests, retention and graduation rates improve (Hermanowicz, 2003).

Enrollment is only one of many variables that influence retention. The University of Kentucky aimed to better understand its current students and engage them with campus life more fully in order to improve overall retention numbers. The campus invested heavily in data analytics by building up a team of 15 data scientists and institutional researchers, who analyze "tiny bits of information" (Straumsheim, 2013, para. 8) derived from student interactions with the university's mobile application, ID card swipes at events, and when students complete short, one question surveys. Along with other initiatives, the data analytics project reportedly increased the university's freshman-to-sophomore retention rate by 1.3 percent. As previously discussed, Purdue University's learning management system-based analytics system matches at-risk students with just-in-time resources. Research done on behalf of the university found that the technology measurably improved student retention, graduation rates, and grades (Arnold & Pistilli, 2012; Tally, 2013), although some have called into question the statistical validity of these findings (see Essa, 2013; Feldstein, 2013; Straumsheim, 2013).

The wide focus of learning analytics represents a shift in thought. Whereas the narrow focus is primarily centered on the idea that analytics can improve *specific* institutional practices and outcomes, the wide focus broadens the approach of learning analytics to consider a much larger goal: learning. There are three aims driving the wide focus: personalize learning, provide tailored feedback to help student self-regulate their progress, and predict specific learning outcomes (Duval & Koskinen, 2014; Mayer-Schönberger & Cukier, 2014).

Personalized learning does not exist solely in digital environments, but the practice is made simpler for instructors to enact and is more closely customized to a specific student when it is data driven. Defined, personalized learning "refers to instruction that is paced to learning needs, tailored to learning preferences, and tailored to the specific interests of different learners" (U.S. Department of Education, n.d.). In order for educational systems to provide tailored learning experiences, students must provide information about themselves and interact with the system on an intimate level to create a wide array of analyzable data points. Once the system is able to build a learner profile, which includes a better understanding of the student's learning style and preferences, it models what the student knows (or does not) and then creates a customized learning experience around a topic or entire course curriculum to fill in the gaps (Bienkowski, Feng, & Means, 2012).

IBM's "Smarter Education Group" is working closely with researchers, education experts, and higher education institutions to develop its PETALS application: Personalized Education Through Analytics on Learning Systems, PETALS "continuously learn[s]" (IBM Research, n.d., para. 7) from how students interact with the system, their success in discrete learning modules, and their responses to interventions in order to craft "personalized and adaptive" learning pathways. In addition to the data students create as they interact with PETALS, its algorithms also pull on historical, anonymized data from over 200,000 students in order to build statistically stronger predictive capabilities based on student profiles (IBM, 2013).

Closely aligned with personalized learning is a secondary feature: tailored feedback. When systems closely track student success and failures, they can intervene in much more specific ways and provide students with detailed reports to help them regulate their learning behaviors and recognize their strengths and weaknesses. These interventions are often initiated by instructors, who know more about the student and her struggles by way of detailed reports, but they can also be created automatically by the system itself. Furthermore, learning analytics technologies that include these features often provide data “dashboards” that visually show checklists, trend lines, graphs, and other statistical measures about a student’s progress, which instructors and students alike can see.

Consider an example of tailored feedback at the University of Washington Tacoma. The institution adopted Persistence Plus, a mobile application that uses information from courses, student performance indicators, and student data, to nudge students to action by monitoring their academic performance and self-reported academic behaviors (Frankfort, Salim, & Carmean, 2012). A student who reported she felt math anxiety, for example, was directed to stress management resources specifically related to math phobia, and just in time–the student received the nudge to action before her next quiz. With regard to student data dashboards, the University of Maryland Baltimore County implemented such a service so that students could have ready access about their activity in online courses. “Check My Activity,” the name of the dashboard, shows students a sum of “any hit, click, or access of any tool or content” (Fritz, 2013, p. 2) within the institution’s learning management system; students are then able to compare their activity with an aggregate sum of activity from fellow students with whom they share a course. Students who viewed the dashboard were nearly two times more likely to earn a C grade or higher than their peers who did not check their activity.

For instructors, one of the many difficult aspects of teaching is understanding what, exactly, influences a student’s level of success. Pedagogical, personal, environmental, and contextual factors all impact the degree to which students master content and are able to apply what they learn. Using statistical measurements like logistical regression and Bayesian probability, learning analytics technologies are able to target specific variables that measurably account for a student’s predicted level of success, which in effect takes some of the guesswork out of *why* students may do poorly in a class and, given their current activities in a course, *if* they will earn a satisfactory grade (Chatti, Dyckhoff, Schroeder, & Thüs, 2012). In many respects, predictive analytics drives personalized learning and the ability of an instructor to intervene in a timely manner when a student is academically at risk.

The University of Phoenix has been at the forefront of learning analytics by developing predictive capabilities into their bespoke online learning environments. Predicting student learning outcomes, wrote Barber and Sharkey (2012), is especially important for the institution, since students take courses on a compact schedule; a single course may only last for five weeks. A typical higher education course that is scheduled over a 12-to-14-week semester enables an instructor to make interventions in a student's progress at a number of different intervals, which is a freedom some University of Phoenix instructors do not have. Shorter courses naturally reduce the amount of time available to instructors to judge student progress and intervene when students are at risk of academic failure. Therefore, predictions about student success are valuable tools, especially if instructors at the University of Phoenix or other institutions seek to make timely interventions or to adjust their own teaching based on predictions of poor student performance.

3.7. Learning Analytics and the Culture of Assessment

American higher education has come under intense pressure by stakeholders to perform at a high level and prove how institutions achieve their respective levels of success by using student data as a bellwether. "Success" is defined differently at each college or university, but stakeholders often use retention and graduation rates to determine the success of an institution. Even though the value of higher education remains high (The Institute for Higher Education

Policy, 1998) and enrollment rates are "steadily increasing" (Delen, 2011, p. 20), stakeholders point to the fact that academic performance is low and undergraduate dropout rates remain high (Caison, 2007; Tinto, 1997). As of 2012, nearly 40 percent of full-time undergraduates failed to graduate (U.S. Department of Education, 2014). Arguing that "better decisions require better information" (National Center for Education Statistics, n.d.), stakeholders increasingly call for more data and better metrics in order to guide students to commencement and to measure the overall cost of an education (Alexander, 2000; Burke, 1998). To appraise an institution and respond to a push for improvements in the academy by powerful stakeholders, a growing culture of assessment has risen within and spread throughout institutions.

As a "primary vehicle" (Kuh & Ikenberry, 2009, p. 26) for demonstrating institutional success, accreditation processes at departmental and institutional levels often require comprehensive reports built on a store of quantitative and qualitative data (Weiner, 2009). Not only is the institution under pressure to compile these reports, but so are the accrediting bodies who review the reports, who as of late have had to respond to their own share of questions about "the basis on which [they make] judgements of academic quality" (Kuh & Ikenberry, 2009, p. 26). Picciano (2012) notes that pressures in regard to student retention and graduation, program-level success, and questions of institutional efficiencies and effectiveness have been "significant" (p. 15) issues in higher education for decades; however, learning analytics may, in effect, change "the very nature" of assessment and potentially bridge the divide between the institution and the policy-maker by providing greater access to raw and analyzed data (Booth, 2012, p. 53).

Some scholars have raised concerns about the connection between assessment and learning analytics technologies, positing that some technological applications could emphasize specific (and highly political) “assessment regimes” (p. 75), like high-stakes testing, over other regimes that directly inform teaching and learning practices (Knight, Buckingham Shum, & Littleton, 2013). It is plausible that pressures from accrediting bodies and outside stakeholders may in fact dictate what data is aggregated by institutional actors to be analyzed by learning analytics technologies. Whether the data is the “right data” and the analysis is the “right type of analysis” is a concern that institutions will need to address and engage in dialogue about when facing outside pressure to use learning analytics.

Another related concern is that pressures to use and report student data from the president down may, to some degree, force higher education to engage in data initiatives and practices that it would prefer not to, like federal and state unit record systems. If such systems were to be built on comprehensive student data profiles, and if those systems determined an institution’s federal and state funding, a college or university would have little choice but to participate. The decision to not participate in data-driven assessment is gradually falling away as more data becomes available and as policy-makers, who are becoming more frustrated with institutional data opacity and ever-increasing institutional budgets, act on their doubts about the value of higher education to society (Ash, 2010; Hebel, 2008; McKeown-Moak, 2013).

Chapter 4. Big Data's Privacy Problems

4.1. Introduction

Big Data presents a number of real and potential benefits. Data-driven services may bring about financial gains, personal comfort and health, efficiency, and–with learning analytics–improved learning experiences. Consider a few beneficial services borne from Big Data initiatives. Progressive auto insurance, for example, analyzes real-time driving behaviors using in-car sensors combined with traditional risk assessment strategies to predict a driver's potential for an accident and to provide tailored, less costly insurance (Stross, 2012). And "big-box" stores, such as Target, log buying behaviors and analyze personal profiles to predict products its consumers may need at just the right time, such as when a woman is pregnant (Duhigg, 2012a; Duhigg, 2012b).

But as quick as we identify the benefits brought about by Big Data, so, too, do we see the ways "large datasets and the use of analytics… implicate privacy concerns" (Tene & Polonetsky, 2012, p. 65). Data-driven practices rely on personal information–in identifiable and deidentified states–and those who become aware of Big Data's influence in their lives are often bewildered by its presence. About this moment of awareness, danah boyd (as cited in Hardy, 2012) remarked that it elicits "a general anxiety that you can't pinpoint, this odd moment of creepiness" (para. 3). It is as if Big Data stimulates our hackles, and alerts us that something is not as we thought it seemed. In the 1970s, society felt similar anxiousness over the rise of massive databanks. Then, the worry was primarily about the government's ability to create secret records about specific segments of the American population. Now, Big Data induces anxiety for similar reasons, but on a much larger scale. The worry has spread and it now concerns any number of organizations and institutions, government or otherwise, who can capture, store, and analyze the private lives of individuals who connect to the Internet and engage with data-driven services.

This chapter's sections address privacy problems related to Big Data first by considering informed consent practices and challenges thereof. Next, I take up surveillance as a dominant concern in the literature, especially given new technical means of aggregating data and information about individuals. I follow this with an analysis of emerging questions regarding transparency and opacity with regard to Big Data practices, and how those who pursue Big Data often gain power over data subjects. The chapter finishes with a conversation about how Big Data attempts to digitally mirror human identity and activity and the problems of doing so; individuals may wish to have their data anonymized to protect against this harm, but as I discuss in the final section, deidentifying data has become especially problematic due to Big Data.

4.2. Informed Consent

As a guiding principle, informed consent continues to frame privacy policy in the United States, but Big Data presents particular challenges to upholding this longstanding standard. Informed consent's importance was introduced in the Code of Fair Information Practices (CoFIP), which was developed in *Records, Computers, and the Rights of Citizens* (U.S. Department of Health, Education, and Welfare, 1973). It greatly influenced the development of the Fair Information Practice Principles (FIPPs), another longstanding and influential privacy guide (OECD, 1980; OECD, 2013; Privacy Protection Study Commission, 1977).

Informed consent generally serves two purposes. First, organizations and institutions who gather and use personally identifiable information must make individuals aware that they are doing so; furthermore, they must also inform individuals about how they will use the information, the purposes to which they will put it, and any rights individuals have. Such rights are commonly expressed as, but not limited to, the right to review and amend the records that contain information about them. Second, individuals should, as an explicit acknowledgement of their understanding of the information practice under consideration and agreement to participate in that practice, consent before an organization or institution can make use of an individual's personal information.

The feasibility of informed consent in a Big Data environment is under question for a number of reasons. First, Tene and Polonetsky (2013) argue that informed consent is burdensome, both for individuals and the organizations and institutions with which they enter into a relationship. They write:

> On the one hand, organizations are expected to explain their data processing activities on increasingly small screens and obtain consent from often-uninterested individuals; on the other hand, individuals are expected to read and understand complicated privacy disclosures and express their "informed consent." (p. 261)

There are two interconnected issues at play. The first of which is a concern regarding the readability of terms of service agreements written in legalese and the ability of a user to interface with agreements that have not been optimized for particular screen sizes or interactive environments. If the terms of service agreement (ToSA) is readable, then the second concern is in regard to whether or not the agreement is understandable and if the individual takes the necessary time to read it in full in order to make an informed decision. As mobile devices like the iPhone become the standard device of choice, the design of ToSAs and their placement within specific environments, like an app store, will continue to rise in importance. Furthermore, ToSAs are so numerous that it is nearly impossible to read every single one with which we come in contact, and we often choose to skim or bypass reading them entirely (McDonald & Cranor, 2008). If we read every ToSA presented to us, we would spend well over 80 hours a year *skimming* agreements and 181 hours *reading* them from top to bottom, and this accounts only for ToSAs for websites (McDonald & Cranor, 2008, p. 563). Given the time commitment, it is economically infeasible to read and consent to all of the agreements we encounter, much less to understand the privacy rights and expectations outlined within each.

Second, Big Data increasingly complicates the flow of data and muddles our understanding of the exact purposes to which our data will be put (Tene & Polonetsky, 2013). Informed consent "takes place against an increasingly complex backdrop…[of] intricate arrangements involving dense networks of platforms and applications, including contractors, subcontractors, and service providers operating globally" (Tene and Polonetsky, 2013, p. 261). Given that the *de facto* standard of Big Data practices is to aggregate and analyze data from a variety of sources, it is nearly impossible to understand how our personal information flows into, outside of, and within the organization or institution with whom we are entering into a relationship for data-driven services. And ToSAs rarely, if ever, detail in complete description how our data and information will be used, which could by design obfuscate information practices, or it could be that the service provider simply does not know how that information will be put to use.

In years gone by when information practices were much clearer and purposeful, informed consent was an adequate method for notifying individuals about the role of their personal information in data-driven services and their rights thereof. But Big Data and our digitally-mediated lifestyles have made the process difficult and nearly unmanageable. While informed consent is still a laudable principle on which organizations and institutions should build policy and design technology to protect one's information privacy, it has been weakened in an era of Big Data.

4.3. Surveillance

When organizations and institutions use data representative of the details of our lives to control, influence, regulate, and govern our behaviors, they engage in unsettling surveillance–or "dataveillance" (Clark, 1987; van Dijk, 2014)–practices we often find creepy (Murakami Wood, Ball, Lyon, Norris, & Raab, 2006). In combination with profiles full of intricate personal details, sensors, geo-location tracking, and metadata analysis can create an accurate representation of our physical whereabouts at any given time and render our "everyday lives increasingly transparent to large organizations" (Lyon, 2014, p. 4), even though surveillance practices are opaque to those they surveil (Richards & King, 2013). New, data-driven surveillance is "unprecedented in human history" (Richards, 2013, p. 1936), and it has led many to rethink how to define surveillance.

Laymen and academics alike are quick to employ Orwell's Big Brother as a defining characteristic of Big Data surveillance. Orwellian surveillance employs a top-down perspective, whereby government power and "targeted, purposeful spying" (Andrejevic & Gates, 2014, p. 185) are exerted through technology (e.g., "telescreens" in Orwell's story) to force individuals to behave in state-sanctioned ways. The worry is that totalitarian governments will use Big Data technologies and practices to monitor and mold citizen behaviors; historically, surveillance has been used to do just that, especially in tyrannical states, but "surveillance is not just for communists and dictators" (Richards, 2013, p. 1938). As the Edward Snowden revelations showed, even democratic countries like the United States and the United Kingdom engage in data-driven surveillance practices to monitor, track, and intervene in the lives of citizens and terrorists alike.

If Big Data technologies and practices were only available to government actors, an Orwellian approach to surveillance would be adequate, but it is not. Others argue that the top-down approach to data-driven surveillance is moot, given that the growing awareness of Big Data may have a panoptic, self-regulating effect (Hier, 2003). Instead, the “liquidity” of post-modern society and the increasingly free flow of information redefine the watcher-and-watched assemblage within a complex socio-technical environment (Bauman & Lyon, 2013). In a digitally-mediated society, liquid surveillance theory argues that this transition from state-sponsored to privately-controlled surveillance practices significantly broadens the scope of observation and increases the opportunity for individuals, organizations, and institutions to methodically manipulate the lives of many.

Some organizations and institutions using Big Data practices may adamantly defend that they are not practicing surveillance in order to sidestep the inherent negative connotations of the term, but they are surveilling. David Lyon (2007) defines surveillance as the “the focused, systematic and routine attention to personal details for purposes of influence, management, protection or direction” (p. 14). By this definition, those who pursue Big Data, especially to learn from and direct an individual’s behavior, are participating in surveillance, although rarely in the tradition of “totalitarian domination” (Richards, 2013, p. 1937). Instead, Big Data surveillance materializes as, for example, convenient behavioral advertising. A primary goal of Big Data surveillance is rarely to target a *specific* individual, but rather to influence the “categorical profile of the *collective* [emphasis added]” (Hier, 2003, p. 402). By that measure, the effect of Big Data surveillance systems is especially concerning since the aggregate effect is plausibly larger when contrasted with legacy surveillance practices that target individual people.

Big Data surveillance is clearly troublesome to many, and while government action in data-driven surveillance is a bona fide concern, there are other significant issues at stake as well. First, Big Data systems use data and information gained from observing individuals to intervene in their lives and, in some cases, make automated decisions for them. In doing so, they deny humans a degree of personal liberty (Lyon, 2014). For example, credit scoring systems often determine an individual's credit worthiness based on her financial assets, debts, and history. At the same time, companies like AvantCredit and ZestFinance argue that "all data is credit data" (Leber, 2013, para. 6), and have begun analyzing social media usage as part of their scoring processes. The fact that credit companies do this sort of social activity-based analysis would come as a frightening realization for many. And since credit scores influence important financial aspects of an individual's life, especially when applying for a mortgage or car loan, people may begin to rein in their behaviors or filter what they publish in the public domain in order to game the credit scoring system to their advantage.

A surveillance system built on top of troves of personally identifiable data "menaces our society's foundational commitments to intellectual diversity and eccentric individuality" (Richards, 2013, p. 1948). In so doing, data-driven surveillance risks creating a wide-sweeping "chilling effect" (Solve, 2006; Stanley, 2012). A chill on personal behaviors may pressure individuals to limit their freedom to act of their own accord, pursue ideas, speak openly, and associate freely with others. Big Data surveillance conflicts with the legal rights and social values of democratic societies, such as the United States, that uphold the pursuit of intellectual ideas and the freedom to express those ideas, no matter their aberrant character.

Second, the digital dossiers that data-driven surveillance systems create and on which Big Data systems rely for analytic predictions risk perpetuating discriminatory and unfair practices. The most statistically powerful, and equally invasive, Big Data systems aggregate as much available data about an individual as they can possibly gather for their analytical purposes. Michael Schrage (2014) of the *Harvard Business Review* details that companies have the ability to filter through massive datasets to a granular level to target the most profitable type of customer based on combinations of their postal code, gender, sexual orientation, ethnicity, and even their expressed degree of happiness in their marriage. Knowing such factors can help companies direct their marketing efforts in order to maximize profit gains.

But Big Data practices that privilege one group of individuals with discounts and preferential treatment over another based on the color of their skin, their ethnic heritage, or their marital status reeks of discrimination and inequity. It may be that surveillance-based profiling and data-driven interventions are "good business" (Richards, 2013, p. 1937) and "value-added personalization and segmentation" (Schrage, 2014, para. 2) might lead to financial gains, but these practices present plausible harms to pluralistic societies in which fair treatment and equal opportunity trump the vestiges of discriminatory systems.

4.4. Transparency, Opacity, and Power

One of the driving goals of Big Data is to render the world transparent by transforming all things possible into ones and zeroes. Yet, the irony is that with transparency comes significant opacity, and only those driving Big Data initiatives have the power to make plain the data-driven processes influencing our lives. Transparency, opacity, and power as they relate to Big Data carry with them their own unique issues, but they are ultimately intertwined.

The gains Big Data will purportedly bring to society require us to participate by disclosing information about ourselves in ways that often go far beyond what we might expect, thereby encroaching further into our personal spheres of privacy. But by removing obfuscations and barriers to our most personal of information, Big Data has more digital fodder to analyze. Often, we have little choice but to play the Big Data game as data-driven systems become more the norm than the exception. And each time we log on, sign up, and interact with such a system, Tene and Polonetsky (2013) astutely point out that we play "a game of poker where one of the players has his hand open and the other keeps his cards close" (p. 255). Continuing with the card analogy, it is also problematic that those holding cards constantly change the game without our knowledge.

Organizations and institutions toil at their Big Data systems and tweak their information practices in order to maximize the data and information they analyze, but the fact that this is happening is "increasingly invisible to those whose data is garnered and used" (Lyon, 2014, p. 4). This is the "Transparency Paradox," as Richards and King (2013) termed it, and it highlights the discord between individuals and those who seek their information. If informed consent worked as it is designed to, and if the CoFIP was used in the spirit in which it was created, those who pursue Big Data practices that ingest personally identifiable information would *clearly* detail how individual information is collected, used, disseminated, and maintained. This, however, is not the case. Companies continue to lock away "the secret sauce" (Tene & Polonetsky, 2013, p. 243) of Big Data algorithms, hide the criteria they use to make statistical predictions about our lives, and nudge us to behave in ways that benefit their ends. In effect, those of us concerned about how Big Data influences our lives are left in the dark. Should we seek some understanding, some light, we must delve far into a Kafkaesque warren of intersecting legal, social, and technical tunnels (Kafka, 1968; Solove, 2004; Tene & Polonetsky, 2013).

By understanding the transparency and opacity issues of Big Data, the power imbalance becomes self-evident. If knowledge is power, then organizations and institutions pursuing a Big Data agenda are gaining a significant amount of control over the lives that they analyze and intervene in. The less individuals know about "the multiplicity of agents and algorithms" (boyd & Crawford, 2012, p. 673) collecting and analyzing information about them, and the more they are confused about the processes by which they can regain control over their private information, the weaker their standing in relation to Big Data practices (De Filippi, 2014). Moreover, insights gained via Big Data into the lives of individuals presents an illusion that the technology knows the human better than the human knows itself. Like a doctor telling a patient her diagnosis, Big Data presents itself as an expert about the construction and future paths of human lives. Such expertise is hard to deny given the claims Big Data proponents make to advancing "truth, objectivity, and accuracy" (boyd & Crawford, 2012, p. 663). As a result, individuals may naïvely submit themselves to Big Data's influence and power.

Regardless of the aura of Big Data as a quantitative, technologically-driven phenomenon, humans–especially "those with access to the data and the processing power" (Andrejevic & Gates, 2014, p. 190)–are the ones in positions of power (Manovich, 2011). This reality brings about epistemological and ontological questions. Big Data as a technological practice is not neutral, and those who create and control algorithms embed them with their own perspective, biases, and politics, as technologies often are (Andrejevic & Gates, 2014; Winner, 1980). In effect, those in power reframe "key questions about the constitution of knowledge…and the nature and categorization of reality" (boyd & Crawford, 2012, p. 665) in their choices about the data they include for analysis and their predictions regarding human behavior. Bowker (2005) observes that "data should be cooked with care" (p. 184), so we must be wary of those who bake algorithms with bias, prejudice, and political motivations.

4.5. Data Doubles

Big Data practices often aim to capture and analyze as much of the human experience as possible, including physical movements, mental processes, and emotional states. In so doing, individuals are taken from a corporeal whole, transformed into binary code, and abstracted into what Haggerty and Ericson (2000) term a "'data double' of pure virtuality" (p. 611). The end goal of creating data doubles is the transformation of "the body into pure information, such that it can be rendered more mobile and comparable" (p. 611).

When data doubles are created from Big Data practices, humans are broken down into new, unique flows of information in a two-step process. First, technical means collect identifiable data about an individual as she interacts with sensors, networks, and interfaces. These processes transform her behavior into a coded reflection, a digital "simulacrum" (Poster, 1996) to be stored in databases. Second, the data is analyzed, compared with the data doubles of other humans, and then infused with predictions, as well as augmented with relevant information. The end product is a composite of raw data and new forms of information that represent the individual not only as she is but also as what she may be.

The purpose of creating a data double may be to inform the individual about whom the double was made (as is often the case with the quantified-self movement), but digital representations of humans are typically done with other purposes in mind. Haggerty and Ericson (2000) write that data doubles, "rather than being accurate or inaccurate portrayals of real individuals,… are a form of pragmatics: differentiated according to how useful they are" (p. 614) to the organizations and institutions seeking to govern, control, and influence human behavior. Lianos (2003) argues that control may, in fact, be a side-effect of data-driven business, and not a primary intention, but it becomes "quietly embedded in institutions as they mediate an increasing range of our choices" (Los, 2006, p. 72) through digital means.

One problem, of many, with data analytics is the mythology that they capture everything about an individual, when in reality that is usually far from the case. "There is no guarantee that the data collected is either comprehensive or representative," write Andrejevic and Gates (2014, p. 191). But as Big Data practices become the normative way of examining, analyzing, and directing human behaviors, they rely on incomplete profiles of those they analyze. In essence, Big Data may be able to draw a "data double" of a person, but it rarely has the full picture. We may be the sum total of our data, as Don DeLillo (1985) acutely described in his fictional account *White Noise*, but the sum is never fully complete. And, the data and information that does become part of data doubles is never truly accurate once it is decontextualized from its source (Los, 2006).

The response from Big Data advocates is easy to foresee. They argue that an incomplete dataset is indicative of a need for more data, or that if something cannot be datafied or measurable, then "it doesn't exist" (Bowker, 2013, p. 170). We must be wary of those pursuing Big Data agendas with "corrupt" data doubles, as the consequences of doing so could lead organizations and institutions to make predictions about and intervene in human lives with harmful consequences.

4.6. De/Re-Identification of Personally Identifiable Information

A standard procedure to protect against data breaches has been to deidentify, or anonymize, personally identifiable information within databases. To successfully anonymize data, database administrators and information technology professionals use specialized techniques to accomplish these ends. Administrators first categorize potentially identifying information, both by examining the data under consideration and by comparing it with other datasets which, when linked, could reidentify individuals. This process may be guided by intuition, based on industry best practices, or done in accordance with organizational policy or law. Once administrators single out problematic information, they modify the database by suppressing (deleting or omitting) or generalizing (altering) the data held within (Samarati & Sweeney, 1998). Others simply aggregate the data in statistical form as a way of providing usable information without the deidentification risks. While statistically aggregating data greatly reduces the potential risk for reidentifying individuals, it inherently limits the utility of the data (Fefferman, O'Neil, & Naumova, 2005).

The problem with deidentification procedures is that database administrators often disclose deidentified datasets into the wild, often without ever following up with those to whom the data was disclosed, if it was used properly, or confirming the fidelity of the anonymized dataset. Paul Ohm (2010) calls this process "release-and-forget anonymization" (p. 1711), and it is a standard practice that is no longer viable in an era of Big Data. In his groundbreaking article on the risks of release-and-forget anonymization, Ohm (2010) lays out a case for why this model for deidentification of personally identifiable information is untenable. Citing AOL's release of search data (see Barbaro & Zeller, 2006), Group Insurance Commission's disclosure of state employee hospital visits in Massachusetts (see Greely, 2007; Sweeney, 2000), and the Netflix prize study (see Narayanan & Shmatikov, 2008), all of which represent instances where individuals in de-identified datasets were reidentified (or potentially could have been), Ohm (2010) argues that these cases “sound the death knell” (p. 1705) for technical deidentification practices and the legal frameworks that support them. What has come to the fore to make reidentification possible is the networking, or linking, of disparate databases and the computational power now available to analyze massive amounts of data. Reidentification is also a more salient issue these days due to the overwhelming amount of personal information and trace data we leave as we interact with a growing web of information systems.

Two linked databases may provide enough data to reidentify a once anonymized individual, but what worries Ohm (2010) is the “accretion problem.” He writes:

> The accretion problem is this: Once an adversary has linked two anonymized databases together, he can add the newly linked data to his collection of outside information and use it to help unlock other anonymized databases…. Because of the accretion problem, every reidentification event, no matter how seemingly benign, brings people closer to harm. (Ohm, 2010, p. 1746)

Accretion is what enabled researchers to link Netflix data with data from the Internet Movie Database and reidentify users of both systems, which led to potential disclosures of users’ sexual orientation based the types of movies they rented and rated (Singel, 2009).

Until recently, most privacy statutes and regulations have protected organizations and institutions from penalty if they anonymized datasets under their control (Ohm, 2010). The regulations assume “that in the absence of [personally identifiable information], there is no privacy harm” (Schwartz & Solove, 2011, p. 1816), so there has been no motivation to seek out more rigorous technical protections against reidentification, because anonymization has proved sufficient under the law. With the stark revelations brought about by accretion, lawmakers and technologists alike are completely rethinking anonymization. At the center of these conversations is a latent question of whether or not protecting personally identifiable information is even possible given the preponderance of data, information flows, and the risk of accretion.

One of the problems of using personally identifiable information as a cornerstone for privacy regulation is that it lacks a clear definition. According to Schwartz and Solove (2011), personally identifiable information has instead been confined to three approaches: the tautological approach, the non-public approach, and the specific-types approach. The tautological approach holds up a standard that personally identifiable information is that which identifies a person; it is distinguished from other approaches by the fact that it remains a non-restrictive standard, but it fails to be useful due to its ambiguity and the threat of accretion. The non-public approach argues that non-public information should be given special protections under certain circumstances and in particular contexts, but this may still leave common personally identifiable information (like one's address) at risk if it is deemed to be public information. Finally, the specific-types approach states that "if information falls into an enumerated category, it becomes *per se* [personally identifiable information] by operation" (Schwartz & Solove, 2011, p. 1831) of particular rules; thus, this approach is much more restrictive than those previously outlined. But definitive restrictions may be too limited in scope or fail to account for situations when linked data of non-restricted information reidentifies an individual. All three approaches have inherent weaknesses.

If through accretion seemingly all deidentified data can be reidentified, as Ohm (2009) argues, and if common approaches to personally identifiable information are problematic, what are the possible solutions to resolve the potential harms to privacy? Tene and Polonetsky (2013), in similar fashion to Ohm (2009) and Schwartz and Solove (2011), put forward the idea that the risks of reidentification must be balanced with the potential benefits organizations and individuals may gain from data-driven services. To this end, they argue that identifiable data should be viewed on a risk continuum: on one end is positioned data that, if reidentified, would cause little to no harm to the individual; on the other end is data that would present significant harm if reidentified. Risk level would be determined by the statistical probability of reidentifying the data. Deidentified data at a higher risk level would receive greater legal and regulatory protections, but would ultimately reduce the utility of the data; data with a lower risk score would maintain its utility and be less restricted. But statistically-derived risk levels alone are not enough. Regulators should also consider "an organization's *intent* and *commitment* to prevent reidentification" (Tene & Polonetsky, 2013, p. 259, emphasis in original) in addition to the risk level of the data under question. Ohm (2009) also argues that emerging legal requirements should focus on "narrow contexts and specific sectors" (p. 1759), instead of developing sweeping legislation that cannot account for the needs of particular categories of organizations and institutions. Taken together, these technical and legal solutions may reduce the reidentification risks inherent to Big Data practices.

Chapter 5. Learning Analytics and Related Problems of Student Privacy

5.1. Introduction

Specific Big Data practices, like learning analytics, share many of the generic problems related to aggregating, storing, and analyzing massive sets of data about identifiable individuals. But due to the ways they are applied and the contexts in which they are situated, all Big Data practices have unique privacy problems. Where learning analytics is concerned, practitioners, pundits, and critiques alike recognize a "growing concern that ethical and privacy considerations" (Johnson, Adams Becker, Estrada, & Freeman, 2015, p. 12) are not advancing at a pace equal to the development of the technology itself, and therefore are not influencing related tools and systems in ways that attempt to resolve the outstanding issues. Of course, this is a common situation for quickly emerging technologies, especially those that show promise to greatly influence social life. But the degree to which higher education institutions are able to track and analyze the minutiae of a student's online and offline activities amplifies these concerns.

In the parts that follow, I outline a number of Big Data problems directly related to learning analytics in higher education. Roughly speaking, these problems are associated with institutional regulations (e.g., policy development), student rights to informational privacy (e.g., informed consent, control over one's data/information), data practices (e.g., acquiring accurate and complete data, data governance), the ethics of learning analytics (e.g., obligations, beneficence, and fairness), student autonomy (e.g., a student's ability to act based on her own wants and interests), and, finally, codes of practice or a lack thereof. In addition to my own reading and understanding of the literature, the areas are informed by Sclater's (2014) substantial synthesis of extant learning analytics research.

5.2. Information Policy

From a policy perspective, the handling of sensitive student data and information has been a sort of "blind spot" (Greller & Drachsler, 2012, p. 50) for colleges and universities. While student information that is explicitly a part of one's academic record is handled by the office of the registrar, subsidiary personally identifiable information is strewn throughout various campus information systems and in departmental records; in effect, the purview of responsibility is spread throughout campus and may come under a variety of policies or none at all. Tracking student behaviors and collecting stores of student-derived data exacerbates this problem to the extent that existing "maps" (King & Richards, 2014, para. 3)–the ethically-grounded institutional policies that guide the caretaking of student information–are no longer helpful given rapid technological change.

Without up-to-date guidelines, learning analytics' development may reach an impasse. Students, parents, and careful observers of higher education are likely to jump to negative conclusions about learning analytics due to privacy concerns, much in the same way they have with other Big Data practices. They may feel that tracking student behaviors is "creepy"; they may feel that intervening in a student's life is invasive. Without policies that carefully direct and justify institutional actions, concerned voices will become louder and harmonious about the dangers to the detriment of learning analytics' progress (Greller & Draschler, 2012; Siemens, 2012). Moreover, failing to address such concerns will do more harm than simply hindering learning analytics; it will negatively affect an institution's reputation and, potentially, warrant the university status as a "bad apple" institution for such practices (Schwartz, 2010).

In response to emerging calls for new institutional policies, some may point to their longstanding institutional review boards and their existing guidelines regarding student privacy. While institutional review boards are actively engaged with emerging concerns related to research participant privacy (students or otherwise), and while the same boards continue to evolve their policies, "learning analytics poses some new boundary conditions" and questions (Pardo and Siemens, 2014, p. 442).

First and foremost, there is a question of whether or not institutional review boards are required to assess the risks of learning analytics projects done in the course of "normal business." Consider, for example, institutional research done using student data. The risk to students is often minimal and, the research is conducted in "commonly accepted educational settings, involving normal educational practices" (Protection of Human Subjects, 2009, §46.101 (b)(1)), which meets one of the six federal criteria for exempt research. Furthermore, institutional research is typically exempt from review in part because there is no intent to publish the research. These exemptions may free learning analytics projects from restrictive policies if their driving purpose is to inform instructional, departmental, and institutional practices.

Second, assuming institutional review boards do take up learning analytics projects for consideration, it may be difficult for them to determine the level of risk to students. One of those risks concerns student identification. Whoever is directing the learning analytics projects, a team of researchers or an arm of the institution, they may not be able to successfully anonymize student data. The institutional review board, along with the researcher, will have to assess whether or not data accretion is likely to occur and determine if there is potential for anonymized datasets to be linked to one another which may possibly reidentify students.

Third, Pardo and Siemens (2014) also point out that learning analytics presents intellectual property (IP) issues. In the process of aggregating stores of student data–especially those derived from learning management systems–institutions may also be subsuming the intellectual work of students embedded in discussion posts, essays, other assignments, and miscellaneous communications (e.g., private messages and e-mails). For example, an institution may use a student's IP to improve its analytic practices or to make assessments of its student body. Arguably, the institution benefits from the IP of the student by using it to improve its processes, which has the potential downstream effect of making the university more competitive in the higher education market.

But, the IP issue extends outside of the institution as well. Often, colleges and universities contract with educational technology providers for specific learning applications and systems. Educational technology companies, they often argue, need the data they aggregate from students to iteratively improve their systems and provide services. Much like the institutional example above, technology providers benefit from data-based student IP by continuously using it as a testbed in which to develop new products and services. It is unclear whether or not institutions are aware that by releasing student data to technology providers, they may enable vendors to keep and analyze student-derived IP in perpetuity.

5.3. A Lack of Transparency

Many argue that learning analytics suffers from transparency issues and that drafting new policies will help to reduce opacity around data analytics and build trust with students and other stakeholders (Polonetsky & Tene, 2014). Sclater (2014) argues that "transparency regarding the purposes to which data is being put, who will have access to it, and how identities are being protected is a responsibility of the institution" (p. 20). Dialogue between a university and its students before, during, and after policies are drafted will help to calm trepidation about potential abuses of data analytics (Richards & King, 2014; Slade & Prinsloo, 2013). Furthermore, transparency's "sunlight" (Brandeis, 1913, para. 1) can illuminate the darkness of learning analytics, especially the secret algorithms and decision-making processes that influence the lives of students (Clow, 2012; MacCarthy, 2014).

There are situations, however, that may run counter to the ends transparency tries to accomplish as a means. Transparency does not guarantee complete student buy-in. Even if institutions are fully transparent about data-driven practices and the benefits and harms that may accrue, such efforts may still not go far enough to assuage fears in those who want guarantees they will not be harmed. Some students may still want to opt out of learning analytics. (I will address *specific* issues related to opting out below.) Laying bare the harms students could experience due to predictive systems runs counter to the norm (Crawford & Schultz, 2014), and it may actually be to the institution's benefit to shroud learning analytics in some mystery in order to reduce some friction points with the student body.

5.4. The Role of Informed Consent

Questions exist as to whether or not students need to consent to participate in learning analytics initiatives and other data capture practices that support them. As is usually the case, students are not informed about the extent to which colleges and universities gather, aggregate, and analyze their private information. Nor do they understand how that analysis figures into their educational experience, especially due to emerging predictive analytics systems. If informed consent was the norm in higher education, students would be made aware of and have to actively consent to data-driven practices that use their information. Taking this further, federally-mandated annual FERPA notifications would remind students of their consent, what they consented to, and if they need to be made aware of any new practices that require further consent. At present, this is not the typical course of action.

Other researchers question whether informed consent is even necessary (see Slade & Prinsloo, 2013). In fact, Land and Bayne's (2004) investigation into student surveillance in learning management systems found that students cared little about having their digital behaviors logged and examined; in fact, they expected it. For these students, and for those that share their sentiment, data-driven student surveillance is legitimate and socially acceptable, and informed consent processes are not necessary.

We can imagine, however, student perceptions of informed consent and institutional surveillance changing as they consider different types of tools for data tracking and the purposes to which that data is put. Today, the learning management system is a common tool in higher education, and students recognize that participating in its environment is required if they enroll in online courses and, increasingly, traditional face-to-face courses. But the growing adoption of RFID-enabled student IDs, biometric scanning devices (e.g., for fingerprints, irises, and palms), and sensors may discomfit students to the extent that they want–or expect–some process of informed consent in order to become aware of and express their preferences with respect to sensitive types of data collection and tracking. Institutional actors who argue that their private lives are "less important than the institution's right to carry out learning analytics without consent" (Sclater, 2014, p. 11) will have to justify their position to students if they seek a process of informed consent to express their privacy preferences.

5.5. Opting Out of Learning Analytics

Providing students opt out opportunities may present institutions with complicated "ethical and logistical problems" (Sclater, 2014, p. 17). Like other data-driven services and associated terms of service agreements, students may not fully understand what it is they are opting out of. For instance, they may not be able to grasp how data-driven practices are embedded within a diverse array of experiences, educational or otherwise, while they are enrolled. If the purported benefits of learning analytics transpire, opting out of such practices may stunt students' learning potential, and the personalized experiences analytic technologies enable would be non-existent.

In contrast to students who opt out, students who opt in may be afforded additional resources and opportunities over those who do not, which highlights an emerging fairness issue. There is a question about whether or not institutions would be obligated to explain the fairness issue to students and clearly spell out what benefits and harms may accrue from both sides, opting in or opting out. It is plausible that some students may recognize this issue and, in effect, feel undue pressure to opt in, regardless of their privacy concerns. Additionally, if some students choose not to participate in data-driven practices, the aggregate dataset on which analytic technologies rely will be less comprehensive and, consequently, the data colleges and universities use for instructional and institutional purposes may be statistically underpowered, incomplete, and potentially misleading.

5.6. Control Over Student Information

Conversations related to information privacy writ large often invoke the idea that an individual's information can be held private if she is in control of it. To the extent that information control plays out in higher education in relationship to learning analytics, some claim that students have a justifiable argument that they should be in control and, in effect, own the data they create (Jones, Thomson, & Arnold, 2014). By putting students in control of their data, they would be empowered to use it (or not) in ways that comport with their values, goals, and expectations of privacy.

There is also an outstanding concern that students, if they controlled a machine-readable files of their data, may use it irresponsibly. Under pressure to payoff mounting student loan debt and due to other unfortunate economic conditions, it is plausible that students may sell their data to data-hungry educational technology ventures to make a quick amount of cash without fully considering the downstream effects on their privacy. If institutions controlled the data and limited access to it, they may safeguard students from themselves and predatory technology companies. Furthermore, student data ownership may lead to situations where third parties, like health insurance companies and future employers, require students to disclose their data as part of applications for services and employment (Reilly, 2013). The behavioral information the data contains may be especially useful for these types of organizations who want to analyze what types of risk a student takes (is she a procrastinator?), if she is a hard worker (did she visit the library more than twice a week?), and whether or not she was a team player (to what extent did she establish working relationships with peers in her online courses?). Again, institutional control of student data would, potentially, mediate these concerns by limiting the extent to which students may act in ways harmful to themselves and their future.

5.7. The Right to be Forgotten

As part of controlling and managing one's data, some have argued that students should be provided the opportunity to purge institutional systems of their data. The advantage of such a right is twofold. For one, enabling students to excise their data profiles, either at a granular level or in whole, would provide them the ability to strip potentially embarrassing, misleading, and inaccurate information that could become "permanent blemishes" (Slade & Prinsloo, 2013, p. 1520) on influential educational records. In addition to intellectual exploration, college provides a time for many students to play with identity formation and establish themselves as adults. The growing surveillant assemblage of interconnected information systems on campus and their ability to capture the personal lives of students also means they may capture potentially humiliating and harmful behaviors and communications, which could haunt them far into the future. A second advantage of a right to delete one's data is that it, in combination with other factors, may help to build trust between students and their institution. By relinquishing control over student data, students may interpret this action as a sign that their university respects individual privacy preferences and, therefore, is more amenable to data analytics practices.

Of course, there is a significant tradeoff institutions need to consider if they provide students a "right to be forgotten" (see Mayer-Schönberger, 2009) by enabling data deletion mechanisms. Should universities build systems that allow students to delete, say, their learning management system tracking data, or if they establish a policy that data of non-active students (e.g., those who graduated or dropped out) will automatically be purged, this removes extremely valuable historical data. Pardo and Siemens (2014) argue that "keeping student data will be helpful for the university to refine its analytics models, track the development of student performance over multiple years and cohorts or simply for internal or external quality assurance processes" (pp. 445–446).

Student deletion of data is just one way that an institution may miss out on valuable information. Over time, individual departments have invested in resources to capture student data, they have established workflows with their systems, and they have built up specific insights and actionable information from that data. If universities adopt campus-wide learning analytics technologies, especially those that supplant established systems, departments may not welcome them with open arms; to this point, research already shows that "many departments [are] reluctant or unwilling to share data necessary for analytics" (Bischel, 2012, p. 16). In effect, institutions may not even be able to provide students a right to be forgotten when departments defensively protect the data they actively curate.

5.8. Data Maximization and Data Quality Issues

For institutions to optimize learning analytics practices, they will have to maximize the data they can analyze. The bigger the dataset, the more potential insight institutions can mine for; the greater the insights, the more competitive they can be in an increasingly competitive market. And given that learning analytics is used to intervene in the lives of students, gaping holes in datasets used to predict learner potential and guide student progress could prove harmful to the student whose life is being molded by this technology.

Many have argued that data quality is one of the foremost challenges to learning analytics (Bischel, 2012; Greller & Draschler, 2012). There are a number of statistical and technical concerns, especially given the nascent character of the emerging analytics technologies and the data infrastructures on which they rely. For example, existing learning analytics applications may improperly predict a learner's at-risk level (e.g., for a poor grade) due to an incomplete dataset; furthermore, systems that capture student behaviors may improperly classify data if they are setup incorrectly or present technical bugs due to interoperability issues. Although data-based systems seem objective, they are human-built technologies and are, therefore, subject to human error.

In addition to technical problems, students may create unique data quality problems. Students aware of their institution's use of learning analytics may try to obfuscate data gathered about themselves and their digitally-captured behaviors by providing false or misleading information. A concerned cohort of students, for example, may have a lead student carry the group's RFID-enabled student IDs around campus in order to spoof RFID readers enabled in campus buildings. It is also plausible that students will try to "game" (Bollier, 2010, p. 6) learning analytics systems by behaving in ways that provide undeserved rewards for specific behaviors. If learning analytics systems keep track of a student's aggregate amount of time in a learning management system-based online course and the instructor uses that metric to grade participation, the student may login and logout after a certain duration of time throughout the week but not necessarily participate in the course simply to get the grade without doing the work.

Even if institutions are able to aggregate student data without quality issues, they will struggle to capture complete and comprehensive datasets. Any data taken out of context will fail to bring with it the "cultural and behavioural [*sic*]" (Sclater, 2014, p. 27) characteristics that frame it. The life of the student, in all of its complexity, may be captured in data snapshots, but those snapshots will always fail to take into account surrounding reasons that help to explain behaviors, especially behaviors that depend on just the right setting and variables (e.g., a student's emotional state after being fired from her job, or a burst of intense focus on her studies after illegally taking

Adderall). This is extremely problematic given the penchant of learning analytics supporters to use systems that use a minimal set of variables to predict learning outcomes; not only does it reduce learning to the most easily accessible metrics, but it often fails to account for the role of affect in the learning process.

Without complete datasets, institutional actors are left to make questionable interpretations of the statistics and data visualizations learning analytics systems present based on the most accessible data–not the most useful or accurate. Due to the nascent nature of data mining in higher education, it is still relatively unclear whether the accessible sets of data are the right datasets to analyze. Also, the instruments, tools, and systems upon which institutions base their interpretations may even output different statistics and visualizations using the same set of data (Greller & Drachsler, 2012). In part, this is due to varying approaches to measurement and technical design, but also due to human influence, varying levels of competency, and the embedding of politics, prejudices, and biases (actively or subconsciously) into learning analytics systems. The problem, then, is that institutions are left interpreting and acting on data that may be misleading or, worse, harmful.

5.9. Data Security

With the increase in student data due to data analytics practices, there is a growing concern regarding data security on campus. Higher education institutions have been subject to massive data breaches in the past, and the comprehensive nature of emerging datasets may make colleges and universities potentially lucrative targets in the future. At the University of Maryland, hackers seized 300,000 faculty, staff, and student records (Svitek & Anderson, 2014); at Indiana University, a bot downloaded a partial set of 146,000 unsecured records of students and graduates (O'Neil, 2014a); and at Maricopa County Community College District, multiple breaches disclosed personally identifiable information of 2.4 million people over three decades (O'Neil, 2014b). Such breaches cost institutions millions of dollars, not only due to the identity protection services they have to offer those affected, but also because data security is a "multiheaded hydra" (O'Neil, 2014c, para. 8); institutions often have to acquire forensics resources (if they do not have them in place), assess and rebuild security teams, and shore up network infrastructures. Arguably, though, the larger cost to institutions is the hit they take to their reputation when the data leaks go public.

Institutions will have to consider fine-grained policies and carefully design technologies in order to reduce disclosures of sensitive student data and information. Sclater (2014) reports that two universities, Oxford Brookes University of the United Kingdom and Charles Sturt University of Australia, implemented such policies that strictly govern data access by roles and permissions levels based on the sensitivity of the data. Such policies will have to work in tandem with technologies that can govern the flow of private student information between systems and users. While higher education institutions use student information systems to this end, it is unclear whether or not these systems are up to the task of managing the rising tide of student data and information captured from data trails, emerging from sensors, and aggregated into data warehouses.

5.10. Third Parties

Part of the need for more informed policy and greater control over information flow, technologically speaking, is to strictly control the data institutions provide to third parties for services rendered to them and to reduce the chance of downstream disclosures to other third parties. Often, colleges and universities enter into contracts with educational technology providers for systems, like learning management systems, and services, like data analytics, to pursue projects and gain actionable information which institutions cannot achieve on their own or feel are more efficient to outsource to companies and contractors. Proponents and critics of learning analytics alike recognize that educational technology companies may, and often do, retain student data produced in their systems in order to improve feature sets and build new technologies. These data practices have raised significant concerns, especially regarding the lifecycle of student data. For example, some wonder how long third-party providers will keep, store, and protect the student data they absorb; perhaps more importantly, there is uncertainty about what restricts a company from selling the data, especially given that it may claim ownership of it. Furthermore, should a provider experience a data breach, there is a question about who is responsible–the company or the university–for damages and what actions are required of them under the law.

These concerns and others have raised the discussion about student privacy and third-party educational technology providers at the national level. Led by the Future of Privacy Forum and the Software and Information Industry Association, the two organizations established a non-binding "student privacy pledge" to hold education technology companies accountable to 12 principles regarding the proper use, maintenance, and security of identifiable and non-identifiable data and information (Student Privacy Pledge, 2015). As of this writing, 125 companies have signed onto the pledge, including the major players in edTech: Apple, Google, Houghton Mifflin Harcourt, and Microsoft. The pledge, of course, is just that: it is not a guarantee that the companies will actually stay true to the principles, and some have already faltered in their commitment (see Singer, 2015).

As useful as the privacy pledge may be in terms of moving the student privacy ball forward, it only applies to K-12 educational institutions. In fact, where higher education students are concerned and where third parties are discussed, there is a privacy blindspot. In addition to the privacy pledge, President Obama's (The White House, 2015) recent announcement of a new "Digital Student Privacy Act" (DSPA) also fails to extend new privacy protections to higher education students (Kolowich, 2015). It is unclear why this is so, as higher education students, like their younger peers, are just as susceptible to targeted advertising and data sales by third-party educational technology providers, which the DSPA protects against (Jones, 2015).

While privacy discussions surrounding third parties and student data tend to focus on educational technology companies, there is concern government actors will revive their interest in statewide longitudinal data systems (Sun, 2014). Interestingly, the fears have shifted away from Orwellian schemes and concerns over Big Brother to the problems that a "proverbial student record" (Zeide, 2014) may create. Zeide (2014) notes that "privacy advocates worry that students' early behavior and performance will follow them through the educational system and into the workplace, where decisions will be based on outdated or irrelevant information" (p. 5). Of course, students have always had permanent records in the form of transcripts, but today's data analytics practices raise this concern due to comprehensive ways they can capture daily life.

5.11. An Obligation to Act and the Harms Thereof

With an increase in accessible data and the growth of data analytics, there is also an emerging question as to whether or not colleges and universities are morally obligated to act on the information they glean from learning analytics and other data-driven projects (Campbell, 2007; Kay, Korn, & Oppenheim, 2012; Willis et al., 2013). If, for instance, an instructor is aware that her students are more likely to succeed in her online course if they login early in the week, what obligation does she have to incentivize this behavior and penalize those who login later in the week? Furthermore, consider a small-scale institutional research project that has statistically proven retention rates correlate with the duration of time a student spends at the library. Part of the data that supports this claim was captured by way of an RFID reader at the entryway of the library that logs a student's entrance by reading the RFID chips embedded within her ID. Does the institution now have an obligation to scale the project up and track a wider swath of the student body's library usage to improve retention rates overall? Kay et al. (2012) argue that institutions and their respective actors (e.g., administrators, faculty, teaching assistants, etc.) may, in fact, have an ethical and legal "duty of care" (p. 20) in both of these scenarios, as failing to act may lead to varying degrees of harm for affected students.

An obligation to act does not mean that subsequent actions are equitable and justifiable. In fact, interventions in student behaviors based on information derived from learning analytics applications may entrench unfair biases and harmful prejudices if the underlying data is suspect or the algorithms wrongly categorize or label students (MacCarthy, 2014). Institutional actors may infer from faulty information that they need to expend time and resources to support at-risk students who, in reality, were incorrectly identified as such; the harm, then, is to the students who truly needed the help but were left unanalyzed. This issue is compounded when inaccurate learning analytics systems automatically intervene in student lives on behalf of the university by restricting students access to courses for registration, piling on additional resources to supplement their identified learning deficiencies, or referring students to "learning centers" and their advisors for unnecessary appointments.

The aim of learning analytics interventions is to mold student behavior in such a way that the end result is better learning outcomes or, for some institutions, improved institutional efficiency. While these are two laudable goals, there is growing concern that the human shaping done by way of statistically-derived interventions may, in fact, lead to some negative behaviors, even if the end goals are met. By nudging students to improve on their learning deficiencies or by labelling students as at risk, their awareness of their inadequacies may be heightened (Swenson, 2014). And while knowledge sometimes is the power that drives an individual to make changes in her life, in the case of learning analytics it may be that students simply internalize these nudges and labels as unhelpful criticisms. For students who lack self-confidence or have experienced unsupportive learning environments in the past, interventions may simply be a technological manifestation of the cruel teacher or unsupportive parent who constantly nagged the student to "do better." And where predictions are concerned, especially predictions that are particularly bleak, students may internalize their statistical score as a foregone conclusion and as more evidence that they "won't amount to much." Any effort by students to turn their predictive scores around may be seen as "fruitless labor" (Willis & Pistilli, 2014, para. 10). These issues are serious and real, and they are particularly relevant to students who already come from disadvantaged backgrounds and for whom additional educational support and resources are especially needed; in fact, these groups are sometimes the same populations learning analytics projects target with the hopes of improving high attrition rates (Ferguson, 2012).

The constant barrage of predictions, labels, and visualizations of a student's progress (or lack thereof) may also produce negative student behaviors as a result of the continual "spoon [feeding]" (Sclater, p. 40) of data and information. While Ellis (2013) argues that students are paying customers and what they pay for is assessment of their work, a Big Data-driven assessment regime may go too far. Instead of promoting hard work, assessment processes embedded in learning analytics may promote student behavior that relies more on data dashboards and nudges to action instead of personal reflection and accountability. Put differently, students might be trained to make education choices solely based on external factors rather than developing their internal critical thinking skills, which is always a goal of liberal education. When the quantified student leaves higher education to enter the "real world" and start a career without the crutch of data analytics, there is a question of whether or not students will be able to succeed on their own.

5.12. Student Autonomy Concerns

Learning analytics technologies present challenges to student autonomy. Scalater (2014) and Willis et al. (2013) note that automated interventions and statistical predictions of future performance risk create a learning environment where students believe they have little choice but to act as learning systems say they should or they will. According to Rubel and Jones (forthcoming), "autonomy includes having the ability to self-govern…to be able to make decisions for oneself, based on one's reasons and one's own values" (p. 9). When technologies, like learning analytics, fail to respect students as agents engaged in "active, creative enterprise[s]" (Benn, 1984, p. 229) in and out of university classrooms, red flags are immediately raised. The primary concern is that learning analytics often hide information away from students (e.g., what data is used in predictive algorithms) that they need to make informed decisions, and learning analytics often fails to present students with a full range of options for future action. Instead, learning analytics forces students to act in ways that may be beneficial for the institution and not the student (Johnson, 2014). According to Johnson (2014), learning analytics acts in coercive, paternalistic ways that encourages conformity to institutional values, not values chosen and pursued by students.

There are strong ties between student autonomy and information privacy, and Rubel and Jones (forthcoming) present three facets of the relationship. First, students may seek privacy as an *object* of autonomous choice. In seeking privacy, students indicate that they wish not to disclose some information about themselves, or they wish to act without the burden of dataveillance, for example. Second, privacy may be a *condition* of autonomy, which is to say that students may not be able to establish a sphere of privacy if conditions preclude them from acting based on their own wishes and values. Finally, full (or at least, maximized) access to information enables students to make autonomous choices, and without that information they are left unable to "interpret their situation in the world" (p. 9).

Homing in on specific learning analytics issues, it is highly plausible that these three conditions are rarely met when learning analytics is deployed across campuses. For example, students may not be given the choice to pursue privacy due to institutional policies and practices that are meant to maximize data aggregation and analysis. Similarly, data-based student surveillance may become so inclusive that it would be impossible to escape the gaze of learning analytics technologies. And like government agencies and data-driven companies, higher education institutions may wish to hide information from students in order to "nudge" them in directions they see fit; under this type of information-restricted regime, students are ignorant of other potential ways to navigate their higher education experience and are, potentially, blinded to alternative ways of viewing the world.

5.13. Seeking a Code of Ethics

The problems of learning analytics and Big Data have many seeking a code of ethics to guide such practices (see Berg, 2013; Ferguson, 2012; Prinsloo & Slade, 2013). Policies and legal frameworks built before the emergence of learning analytics often fail to address specific ethical issues (Prinsloo & Slade, 2013). As such, practitioners are left to use their own decision-making criteria to decide whether or not certain data practices are morally just and appropriate; however, the slow rate of progress with respect to ethical decision-making guides is problematic given the rapid development of technology (Swenson, 2014). Individual approaches to issues such as student privacy, data security, and intellectual property may be blurred by personal bias and self-enhancing goals, and not driven by concerns for student interests. In order to resolve this problem and drive the construction of institutional policy, various codes of ethics have emerged.

In the UK, the non-profit organization Jisc has been leading conversations regarding learning analytics with their "Effective Learning Analytics" research series (see Jisc, n.d.), from which their "Code of Practice for Learning Analytics" developed (Sclater, 2015). The document outlines eight unique areas that institutions should respect to ensure "that learning analytics is carried out responsibly, appropriately, and effectively" (p. 1). The eight areas are summarized below:

- Responsibility: Institutions must choose who is responsible for the effective use of learning analytics.
- Transparency and consent: Institutions must define the scope of data collection (including data sources and types), be transparent about how it will be used and to what ends, and students should be given the opportunity to consent to data practices, except under special circumstances.

- Privacy: Student data should only be used by those in the institution who have a legitimate need to do so, and anonymized data should be secured–as much as possible–against re-identification risks from metadata analysis and data aggregation.
- Validity: Data and algorithms used for learning analytics purposes should be "understood, validated, and reviewed" in order to minimize data inaccuracies and to understand the implications thereof.
- Access: Students should have access to data used for learning analytics and the processed data created by learning analytics in a portable format.
- Action: Institutions should clearly state when analytics will prompt interventions and if students are obligated to act on the analytics.
- Adverse impacts: Learning analytics technologies must be used in such a way to reduce harms to students due to, inter alia, improper labelling, institutional preferences, discrimination, and power imbalances.
- Data stewardship: Data should be minimized to only useful levels, used in accordance with applicable laws, and retained for well-defined periods of time; furthermore, students maintain a right to be forgotten.

The Jisc "Code of Practice" is geared towards institutional uses of learning analytics technologies, which separates itself from the Asilomar Convention's (2014) code of ethics. The Asilomar Convention, in contrast to Jisc's ethical code, was established primarily by educators, researchers, and ethicists with the aim to develop "an ethical framework to inform appropriate use of data and technology in learning research [emphasis added]" (Stevens & Silbey, 2014, para. 1), not institutional practice. As such, it is concerned more with advancing the learning sciences through research and data sharing. The Asilomar framework adopted and reworked principles from the CoFIPs and the Belmont Report of 1979 to establish general tenets to guide ethical research in the area of learning analytics and data-driven education. For the purposes of comparing this framework with that of Jisc's, the principles are summarized below:

- Respect for the rights and dignity of learners: Data practices and results of data-driven research about student behaviors should be made public while respecting student privacy, especially given the risks of collecting and analyzing identifiable student data; furthermore, researchers should provide appropriate and useful informed consent mechanisms.
- Beneficence: Researchers should maximize benefits of data-driven research while minimizing possible harms.
- Justice: Findings from research should be used to the benefit of all learners and with special effort to reduce inequalities around learning opportunities and attainment.
- Openness: Research is a public good that should be inclusive, transparent, and open to criticism to improve rigor and validity.
- The humanity of learning: Educational technologies should be used to improve and not degrade the human experience of learning.
- Continuous consideration: Ethical conversations in the area of data-driven education should be ongoing, especially given the relative newness of the technology and related practices.

There is at once a clear contrast and an overlap between the Jisc and Asilomar ethical frameworks. For one, the obvious differentiating aspect is the orientation of the Jisc "Code of Practice" to fit institutional needs (e.g., actor responsibilities and policy considerations) versus the emphasis on research interests (e.g., data sharing and the public value of science) within the

Asilomar Convention's principles. Nonetheless, they share a focused interest in creating and improving data-driven technology and practices using ethical reasoning to support their aims. Each framework recognizes that student privacy is one major consideration to be mindful of, but they also demonstrate an awareness of other issues, which may be less visible, namely beneficence, justice, and humane concerns.

The Jisc and Asilomar frameworks are living documents; the former is still under revision, and the latter recognizes that ethical conversations are always in flux, especially given the fluidity of socio-technical environments. But even though they continue to be reworked and modified, they have already proven to be useful documents, both in sparking scholarly conversation (see Beattie, Woodley, & Souter, 2014) and informing institutional policy (see The Open University, 2014). In their present form, they act as a solid starting point for institutions looking for guidance to inform their practice with learning analytics technology.

Chapter 6. The Research Problem, Approach, and Methods

6.1. Introduction

The previous chapters detail the rise of student data gathering and use, especially with respect to the emergence of Big Data practices. But, as I discussed in section 1.2., very little of the research empirically addresses how institutions are addressing student privacy concerns as they develop Big Data socio-technical practices, like those that look to take advantage of learning analytics' purported benefits. I take up this mantle in the remaining pages. The following sections within this chapter outline the approach I adopted for my empirical study, my research paradigm, and the particular methods I used to gather and analyze data.

6.2. Research Motivations

The review of the literature indicates that there is very little practitioner and scholarly understanding of how institutions address student privacy issues while building capacity for learning analytics. There has been a rise in scholarship related to conceptual and cautionary concerns about student privacy and emerging data practices, but very little of it can be categorized as empirical, "on the ground" research into institutional work related to the technology. Due to this gap, we do not know how institutional actors perceive such problems; therefore, we are unaware about how these perceptions impact their practice in relationship with learning analytics, nor do we know the effects of these problems on institutional capacity building exercises. My study addresses this gap.

In addition to the gap in the literature, this project is motivated in part by my professional interests. For over 10 years, I have participated in conversations about educational technology, built educational technology systems, and taught in face-to-face and online courses where I have used educational technology in support of my instructional aims to enhance learning experiences. My work afforded me the opportunity to see successful, empowering uses of technology in the classroom, in addition to applications of educational technology that failed completely and, in some cases, negatively affected the learning environment. I recognize that educational technologies–and the data upon which they rely–are powerful and various actors yield them as a means to various ends, some well-intended and some not. Put simply, educational technology holds my interest due to its many facets; thus, it motivates my research agenda, including this project.

There are a number of potential gains we may achieve by pursuing research in this area. First, it is clear that learning analytics introduces new data practices and student privacy concerns that were previously unknown, even though the technology and the practices that motivate it build on over 170 years worth of student information use in higher education. Practitioners are often working in the dark, with one hand in front of the other as they encounter privacy problems. This research may help them to consider and resolve student privacy problems before they even begin building capacity for the technology. Second, while the conceptual work about student privacy as it relates to new data-driven educational technologies is useful, it may be difficult for practitioners to envision how these issues materialize in day-to-day practices; empirical work helps to bridge that divide. Finally, discussions are beginning to emerge related to larger policy questions related to student privacy due to state-of-the-art technologies, like learning analytics (see Anderson, 2015; Garcia-Kaplan, 2015; Herold, 2015). Policy revisions and new constructions could benefit from empirically-driven stories that detail institutional practices, those that do good with those that present harm.

6.3. Research Approach

A socio-technical approach rooted in the tradition of social informatics informs this study, which is especially valuable for understanding issues of privacy (Waldo et al., 2007). To be clear, what I mean by "socio-technical" is the interrelation of social and technical elements in specific social contexts (Kling et al., 2005; Mumford, 2000; Sawyer & Eschenfelder, 2002). Such an approach affords researchers, like myself, the opportunity to investigate particular technologies and their connection with "specific individual and institutional arrangements among people and [the technology's] larger social milieu" (Kling et al., 2005, p. 14) in order to examine particular social effects resulting from their use.

At a broad thematic level, the drive to understand technological consequences on human life motivates most social informatics research. Due to this aim, social informatics is often a sense making exercise, in that research within this tradition seeks to untangle the complex ways in which people and technology relate. Often, social informatics researchers aim to characterize how actors of all kinds create, modify, use, and dismiss technologies with the intent of understanding how technological design shapes social life (Bijker, 1995; Mackenzie & Wajcman, 1985; Sawyer & Eschenfelder, 2002). Other goals include, inter alia, understanding how technologies impact agency (Orlikowski, 1992), re/structure relationships (Sawyer, Crowston, Wigand, & Allbritton, 2003), and support or deny values (Friedman, 1999). And since implementations of technology often favor some groups of people to the detriment of others (Wellman, 2001), research stemming from this tradition considers the political elements of technology (Winner, 1980) and the ways in which they are configured to achieve strategic and purposeful ends (Fleck, 1994). Research that considers the politics of technology provides opportunities to see technological artifacts, systems, and practices as expressions of human power and enables critiques of technological arrangements that disenfranchise and potentially harm certain groups of individuals.

The "methodological pluralism" (Sawyer & Eschenfelder, 2002, p. 437) of social informatics allows researchers to employ a vast array of methods and theoretical approaches in socio-technical research. There is no requirement to use a particular type of quantitative or qualitative method, nor is there a mandate to use a particular set of formal theories or build substantive theory. "Best-fit" is the modus operandi of researchers in the social informatics tradition. This freedom has allowed me to adopt a theoretical approach that is well-matched with my socio-technical orientation.

Helen Nissenbaum's (2004, 2010) theory of contextual integrity guides this study by providing a heuristic for analyzing and assessing privacy problems in specific contexts. Both social informatics and contextual integrity recognize that the characteristics of a given context are often fluid and debatable; therefore, researchers must set and justify contextual bounds "through either *a priori* depiction or *post hoc* description" (Sawyer & Eschenfelder, 2002, p. 436, emphasis in original). Often, this is more an "exercise of discovery" (Nissenbaum, 2010, p. 134) and less a clear-cut application of readily accessibly contextual definitions. For Nissenbaum (2010), contexts are "characterized by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)" (p. 132); these elements constitute contexts as "abstract representations" (p. 134) of structured social settings.

Contexts are often interpretive, so it is my responsibility to be clear about what contexts I am studying to remove ambiguity. To this point, I have focused the conversation on the use of student information and data within higher education by empirically studying specific institutions that were building capacity for learning analytics. Thus, my contextual approach works at two interrelated, layered levels. This is what Nissenbaum (2010) calls contextual nesting. My empirical study focuses in on particular institutions (to be detailed in full later on)–call this the micro-contextual level. These institutions, however, also exist within the larger contextual sphere of higher education–call this the macro-contextual level.

Since specific higher education institutions fall within the macro-contextual level, they more likely than not that share the same types of roles, activities, norms, and values. Opp's (2001) work on norm theory explains this contextual homogeneity across institutions by arguing that actors exist in an established "integrated social network," which governs actor behavior and ends-oriented actions. For example, each institution employs administrators, faculty, and staff who engage in role-specific activities, which generally work towards advancing certain values, such as the mission of the institution and educating students in ways that prepare them for professional positions and personal growth, among a variety of other purposes (see Bok, 2006). The norms of the individual institution "prescribe and proscribe acceptable actions and practices" (Nissenbaum, 2010, p. 133) that guide employee behaviors and direct projects in support of specific ends.

At the micro-contextual level, higher education institutions vary in unique ways. For example, institutions will employ different technologies to achieve particular ends, for example to increase retention rates. This is an example of a variation on contextual activities. Similarly, even if institutions share similar activities, it does not follow that they necessarily use the same actor roles or even subscribe to the same norms, nor do they share the same values. To this point, Scott (2008) argues that we should expect normative variation, especially at the role level. Therefore, micro-contextual inquiry is especially useful for analyzing socio-technical assemblages with respect to how technologies impact social behaviors, structures, and values in particular situations.

Where privacy is concerned, contextual integrity "accounts for a right to privacy in personal information… in terms of appropriate flow" (Conley, Datta, Nissenbaum, & Sharma, 2012, p. 772). Appropriate flow concerns the "transmission, communication, transfer, distribution, and dissemination" (Nissenbaum, 2010, p. 140) of information, which is regulated by informational norms. Informational norms are context-based. So, informational norms that apply to the macro-context of higher education, for example, may not apply to secondary education. Informational norms require an understanding of actor roles, with special consideration to senders (*from* whom information is sent), receivers (*to* whom information is sent), and subjects (*about* whom information is transmitted). Furthermore, informational norms encompass informational properties, including relevant attributes, types, and the nature of information. Finally, informational norms are defined by transmission principles that constrain flow. Transmission principles are, for example, sometimes determined by agreements that determine bidirectional information sharing; by commercial relationships that compel information disclosures for services; by earning access to information by way of entitlements or rewards, and so on (Nissenbaum, 2010). Sometimes transmission principles are codified in policies and laws (e.g., FERPA), other times they exist as informal social pacts.

Informational norms serve as telltales for recognizing when privacy invasions are present. "[Privacy expectations are] preserved when informational norms are respected," writes Nissenbaum (2010), "and violated when informational norms are breached" (p. 140). More specifically, new socio-technical systems violate contextual integrity and raise privacy red flags when information flows are modified. In this sense, contextual integrity is maintained when the mélange of actors, information attributes, and transmission principles remains constant; it breaks when parts of it change or are removed.

The theory of contextual integrity includes both descriptive and normative components. It is descriptive in that it enables researchers to audit a context and identify a particular privacy threat. It is normative in that it provides researchers the necessary background to make assessments of whether the threat is good or bad and to make a value judgement about it. Think of doing contextual integrity investigations as a two-step process. First, researchers must take stock of a context, which includes understanding the context's accepted and justified norms and values that support information flows. Second, using a rich understanding of the context under consideration, researchers address how new information practices–the ones that raised privacy flags–conflict or complement with established norms; in doing so, they can elaborate on emerging "moral and political" (Nissenbaum, 2010, p. 72) issues affecting privacy.

The descriptive component of the theory of contextual integrity is informed by a multi-part decision heuristic. I used the heuristic to frame some of my interview questions (the primary data-gathering method to be discussed in full later on in this chapter), and I deploy the framework in section 8.2. to analyze the privacy issues that emerged in my findings.

6.4. Justifications and Problems with a Contextual Approach to Privacy

While I find Nissenbaum's (2010) contextual approach to privacy complimentary to the needs of my research, I understand that others may not, especially given the plurality of privacy theories in the extant literature. As such, I need to justify why contextual integrity fits my needs. I will, in the following, provide a brief description of approaches to privacy along with a fair accounting of their general weaknesses in order to build a case for my inclusion of contextual integrity herein.

As a “chameleon-like word” (BeVier, 1995, p. 458), privacy’s definition changes in order to capture a variety of interests and, therefore, researchers consider it with an equally diverse array of theories. There is no “unitary definition of privacy” (DeCew, 1997, p. 61). In part, the variety of strategies that have emerged with respect to privacy are due to a failure among the courts to legally define it (McCarthy, 2015). But scholars also bare some responsibility: they cannot agree on a “clear idea” (Thomson, 1975, p. 312) of what it actually is, yet “disputed boundaries” (p. 313) have solidified around it. According to Daniel Solove (2008), most theorists conceptualize privacy using a standard method, which attempts to “articulate what separates privacy from other things and what identifies it in its various manifestations” in order to “define [its] unique characteristics” (p. 13). Solove identified six common classifications of privacy that the literature has established using this method: the right to be let alone, limited access to self, secrecy, control over personal information about oneself, personhood, and intimacy. I take each in turn briefly below, and in doing so I summarize Solove’s (2008) research.

Quoting Thomas Cooley (1888), Samuel Warren and Louis Brandeis (1890) wrote that the “protection of the person” requires the “right to be let alone” (p. 195). Since the publication of their pivotal article, the right to be let alone has become a dominant privacy approach. Responding to emerging technologies of their time, namely the Eastman Kodak snap camera, Warren and Brandeis recognized that photographers could take pictures of subjects without their knowledge, and, thus, capture their subjects in compromising positions. Furthermore, the photographs could further “fuel” (Solove, 2008, p. 16) the human-interest stories that had become widely-popular in newspapers. Warren and Brandeis (1890) argued that photographic technology and society's interest in the stories it could tell would invade "the sacred precincts of private and domestic life" (p. 195). So, the right to be let alone is interpreted as the right for individuals to protect themselves from intrusion into their lives and, ultimately, to seclude themselves from others.

Privacy as defined as limited access to oneself shares qualities with the right to be let alone, in that it promotes concealment and separation from others–both forms of seclusion (Solove, 2008). Limited access accounts of privacy promote the idea that the individual has the right to choose how much of her thoughts, feelings, and activities should be made public (Godkin, 1890). Sometimes limited access to oneself is a choice one makes, other times it happens by accident, compulsion, or without conscious thinking (O'Brien, 1979).

To this point, the definitions of privacy have included, among other things, concerns about one's physical whereabouts, personal relationships, activities, and psychological state. Yet, privacy-as-secrecy is focused solely on informational issues, which is to say that it concerned with "the concealment of personal facts" (Solove, 2008, p. 22). Privacy protects individuals from disclosure of information about themselves to others, which could be used by second and third parties to their disadvantage. The secrecy account of privacy shares a close relationship with limited access accounts in that secrecy is never total: an individual may reveal facts about herself to some but not others.

Privacy as control over personal information is one of the more "predominant theories" (Solove, 2008, p. 24). This conception, as Alan Westin (1967) argues, defines privacy as a claim by individuals to determine "when, how, and to what extent information about them is communicated to others" (p. 7). It emphasizes not that information is absent in others' minds, but that the locus of control over the information rests with the individual and, therefore, the individual sets the terms under which others can access and use the information (Fried, 1968).

A key concern shared by Warren and Brandeis (1890) was the protection of the person in terms of her character, reputation, and identity. Scholars pursuing personhood privacy, as termed by Freund (1971, 1975), argue that privacy invasions negatively affect personhood when they demean individuality, affront dignity, or assault personality (Bloustein, 1964). When personhood privacy is maintained, it "protects the individual's interest in becoming, being, and remaining a person" (Reiman, 1976, p. 44) by promoting autonomous living.

Solove's (2008) final classification of privacy theory, intimacy, "recognizes that privacy is essential not just for individual self-reflection, but also for human relationships" (p. 34). Solove (2008) goes on to explain that individuals value privacy in order to control degrees of "self-revelation" (p. 34) that determine what types of relationships should emerge (e.g., friendships, intimate lovers, etc.). To Julie Inness (1992), intimacy is the "common denominator" (p. 56) across many sorts of privacy, as it motivates how and why people disclose personal information, provide or limit access to themselves, and make certain decisions.

While each of these six classifications of privacy have and continue to do important work in pushing the privacy conversation forward, they have inherent weaknesses. For a full accounting of their individual deficiencies, see Daniel Solove's (2008) work. But generally speaking, they tend to be too narrow or overly broad regarding what is and is not private. About this, Solove (2008) writes:

> When a conception of privacy is too narrow, it ignores important privacy problems, often resulting in the law's failure to address them. On the other hand, privacy conceptions

> that are too broad fail to provide much guidance; they are often empty of meaning and have little to contribute to the resolution of concrete problems. (p. 39)

The standard method of defining privacy by associating it with specific characteristics and rights naturally includes some things to the detriment of others. In effect, whatever the approach to privacy, it will struggle to find the right–and agreed upon–balance. What is needed is a "bottom up" (Solove, 2008, p. 40), "pluralistic" approach to privacy that encompasses a wide variety of interests, characteristics, and rights rooted in a given context.

A contextual approach to privacy does not, *a priori*, establish what aspects of privacy are valuable and relevant. For instance, this approach does not allow me as the researcher to state that intimacy is *the* way to look at student privacy issues related to Big Data practices. Instead, privacy-as-intimacy has to prove its relevancy in my context under study. An advantage of a contextual approach is that it allows for clusters of privacy characteristics to materialize. It could be that intimacy is an especially relevant aspect of, again for example, student privacy. It may also be that information control is *also* apropos to student privacy concerns. One does not preclude the other, and, in fact, they may work in tandem.

Contextual approaches to privacy problems also take into consideration temporal aspects, which unbending universal approaches to privacy tend to leave out. About this, Solove (2008) writes that "we need not demand that a theory of privacy be impervious to history[….] The matters we consider private change over time" (p. 50). Information practices, roles, norms, and associated values ebb and flow together as a socio-technical blend. Contexts naturally reflect this ever-evolving amalgamation of social and technological characteristics. And while there may exist elements of privacy that remain relatively static throughout years of history, they are always in flux due to the fact that they are embedded within changing contexts. "Privacy is a condition we create, and as such, it is dynamic and changing" (Solove, 2008, p. 65).

A final advantage of a contextual approach to privacy is that it refocuses attention on specific privacy problems. By examining specific problems in a given context, researchers take a more pragmatic, and arguably more useful, route towards resolving privacy issues. While privacy theory abstracted from a context may guide policymakers, for instance, it will often fail to capture the nuance of a privacy issue because of its distance from contextual conditions. If, instead, researchers and theorists start with an examination of privacy problems as they exist in specific contexts, the recommendations and abstract generalizations that emerge from these studies would be descriptively richer, grounded in real life, and helpful for others who are left to settle the social, technical, and legal aspects of privacy matters.

Like all privacy theories, a contextual approach to privacy has its share of criticisms. First, there is the concern of how to define a context. Nissenbaum (2010) pulls on established sociology to define contexts as "structured social settings with characteristics that have evolved over time (sometimes long periods of time) and are subject to a host of causes and contingencies of purpose, place, culture, historical accident, and more" (p. 130). What bounds a particular context, however, is hard to define. Researchers who use contextual integrity as their guiding framework must be able to draw those bounds clearly and ably defend them. Nissenbaum (2010) recognizes that contexts may be informal, formal, nested, in conflict with each other, overlapping, and so on; however, she provides no method for defining a context, except to say that it is an "exercise of discovery" (p. 134).

Related to this concern is a secondary issue related to the facile nature of context choice. It may seem relatively straightforward to say, for example, that the context under study is public higher education. Within education writ large, this contextual choice seems relatively straight forward with respect to setting contextual bounds. However, it masks the fact that there are important contextual factors *within* certain segments of higher education that may have more of a defining role than the obvious bounds.

Consider, for example, the importance of austerity conditions in public higher education. Public colleges and universities continue to face a decline in funding from their parent states; in response, institutions have raised tuition rates, which has decreased the affordability of a college education (U.S. Government Accountability Office, 2014). One socio-technical response to these conditions has been to increase institutional efficiencies by adopting data-driven systems in order to identify wasteful, redundant, or unnecessary practices, programs, and employees.

If the contextual bounds fail to take into consideration and recognize important factors, like austerity measures and the role of technological initiatives, then the resulting empirical product will be flawed. The inverse is also true. If a researcher uses these factors as the defining characteristics of a context to the detriment of other less dominant, but still important, characteristics, then her findings will also be missing crucial elements.

To resolve the problem of defining one's context, the researcher should take into consideration the following recommendations. First, researchers need to clearly bound their contexts. Pragmatically speaking, this is useful for keeping a study in scope. Descriptively speaking, this helps readers of one's research understand what is being studied. Second, researchers should use research methods that allow for the emergence of important aspects of privacy problems and relevant contextual factors. *A priori* deductions about privacy and a particular context are very likely to miss important characteristics and conditions that help to make sense of the privacy problem and define the context under study.

A second problem with a contextual approach to privacy is the perception that it tends towards the descriptive; that is to say that it takes stock of privacy problems and surrounding conditions without allowing for normative evaluations. It could be, according to Alan Rubel (personal communication, December 3, 2014), that contextual inquiry may simply identify that a new socio-technical practice is compatible with bad–but entrenched–values. And if new practices do not trip red flags in heuristics like Nissenbaum's (2010), then researchers will not be motivated to critically review them when using Nissenbaum's framework.

Helen Nissenbaum's (2010) framework of contextual integrity accounts for this concern, however. By her own admission, her framework has "blind spots" (p. 161) related to entrenched practices. However, when a new socio-technical practice raises a prima facie privacy concern, an opportunity to take stock of contextual values and practices presents itself, which enables researchers to make grounded "moral justifications" (p. 166). These justifications take into consideration how entrenched norms, in comparison with the novel practice, create moral quandaries or deserve their established standing, and vice versa for the novel practice. Nothing precludes researchers and theorists from making normative assessments of entrenched practices using a contextual approach. In fact, those assessments are made much stronger because of grounded evidence.

6.5. Research Methodology

Epistemological viewpoints, ontological beliefs, and axiological alliances all determine a researcher's paradigm for empirical inquiry and influence her methodological choices. Three dominant research models exist: pre-positivism, positivism, and post-positivism; the latter is also termed the "naturalistic paradigm" (Lincoln & Guba, 1985). For the purposes of this study, I adopted a naturalistic approach, as it was a good fit with my constructivist grounded theory and interpretive case study methodology.

Naturalism

The ontological viewpoint of naturalists is that there are multiple, socially constructed realities, which are so complex and unique that scientific methods of control and prediction often used by positivists are often untenable. The epistemological viewpoint is that the object under study (e.g., human interactions) is constituted by the interaction between the researcher and the object: the researcher influences the object, the object influences the researcher, and both, then, are co-constructed (Lincoln & Guba, 1985). Naturalists reason that if there is no singular reality (the ontological viewpoint), and if knowledge created is co-constructed (the epistemological viewpoint), then one would not expect to create a generalizable body of knowledge. Naturalists argue, then, that explaining specific processes and facts (an "idiographic body of knowledge") on a case-by-case basis is the more realistic aim of research (Lincoln & Guba, 1985, p. 38).

Concerning values, the axiology of naturalism is that it is "value-bound" (Lincoln & Guba, 1985, p.38), which is to say that values are wrapped up in the entirety of the research process. The following statements represent the value-bound nature of naturalism:

1. Research projects are influenced by the researcher's values;
2. research is influenced by the paradigm that frames inquiry;
3. where methods and guiding theory are concerned, their inherent values influence research;
4. values inherent in the context under study influence the research;
5. and all influencing values must be compatible, for value dissonance creates questionable "findings and interpretations."

Together, the ontology, epistemology, and axiology of naturalism form the key tenets of naturalistic inquiry. And if the methods one chooses creates dissonance with these tenets, readers should question the project's research design. The following parts of the section address the relationship between the tenets of naturalistic inquiry with constructivist grounded theory and interpretive case studies, which guided this project's design.

Grounded Theory

In response to the "Great Man" theories of Max Weber, Émile Durkheim, and others that dominated qualitative research in their time, Barney Glaser and Anselm Strauss (1967) sought out new methods and explanatory processes that pivoted towards generation of theory instead of verification of what had already been deductively reasoned; their creation was grounded theory. Grounded theory is the comparative analysis of discrete data to generate a conceptual understanding–a grounded theory–to illustrate "what is going on" (Glaser and Strauss, 1967, p. 23) in a study. It is naturally inductive; that is, whatever abstractions researchers make are grounded in data. While Glaser and Strauss are the forefather of grounded theory, many variants have emerged.

There exists different "schools" (Bryant, 2009, n.p.) of grounded theory, and so I need to explain and justify which one I adopted for this study. Generally, grounded theory schools are demarcated by their paradigmatic roots, methodological approaches, and points of emphasis. Denzin (2007), in fact, has found at least seven unique variants of grounded theory over four decades of its development. While it is unnecessary to dissect each of the seven variants of grounded theory here, it is useful to compare its more positivist, or what Charmaz (2014) terms its "objectivist" characteristics with its naturalistic arm.

Critics of Glaserian and Straussian grounded theory maintain that a "cloak of objectivity" shades their respective approaches and continues out-of-date positivistic traditions (Charmaz, 2014, p. 321). What marks Glaserian grounded theory as positivistic is its treatment of data as value-free abstractions from "the historical, social, and situated processes of their production" (Charmaz, 2014, p. 237). Glaser maintains that data are like fruit: the researcher's job is simply to pick them, describe them, and theorize from them. The researcher's description may be "perspectival" (Glaser, 2001, p. 48), but the data will always remain outside the influence of the researcher.

In contrast with the positivist approaches to grounded theory stands constructivist grounded theory, which is aligned with naturalistic inquiry (Bryant 2002, Charmaz 2000). Like its name hints at, constructivist grounded theory "uses both data and analysis as created from shared experiences and relationships with participants and other sources of data" (Charmaz, 2014, p. 239). Data and the process by which it is discovered are co-constructed in a particular place and time–a context–between the researcher and her research subjects. I adopted constructivist grounded theory for this project, which is the first foundational component of my research methodology.

According to Charmaz (2014), "constructivists study *how*–and sometimes *why*–participants construct meanings and actions in specific situations" (p. 239, emphasis in original). Through interpretation of their data, researchers in the constructivist tradition make meaning and sense of the data. Since researcher interpretation is key to theory development, the intermediate (data) and final (written work, including theory) products are built "under preexisting structural conditions, arise in emergent situations, and are influenced by the researcher's perspectives, privileges, positions, interactions, and geographical locations" (Charmaz, 2014, p. 240). The constructivist will also take into consideration the multiple realities of the study's participants, as well as that of her own. Participants have a unique worldview, and they develop "meaning and action" (Charmaz, 2014, p. 241) accordingly. It is the researcher's goal, then, to arrive at an interpretation of a participant's worldview.

The idea of encountering varied worldviews aligns with ontological perspective of naturalistic inquiry and constructivist grounded theory. When researchers encounter multiple realities of their research subjects, they are often left to make interpretations in order to reconcile different ontological perspectives. Theory developed from interpretations "assumes emergent, multiple realities; indeterminacy; facts and values as inextricably linked; truth as provisional; and social life as processual" (Charmaz, 2014, p. 231). One of the primary goals, then, of grounded theory research is to develop an understanding of how participants view reality, construct their own, and act within it.

Researchers enter in to the lives of their participants, and therefore need to be aware of and reflexive about ways in which they may influence the project. "Reflexivity," writes Charmaz (2014), "includes examining how the researchers interests, positions, and assumptions influenced his or her own inquiry" (p. 344). Speaking about axiology, Creswell (2013, p. 20) writes, "all researchers bring values to a study, but qualitative researchers make their values known in a study." The constructivist should lay bare as appropriate her own axiology, so as to confront how those values may influence her interpretations of the study and its final constructions.

I recognize that I bring my own history, values, and knowledge to this project. Without getting deeply autobiographical, I am a part of a generation that is tightly associated with the development and use of the social web. As a part of this generation's shared experiences, I have participated in online, data-driven social platforms–but I have also rebelled against them. I recognize that applications based on personally identifiable data create current and future privacy problems, and so I have purposefully left some of these online spaces, covering my tracks as best I could. These actions represent how I value information privacy, which is a key concern in this project. Furthermore, my understanding of privacy is informed by a significant amount of reading, writing, and deliberating on issues and theory related to personal and informational privacy, both in and out of the classroom. In particular, the theoretical reading I have done informs my life and my work, and I recognize that I am not entering into this project as a "blank slate."

I reflexively tried to neutralize my own personal history and values by writing first-person narrative journal entries. By writing in a stream of consciousness about my reflections and insights, I was able to bring to the fore potentially influential personal values and perspectives that could adversely color my findings. Noting these issues allowed me to take stock of how they could influence my work. I journaled after holding conversations with my research participants and during data analysis procedures.

With respect to the role of past reading and learning, I cannot account entirely for its influence on this project. Echoing Cutcliffe (2003), "I can only ever know a relatively small portion or percentage of my conscious; I can only ever attain a limited degree of self-awareness" (p. 140). With that said, I judiciously integrated literature and pursued conversations related to the themes of this project when I felt that they would prove useful for improving my understanding of emerging findings and the context under study.

Interpretive Case Studies

Constructivist grounded theory (Charmaz, 2014) is especially well-suited to interpretive case study designs, which is the second foundational component of my research methodology.

Charmaz's approach to grounded theory emphasizes that construct building and, ultimately, theory building is a result of participation by the researcher in "shared experiences and relationships" with the data. As a result, her approach to grounded theory "lies squarely in the interpretive tradition" (Charmaz, 2006, p. 130) of interpretive case study designs.

Interpretive case study designs allow researchers to focus their attention on issues of information system development, adoption, use (or non-use) in specific organizational settings (Darke, Shanks, & Broadbent, 1998). Walsham (1993) emphasizes that interpretive case studies allow researchers to understand the "context" and "processes" (p. 4-5) of information systems, as well as how each influences the other. Importantly, interpretive case study designs make epistemological and ontological assumptions.

Interpretive researchers make a "primary presumption" of constructivism; that is to say that an individual's understanding of reality and the ways in which knowledge is created and acquired is dependent on her "participation in social processes" (Orlikowski & Baroudi, 1991, p. 13). At the heart of interpretive case study research is the researcher's aim to understand "human sense making [through] language, consciousness, shared meanings, documents, tools, and other artifacts" (Klein & Myers, 1999, p. 69).

When researcher's adopt an "outside observer" (Walsham, 1995, p. 78) role, which I employed for this study, they are able to develop a rich understanding of participants' perspectives. According to Walsham (1995), an outside observer role places the researcher at the fringes of the organization under study; as such, she is seen as having less of a "direct personal stake in various interpretations and outcomes" (p. 77). Since the researcher is not viewed as an influencer or threat by the research participant, data collection may be richer due to forthright conversations. However, unlike in ethnography, the researcher's relative abstraction from the organization reduces her opportunity to capture some action, decisions, and negotiations from "the inside" (p. 77).

Interpretive case studies are best-matched to data collection and analysis methods that support an "iterative process" (Walsham, 1995, p. 76), like grounded theory. Iteration in both areas allows the researcher to be flexible and responsive to new opportunities and leads for data collection; similarly, views on emergent concepts stemming from analysis are also iterative in that they are initially constructed and refined as the reacher forms a gestalt understanding of the data. Eisenhardt (1989) suggests that iteration is key to theory building in interpretive case studies, as it allows for confirming emerging constructs and reconciling contradictions. The following sections detail how I iteratively selected (section 6.7.), collected (section 6.8.), and analyzed (section 6.9.) data for this project.

6.6. Data Sources

My study used multiple sources of data, which is common for grounded theory studies. Barney Glaser (2001) famously wrote "all is data" (p. 145), which has become a common dictum in grounded theory research. "What is going on in the research scene," explains Glaser, "is the data whatever the source, whether interview, observations, documents, in whatever combination." This inclusive approach to data allows the researcher to case a wide net and remain open to and aware of the potential conceptual value to be gained from multiple types of data. It is often the case that grounded theory studies are framed by one particular type of data, say interviews. And while many a grounded theory has been developed on the back of interview data alone, an openness and willingness to seek out other relevant data different from the primary type of data can potentially add new explanatory information to a study by providing new and different data to constantly compare "The richer the range of data," writes Holton (2008, n.p.), "the greater the potential for producing multivariate theory."

For this study, I used a multi-site, interpretive case study design. At each case, I interviewed participants as my primary source of data, but I also incorporated publicly available information and privately provided documentation from my participants. The following sub-sections provide an overview of my case sites, interviews, and documentation with special regard to selection and inclusion criteria; in section 7.2., I provide a detailed description of the data that informed my findings.

Case Study Design and Case Sites

In his oft-referenced text, *Case Study Research*, Robert Yin (2009, p. 18) defines a case study "as an empirical inquiry [into] a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident." Case studies are exemplary, he writes, when the cases under study are interesting and when their issues are important theoretically, practically, or in terms of policy (Yin, 2009, p. 185). As such, case studies help to describe specific issues, in specific places, and specific times in order to inform a wider public about a particular problem.

As Yin (2009) politely alludes to, not all case studies are interesting or appeal to wider audiences. An internal case study of a governmental department's workflow, for example, may be especially relevant and useful for that department's administration; however, the study may provide little value to a general audience due to already existing "workflow" studies and applicable theoretical arguments. Thus, it is important to conduct exemplary case studies that have the potential to break new conceptual ground, inform practice, or effect policy decisions. The case sites I included in this project are exemplary in that they illuminate how institutions are working through issues of student privacy as they develop new practices, technologies, and related policies.

It may be simple for a researcher to pin down an issue to study within a case study, but less so to define, exactly, what the case is. In fact, this has "plagued" (Yin, 2009, p. 29) researchers due to the decisions one has to make about what bounds a given case and the lack of a clear and guiding definition (see Ragin & Becker, 1992 for wide ranging interpretations of a case). For some, a case may be an event, for others a definable group of people, such as a familial unit. The onus is on the researcher to clearly define the unit of analysis–the case or cases–in such a way that the reader understands what a unit is and how it will inform the study's research questions. Moreover, the definition of the unit must be comprehensive, including the types or roles of individuals considered as data subjects, their spatial characteristics (*e.g.*, a neighborhood or a specific organization), and their temporal relationship to the issue under study.

For the purposes of this study, my main unit of analysis began at the institutional level, which is to say that one unit is one college or university. Below this existed a sub-unit level, which included institutional actors based on specific roles, responsibilities, and their relationship to the learning analytics project (I provide detail about these roles below).

I selected two case sites–two unique higher education institutions–for this study. Each was building capacity for learning analytics and had invested significant resources to see their efforts come to fruition. My decision to select these institutions came after completing a pilot study of one of the institutions and discovering that the first case had a working relationship with the other, especially where shared resources were concerned.

Since the institutions were closely connected in their respective development of learning analytics, there were certain pragmatic advantages to my choice. In some respects, it was easier to identify the important actors at each institution, as some involved in learning analytics projects had cross-institutional relationships. For example, instructional technologists at one institution were in contact with their peers at the other institution. Furthermore, the institutions shared the same technical infrastructure, including one data warehouse and a single installation of one type of learning analytics technology. To an extent, this also meant that the institutions shared documents.

In other respects, the institutions diverged from each other. The institutions took different tacks when dealing with specific processes, policies, and culture building around learning analytics. And even though they had collaborated on one type of learning analytics technology, they also pursued different learning analytics projects on their own. The connection between the two institutions and the unique variations between them led to a rich comparative case study analysis.

I have given my case sites pseudonyms for this research project to protect the identity of the institutions and the participants of this study who work for them. I generally describe each institution below using details from the Carnegie Classification of Institutions of Higher Education.

My first case site was Hammond University, a large, four year, public institution that serves primarily residential students. Some of its programs are offered fully online, but most require regular visits to campus for courses. Its student population is made up mostly of undergraduates, but it does grant masters and doctoral degrees. My second case site was Saint May State University, a four year, public institution smaller in size than Hammond University. Its student population is mixed between undergraduates and graduates, but the former makes up the bulk of its size. The institution grants both bachelors and masters degrees, and some programs at both levels are offered fully online, although the Carnegie Classification system still ranks it as a primarily residential institution.

Interviews

Interviews, and the digital audio recordings produced from them, were my primary data source for this project. The following sub-sections addresses what an interview is, the goals of interviews, and some ethical issues. Below, I also discuss my selection criteria for interviewees, what participant roles and responsibilities were pertinent to my study, how I accessed my interviewees and elicited their participation, and, finally, the procedures I used to collect data during my interviews.

Interviews are just that: a sharing of *views* on a given topic *inter* (between or among) individuals. They provide participants an opportunity to engage each other in a focused conversation and to talk about their unique understanding of and perspective on their world. Dialogue encourages interview participants to "convey their situation from their own perspective and in their own words" (Kvale, 2007, p. 11). In some interviews–but not all–an overarching goal is for participants to work together in the interview to validate, assess, and build knowledge; as such, this approach to interviewing considers conversations as "site[s] of construction" (Kvale, 2007, p. 21).

Interviews present ethical dilemmas related to power and exploitation. With regard to the former, the amount of power an interviewer has and the extent to which she can use it to control the interview is lopsided in her favor (Gubrium & Koro-Ljungberg, 2005). It is through discourse that Foucault (1995, 1998) sees power being exerted, transmitted, and produced, and so there are significant power problems using interviewing methods. In the case of the interview, the interviewer sets the framework for the conversation (its topical matter), dictates its momentum (by beginning, modulating, and ending the interview), deliberately manipulates it (by selectively choosing specific questions), and exercises authority over interviewees (by limiting their agency in the conversation) (Wang & Yan, 2012).

The second ethical dilemma is exploitation, and it is specifically related to grounded theory-style methods. The researcher picks and choose the relevant bits and pieces without thinking about them as a part of a whole story, which Dey (1999) argues is a form of interviewee exploitation. By focusing on developing an "analytic story," the researcher prioritizes pieces of the interviewee's story about the participant's lived experiences for the project's benefit, not the interviewee's benefit.

Researchers reduce the effects of both the dilemma of power and exploitation using a constructivist approach to interviews (Charmaz, 2014). An "egalitarian" (Charmaz, 2014, p. 87) approach to discussion reduces issues of power when interviewers and interviewees collaboratively exchange ideas, reflections, and analyses to build an understanding of the latter's worldview. Where exploitation is concerned, this issue is lessened when the researcher returns back to the interviewee post-interview, explains her findings, and establishes through dialogue an understanding of the value of the interviewee's participation in the project; section 6.10. addresses the concept of resonance, which in part requires having this conversation with interviewees. By doing this, the researcher shows respect for interviewees' experiences (Blumer, 1969) and validates their thoughts, ideas, and stories as important.

Another controversial issue with regard to interviewing is selecting the number of participants. Researchers continue to argue what the optimal sample size is for an interview-based study, suggesting dissimilar ranges and criteria that ultimately depend on one's methodological orientation (see Creswell, 2013; Francis et al., 2010). For example, Bernard

(2000) argues for 36 interviews for an ethnographic study, Morse (1994) suggests at least six interviews for phenomenological studies, and Bertaux (1981) recommends at least 15 interviews for any type of qualitative research. But a shift has occurred in qualitative research that deemphasizes pre-selecting an amount of interviewees towards letting the emergent findings indicate if more interviews are necessary (Beitin, 2012).

It is challenging for researchers to choose the optimal number beforehand, given that the number of interviews has little to do with arriving at grounded findings. Too many interviews present ethical issues (e.g., waste of resources and time) (Francis et al., 2010); create an overabundance of data, which may stifle deep, "penetrating" (Kvale, 2007, p. 43) analysis that fewer interviews would allow; and do not necessarily "guarantee" (Charmaz, 2014, p. 107) original or better research. Too few interviews reduce the transferability of the study's findings into other contexts, reduce the validity of whatever theoretical themes are developed due to a lack of evidence, and reduces the respectability of the study (Charmaz, 2014).

Instead of predefining a sample size, researchers now consider saturation to be the "gold standard" (Guest, Bunce, Johnson, 2006, p. 60) for determining the amount of interviews to pursue over the course of a study. Colloquially, saturation is an intuitive feeling that the researcher "has learned all there is to be learned from the interviews" (Johnson & Rowlands, 2012, p. 108). But if a hunch is enough of a gold standard to defend a sampling size, then interview-based qualitative research would be in dire straights. Fortunately, sampling for saturation, or theoretical saturation more specifically, has strong roots in the grounded theory tradition (Glaser & Strauss, 1967). I describe theoretical sampling and saturation in-depth in section 6.9., but I will shortly say here that sampling for saturation is motivated by the "evolving theory" (Draucker, Martsolf, Ross, & Rusk, 2007, p. 1137) and what a researcher learns from the stories of her subjects.

After the question of how many interviewees comes the issue of who to interview. Perhaps more so than in other types of qualitative studies, the initial interviewees in a grounded theory-based study are more important simply because the early emergent concepts may lead the researcher in a particular direction moving forward. Therefore, it is important that the researcher chooses the initial set of interviewees carefully based on particular criteria; I have done just that using criterion sampling procedures as my guide (Creswell, 2013). Below, I describe my initial participant criteria, which are based on particular types of roles and responsibilities that relate to learning analytics projects. Once an interview nears completion and the interviewee has an understanding of the pertinent questions to the research problem, the researcher can also ask the participant for a "nominated sample" (Morse, 2007, p. 236), or recommendations for other potential participants who met specific criteria. I asked each of my interviewees to nominate potential institutional actors whose work somehow informed learning analytics projects on their respective campus.

In order to capture rich data from a wide swath of institutional actors who were either directly or tangentially involved with learning analytics at my case sites, I cast a wide net to select my interviewees. I describe in the following sub-sections the types and roles of interviewees I was able to gain access to using criterion and nominated sampling.

Data scientists

Data scientists are individuals who build algorithmic models for learning analytics technologies. These models describe and predict student success based on what is known about the individual. For example, socio-economic information, academic histories, biographical variables, and behavioral data are all sources of raw data and processed information that a data scientist, either on her own or influenced by others, may include in a learning analytics model.

Instructional technologists and designers

Instructional technologists and designers are intermediaries between those who build and administer the learning analytics system and the end user, primarily faculty. These individuals instruct the faculty member on what analytics models are available, what information is included in each model, and how a model, when used in their course, could aid their teaching. Once a model is chosen, they work closely with the instructor to monitor the model's outputs and determine if its statistical descriptions and predictions are providing accurate or misleading information.

Campus legal counsel

Campus legal counsel ensure that learning analytics technologies and related practices abide by campus, state, and federal policies and laws. In close working relationship with individuals like chief information officers, the institutional review board, project leaders, data scientists, and others, campus legal counsel reviews what if any legal issues learning analytics may present.

Registrars

Registrars are authorized individuals responsible for student records. They collect, record, maintain, and report student data within the guidelines put forth by the institution and in accordance with state and federal laws. Part of their duties are to administer student information systems–systems which store sensitive data about individual students–in order to maintain a level of security and provide access to authorized data records. The registrar may also work with other institutional actors to enable access to and sharing of data more broadly throughout campus.

Information officers

Information officers on college and university campuses provide leadership and direction for the campuses information technology ecosystem in support of the research, teaching, and learning aims of the university, as well as its administrative needs. In tandem with relevant committees, information officers develop governance strategies to guide information technology development and write campus policies to promote appropriate use. They are also responsible for developing policies and guiding processes for protecting personally identifiable information (such as student records, medical records, employee records, etc.), as well as assessing risks and managing the impact those risks may have on the institution.

System administrators

System administrators include individuals at the campus level who operate the physical infrastructure and applications of learning analytics related systems. Their responsibilities include system installation, maintenance, configuration, and customization of a given system to meet the needs of the end user. Additionally, they may be tasked with hardening a particular system's security using hardware and software upgrades, as well as by suggesting policy to information technology leadership to maintain a safe and securing computing environment.

Instructors

The instructors category of actors is inclusive of all full-time faculty, part-time instructors, and teaching assistants who have a responsibility to educate students in a for-credit university program. Instruction may be done in a face-to-face environment, fully online, or a computer-mediated hybrid of the two using some form of learning management software.

Student support professionals

Student support professionals work on staff to provide resources, advice, and an instruction to individuals and groups of students. In that sense, this group includes advisors, but it also encompasses actors who direct specific initiatives to help target students, like minorities or first generation students.

Documentation

Documents can "corroborate and augment" other evidential data by confirming information (e.g., names, dates, etc.), adding new details to known information, and providing inferential evidence (e.g., in order to trace a network of communication and establish new leads and questions) (Yin, 2009, p. 103). And by using documentation in addition to other qualitative and quantitative data sources, it allows researchers to increase the depth of their findings (Rothbauer, 2008).

I planned to augment my interview data with a large collection of documentation, but I was unsuccessful in obtaining it. In an initial pilot study related to this project, I made contact with a participant who had agreed to provide a cache of documents related to learning analytics and capacity building on her campus. Unfortunately, regardless of multiple requests for access, she never produced those documents. So, the documentation I gained access to was solely public in nature and available on institutional websites; nonetheless, the documents were still useful. Generally speaking, the final cache of documents I obtained described learning analytics technologies and projects, as well as perspectives from campus administrators.

6.7. Data Collection Procedures

I detail in the following sub-sections the procedures I took to gain access to my case sites, research participants, and documentation. I also explain how I transformed my interviews into analyzable data.

Site access

There were no special procedures to follow or permissions to get in order to gain access to my case sites. When my university's institutional review board (IRB) examined my application, they found no reason that justified requiring consent from institutional administrative personnel before beginning data collection. Therefore, I was able to contact my interviewees directly without any intermediary action.

Interviewee access

To gain access to my interviewees, I followed a three-step process. First, I identified potential interviewees by reviewing the institution's publicly accessible directory. Doing so enabled me to review individuals based on their role. When the directory provided ambiguous information, I further researched the potential interviewee by reviewing her department's website; department websites provided adequate information to choose whether or not to pursue the potential interviewee for inclusion in the study. When individuals met my inclusion requirements, I added them to a spreadsheet, which included their name, title, department, institution, and publicly accessible phone number and e-mail. To protect the identity of potential participants, this spreadsheet was encrypted, required two separate passwords to access it, and was only accessible on my personal computer.

Second, I e-mailed individuals who met my inclusion criteria an interview request message. Appendix A includes the contents of this e-mail. When potential interviewees responded, I took note of whether or not they were willing to participate in the aforementioned spreadsheet. In cases where individuals declined, I thanked them for their message and asked if they could nominate other individuals for participation; if they responded to this, I took note of those individuals, vetted them based on the inclusion criteria, and sent them an interview request e-mail. Sometimes potential interviewees did not respond to my initial request. In these cases, I followed up with two messages before I stopped attempting to contact them.

When potential interviewees accepted my invitation for an interview, we corresponded via e-mail to set up a time and location to meet. See Appendix B for the contents of that e-mail. There are two pieces to point out from that e-mail. One, I allowed potential interviewees to make the choice of the meeting location. I did this to optimize their level of comfort, enhance the security of the conversation, and protect their privacy per recommendations in the literature (see Denzin & Lincoln, 2003; Seidman, 1991). Providing interviewees choice in this matter was especially important given that I was unfamiliar with the surroundings of their institution. Not all interview locations are optimal, however. So the second point I need to make is that even though interviewees had final choice in the matter, I provided guidance by explaining some of the characteristics of an optimal interview location based on other recommendations in the literature (see Gillham, 2000; Herzog, 2012).

I completed most of my interviews on site at the two institutions with two exceptions. In these cases, the interviewees asked that the meeting be held over the phone, and I obliged. In one case, my schedule did not allow me to get to the case site at a time that worked for the participant; in the other case, the interviewee preferred a phone interview.

Interview procedures

All interviews, with the phone conversations as the notable exceptions, followed the procedures I detail below. Upon arriving at the participant's location of choosing, I briefly described the purpose and scope of the project, how they fit the inclusion criteria, and then we reviewed the IRB-approved consent form together. See Appendix C for the contents of the informed consent form. At this point, I gave each individual a paper copy of the consent form and prompted them to read it and ask any questions. Most participants had already reviewed and signed the online consent form, which I provided access to in the interview setup e-mail. For those that had not, they reviewed and signed the form before the interview began.

At this point, I reminded participants that I would digitally audio record the interview. Although this procedure was detailed in the consent form, I summarized how I would use the recordings of the interview, store them securely, and transcribe them for later analysis. I also reminded participants that I would give them and their institution pseudonyms and remove identifying characteristics completely in project documentation and dissemination of findings as needed. I provided interviewees the opportunity to participate in the conversation without the use of the digital audio recorder, but no participants took up the option to do so. Minus the phone interviews, which could not be recorded due to technical issues and a lack of IRB approval to record in this way, and one follow-up interview, I recorded all interviews. For interviews I did not record, I took comprehensive field notes.

Interviews generally lasted between 30 and 60 minutes. Each interview followed a similar script; see Appendix D for a copy of the script. With that said, interview questions did change based on the interests of the interviewee, her experiences, and the natural evolution of the study as thematic elements emerged after coding data and due to seeking theoretical sampling (I will say more about this in the following section).

When the interviews came to a close, I finished the conversation with three remaining questions. The first was to elicit topics, concerns, or ideas about the conversation's subject matter that the interviewee felt needed to be discussed but had not. The second was to test for participant reactivity. Participant reactivity includes the Hawthorne effect and can result from demand characteristics (McKechnie, 2008); in both cases, participants change their behavior based on their participation in research and due to what they believe the researcher expects from them. I asked participants, "during our conversation today, did anything prompt you to respond differently than if you were speaking with a close, trustworthy friend?" No participant indicated that they had. Regardless of theses efforts, reactivity is often inevitable and often difficult to identify; it is not entirely clear whether or not I successfully controlled for its effects. The final question I asked participants was for them to nominate other individuals whose experiences and insight would add value to my research.

After the interview's completion, I sent an e-mail of thanks to my interviewees for their participation. In the message, I provided my contact information, prompted interviewees to ask any questions about the project and their participation in it, and attached a duplicate copy of their signed consent form for their records.

Recording, note taking, and transcribing

Digital audio recordings served two distinct purposes for this project. First, using a digital audio recorder in the interview process enabled me to conduct the conversation uninhibited by the pressure to note whatever seemed to be the more salient points of the conversation. Dialogue flowed more freely, and by not taking comprehensive notes, I removed a potential distraction. Second, by capturing the nuances of speech–hesitancy, emphasis, tone–my recordings brought to light important insights that I would have lost in notes alone.

Once the interviews were conducted and recorded, I transcribed the audio files using an outside company. In some cases, using outside transcription companies can cause concerns among IRB representatives. However, I used the services of a company that my home institution had previously vetted; therefore, my IRB had no issues with contracting out this part of my research to a third party. I reviewed all transcriptions for accuracy by listening to the audio files while reading the transcribed text. This proved necessary, as the transcriptions were not always accurate. I reordered transcriptions until they were accurate. Transcriptions proved to be useful sources of data, as they were well-matched to constant comparative methods of coding, which I detail in section 6.9.

Documentation

The documentation I included in this project helped to establish greater contextual understanding of the institutions under study and add background information to the stories my interviewees told. However, I felt that the public-facing documentation was too far removed from the sense making of my interviewees, so I used it a resource separate from codes that emerged from data analysis. I describe what documents I acquired in the meta description of the data in section 7.2.

6.8. Data Analysis Procedures

Grounded theory procedures have proven to be “extremely useful” (Urquhart, Lehmann, & Myers, 2010, p. 358) in accounting for and describing processes and action related to information systems in particular contexts (Myers, 1997; Goulielmos, 2004). As a qualitative research toolkit, grounded theory provides researchers very specific procedures for their use. And since their creation in the late 1960s, researchers–especially those concerned with the validity of qualitative research methods–have tested, evaluated, and refined these tools, enhancing their credibility along the way. In the sub-sections that follow, I describe the five core grounded theory procedures I employed to analyze my data. These procedures include: coding, the constant comparative method, theoretical sampling, memoing, and saturation. I also discuss the role of technology in my data analysis procedures.

Coding

Grounded theory is an inductive process that requires the researcher to enter into an area of study with little else besides a guiding research problem. But through coding, or the segmenting and labeling of data, researchers begin to develop ideas about what is going in their studies (Becker & Geer, 1960). In doing so, they build substantive and theoretical constructions.

By open coding (also known as initial coding), the researcher considers what the data is telling her about the area under study, what seems important or most relevant, whose point of view the data comes from, and if discrete pieces of data naturally relate to one another (Charmaz, 2006). Open coding is a three-step coding process, starting first with word-by-word codes, then line-by-line codes, and finishing with incident-to-incident codes. As it sounds, word-by-word codes use individual words as distinct "slices of the data" (Glaser & Strauss, 1967, p. 65), which helps the researcher's sense making of the "flow" (Charmaz, 2006, p. 50) and "meanings" of content. Line-by-line coding focuses the researcher on each line of text as unique and separate from the previous and the following. Incident-to-incident coding, in contrast to the previous open-coding procedures, relies on comparisons of incidents, or discrete datasets. For example, a researcher could compare interview transcripts against each other to make close comparisons of interviewee responses to a particular question.

Focused coding–the second phase of coding–differs from open coding both in terms of process and intent. Because, at this stage, the researcher has greater thematic understanding of "what is going on" in the data, she can abstain from strict word-by-word, line-by-line, and incident-to-incident processes previously employed to uncover the relevance of the data to the research problem under study. Focused coding is “directed, selective, and conceptual" (Charmaz, 2006, p. 57), as such there is no need to consider every datum as relevant; instead, the researcher focuses in on what has emerged at this point in terms of frequency and analytic significance. When a researcher either reenters into her past data or gathers new data with focused codes in mind, she seeks to explain the code’s significance by sifting and winnowing further to uncover its characteristics and place codes into categories. The result of focused coding pushes a researcher’s data from description closer to theory.

Finally, the third phase remains: theoretical coding. Glaser (1978) presented theoretical coding in his text *Theoretical Sensitivity*, introducing researchers to the idea that categories of codes need to be examined for potential theoretical relationships. In the text, Glaser presents 18 “families” of theoretical codes, each with its own properties. For example, the “Six C’s” orients the researcher to examine causes, contexts, contingencies, consequences, covariances, and conditions. Charmaz (2006) argues that Glaser’s families of theoretical codes “must earn their way” (p. 64) into the researcher’s analysis; that is to say, they should not be forced upon the data. However, any of Glaser’s families of codes may provide the researcher a “sharp analytic edge” (Charmaz, 2014, p. 151) as she works through the theoretical coding process.

For this study, I followed open, focused, and theoretical coding phases as I analyzed data. There was, however, a notable deviation. In the pilot study, I attempted word-by-word coding. I found this tedious, and it did not evoke useful codes as much as line-by-line coding had. Therefore, I refrained from using it for this study. Instead, I primarily employed line-by-line coding in the first round of data analysis; subsequent rounds of coding were successful using focused coding methods.

The constant comparative method

While the coding processes of grounded theory may assist the researcher in generating an understanding of the area under study, it is the constant comparative method that lends credence to the codes and, in part, validates the researcher's conceptual and theoretical findings. The constant comparative method requires the researcher to take codes and compare them among other codes. Furthermore, the constant comparative method pushes the researcher to compare categories and memos (I describe memos in full below), which are supported by relevant codes. If the codes are supported by new data upon comparison, then they are relevant to the study.

Constantly comparing codes throughout the study strengthens their validity, builds in new characteristics, and develops a greater understanding about the characteristics of the code. The goal here is not to examine codes to see if they fit characteristics of extant theories, but instead the point is to determine the code's relevance in conceptual categories and the larger empirically-based story (Charmaz, 2014). In this project, I primarily compared codes to codes to determine overlaps; this also helped me build categories of codes with explanatory characteristics. I also compared within and between memos and categories to fill out my understanding of the emerging data and develop theoretical concepts.

Theoretical Sampling

Grounded theory has two foundations, the first is the constant comparative method and the second theoretical sampling (Holton, 2007). Theoretical sampling serves as a guide to researchers, prompting them to be "theoretically sensitive" (Glaser & Strauss, 1967, p. 46), to think conceptually, and fill gaps in their emergent categories (and related theory) through systematic data collection. Researchers use theoretical sampling to seek and collect answers to "targeted questions" (Morse, 2007, p. 240), questions they feel need to be answered in order to compare characteristics, relationships, and contradictions among categories in order to move towards saturation.

Before theoretically sampling, a researcher may feel stuck, or at a loss for a direction to pursue new data and create fresh findings. Researchers use theoretical sampling to think critically about some of the missing pieces in their emerging categories. A researcher may memo before theoretically sampling, reviewing in a reflective way how her categories have progressed and ask herself about what else is needed to build in more categorical properties. Charmaz (2014) states that theoretical sampling helps researchers create new questions about emerging categories and define their properties.

I used theoretical sampling specifically to target my participants' perceptions of student privacy. For example, early on the study I asked a general question of how participants defined student privacy. As my interviewees brought up various aspects of student privacy (e.g., privacy as a form of information control, ownership as a way to protect one's information, etc.), I reformed my general question and asked more targeted questions to theoretically sample for data related to these specific constructs of privacy. Theoretical sampling was not useful for all parts of my data collection, however. For instance, it was not necessary to theoretically sample for contextual details of the two institutions, as those details were based in fact not in theory. But, for codes and categories related to privacy that were more abstract and based on the interviewee's perspective, I felt theoretical sampling was a good analytical tool to use.

Memoing

Coding slices the data, breaking it into discrete pieces that need to be put back together to tell a story. Categorizing codes helps with this, for sure, but joining related codes still fails to tell *exactly* what is going in these groups. Enter memoing, the critical "intermediate step" (Charmaz, 2006, p. 72) after coding and before drafting.

Memoing is the researcher's process of writing in a narrative fashion the thoughts, questions, and future directions of her research as she immerses herself in the data gathering and analysis process. This narrative, in essence, is a conversation with the data. It stores the leads one finds in the data and interrogates the gaps one discovers. Combined, the memos a researcher creates develop an archive of the study's process, which is essential for answering questions related to the credibility and validity of the study.

Memos take on different forms depending on where the researcher is in the grounded theory process and her progress in the project as well. Individual codes, for instance, may get their own memos that describe *why* the code was used and *what* it does to help explain the data. Researchers may memo categories, as well, especially in their nascent stages in order to work through suspicions; and, at later points in data analysis, memos can help to raise categories to a "conceptual level" (Charmaz, 2006, p. 73). This writing process elevates the data, gives it its structure, and eventually becomes "the core" (p. 94) of a grounded theory analysis.

Reaching saturation

The question of when to stop seeking new data is an important one. According to Charmaz (2006) and Glaser (2001), the researcher ceases collecting data when her conceptual categories yield no new defining characteristics using the constant comparative method. It may be true that the category continually emerges in the data, but theoretical saturation has not been reached if the category presents new insights about its defining attributes. Researchers employing grounded theory argue that theoretical saturation of categories is the aim of these types of studies (Charmaz, 2006); yet, according to Morse (1995), the literature is sprinkled with *claims* of saturation without *proof*. The onus, then, is on the researcher to be transparent about how she reached saturation and comprehensively describe the core categories when writing up the study.

Two indicators characterize how I knew I reached saturation in this study. First, I experienced saturation across major categories when no new salient characteristics emerged from interviews. Proof of saturation for this indicator is represented by the rich description I provide in the findings, which is supported by the fact that I was able to interview a wide variety of actors involved with learning analytics initiatives across the two campuses and capture a large amount of data through in-depth interviewing. Second, a pilot study informed initial category building, but at that time had not reached saturation. However, the categories I developed in the pilot did indicate the relevant areas that needed more data. This study was able to build on the work of the pilot and establish greater depth of understanding within each relevant category which led to saturation. There always exists an opportunity to add and analyze more data, but I reached a point in my study when I believed more data would not have added any value to my understanding of the emergent findings.

The role of technology in grounded theory data analysis

Grounded theory researchers often employ qualitative research software to quantify and visualize emerging codes and build core categories more easily, and sometimes with greater rigor, than when it is done via simple documentation. Using software allows researchers to see particularly frequent codes and groups of codes throughout an entire dataset. Special applications also allow researchers to compare the frequency of certain codes and examine them for relationships, as well as create a history of a category's emergence by taking snapshots of its development over the course of a project.

What is particularly useful about qualitative research software is that it aids memory recall of codes, categories, and important memos that can become lost as project documentation amasses. Using software strategically allows the researcher to also create explicit links between products of data analysis, which enables researchers to see a network grow between various data products and helps to further solidify connections among important emergent concepts.

There are problems with qualitative research software, however. Sometimes, researchers fall into a trap: the software leads them to believe that the quantity of a code or category is more telling than its qualities, or she may think that relationships exist when they are in fact spurious. And Thomas (2011) argues that the links made between codes, memos, and data in qualitative research software may "fossilize" (p. 128), thwarting the constant comparative method and the flexibility grounded theory encourages of researchers to move in and out of data. Furthermore, these effects may lead the researcher to emphasize description using the quick quotations that the linked data provide, which is in tension with grounded theory's focus on explaining the characteristics and qualities of the codes and their conceptual power.

Coding using qualitative research software is also simple, which could ultimately encourage a sort of coding tunnel vision at the expense of memoing. The problem is that coding without memoing results in poorly constructed categories and inhibits theoretical sampling. Memos pull codes together, establish their relationships, and frame the category's structure through narrative description. "Memoing," writes Thomas (2011), "is where the action is" (p. 139) in grounded theory, and the researcher must remain vigilant against the pull of the software to code at the expense of memoing.

Finally, qualitative research software brings with it an amount of clout that some reviewers and readers of research translate as validity. Often, researchers cite in their methods that they used software to assist in their data analysis procedures, leaving it up to the reader to interpret what exactly that means. Readers should respond to the text with a "so what?" But, Thomas (2011) reports that when the researcher simply states that she has used software, it may "blind" her readers with an undue sense of rigor. Therefore, it is entirely on the shoulders of the researcher to explain the value of using software and transparently describe how it was employed to aid the study.

For this project, I used three software applications to support my data analysis: NVivo, Numbers, and MindNode. NVivo is designed specifically to support researchers in qualitative research, Numbers is a spreadsheet application for Apple computers, and MindNode is a mind mapping application for organizing concepts. I coded interviews first by using NVivo for the pilot study that informed this project, but I transitioned to hand coding paper transcripts later on. The reason for my transition was that, as Thomas (2011) warned, I found myself coding at times without properly reflecting on the value of the code. For the first round of hand coding, I would listen to the audio and code any striking parts of the interview. For the second round of coding, I read the transcript and coded line-by-line. After I coded individual transcripts, I entered the codes into a Pages spreadsheet. The spreadsheet worked as a master index of codes, where the code was attached to each interviewee's pseudonym, the transcript page on which the code occurred, and the frequency of the code. See Appendix E for a table of relevant codes. After hand coding each transcript, I drafted memos about emerging categories. Sometimes these memos were solely related to the individual interviewee, while at other times they acted as iterative drafts of emerging categories and theoretical concepts. See Appendix F for a table of memo titles and summaries.

I used MindNode as a way to visualize emerging categories. MindNode is visualization software that enables users to create mind maps of ideas using parent and child relationships. See Appendix G for an example mind map used in this project. The codes served as the basis for the creation of the parent and child relationships, and they helped establish important relationships among characteristics of a category.

6.9. Evaluative Measures

In order for a grounded theory study to instill confidence in readers, researchers must make clear distinctions between "process and product" (Charmaz, 2014, p. 336). Understanding *how* a researcher arrives at her findings (the process) is often just as important as the findings themselves (the product), since knowing the steps taken and effort put into the project helps to raise the believability of the findings and the respectability of the researcher. To this end, Charmaz (2014) suggests five criteria for assessing grounded theory studies: credibility, originality, resonance, usefulness, and dependability. I have used each of these criteria as evaluative measures in this project in order to increase its rigor.

Credibility

A concern regarding grounded theory studies is their credibility, which is to say that readers need to determine if the findings and the related emergent theory are trustworthy and believable. To achieve credibility, the researcher must first show "intimate familiarity" with the topic under study (Charmaz, 2014, p. 337). Researchers can prove their familiarity by demonstrating a depth of knowledge about the area in the detail of their theoretical categories and by contextualizing their study. Furthermore, the researcher must demonstrate not only knowledge of her study's area, she must also prove a certain understanding of the relationships the study has to other parts of the world by tying it to surrounding contextual elements, like social, political, and technological factors. Second, the researcher must prove that the evidential data is sufficient to explain the findings and has been systematically gathered. Readers judge sufficient data based on the "range, number, and depth of observations" the researcher makes; whether or not the data has been gathered systematically is judged based on proof the researcher made purposeful "comparisons between observations and between categories" (Charmaz, 2014, p. 337). Finally, a study's credibility is also judged by the "logical links" the researcher makes between the data and her theoretical claims (Charmaz, 2014, p. 337).

I employed a number of techniques in order to achieve credibility. First, the historical analysis of student information use and related privacy issues in higher education provides my readers a useful foundation on which to build their understanding of modern concerns via the extensive literature review. This background material should build confidence in my readers regarding my understanding of the literature. Second, this chapter transparently details my data collection and analysis procedures, which should leave my readers, if successful, with few questions about my methods. Third, my findings, which follow this chapter, demonstrate my inclusion of contextual factors in my rich descriptions and grounded theoretical claims. Finally, readers will see direct ties between the claims I make in my discussion with my findings.

Originality

Readers can judge originality in part by considering its "freshness" (Charmaz, 2014, p. 337), which is to say that the project is judge by whether or not it adds to new knowledge. To demonstrate fresh work, the researcher can elicit the extant literature, pull together relevant work for comparison, and carefully demonstrate how the current study extends, criticizes, or negates past work; it may also blaze new areas of research, but it still must separate itself from what is already known. The onus is on the researcher to position her study in such a way that makes plain its uniqueness in comparison with related research.

Originality is also evaluated based on the study's social and theoretical significance. Studies need to demonstrate that they have something to offer society, either in terms of helping individuals understand their day-to-day lived experiences or through concepts that explain social processes. Readers should base their judgement of the study's originality on the extent to which it informs a broad audience and helps them to comprehend their world.

In part, the literature review demonstrates that little is known about how institutional actors perceive and address student privacy issues as they build capacity for learning analytics. While this gap in the literature presents opportunity for original research, I argue that the social significance of studying student privacy issues is the more important and original aspect of this research, especially given the growing concern about privacy issues throughout society.

Resonance

Readers can evaluate a grounded theory further by addressing its resonance, or the extent to which the study captures the "fullness" (Charmaz, 2014, p. 337) of the studied experience. To achieve resonance, the reader must consider the categories the researcher has presented, interrogating them for weaknesses or missing elements. If there are obvious gaps in the categories or clear leads left unexamined, the study's resonance is reduced.

Next, the researcher is responsible for bringing to the surface the in-process sense making of her participants, or what Charmaz (2014) calls the "liminal and unstable" (p. 337) meanings that participants reveal. To stick to what is known or what is established and concrete may overshadow the grey areas of a study which bring to light not-quite-certain participant worldviews. The researcher must prove to the reader in her writing of the study that these areas have been examined, or at least considered. This can be done by critically examining relevant debates and opinions opposite to the participant's.

The researcher is also responsible for drawing explicit links between the individual lives of her participants to "collectivities" (Charmaz, 2014, p. 338), institutions, and organizations when the data indicates it is relevant to do so. Individuals, especially those in an organizational study, have affiliations with and direct ties to larger groups of people who are a part of an even larger institution. This creates a network of connections the researcher could investigate.

Finally, resonance is judged in part by the degree to which the participants recognize themselves in the study, make sense of its outcomes, and are provided "deeper insights about their lives and worlds" (Charmaz, 2014, p. 338). Only the participants can judge this element of resonance; therefore, the researcher is responsible for engaging those who participated in the study or like individuals in a discussion about the results of the study and an explanation of its theoretical findings. Sometimes this is called "member checking" (Creswell, 2014, p. 201).

From my perspective, I achieved resonance by constructing comprehensive core categories inclusive of pertinent and useful data borne from my participants' sense making during interviews. I feel confident about my success in this area, in part, due to the careful attention I paid to iterative coding and memoing throughout the project. I also achieved resonance by connecting my participants' perspectives to relevant institutional-level data. Additionally, I contacted six participants and asked them to review the findings, only two agreed. I held informal conversations with the two participants after they reviewed the findings; they expressed general agreement with the findings, especially when they were able to identify their voices and perspectives in my writing.

Usefulness

Another evaluative area for a grounded theory study concerns its usefulness (Charmaz, 2014). Readers can judge usefulness by examining if the researcher's findings are something people can use in their everyday lives and if it can lead to related studies in other substantive areas. Regarding the latter, usefulness is equivalent to the transferability of the research.

One of the overarching goals of grounded theory is to create conceptual understandings of a substantive area in order to inform practitioners. With this in mind, researchers aim to develop interpretations of the practitioner's lived experience that can inform their everyday work and life situations. Like the resonance criteria, the researcher can return her interpretative findings of the research to the participants to discuss with them the outcome and determine from their response if it may be useful to them and others like them.

Next, readers should judge the usefulness of the study based on its potential for sparking research in other substantive areas. While the theoretical findings are borne from a specific area, their reach may inform other areas of related research. The researcher, then, should propose where areas where her findings may be informative and suggest future research for others to investigate.

With regard to usefulness, I believe that the findings (section 7.) and, especially, the discussion (section 8.) within this work are immediately useful for those I interviewed and practitioners like themselves. Participants often remarked after interviews that they looked forward to reading this work because they believed it was necessary and would become useful for their work with learning analytics. I would also argue that the theoretical findings herein provide other researchers solid leads to follow for their own research agendas, which can inform our collective understanding of student privacy, emerging threats to it, and ways in which to possible meet and resolve new concerns. Specifically to accomplish this end, I suggest more opportunities for research in section 9.2.

Dependability

The final step in evaluating the study requires the researcher to prove that her study is dependable. Who makes that determination is important. Unlike the resonance and usefulness criteria which are in part determined by the participants of the study themselves, dependability should be judged by qualified "external auditors," usually qualified colleagues who are not research leads in the study (Lincoln & Guba, 1985). The auditor looks closely at the procedural methods of the study to see if they support the end result. Temporally, auditors can and should do "quality management" (Flick, 2007, p. 135) checks at any point in the study's progress.

The dissertation process has built into it quality management checks. For my program, faculty members within my department and faculty members a part of my dissertation committee reviewed parts of this project as it was written in mastery demonstration papers (an equivalent to some programs' preliminary exams) and during my statement of intent and proposal defenses. Furthermore, my dissertation advisor reviewed chapter drafts prior to the defense of the dissertation. Additionally, I presented parts of this research at scholarly conferences, where I received constructive criticism and affirmation regarding the direction of my work. In this sense, the project was audited by qualified individuals many times over.

Chapter 7. Study Findings

7.1. Introduction

This chapter describes findings that emerged from the data. I begin with a meta description of my data to help my readers understand what informed the relevant categories of data. The sections that follow describe the categories of data that proved themselves important through iterative coding. Generally, these sections consider relevant contextual characteristics that help us understand motivating factors behind, characteristics about, and effects borne from learning analytics initiatives at both case sites. Next, I describe particular privacy issues actors identified. I follow up this theme with conversations related to data governance, a lack of institutional policies, and interpretations and criticisms of the Family Educational Rights and Privacy Act (FERPA), the federal law that governs student privacy and sets how students should be protected with regard to data-driven projects like learning analytics. The following theme details the potential negative effects of pursuing all the data institutional actors can get. Finally, I turn to categories that detail aspects of transparency and information control my participants expressed as important to privacy issues and learning analytics.

7.2. A Meta Description of the Data

The data the informed this project was multi-faceted due to the multiple data types I reviewed. The findings are supported primarily by my analysis of interview transcripts and field notes, but other media and documentation provided supporting contextual evidence as well. The following subsections provide a meta description of the data to express to readers the depth and breadth of the data I analyzed for this project.

The findings are built primarily on the analysis I completed on transcribed interviews. In total, I conducted 20 interviews with 19 participants from the two institutions; one participant contacted me directly with follow-up information, so I interviewed him twice. I conducted 17 interviews face-to-face; the other three interviews were done via phone due to scheduling issues that did not enable meeting in person with the participants. For the three phone interviews and the one follow-up interview, I wrote and analyzed comprehensive field notes. I transcribed the rest of the interviews or used an outside company approved by my institution's institutional review board. Transcriptions resulted in a textual dataset of over 600 pages.

Publicly available media also informed my findings. For both institutions, I searched for texts and multimedia documentation that could fill in my understanding of existing learning analytics initiatives and the technologies they used. My searches discovered three narrative handouts, three institutional reports, five unique websites, and two sets of presentation slides. As I previously mentioned, I did not code this documentation. However, it did inform my contextual understanding of the institutions and their respective projects.

Data analysis resulted in a significant amount of codes, which I iteratively reduced based on their relevance and fit to particular categories. During a pilot project that informed this research, over 330 relevant codes emerged that related to five unique categories. Those codes provided some analytical focus for what was important to this project. Over 300 codes proved relevant to the findings herein; I used those codes to develop the eight categories I present in the following sections.

7.3. Making Sense of the Contexts

Establishing emergent privacy issues that relate to learning analytics, especially through a contextual lens, requires a greater understanding of the two institutions under study. I have structured this section to reflect the array of notable contextual characteristics that Nissenbaum (2010) deemed important in order to highlight data that helps make sense of my contexts. The point of this structure is not to provide a comprehensive audit of all learning analytics technologies, roles, activities, norms, and values in the contexts; instead, the purpose is to bring to the forefront relevant pieces of data in these sections when they interact with the technology of learning analytics and the flow of student information. Furthermore, by detailing contextual data, we can interrogate informational norms and contextual integrity with respect to student privacy at a later point.

Learning Analytics Technologies

Based on a pilot study, I was aware that both Saint May State University and Hammond University were building capacity for learning analytics technology. Furthermore, I knew they shared a single installation of a learning management system-based learning analytics application; I will refer to this application as "Spotlight" going forward. While interviewing institutional actors at Saint May State University, I also discovered another learning analytics initiative aligned with an advising unit; I will refer to this separate learning analytics technology as "Lighthouse."

Both Spotlight and Lighthouse sourced data from their respective institution's student information system, although in different ways. Spotlight was connected to a data warehouse that extracted and stored information from the student information system, whereas Lighthouse imported student information system data straight to the learning analytics application. In addition to student information system data, the technologies accessed other data to inform their predictive capabilities.

Spotlight was able to predict students' final grades in a course based on their activity and grades within the learning management system, academic preparation data from standardized test scores, and profile information extracted from the institution's student information system. Only instructors, system administrators, and project leaders could see the predictive scores, related analytics, and limited student profile information at the time of the interviews.

Lighthouse was able to predict the likelihood of students returning for their sophomore year and their probability of finishing their freshmen year with a successful GPA. These predictions were based on a standardized survey, which students took after enrolling for their freshman year. In addition to its predictive capabilities, Lighthouse measured students' readiness for college and assessed their academic risks using non-traditional assessments (i.e., assessments not related to academic readiness). Lighthouse presented this information in a profile that was, at the time of interviews, only visible to advisors; however, students could request access to their profile.

Roles and Activities

Data analysis revealed important themes with respect to institutional actors' roles and activities as they related to learning analytics initiatives. Specifically, many roles experienced evolutionary changes, while others experienced transformations; some roles were even newly formed. As a result, roles experienced activity changes to accommodate the needs of learning analytics projects.

Based on the data, I placed the actor roles into four groups: policy, leadership, development, and academic. The policy group consists of members of campus legal counsel, registrars, information officers, institutional researchers, and actors who work on institutional review boards. The leadership group includes academic deans and administrative personnel in academic technology. The development group encompasses data scientists, instructional technologists, and information systems administrators. Finally, the academic group comprises advisors and instructors.

As the names indicate, each group assumed an area of activity related to learning analytics. For instance, the data revealed that the policy group was responsible for creating institutional policy, making interpretations of federal law (primarily FERPA), and advising institutional actors with regard to their legal obligations as they build capacity for and begin to use learning analytics technologies. The leadership group provided the vision and direction for learning analytics initiatives, as well as worked to, as they said often, "get the right people at the table" to secure buy-in and input from other institutional actors. The development group worked to establish the technical infrastructure for learning analytics, including building and implementing predictive models, connecting student information systems to data warehouses, and working with vendors of learning analytics technologies to resolve issues and apply patches. Activities within the academic group centered on using learning analytics by instructors and advisors to help students become academically successful.

Changes to role activities emerged in response to institutional capacity building for learning analytics. Most notably, these changes occurred in different degrees for registrars, institutional researchers, academic advisors, and instructors. The effect of learning analytics on role activities is notable because it impacted how participants viewed their role in the institution and their workplace practices.

Consider the case of one of my participants who was the registrar at Hammond University, the larger public institution. To him, learning analytics and other data-driven educational technology systems held the potential to improve classroom experiences and "drive decision making on campus." The registrar's role in the past, both at his institution and as defined by his professional peers, had been to protect student information and limit access to it. But he saw his role shifting. As he said, "I have a responsibility to get the faculty members data to improve their programs, to look at new teaching methods and evaluate that program and that teaching method to see if it is successful." Beyond working with faculty, he saw his role and his office as acting as a "hub among spokes" to enable greater access to student data and education about how to use it effectively.

Though the registrar's role was evolving with learning analytics, for the institutional researchers the degree of change was more transformative and disruptive. In their work, these participants often had access to student data and statistically analyzed it to inform policy and institutional decision making. From their point of view, information officers, administrators, and others within and outside of the institution were pushing a learning analytics agenda that ignored or was ignorant of the analytical work their office already did. These actors, according to the institutional researchers, "seem to want to take credit" for analytics, but "they have no idea what they're doing." In response to the growing interest in learning analytics, one institutional researcher ironically stated that she was "happy to call everything that we do with student data learning analytics," even though she saw no difference between institutional research and learning analytics.

For an academic advising unit at Saint May State University, the smaller institution, a learning analytics initiative had greatly impacted advisors' day-to-day workload. In addition to its predictive capabilities, Lighthouse enabled early-warning alerts, which allowed instructors to flag students whose academic progress needed interventions by advisors. When instructors set flags, the system messaged advisors that they needed to contact the flagged students and then setup an appointment within 24 hours. In the past, these types of interventions would have been less frequent due to the more cumbersome nature of the instructor having to communicate the issue manually with the advisor, but Lighthouse streamlined the process. Over a two-year period, the faculty participation rate with the system increased nearly 700%, and the number of alerts sent to the advising unit jumped over 600%. In addition to responding to the interventions, the unit facilitated the testing for all incoming freshmen students in order to establish a baseline for the predictive scores. With a small mix of part-time and full-time employees, the workload and responsibility for those within the advising unit changed significantly due to learning analytics technology.

Learning analytics impacted the instructional practices of some instructors with whom I spoke. One instructor was keen on using Spotlight, but the system's reliance on a specific grade book structure, particular types of assessments, and other data would have required him to adjust too much of his instructional approach. "[Spotlight] required you to use [it] in specific ways that I found took away flexibility from how I like to manage my classes," he said. After reflecting more about these limitations, he commented that highly-structured classes with large student populations would naturally align with learning analytics, but at his institution, the "face-to-face personal interaction culture" did not mesh with this type of system and the instructional methods his peers employed.

Changes to role responsibilities were most obvious at Hammond University. For one interviewee, his roles and responsibilities changed entirely. He was initially hired to evaluate learning technologies and related projects, but administrators transitioned him to focus solely on learning analytics shortly after his arrival on campus. For another participant, he was hired into a new information officer position partly in response to institutional needs highlighted by learning analytics and other data-driven projects. In this role, he saw himself as "a cross between a chief data officer and… a chief analytics officer," and his responsibility, as he put it, was "to help provide better and more consistent access to institutional data."

Norms

On the surface, the data did not reveal any significant deviations from norms that one would expect at public, higher education institutions. In that sense, descriptive norms, or norms we would expect, at institutions like Saint May and Hammond remained stable; however, injunctive norms were emerging. Injunctive norms are expectations of how individuals should act, and the data signaled that actors felt an obligation to employ data-driven technologies, like learning analytics, at their respective institutions.

When asked what was motivating learning analytics at their campus, many interviewees indicated a growing cultural expectation for higher education to use data and technology to improve educational practices. The trunk of this expectation, they said, was rooted in observations that Big Data practices in other sectors (e.g., online commerce, personalized searching, and social matching in social media) could and should be applied in higher education. As one academic dean said, there was a growing "culture of technology-enhanced learning" driving interest in learning analytics. In combination with a "sense of why aren't we putting [Big Data analytics] to good use?," the pressures to use learning analytics have risen to the point that the technology ought to be utilized to improve post-secondary education.

Interviewees reported that a motivating reason to use learning analytics technologies had to do with accountability and efficiency pressures. "[There is] lots of pressure on higher education," said an interviewee with respect to gaining institutional efficiencies with less financial support. In order to do more with less, a number of respondents felt that analytic technology, like learning analytics, was an easy target. For example, one advisor said the following:

> We have less people power. We have less money. We have less resources. [And] we're trying to do more each as individuals with less capacity. And, you know, people almost immediately look to technology to see if there is a way that will allow us to serve students either at least as well as we were, [or] without degrading our capabilities.

Participants were uncertain whether learning analytics would actually resolve institutional stress points, but their perspective was that outside pressures to use a technological fix were motivating their institution's interest in learning analytics.

Values

The teleological orientation of the two institutions I studied and the actors I interviewed did not reveal significant changes in their values, or sets of "goals, purposes, and ends" (Nissenbaum, 2010, p. 134) with respect to learning analytics at a macro level. What I mean by this is that the generally accepted aims we expect from public institutions of higher education did not deviate from the norm at either Saint May State University or Hammond University in any major way. For me to make a claim that their respective values had changed due to learning analytics would require substantial evidence. For instance, there would have to be proof that major academic initiatives were changed or redirected in order to acquire more student data for analysis, or that the purpose of an institution was redefined due to data-driven education. This type of evidence did not, however, materialize in the data. With that said, the data did indicate that there were interactions between micro-level values and learning analytics at both institutions. Micro-level values, as I define them, do not impact the institutional mission or direction in any significant way, but they do motivate–to varying degrees–the practices of institutional actors.

Interviews with actors at Saint May State University showed value interactions with learning analytics in two ways. One, interviewees in advisory roles to students stressed a "student-centered" culture that could conflict with learning analytics technology. Two, interviewees in an instructional role stated that learning analytics did not mesh with institutional classroom structures and instructional choice. I detail both of these interactions in the following paragraphs.

Through anecdotes and further explanation, interviewees discussed how a "student-centered" culture enabled students, faculty, and staff to develop close interpersonal relationships. Instructors and advisors also conveyed that they often aim to understand students and their needs at a "holistic" level. This culture, they said, could conflict with some learning analytics systems, especially those that are limited in the type of data they can capture about a student and the ways in which they present data and information to institutional actors. One advisor said the following:

> I'm so cautious about a one-size fits all [learning analytics system]…. What we do is provide a buffet of things, and we make it our business to know the types of students, the types of services, ask probing questions, and make a list of things that could work [for students]…. [It's] hard to do that with analytics. A probing question may [help] them discover something that they didn't know about themselves before.

Advisors used Lighthouse, but only to inform themselves about student needs and communicate with students regarding resources and services that could best support their path towards academic success. While they could have used the predictive scores to come to quick conclusions about individual students, they indicated a heightened awareness of the negative effects this could have on student choice; instead, they only used the data as a starting point for conversations. Section 7.11. addresses these potential negative effects in detail.

For Hammond University, its use of learning management system-based predictive learning analytics matched well with its own institutional culture and desire to poorly performing retention and time-to-degree measurements. A part of this problem included getting undergraduate students through common pre-requisite courses with large enrollments, such as Introduction to Biology. In order to address these problems, they looked to Spotlight to predict which students would be academically at-risk (e.g., there was a possibility of failing the course) early on in the semester. The goal was to intervene in the student's low progress in courses to ameliorate failing or low grades. In that sense, learning analytics was well-matched to the goals of the institution, which worked towards graduating students on time.

With that said, Hammond University did encounter value conflicts with regard to how Spotlight could create student profiling concerns. Spotlight's initial design was not to limit access to student profile information, but members of the project at Hammond University deemed that instructors should not be able to see sensitive profile information, especially related to race, gender, and academic preparation. A data scientist explained the issue this way:

> The way that [Spotlight] is setup, it could drill down to a point where faculty could see certain parts of [the student profile], but for privacy issues it's basically all locked down…

> and probably for the foreseeable future until a lot of issues sort of get sorted out. Because there's fears of profiling.

Hammond University's choice to limit access to student profile information within Spotlight affected Saint May State University, since they shared the same installation of Spotlight; Saint May, according to actors at Hammond, did not express a concern about the information. I interpret this difference in approach to the student profile information as a value discrepancy between the two institutions with regard to this specific learning analytics application.

7.4. Raising Privacy Flags

Research participants recognized that student privacy concerns were emerging as they built up capacity for learning analytics. These moments of recognition are captured by, as one interviewee put it, the concept of "privacy flags." My analysis of the interview data indicated that two types of flags existed: policy flags and intuitive flags. On one hand, participants raised policy flags when learning analytics presented privacy issues that interacted with institutional policy and federal law. Intuitive flags, on the other hand, went up when participants felt instinctual responses to learning analytics practices that presented privacy problems. The following sub-sections consider two specific policy flags and three intuitive flags.

The First Policy Flag: Needing Legal Advice

Actors were unclear as to what their institutional obligations were with respect to student privacy due to the "bleeding edge" nature of learning analytics, which raised a privacy flag related to institutional policy. To gain clarity, those directly involved with the leadership and administration of learning analytics projects sought the advice of their legal counsel and institutional review boards. According to one administrator, he was not sure if his project needed institutional review board approval, an exemption, or a complete review. Another participant summarized the ambiguity surrounding what governed learning analytics projects as this: "The question is are some of these learning analytics [initiatives], are they research projects? Are they administrative? We're trying to improve our own courses; that doesn't fall in the realm of research." Participants at Hammond University explained that upon deliberating with their legal counsel and institutional review board, they discovered that learning analytics projects worked under specific FERPA exceptions, with some limitations.

The most pertinent, and oft cited, FERPA exception concerned §99.31(a)(1)(i)(a). The exception allows institutions to disclose student information as long as "the disclosure is to other school officials, including teachers, within the agency or institution whom the agency or institution has determined to have legitimate educational interests." The freedom of this exception allowed institutional actors, and those with "school official" status (e.g., technology vendors like Spotlight and Lighthouse), to collect and use student data for learning analytics purposes, as long as those with access had "legitimate educational interests." When programs meet the requirements of this exception, institutions and their representatives do not need to inform nor gain consent from students about uses of personally identifiable data.

Even though the law clearly details the legitimate educational interests exception, the leadership involved with learning analytics programs still pursued guidance from their institutional review boards to see if there were other privacy obligations of which they should be aware. Institutional review boards, participants stated, were concerned with whether or not learning analytics projects on their campus qualified as research. If they did, they would fall under stricter regulations in order to protect research participants–the students. To qualify a learning analytics project as research, data collection, use, and analysis would need to be done with the intent to create "generalizable knowledge" through dissemination of results (see 45 CFR §46.102 for more information). As this was not the motivation driving learning analytics initiatives, institutional review boards, as one participant stated, categorized them as "institutional improvement and program evaluation" and "tool development" projects. As such, they were not affected by institutional review board policies and the review board gave learning analytics projects a "clean bill of health" to continue their development without requiring additional privacy protections.

The Second Policy Flag: A Need to Know

Learning analytics systems, by the nature of their design, often capture more private, personally identifiable nature about students than was possible with past educational technologies. Moreover, these systems also aggregate sensitive data from student information systems and, possibly from other data sources as well, to build towards analytic ends. As a result, learning analytics technologies hold the potential to provide greater access to student data to a wider swath of actors within and outside higher education institutions. To my participants, especially those involved with Spotlight at Hammond University, greater access to student data raised a policy flag.

Participants often argued that access to student data was determined by a "need to know" policy. This policy was a specific interpretation of FERPA's "legitimate educational interest" clause (§99.31). A member of Hammond University's legal counsel explained the policy this way:

> FERPA has an exception… [that says] school officials have a legitimate educational interest in the information. Now that doesn't mean anybody on campus can access all information. And you may have a legitimate educational interest in one project but not

> another…. So, there's a tendency to think that just because you work at the university you have the right to see everything. That would be the furthest thing from the truth.

For institutional actors to earn access to certain types of student data and information, they would have to "make the case" for needing to know that information.

With regard to Spotlight at Hammond University, administrators of the learning analytics initiative made a proactive determination that instructors did not have a need to know all of the student profile information the system displayed, such as a student's race. About this, one participant said:

> Our argument is that an instructor does not need to know a student's race in order to tell you, like, that you're not doing as well as other people…. That's a concrete privacy issue. FERPA says an instructor does not need to know a student's race or ethnicity to instruct them.

As I detailed in section 7.3., the administrators in charge of Spotlight directed the vendor of the technology to essentially hide this unapproved student information from an instructor's view.

While participants at Saint May State University did not use the words "need to know" to express a policy related to student data access concerns, they employed a similar solution to their peers at Hammond University. For Lighthouse, Saint May State University determined that instructors did not need and should not have access to student profile information that was imported from the student information system into the learning analytics interface. This was complicated, however, when analytic models used the information for their predictions but kept it hidden from the view of institutional actors.

Participants did note that although instructors did not have a need to know all possible fields related to a student's profile information, the "edges" were "blurry" with regard to predictive models given changes in information visibility. In this case, student profile information was not limited when included in the process of designing predictive models, because instructors did not have access to the raw information–even though they would have access to the processed information as a predictive score.

The First Intuitive Flag: Changing Norms and Finding Balance

Since learning analytics initiatives were not strictly regulated under federal law or the rules of the institutional review board, there existed significant flexibility for institutional actors to acquire and use student data within the limits of the need to know policy. Regardless of these freedoms, many of my participants expressed a sense of paternalism to protect student privacy from unwarranted violations and potentially harmful disclosures. However, their perception of changing norms related to personal transparency and expectations of privacy among students worked against this belief.

Participants expressed frustration that their efforts to protect student privacy were not appreciated by the students and were potentially unnecessary. One interviewee expressed this sentiment, saying:

> And then there's part of me that thinks, well, we worry a lot more about this than the students do. You know, we're worried about revealing their race or their gender or their prior [grade point average] and all this stuff that might be important in some kind of learning analytics project. But they're putting those things out on Facebook. You know? Things, that, you know, we're trying to uphold laws and stuff, and they, sometimes the students seem less concerned about the privacy than, I think, we are about it.

The reference to Facebook was common. Interviewees often invoked specific social media companies and adoption thereof by students as an indicator of changing norms towards personal transparency and openness over individual privacy.

Interviewees said that students did not "see privacy the same way" that they did as representatives of higher education institutions. One registrar discussed a new initiative in her office to release student photos in online class rosters, which raised privacy concerns among her peers, but when she discussed the rosters with students they mocked her. "When I met with students," she said, "they laughed at me. They said, 'this is what you guys talk about? This is what you worry about, really?'" The registrar observed that pictures of students are "everywhere, they put it everywhere," and "they don't care [if it is accessible]; many don't, I shouldn't say all. But many don't care." Similarly, a data scientist felt that "students… could [not] care less if you're tracking them. They just don't care. They've grown up in a world where that's normal."

Given their understanding of changing norms related to information sharing, a few participants questioned whether it was time to rebalance their student privacy obligations to be less protective of student privacy. An instructional technologist expressed that a failure to reassess his institution's paternalistic approach to student privacy could be "dangerous." He said, "We have to be careful, but at the same time not overly so." Continuing on a path that conservatively addresses student privacy, he argued, "stands in the way of innovation." His peer in advising argued that institutions should start out conservatively with data-driven projects like learning analytics but be willing to push the boundaries of what is acceptable and useful regarding data practices if the institutional community–students included–agreed to do so.

The Second Intuitive Flag: Encountering Big Brother

My participants raised a second intuitive flag concerning learning analytics technologies and surveillance capabilities, which they characterized as being “Big Brotherish.” Interviewees elicited George Orwell’s famous character and described the technological affordances that enabled surveillance practices as “creepy,” “scary,” and “freaky.” The core of their anxieties centered on the newfound ability institutions gained to track student behaviors, both in physical and digital spaces. At Hammond University specifically, several participants indicated a greater institutional awareness of this concern, citing conversations where their peers expressed “angst” about learning analytics data use and a desire to present the institution as “anti-Big Brother.”

Participants recognized that a core tenet of learning analytics was the technology’s ability to observe what students did and where, especially when applications, like Spotlight, are embedded in learning management systems. That this feature set was common to this genre of technology and central to its efficacy did little to assuage their concerns. To one participant, the fact that learning analytics enabled institutional actors to surveil students created a “path” for even more concerning practices. He said:

> It gets creepy for me when we start talking about the predictive analytics that are based on the student's behavior in a learning management system. Are they really not going to the library? Are we going to scrape their cookies from their computers when they log into the system and see what other websites they’ve been looking at and how much time are they spending? Things can go really, really bad.

To this point, a number of participants commented that they felt a responsibility to limit data aggregation and more invasive surveillance practices, because, as one administrator put it, “the more [data and paths] you add in, the more it starts to feel… like Big Brother.”

Some participants expressed that they could limit Big Brother concerns by drawing boundaries around the ways their institution accesses data and the types of information to which it has access. For instance, a registrar argued that he sees information and data differently depending on whether or not a student actively provides it or the institution acquires it using tracking mechanisms. In explaining this point, he said, "if [students] provide information… they've given us that data, versus [students who] have an ID card and [the institution is] tracking every time you're doing something; I see those as different." Participants also argued that there is a set of data that has been traditionally considered to be academic information, which is "fair game" for learning analytics purposes. This "classic array" of information refers to academic performance, such as grades, and information students provide on their application for admission to the institution. Data and information outside of this boundary elicited Big Brother concerns.

Participants pointed to data deidentification as a way to reduce Big Brother worries. From their perspective, legal counsel participants suggested that deidentification of learner data would relieve the institution from considering outstanding privacy issues and concerns related to policy and law; moreover, it would allow for more freedom with the data, such as potentially allowing the institution to release data to third parties. The problem with deidentifying data was, as an information officer recognized, that it weakened the efficacy of learning analytics. When reflecting on deidentification opportunities, she had this to say: "This is about success. In my view, it boils down to analytics [being] used to help someone be more successful, to graduate, to learn. So, you can't decouple that." Instead of deidentifying data, others argued that surveillance concerns could be lessened by simply using the data with forethought. As one administrator put it, "I think the rubber meets the road in terms of how you use the data… we've always, in a sense, 'tracked' students." If data use was appropriate and justifiable, as some participants argued, then institutional actors and students, they assumed, would have fewer privacy concerns, but the data would have to be used with "benevolence."

The Third Intuitive Flag: All the Data We Can Get Begets Privacy Problems

Learning analytics, my participants often indicated, represented a style of Big Data technological practice. Consider the following interaction with an information officer when I asked her if she would pursue data outside of the learning management system and student information system for analytic purposes:

> Absolutely… I mean there are things that you can correlate together from residence halls, from how people are using their time. Can you associate that with greater learning? Timing of things, when are people doing certain things. Is it all last minute? Is it orderly? Does it occur in relation to assignments? There is so much where you can establish correlations, that in my view, this isn't just the two systems, it could be many other systems.

This perspective pushes forward the idea that an institution's learning analytics initiatives, and learning analytics in general, would benefit from capturing more student data from a wider swath of sources. Similarly, a data scientist enthusiastically commented that he would like access to data from "any technology that a student ever touches" on campus to import "as much data as possible" into algorithms.

The motivations behind pursuing greater amounts and sources of data stemmed from a belief that, as an interviewee said, acquiring "any and all the data we can get" can "add more value" to learning analytics. As one information officer put it, "the more data points we can [get] … the better predictions we'll make." More powerful learning analytics, some participants argued, would augment personalized learning platforms and help create more effective interventions, which rely on stronger predictive models and correlations. Specifically, data that enables institutional actors to "see" how students interact with campus services and the wider campus community (e.g., involvement with student organizations, student employment activity, etc.), may help them understand more of the "informal learning that goes on on campus."

Participants raised an intuitive flag, in this case, when they reflected on the privacy implications of pursuing all the data they can get. For instance, one of the problems interviewees recognized is that they felt there were no regulations limiting their vision for data aggregation and future practices; and to one participant, "that's where you can really start thinking about some scary things," such as how individuals may use large troves of data to profile students. Without defined regulations, institutional actors would be left to make ethical decisions based on their own moral compasses. Relying on personal ethics to use learning analytics data appropriately, however, was seen as problematic. Instead, one participant argued that those pursuing learning analytics needed to abide by a code of ethics; however, even that had issues. About this, an interviewee stated, "We all know that, just because there's an ethics code established, not everyone is going to adhere to it…. Do no harm. But we all know there's going to be people that don't do that."

Some participants argued that data collection should be restricted. Data restrictions, participants argued, should be based on a "rationale line" determined by "reasonable hypotheses" supported by the literature. Instead of accepting a Big Data ethos that encourages trawling for data and any available insights, these participants felt that more directed learning analytics practices would be more practical and less laden with potential harms. They argued that reducing the need to gather all available data and providing justifications for the data institutions do analyze would reduce privacy concerns.

7.5. Needing a Data Governance Strategy

A sharp contrast emerged between Hammond University and Saint May State University with regard to data governance. Both institutions lacked data governance strategies, which is to say they did not have institution-wide policies and procedures to protect student data and enable its use for learning analytics and other large-scale data-driven projects. At the time of the interviews, actors reflected that anything related to data governance was often done in an ad hoc fashion; furthermore, actors involved with data governance decisions were often limited to those seeking access to data and data "gatekeepers," such as registrars. The contrast was made clear when I asked if actors knew of or were actively involved in new processes to establish data governance on their respective campuses. For Saint May State University, the answer was often "no." However, actors at Hammond University, on numerous occasions, expressed that a lack of data governance was a significant problem with regard to data integrity and student privacy. And that due to this perceived problem, they were actively building a data governance strategy.

The sub-sections below address data governance at Hammond University, unless otherwise noted. Specifically, I consider themes related to motivating factors spurring on data governance. I then discuss related problems that interviewees implicitly or explicitly tied to a lack of data governance. Finally, I consider some of the data governance goals Hammond University actors expressed and the actions they had taken to resolve outstanding issues.

A Million Questions

Building capacity for learning analytics technologies highlighted for actors at Hammond University how little they knew about their current data practices and what would be required to push forward with more expansive data-driven initiatives. As the university's chief information officer said, “We have a million questions that come up, and we don’t really have a good place to go for answers.” Regardless of the fact that the institution had experience with data warehousing, that experience did little to guide new data practices, especially those, like learning analytics, that include new sources of sensitive data.

The questions challenging actors at Hammond University surrounded a growing interest among academics and institutional staff to use student data for research and evaluative purposes. Due to this, there was uncertainty as to how interested parties should be granted access to data, since there was no explicit policy. Furthermore, actors had outstanding questions regarding whether or not boundaries should be drawn around student data to limit potentially harmful outcomes, especially given the existence of digital breadcrumbs and the ability to track student movements online and on campus, for example but using student ID card swipes. To this point, a technology administrator said, “that could have real privacy implications, [and] it’s unclear who’s in charge and who’s responsible.”

The questions participants had with regard to data governance, they commented, reflected profound challenges to current data practices. New technologies, data uses, and pressures to pursue mechanisms that support data-driven decision making stressed existing policies and practices. Additionally, as an advisor stated, information technology leadership on campus is pursuing new “governing bodies [and] new committees…. at just the right time when we’re maturing in other aspects of our [information technology] and data thinking.” In that sense, the fact that important policies and other data governance mechanisms were missing was not seen as a dire situation, given the institution was in the process of developing governance mechanisms and policies to, in part, protect student privacy.

A Known Issue

Actors at Hammond University identified problems with which they were struggling, some well-understood and others not. Of those that they were aware of, data stewardship proved to be one at the top of their minds, especially among data scientists and instructional technologists who used student data in learning analytics technologies. When I asked a data scientist what issues he was facing with regard to data governance, he explained, "data stewardship… it is a known issue. We knew we were going to come up against it." To him and others, data stewardship agreements enabled them to use student data in their projects, but these did not exist in any useful way. That is not to say that data agreements did not exist, but instead that such agreements were created on a case-by-case basis or that they existed at multiple levels, such as at individual colleges or departments. As a result, data stewardship issues were fleshed out in "the small silo of individual initiatives," but not in a more generalizable ways that can inform a larger-scale project like learning analytics.

Previously, I detailed in section 7.5. how interpreting FERPA allowed for and limited certain uses of student data. This sub-section moves away from the content of the interpretations to the question of who made the interpretations, which proved to be a lesser understood but still problematic issue for actors. Often, participants called those who made interpretations and held control of data "gatekeepers," "czars," "stewards," or "custodians" of data. And they included a range of institutional actors, including members of legal counsel, registrars, staff on the institutional review board, and those who held custody over special troves of data, like human resource managers.

At Hammond University, the registrar was often invoked in these conversations and described as *the* "gatekeeper" over student data. When I asked him about his role, he responded that he saw it differently than his peers, saying:

> "Gatekeeper" to me makes me think you cannot come in unless you have the secret password…. I think it's my responsibility to put [student data] out there safely, responsibly, and make sure the right people have access to it so that we really improve the educational experience of students.

As his remarks indicate, stewardship of student data differs from gatekeeping in the sense that the former increases access to data, while the latter restricts it. "Traditionally," he said, "[registrars have] been seen as gatekeepers," which interpret FERPA strictly and, as a result, restrict access to student data. As a steward, he takes a more liberal approach to FERPA that, in his mind, still upholds privacy responsibilities but also does not restrict the flow of data to those who can benefit from it.

Regardless of the registrar's perception of his role, other participants characterized him as a gatekeeper. To these individuals, the registrar had the "ultimate say" in whether or not data could be used for learning analytics. And given the lack of a guiding policy with regard to student data use, it was easier for them to go "straight to the top" to the registrar to determine whether or not his interpretation of FERPA allowed them to make use of student data and under what limitations.

Participants sought permission and guidance from the registrar-cum-gatekeeper in part because he held dominion over student data and made the ultimate decision regarding FERPA, but also because they felt disinclined from making their own interpretations of FERPA. For instance, one participant pointed out that he "offloads" privacy concerns to "the very protective people," such as the registrar, because they understand FERPA and work in a decision-making capacity with regard to the law. Another argued that the interpretive work around FERPA with regard to student data use was so political that she tries to remove herself from those conversations. As a final example, one advisor observed that her colleagues were "afraid of FERPA" because they feared a misguided interpretation of the law could spark a lawsuit that would find them personally liable.

Growth Areas

Given existing gaps and problems, actors at Hammond University, especially those in a leadership role, expressed "growth areas" where work needed to be done to resolve outstanding issues related to data governance. One area of development included the addition of a new role to lead data governance on campus, and another included increasing training and awareness related to student data practices.

The breadth of problems related to data governance on campus motivated leadership to seek help in this area by adding new personnel. About midway through interviews, participants alerted me to a new hire, an additional information officer to help with "focused efforts" on data governance, especially with regard to increasing access to institutional data. I was able to interview this new information officer, and I asked him what he was tasked to do in the area of data governance. He explained that there were "four pillars" of data governance the institution had tasked him with: (1) establishing decision rights around data access; (2) developing a stewardship program to classify data; (3) maintaining privacy and security compliance; and (4) working on a technical architecture to support data-driven programs.

In addition to the four pillars of data governance growth, actors in leadership roles also expressed that education around data responsibilities and training with regard to FERPA and other privacy policies would become an important consideration. Training programs already existed in relationship to the student information system, but participants argued that more needed to be done. The increasing amount of data-driven programs that use student data, like learning analytics, and access to various types of student data required broader programs less focused on particular systems and more focused responsible data use, understanding of institutional policy, and awareness of FERPA's rules.

7.6. Educational Records in Flux

With the increase of personally identifiable data due to learning analytics, I asked my participants if the data Spotlight and Lighthouse aggregated, analyzed, and produced complicated the definition of an educational record. Data analysis indicated that interviewees expressed a variety of views in this area. While it may seem that defining what an educational record is would be a straight forward exercise, learning analytics introduces new challenges with regard to federal student privacy law.

The following themes consider various participant interpretations of what an educational record is, especially given the emergence of learning analytics on these respective campuses. I follow this thematic description with a conversation concerning rights student have with regard to educational records and conflicts therewith due to learning analytics technologies and data-driven education. Finally, I address participant expectations related to whether or not students would even be motivated to pursue their right to access educational records that include learning analytics data.

The Definition of "Educational Records"

A seemingly simple question–what is an educational record?–garnered two distinct definitions; I categorize the first as a strict definition, while the second is interpretive. The strict definition, for participants who promoted such a view of educational records, hinged on FERPA and how the term was defined in law. FERPA §99.3 defines educational records as records that directly relate to a student and are maintained by the educational institution (or other parties acting on their behalf in an official capacity). There are six specific exceptions that apply under certain circumstances, but they are not relevant to the current conversation.

When asked to define educational records, a registrar and a campus lawyer at Hammond University quickly cited the FERPA definition. The exchange with the lawyer went as follows:

> *Kyle*
>
> That was one of my questions, which is how do you define an educational record. And it sounds like…
>
> *Lawyer*
>
> [Interrupts] We don't define it. FERPA defines it. It's pretty clear. FERPA says it's essentially anything about a student while they're matriculated here at the institution from the time they matriculate.

Similarly, the registrar referred to the "exact definition" provided by FERPA. To these two participants, the definition was clearly stated in federal law, and the law was what they followed. By adhering to the law's definition, these actors believe all personally identifiable data, excluding that which falls under FERPA's exceptions, became a part of a student's educational record.

In contrast, other actors, including the registrar at Saint May State University, interpreted the legal definition differently. These interviewees drew a boundary around educational records, suggesting that educational records concerned data and information "directly related to the student's education." The actors believed that only when the data had a clear relationship to a student's educational experience and learning did it then count as an educational record. When asked if other data used for learning analytics, like digital breadcrumb data borne from student interactions with other systems or tracking data gathered from student ID usage, was also considered to be a part of an educational record, they felt it was a different type of record outside of FERPA's definition and protection. As a specific example, actors at Saint May State University felt that the predictive scores that Lighthouse produced about students were not academically related since, as an academic dean put it, they had no relationship with a student's "behavior in the classroom [or] the educational environment."

Data Rights and Conflicts

Questions about counts as a record is important because it implicates student privacy rights protected by FERPA. For instance, if learning analytics data is considered, per the strict definition, to be a part of a student's educational record, some participants expressed concern that their institution would not be able to honor FERPA §99.10, which refers to a student's right to inspect and review her record.

About this conflict, a data scientist was adamant that while students should have access to the learning analytics data they create and that which is created about them, "there is no way in the system for them to see it." According to §99.20 of FERPA, a student has the right to if she believes it is "inaccurate, misleading, or in violation of [her] right to privacy." So, if students cannot see data about themselves, then they are limited in their ability to amend their educational record. The data scientist, who was familiar with the design of Spotlight, explained the technical aspects of the issue this way:

> You can't go in and say "remove me from [Spotlight], remove all my records." It just can't happen. It is not scalable. Now, if it's right or wrong is another story…. The technical infrastructure is so complicated [that] to be able to isolate a record and remove it, [there are] so many dependencies on the system.

Not allowing students access to their records may not be a legally acceptable situation. A lawyer for the data scientist's institution argued that the technical issues were not a "legitimate excuse for the institution" not to respect student privacy rights provided by FERPA. "If people have an end product about them," he said, "we generally have to provide [the data] to them." According to the lawyer, there were exceptions, however, that contradicted this statement. First, if the identifiable data could not be separated from data about other students, then the institution would not have to provide it as part of an educational record. Second, if the data access and retrieval problems were such that no student would be able to gain access to their data when they requested it, then, he believed, the institution would not have to provide the data.

Reviewing Student Records

Regardless of the technical limitations and possibly legal requirements, participants questioned why students would want to access, review, and possibly amend their educational records inclusive of learning analytics data in the first place. For instance, if a student felt that learning analytics data, as part of her educational record, violated her privacy rights per FERPA §99.20, she might ask to have her data removed. But to an instructional technologist, that seemed "scary," because students, he argued, would remove themselves from a data-driven process meant to help them with their educational progress. Beyond this specific concern, participants wondered openly what would motivate students to examine learning analytics data.

Participants questioned why students would want to see analytic products about themselves, specifically metadata describing their digital breadcrumbs and the underlying predictive models. For instance, a registrar said, "I don't know if it would be feasible to show them the predictive models, [and] I don't understand why they would care…. I just think students would be more interested in the raw data" and how instructors ultimately assessed their learning.

Here, so-called "raw data" refers to learning inputs (e.g., assignments) and assessment outputs (e.g., project grades) that are untouched by learning analytics platforms. Another participant questioned if students would want to see metadata of their digital activities. "They're going to want to see 'what is it saying about me?,'" she said, explaining that what students will want access to is the final product, like a predictive score, and not the data that informs it. In that sense, my participants did not agree with the perception of some learning analytics advocates that students would want or benefit from systems that supported quantified self practices.

7.7. Questioning the Big Data Ethos

While a number of participants expressed that obtaining "all the data they can get" for learning analytics would be optimal, others recognized that this Big Data ethos was problematic. In fact, as one participant said, it was creating a "tail wagging the dog" effect, whereby the push to adopt data-driven educational technology was affecting practice and the law that governs it in ways that were consequential for student privacy. Two primary themes emerged in this area. First, the desire to access, aggregate, and analyze various troves of data about students from on and off-campus sources brought up important questions regarding the legitimacy of their inclusion in algorithms. Some participants argued that trawling for any and all data might not be justified or even necessary. Second, participants pointed out how pursuing more student data as a Big Data practice stressed the intent of FERPA and its student privacy protections. They worried that the law was inadequate and susceptible to politically-motivated amendments that work against its purpose to protect student privacy. I provide detail for both of these themes below.

Questioning the Legitimacy of Including All the Data for Learning Analytics

When discussions turned to considerations about new data types, sources, and opportunities for analysis, participants expressed some concerns about Big Data's popular ethos that seeking more data for analysis is the most advantageous pursuit. As previously discussed, this driving philosophy was also quite dominant among some participants who saw massive data aggregation as a key practice for learning analytics; however, others questioned this approach as problematic. Seeking all available student data highlighted, as one participant put it, uncritical thinking that exposed institutions to both practical and ethical concerns.

Aggregating all possible student data could "swamp" institutions both in terms of the amount of analyzable data and their ability to capitalize on the insights it could contain. At Saint May State University, an instructor expressed that his institution did not have the capacity to handle any data project beyond what they were already doing; they were already struggling to establish business intelligence systems as it was. If his institution continued to push forward in Big Data style with learning analytics, he said, they would "find themselves overwhelmed with data and the inability to function." To this point, an institutional researcher opined that more data would require her and her peers to "wade through" all of it to make sense of the data, conveying that it would limit their ability to make timely data-based decisions and recommendations.

Another participant was adamant that an "all the data we can get" approach should be tempered, as it promoted aggregating sensitive data and prompted decision making without asking specific questions of the data. "Frankly," the advisor said, "[we need to] be conservative first, you know, and see how things go, because the power of iterative assessment would be our best friend in these kinds of things." But asking questions before aggregating data runs counter to the spirit of Big Data, and participants argued that their institutions should respect the scientific tradition and its emphasis on research questions. To this point, a participant said:

> To me, any analysis is only as good as the data that you put into it. And if you're just putting in what's available and deciding those things are important, [that] doesn't seem to be basing the data that they're interested in on previous research and that doesn't show that that stuff is important.

The Big Data approach especially made participants with a research background apprehensive, since their perception was that it did not promote standard principles of scientific research, such as stating hypotheses, asking research questions, and using previous research to inform the design of analytic algorithms.

The validity of some data types was especially concerning to a few participants. One argued that popular data for learning analytics, like timestamps, may not be the right proxy for engagement or other behaviors; yet, these types of proxies were often included in learning analytics technologies without justification. “Data is not infallible,” she commented, “I’m very skeptical that the magic bullet is going to be found in data for learning analytics.” Instead, another interviewee argued, “We don’t need the entire ball of wax. We can actually be much more strategic and explore the predictive power of different variables and be successful,” and in doing so not “dip into the space where you might involve privacy issues.”

Lag, Loopholes, and Swallowing the Rule with Regard to FERPA

The predominance of FERPA in conversations regarding student privacy and learning analytics brought to the fore particular concerns regarding the law's ability–or lack thereof–to meet the emerging technology's challenges. Outside of lawyers and registrars, who were most familiar with the law and its amendments, participants in other roles stated that the law was "really old" and "built in a time prior to this space," arguing that its construction could not hold up to technological advancements and new data practices.

The incompatibility between learning analytics and FERPA's compliance requirements was one important problem. As mentioned in section 7.8., there was a concern regarding an institution's inability to isolate and retrieve individual student data from learning analytics technology for students to review. Since participants argued that the technology could not comply with FERPA in this regard, one instructional technologist felt there would be "political backlash" from colleges and universities if the Department of Education tried to enforce compliance. Instead, he believed, FERPA would simply have to change to reflect the needs of data-driven education, saying, "This is a new reality…. Learning is becoming more and more mobile. It's becoming more independent. It's becoming more personal. That all leads to more digital exhaust. It's not an issue that's going away."

Conflicts between learning analytics systems and the law highlighted to my participants how technology often outstrips the laws that govern it. To this point, an information officer expressed that "you're always going to have a lag" between law and technology, and the expectation was that, in the case of learning analytics, the former will adapt to the needs of the latter. If FERPAs stays as-is, it will simply be "ill-equipped" to handle the needs of modern technological development in higher education.

Beyond the incompatibility issues, participants stated that FERPA provided higher education freedoms to use student data at the expense of privacy, and institutions needed to make local decisions to fill in the privacy gaps. In regard to this, one advisor had this to say:

> I think FERPA's too broad. And, it allows the institution, frankly, too much wriggle room for what it can or cannot do or should or should not do. I think that either our laws, our

> policy, or both need to be more robust around this landscape when we start talking about getting data on students…. And, if the law's going to take forever to catch up, we can do the thinking and we can put in place institutional policies that have the same weight here as law.

In fact, the recommendation to shore up the perceived "gaps" and "loopholes" in FERPA with institutional policy was a common one. Yet, participants often shared that their institutions had had not followed through on this oft-cited suggestion, regardless of the fact that they saw it as their institution's "responsibility" to be protective of privacy in areas where the law was weak.

A particular area where actors considered the law to be inadequate and in need of institutional protections concerned the difference between student-provided information and system-observed behavioral data. An institutional researcher argued that FERPA's construction lent itself towards protecting data that students provide to institutions, such as on admissions applications and for institutional directories. But it was not "very explicit or well-positioned to deal with," much less protect, observed behavioral data, which includes digital breadcrumbs captured by analytic technologies as students interact with systems.

Participants familiar with the FERPA discussed recent amendments to the law. These participants noted that the changes, in their opinion, increased privacy protections with respect to emerging data types, like biometric data. However, they were also worried about other amendments that, generally speaking, increased the amount of student data and the rate at which it flows to actors within and outside the institution. Where actors within the institution were concerned, a lawyer said, "I mean it's weird. [The amendment] makes it easier for people to get access, but it complicates the university's rule because student data is flying out at a rate faster than most people are usually used to or comfortable with." Where actors outside the institution were concerned, participants expressed that a recent change to FERPA shifted the law away from privacy in an effort to promote accountability measures in ways that, as one participant put it, were "embarrassingly incorrect."

In combination with incentives built into the 2009 American Recovery and Reinvestment Act, one amendment to FERPA enabled states to create longitudinal databases and forced public higher education institutions to participate, which was particular concerning to participants. Differing opinions emerged among interviewees about the value of longitudinal databases, but they were uniformly troubled about how such databases would impact student privacy. A registrar conveyed a sense of conflict, saying that there "could be good intentions" motivating the aggregation and analysis of longitudinal student data, but state actors would need to justify what data they collect and the purposes to which they put it. But from a lawyer's perspective, longitudinal databases were antithetical to the purpose of the law, saying it "more than swallows the rule." He said:

> We have to provide for it…. We have no control over what [the states] are going to do with the data. And we're telling students it's getting shipped off, so it is what it is…. I know some people are really happy about it; there are some people who aren't really happy about it…. It is an exception, so there's not a lot we can do.

In this case, FERPA's amendments increased data practices and enabled greater data flows, while at the same time limiting what exactly an institution could do to protect their students from potential downstream effects.

7.8. Telling Students and Being Transparent

The role of transparency in decreasing privacy concerns thematically emerged in my discussions with participants often and in surprising ways. They often characterized transparency, as one actor put it, as a driving force that would make learning analytics practices understandable to students. The actors with whom I spoke characterized transparency as an act of broadcasting information about learning analytics to students. Conversations with participants brought forth a number of themes regarding transparency and learning analytics that directly and indirectly addressed student privacy concerns. Predominantly, interviewees hoped that transparency would allay students fears and build trust with students. Actors also described a number of transparency strategies that actors thought could accomplish related goals; however, these were often aspirational. Very little evidence emerged from conversations indicating that the institutions were systematically pursuing means by which they could achieve transparency about their learning analytics initiatives.

Telling the Students

Students were largely unaware of learning analytics initiatives of their respective campuses. Their lack of knowledge stemmed from the fact that actors at Saint May State University and Hammond University use the technology primarily as an informative tool to inform instruction and advising. Participants believed, however, that students would and should become more aware of learning analytics in the future. By telling students about learning analytics, actors believed they could address emergent privacy concerns, get student buy-in, and justify data-driven practices.

Interviewees indicated that as capacity for learning analytics grew and developed a higher profile on campus, student awareness would increase and student privacy complaints could increase. An administrator believed that student perception of learning analytics could go one of three ways, saying:

> I think it's going to be interesting. I would suspect that there would be a split, where I think a large majority of students will go, "Ah, cool." And I think there will be some students who will go, "Huh, well, yea, alright, whatever. I don't know." And then I think there will probably be a minority of students who will feel somehow threatened by the process.

This participant's perception aligns with others who expressed that students might not care about privacy to the extent that they do. But to allay the concerns of the minority group and others who developed privacy concerns, he felt it was "important that the messages are clear as to what's happening and what's not happening" with regard to learning analytics and student data use. Similarly, others argued that transparency could be a useful strategy to "counteract" concerns and "iron out" problems as they emerge.

Actors expected that transparency about learning analytics would also encourage students to buy-in to data-driven practices. To this point, a lawyer said, "If you tell people that you're using [learning analytics], people are more inclined to be like 'Ok, cool, that's kind of fun." By sharing with students what the technology is, what the motivating purposes behind its adoption are, and how the institution feels it is beneficial for improving the educational experience, the argument often went, students will "accept the reasoning for it," even if the benefit was more for the institution and future incoming classes.

Participants argued that being transparent about learning analytics would provide an opportunity to justify why the technology was useful for the institution and an important part of the educational process. An information officer asserted that transparency should be proactively used as a message to prospective and current students alike. He argued that by doing so, his institution's instructors would communicate their intention "to be the best educators [they] can be" and that "[Learning analytics] is one of the things that we do to… make that happen." Using this rhetoric as well, other actors argued that students need to be made aware of the fact that they are participating in a "larger effort" to help improve education and student learning outcomes.

Establishing Trust

Participants presented transparency as a method by which they could establish trust with students regarding learning analytics and other data-driven practices. In interviews, actors articulated their belief that students do not trust the institution until it has proven itself worthy of their confidence and believe that the institution will use their data responsibly towards worthy ends. About this, one information officer had this to say, "We need to be transparent. Students don't know us from Adam. You know? We're just a bunch of old people who are holding on to their stuff, and they don't trust us." Transparency, then, was seen as trust-building exercise.

Participants understood that a misstep by the institution with student data, either due to a blatant privacy invasion or a data breach, could negatively impact the trust students had with their university. Individuals in instructional and advisory roles often emphasized that they worked hard at establishing trust in their relationships with students; they did so to glean information from students and better serve their needs. To this point, an advisor said, "Once that kind of bond of trust initiates, [students] start providing data and information about themselves automatically." But not everyone valued trust in the same way.

While advisors and instructors argued that trust-building was valuable for their roles, they admitted that some of their research-oriented peers did not consider trust with students to be as important. And so they felt that a lack of trust between students and institutional actors created "critical issues" deserving of institutional effort. The institution, they argued, needed to "constantly show" that it cared for the welfare of its students, was dedicated to their education, and could provide students resources to help them succeed. A transparency agenda, they felt, could help them do just that.

A number of interviewees said that being transparent about learning analytics would seed important, educational conversations with students as well. These conversations would, they hoped, help to shore up "faith boundaries" around privacy expectations with regard to appropriate data use. Participants did not express specific strategies they would employ to be transparent and develop the conversations, but they valued dialogue between institutional actors and students since it held the potential to bring to the fore privacy questions and considerations formerly unknown. Also, an information officer felt strongly that transparency around student data tracking would create learning opportunities centered on information privacy. "By having these conversations," she said, "[we] are helping them learn."

Besides its usefulness in helping to establish trust and educating students, actors also expressed that they were obligated to be transparent and it was a responsibility they could not ethically shirk. When asked if it would be fair for an institution to use personally identifiable information and data about students without their knowledge, a participant strongly responded that such practices would be "unethical, unfair, and, frankly, unprofessional." In fact, others argued that there was little reason to be less transparent about learning analytics than in other areas of institutional practices, like information policies and instructional initiatives, where transparency in one form or another is expected.

Points of Transparency

In addition to suggesting a general strategy of transparency, participants offered up suggestions of specific points in time at which students should be informed of learning analytics. The suggestions were not developed, systematic approaches. Instead, they were proffered as options that seemed to promote an agenda of transparency. The points of transparency were simultaneously relational and temporal. Points of transparency are relational when they inform relationship building among students, their institution, and its actors (primarily instructors); they are temporal when transparency happen at specific times.

Relational points of transparency occurred, primarily, when students begin to create a relationship with the institution, such as during the admissions process and when students matriculate. The institution, participants felt, entered into a type of contract with students that required it to be more forthcoming with information to maintain a healthy relationship. This could be accomplished, a participant said, by simply informing the student using additional documentation or by using a "terms of service approach." Others offered the opinion that instructors should inform their students of how they used learning analytics at the beginning of the semester as "the first step" in developing the student-teacher relationship. To this point, a participant said:

> Tell your students that this is what you're doing in class. It makes total sense for them to understand that this is part of the curriculum. Say, "this is what we're doing. For me to help you best, we're going to employ this sort of learning analytics approach or this particular model."

Telling students about using learning analytics in the classroom could be an extension of other justificatory, transparent practices, such as sharing why an instructor employed a teaching method or why she chose particular learning outcomes.

Matriculation and the start of a course mark important times when students could be made aware of learning analytics practices, among other just-in-time moments. For instance, if predictive analytics indicate that students have an insufficient academic background or skill set that will, probabilistically speaking, limit their success in a course, students could be made aware of this information at the time of course registration. Others suggested that students could be informed of learning analytics practices at the beginning of each semester, especially if institutional data gathering, use, and analysis changed in a way that could present privacy concerns. A participant said that semester notifications honor "what we told students we would do when we collected information from" them and keep them informed of changes thereof.

7.9. Self-Fulfilling Prophecies

Profile information, scores distilled from student surveys, and predictive analytics developed from student data could create negative "self-fulfilling prophecies." Interviewees expressed that the information learning analytics presented to faculty, staff, and students could create situations where individuals internalize it without critically considering it. In the following sub-sections, I detail what a self-fulfilling prophecy is and how it could emerge. Additionally, I discuss what interviewees did to protect against self-fulfilling prophecies.

Self-Fulfilling Prophecies and Concerns Thereof

Interviewees suggested that self-fulfilling prophecies could emerge when students or institutional actors assimilate learning analytics information without critically analyzing it. Self-fulfilling prophecies could occur when institutional actors review a student's profile information, including a student's college preparation scores, a history of her grades, or profile information about her interests. Self-fulfilling prophecies could also occur when these same actors look at a student's academic scores, non-cognitive abilities (e.g., so-called "soft skills"), and predictive measures of success in the classroom, in a program of study, or in college writ large. If institutional actors do not review the information judiciously, the could develop pre-conceived ideas about what a specific student is capable of and treat the student as a lost cause.

The larger concern about self-fulfilling prophecies participants expressed concerned students seeing learning analytics predictions. In this situation, interviewees were worried that students could make quick conclusions about themselves with negative consequences. The negative outcome could be that students believe they have little control of their future and the choices they make given predictions of academic failure or less than desirable learning pathways (e.g., not predicted to be successful as a pre-med major).

Actors at Saint May State University were more vocal with regard to concern over self-fulfilling prophecies. An advisor at the institution described her worries and that of her peers with a scenario, saying, "A student goes into their [learning analytics profile], you know, a first month freshman goes into their [profile] and says, 'Oh my gosh, I'm low in academic engagement,' and doesn't read any further. Do they just give up or do they keep going?" Others on campus did not want to create situations where students would "struggle" with data that indicated they had a low chance of academic success. As a result, the campus put a "remarkably tight lid" on access to the learning analytics system to limit the potential of self-fulfilling prophecies.

Saint May State University participants were also concerned that instructors could unwittingly enable self-fulfilling prophecies by giving information from learning analytics technologies to students without providing important contextual information, such as how predictive analytics come to their statistical conclusions. An advisor explained his institution's decision to limit instructor access this way:

> It goes back to that self-fulfilling prophecy. We are a faculty-driven advising model campus. With that, we would need to be able to provide intensive training and conversations with faculty who are advising students. And we had some concerns that if we didn't do it properly that the information would get in the wrong hands. It could be given incorrectly to the student, therefore providing that self-fulfilling prophecy.

Furthermore, participants discussed how if advisors or instructors had access to the information they might begin to make poorly informed assumptions about a student's potential. About this, an administrator posed this question: "If instructors or advisors saw somebody having a low score or something like that, then would they be treating that student differently than otherwise?" Here, the apprehension was that making quick conclusions would limit the resources and support instructors provided to students, as well as produce situations where instructors assessed students unfairly and hastily.

Protecting Against Self-Fulfilling Prophecies

Participants argued that students should be given choices and personal freedom while attending university, and not feel restricted by predictive analytics. A participant at Hammond University explained the potential harm this way:

> It can also, I think, harm the students if we say based on your profile and this analytics report, you really shouldn't do this major. Because, 83% of the students who have your profile and try to pursue this path have not done well and end up with an X, Y, and Z. That could really scar and screw people up and also shut doors to possibilities for students when we shouldn't be shutting doors.

A participant explained that instead of reducing educational options, his institutions should try "to give students the choice to make decisions for themselves and to hopefully make good decisions. They have every right to make bad decisions and sometimes they do." Given their concern for these issues, some looked to proactively protect against them.

At Saint May State University, participants explained how they were trying to reduce the risk for self-fulfilling prophecies. When students took the standardized survey, which served as the foundation for Lighthouse's analytic scores and predictions, a dedicated advising unit on campus informed students about the system. It did so during individual meetings with students, at small group meetings, and at a presentation to the campus' student senate. Advisors explained that the intention of these sessions was to remind students that the scores only represent "a moment in time," that they have the power and the institution has the resources to help them achieve at a level beyond what the predictions say they can. With regard to the Spotlight system, participants at Hammond University and Saint May State University did not express a proactive plan to inform students of predictive scores; however, they did state the importance of doing so.

7.10. Information Control and Ownership Rights

Participants expressed that students should be given the right to determine for themselves if and how institutions could make use of their data and information for learning analytics. After data analysis, two major themes emerged that help us understand participant perspectives in this area. The first theme considers what participants mean by "information control" and why they valued it. Control, they argued, empowered students to make important decisions regarding data access, use, and analysis. While this viewpoint was expressed as a core tenet of student privacy, there were few mechanisms by which students could express this right. Interviewees suggested a number of ideas on how to promote student data and information controls in the future, but even those had pragmatic limits or conflicted with institutional goals. The second theme details the relationship between information control and information ownership. Participants also discussed how student control over data could be interpreted as an ownership right. Yet, as they pointed out, ownership–like control–may need to be limited for a number of reasons.

The Power to Make the Decision

When conversations turned to student rights related to learning analytics initiatives and their use of student information, participants expressed a unified opinion that students should have control over their data and information. As one participant put it, "I think it's very critical students should have the ability to protect their [privacy] rights. If they want to be more protective, we should give them that opportunity." Control, they argued, began with the institution informing students of the data they collect and the purposes to which it will put it. Participants thought students expressed this control by consenting or not consenting to the information practices the institution describes. Regardless of their opinions on the feasibility of informed consent, participants viewed such processes as the gold standard, since they give students "the power to make the decision for themselves as to whether or not they want the institution" to use their information in particular ways.

At the time of the interviews, informed consent processes were more of a long-term goal than a present practice. In fact, there were no informed consent procedures as such. Participants at Hammond University told me that they did inform students that they were participating in classes equipped to use the Spotlight technology, but students had no choice in the matter. The same was true at Saint May State University. With regard to the Lighthouse initiative at Saint May State University, participants discussed how students were verbally informed of the purposes of the technology before and after they took the necessary surveys that underpinned much of the analytic scores. If students preferred not to participate in the testing, they could do so, but they were not actively informed of this option.

Individuals pointed out that informed consent processes were important, because they gave students an opportunity to understand what data-driven activities would occur with their data and the reasons for which their data was needed. In that sense, participants argued that their institutions could treat informed consent as a type of "terms of service," which was briefly discussed in Section 7.10. Terms of services agreements would require universities to alert students of when changes occurred to information flows and when new consent was needed. An advisor and administrator argued that students should be "offered the opportunity to elect in or elect out at any point" in their academic career; this option would be a part of the terms of service agreement students would sign during the admissions process.

Other participants pointed out the limitations of informed consent processes. Informed consent processes that allowed students to opt in or out of institutional data gathering and learning analytics practices presented two specific problems. The first of which was that students who opted out could create data gaps. Data gaps could either weaken algorithms or limit a university's ability to serve the student due to a lack of information about her. The second problem was that informed consent added "administrative and technical" complications. Participants who expressed this concern did not see how to logistically fulfill informed consent mechanisms without stressing existing resources or needing new technologies. An educational technologist used a scenario in which a graduate wanted to remove her data from a university's archives to discuss this issue:

> For example, if I've graduated, I've been gone for 10 years, and then I've read something about data privacy and I get really incensed and come back to the university and say, "I want all of my digital records exhumed." The cost associated with that would be too much, because at that point everything has been archived. So, you have to go back and bring back every single semester, and then extract anything for a student, and then fix any dependencies. Just technically, it is not feasible.

In this case, the institution would lose historical data, which could inform algorithm design, and the resources and time involved may be exorbitant. As such, it may be the "perfect situation" to allow students to opt out, but it may also be impractical.

It is Their Data

The common argument that students should be in control of personally identifiable data and information often stemmed from a belief among my participants that students should have an ownership right to personally identifiable data and information. An institutional researcher put the argument this way: "It's not our data, it's their data. We're the custodians." Part of the belief that students owned their data was legal, as some interpreted FERPA's definition of educational records and related privacy rights to include ownership rights. Others, however, felt that the data and information students bring to a university and create during their tenure is a part of the product they pay for. A participant explained this perspective, saying, "students are paying us to educate them. Because they're paying, I think that they should have that right…. Students can choose where they want to go. [As a student,] I would want to have the right to use the data how I wish."

There were outstanding conflicts with an ownership perspective, however. Data ownership rights, some argued, could limit an institution's practices or create conflict with students. At one institution data ownership conversations had advanced all the way to the president, and, reported one participant, got heated. According to this institution's president, the university owned the data. The participant said, "Frankly, he got sick and tired of the argument, I think." Part of the frustration stemmed from the fact that the data could be used by the institution to create more efficient practices and cut costs; potentially losing control over student data may limit the institution's ability to do so. It would be better, a participant argued, not to give students the ability to remove data from the institution's control, as it could close down an important information "pipeline" that is used in students' "best interest" and to run the institution. Others pointed out that negotiating data ownership rights could become confrontational. Students may contend that data about them is theirs; institutions may argue that the data was created in their environment with their systems. "Now there's a line in the sand," an information officer explained, "[and] that's a harder negotiation to go through than [saying] 'here is what we are doing with the data.'"

Chapter 8. Discussion

8.1. Introduction

The thematic findings in the previous chapter serve as a starting point in this chapter in order to develop a broader conversation about student privacy and learning analytics. To that end, I have divided this chapter into three sections. First, I review Nissenbaum's (2010) decision heuristic before employing it to make a multi-step assessment of contextual integrity with regard to my case sites. I argue in my final assessment that contextual integrity has indeed been violated. Next, I address a weaknesses in using contextual integrity as a privacy framework and suggest that we need to consider micro-level contextual issues. In the final section, I recommend that student privacy should be conceptualized as control, and I argue that to support such an approach to student privacy requires institutions to reestablish informed consent processes, build privacy dashboards, and develop a technical identity layer that can respect student privacy preferences.

8.2. An Assessment of Contextual Integrity

Evaluating information privacy concerns and contextual integrity requires a comprehensive amalgam of evidence. It is no small feat. In order to make such judgments, researchers need to seek all types of support for their claims, including historical, empirical, and conceptual. But even after data is gathered and analyzed, the heuristic requires interpretive skill to understand the interrelationship among contextual data and uncover if emerging technological practices stress existing privacy expectations. To this point, I have developed a useful amount of data to support contextual integrity claims regarding student privacy and learning analytics, which I can augment with interpretive insight. In the subs-sections that follow, I use my historical framing, understanding of the relevant literature, and empirical findings to evaluate the status of contextual integrity at my two case sites, Saint May State University and Hammond University. In so doing, I follow Nissenbaum's (2010) decision heuristic, which requires researchers to build evidence for their final evaluation of contextual integrity, and I conclude that a contextual integrity violation has occurred.

The Decision Heuristic

Nissenbaum's (2010, p. 181) decision heuristic provides a series of steps to determine if new socio-technical systems invade privacy in a particular context. For review, I provide the heuristic below:

1. Describe the new practice in terms of information flows.
2. Identify the prevailing context.
3. Identify information subjects, senders, and recipients.
4. Identify transmission principles (i.e., contraints on information flows).
5. Locate applicable entrenched informational norms (i.e., contextual expectations of information flows) and identify significant points of departure.
6. Make a prima facie assessment. A breach of informational norms yields a prima facie judgment that contextual integrity has been violated because presumption favors the entrenched practice.
7. Evaluation I: Consider moral and political factors affected by the practice in question.
8. Evaluation II: Ask how the system or practices directly impinge on values, goals, and ends of the context.

9. On the basis of these findings, make a claim of contextual integrity (i.e., determine if informational norms have been violated).

For organizational purposes and to make clearer, succinct arguments, I combined step four with step five; and to the same end, I combined step seven with step eight. I address all nine steps in the sub-sections that follow.

Information Flows

Findings revealed that emerging learning analytics practices were altering existing and creating new flows of personally identifiable student data and information; however, the resulting flows of information differed depending on the specific learning analytics system in a number of ways. Both systems made data more visible, accessible, and fluid to a wider swath of institutional actors, in potential and in practice. The first change to information flows concerns its visibility.

Student profile information stored in student information systems is a fine example of an information flow that underwent a significant change in its visibility. Before Hammond University and Saint May State University adopted learning analytics systems, student information flows were less visible or even hidden to students and most institutional actors. At both institutions, this information was usually restricted to the registrar's domain, with exceptions for special projects. Never was it released *en masse*. Yet, the learning analytics initiatives made this possible, as each system exported student data and information into individual student profiles for institutional actors beyond registrars to access, if they had the right privileges. The findings revealed that some institutional actors felt that this newfound visibility of student information was unsettling, potentially harmful, and in conflict with institutional policy and federal law.

System log data, or digital breadcrumbs, was another example of newly visible information flows. Spotlight analyzed the digital breadcrumbs "dropped" by students within the learning management system in which it was embedded. And these breadcrumbs, while constantly created as students interacted with the system, existed as hidden data. But with Spotlight, this hidden data flow became visible to instructors. But seeing where students had been, what they had done, and for how long within a digital classroom in the learning management system was, to many participants, teetering on surveillance. Making this type of hidden information flow visible was disconcerting, albeit potentially insightful for understanding learning processes.

The visibility of information flows was not the only important change: new flows also emerged. With regard to the Lighthouse system at Saint May State University, a new flow of student information was necessary in order for the institution to take advantage of Lighthouse's analytic affordances. Saint May State University created a new information flow based on the pre-enrollment student survey that underpinned retention and GPA predictions, as well as scoring students on their non-cognitive abilities. The survey was augmented by flows from other systems and institutional actors, like instructors and advisors who added comments about individual students.

Learning analytics initiatives at both institutions also revealed the fluidity of student information. The Spotlight and Lighthouse projects worked using restricted flows of information among a small network of systems (e.g., learning management systems and student information systems). Responses from interviewees suggested that opening the taps of institutional systems could reap a greater collection of student data and information, which could prove beneficial for longer-term learning analytics projects.

But thinking of student information as flowing is limiting; we also need to consider pools of information and data as well. Instead of focusing on whether flows of information are changing, we may be better served by looking at how institutions are aggregating all the data they can get in a pool–or an archive–for later analysis. Pools of data are analytically powerful for institutions, as historical data enables data scientists to get access to large amounts of historical data, which is useful for building better predictive models. Researchers know little about data pools, as it is an emerging concept. And practitioners do not know how to manage or employ large data pools effectively; the findings suggest that this is because institutions are just now figuring out their data governance responsibilities. But if we are concerned about the emergence of digital dossiers on campus, we need to start looking closely at information and data pools.

The Prevailing Context

As discussed in section 6.5., it is challenging for researchers to place boundaries around contexts. Nissenbaum (2010) tends to define contexts at a macro level, which is to say that she targets visible and understood institutional structures (e.g., particular types of hospitals and schools). This approach proved sufficient for this study, as it was straightforward for me to define the context as public higher education. My two case sites represented fairly typical non-profit, public, higher education institutions. Nothing in the data indicated that characteristics about either institution deviated from what a person familiar with higher education in the United States would expect. By that I mean, as far as the data indicated, the institutional structure, practices, goals, ends, and values of both universities were standard.

Yet, notable variations did emerge at micro contextual levels with regard to student privacy. Actors at both institutions expressed that protecting student privacy was something that they should do, but I saw variation in privacy protection methods and the perceived degree to which student privacy should be protected among different roles. For instance, registrars, institutional researchers, and lawyers, who understood FERPA and worked with student data in their work practices, were less concerned with the privacy issues inherent to learning analytics. They understood their institutional policy and legal obligations; furthermore, they were not concerned about how they interacted with student data as part of learning analytics practices as they had a legitimate reason to do so under the law. In contrast, data scientists, academic administrators, and information officers were more concerned with the student privacy implications connected to learning analytics. Their concern was driven by their more frequent use of and involvement in the development of learning analytics systems; furthermore, they believed there was a lack of privacy standards to guide their work. For a final example, participants in advisor roles were most vocal about the autonomy issues the technology presents, because they valued the student-advisor relationship and how it represents an exercise in individual path finding–something predictive analytics challenged.

Information Subjects, Senders, and Recipients

The increase in student information flows resulted in changes to information subjects, senders, and recipients. Actors interviewed at Saint May State University and Hammond University say the driving goals of learning analytics technology is to improve teaching and learning practices as well as institutional programs. Achieving these goals pushes institutions to put student information to use in previously unknown ways that change subjects, senders, and recipients.

With regard to subjects, new learning analytics practices target students. But, historically speaking, this is nothing new for higher education. The historical evidence shows that institutions have long sought information about students for educational and institutional needs. However, it is fair to say that learning analytics at both institutions placed a new emphasis on information disclosures and data grabs that could obtain information from students, whether they were aware of it or not. The question, then, is not if the subjects of the information practice changed, but whether the subjects became more aware of such practices. The answer here varies.

The findings suggest that students at Saint May State University were more fully informed of learning analytics practices than students at Hammond University. Students awareness was especially high with regard to Lighthouse, because, as I reported in section 7.9., advisors spoke directly to individual students and presented to the student senate how they were using the technology. However, there is little evidence as to whether students in classes that employed Lighthouse, either at Hammond University or at Saint May State University, were aware that their digital movements were tracked and that their personal information was included in predictive algorithms. Although I did not ask them directly, participants involved in the Spotlight project did not indicate if students were systematically made aware of their participation in this learning analytics initiative.

Identifying senders of student information in learning analytics leads to a paradox. It seems, at first thought, that the *subjects*–the students–of learning analytics are also the senders of their own data and information. That runs counter to the norm, where institutional actors send and retrieve information about students. But Spotlight and Lighthouse relied on students sending information, or more precisely, creating information about themselves for the systems to analyze. Where student profile information was concerned, the registrars still filled a traditional "sender" role by exporting student information for the systems to import for analysis. Other than that, there were no notable changes to senders of information.

Finally, it is important to document the changes in recipients of information due to learning analytics practices. Both systems created new opportunities for a wider group of institutional actors to access to student information; however, both institutions also created access restrictions to limit who actually got to see information derived from learning analytics. At Hammond University and Saint May State University, participants reported that, in addition to instructors, only actors supporting the development of Spotlight could see information borne from the system. With regard to Lighthouse at Saint May State University, the situation was similar; advisors were the main recipients of Lighthouse's analytic findings, but administrators also could review aggregate statistical reports.

Secondary recipient groups of learning analytics information represent types of individuals who were not intended, either by technological design or policy, to receive information but did so anyways. For example, it was not Saint May State University's intent to make Lighthouse's personally identifiable analytic findings generally accessible to students. But if students requested access, they could get it. Furthermore, participants at the university also indicated that there was interest among institutional committees, such as academic appeals committees, to get access to analytic findings and personal communications (e.g., instructor and advisor notes) within the system. Assuming that other institutional actors saw value in the information both Spotlight and Lighthouse gleaned from students, it is fair to say that other secondary recipient groups could emerge in the future, unless the institutions develop strict policy to limit downstream information access.

Transmission Principles and Informational Norms

Transmission principles restrict information flows by governing which "terms and conditions" (Nissenbaum, 2010, p. 145) apply to particular flows. As I discussed in section 6.4., transmission principles are often codified in laws, policies, and contracts; they can exist as implicit expectations in certain social situations (e.g., confidentiality, reciprocity) or by way of explicit agreements (e.g., contractual negotiations for information exchanges). Furthermore, transmission principles are governed by context-specific informational norms. The question at hand in this sub-section is what, if any, transmission principles and informational norms needed to be created or altered to govern learning analytics practices at my case sites. Research findings indicate that participants relied on existing transmission principles and informational norms for learning analytics initiatives, and for many this was problematic.

Participants relied heavily on FERPA to understand how their respective institution should protect and use student information. The law informed their "need to know" policy, which dictated what types of student information actors could see within learning analytics systems. Even though interpretations of the law dictated some access restrictions, FERPA provided freedoms to institutions to use whatever student data and information they had control over in learning analytics algorithms. Institutions simply had to justify data practices as part of an evaluation project or for institutional improvement.

Even though the institutions had great flexibility to use student information as they wished, some participants enacted ad hoc policies to limit information flows. Ad hoc policies emerged based on some participants' personal intuition that freely flowing student information for or derived from learning analytics was potentially harmful. This phenomenon was most obvious at Saint May State University when an administrator and advisor determined that students and instructors should have limited access to Lighthouse due to their concerns over self-fulfilling prophecies. These concerns resulted in creating new user roles for the system with restrictive access rights. While these ad hoc policies assuaged some participants of privacy concerns, many participants recognized that creating policies on the fly was not adequate and more needed to be done to develop institutional policy. But at the time of the interviews, this type of transmission principle was not established.

With regard to informational norms, participants expressed that information flows between students and the institution (and its representatives) were required as part of relationship-building and sustaining processes. Not only did the institution need students to disclose their information to start a relationship, but they also needed it to provide services and programs to meet students' educational and personal needs while they were enrolled. Limiting the flow of student information could stunt the student-institution relationship, limit institutional actors in their attempt to educate students, and stymie data-driven projects to improve the institution.

A Prima Facie Assessment

The framework of contextual integrity instructs researchers to take stock of the relevant contextual factors and make a prima facie assessment of privacy invasions before moving on to more advanced and informed assessments. As a reminder, contextual integrity is maintained when contextual expectations of information flow remain stable; but when these expectations are contravened, it is an indication that contextual integrity has been violated and a privacy invasion has occurred. Specifically, contextual integrity violations arise when new socio-technical practices change the attributes of information; include new senders, subjects, and receivers of information; or change informational norms.

On the surface, learning analytics does not seem to violate contextual integrity. Evidence from the findings indicate that it represents an evolutionary step in the use of student information to inform educational practices and institutional processes, and not unlike what fits within a history of student surveillance in higher education. But, too many prima facie threats to contextual integrity have emerged.

Learning analytics threatened contextual integrity at my case sites in three particular ways due to surveillance concerns, emerging dossiers, and problems with a feckless privacy law. Importantly, these threats were recognized by participants who saw learning analytics as a paradigm-shifting socio-technical practice that challenged many outstanding student privacy protections. I will address the three threats to contextual integrity below.

First, nothing raised privacy concerns more among participants than the fact that surveillance practices were embedded in learning analytics and that technology increased the potential for future surveillance practices. Yet, participants were without recourse to rectify these issues due to the fact that tracking technology was part and parcel to learning analytics, especially where Spotlight was concerned. Unlike information visibility issues within Spotlight's user interface, which participants were able to fix with help from the vendor, breadcrumb tracking was integral to the system's predictive capabilities, rendering such fixes out of the question. While these affordances of the system were worrisome, it was what they potentially represented for the future that was even more troublesome. Future surveillance practices, especially those built on top of even greater sums of sensitive data, marked a particular challenge to participants at my case sites. Future visions of student surveillance, which included RFID and network activity tracking, discomfited them, as the visions challenged the amount of surveillance they felt was necessary or prudent to improve student learning.

The present and future surveillance practices represent threats to contextual integrity in several ways. First, they take hidden digital breadcrumb data and make it visible, which changes the attributes of the data. Second, using breadcrumbs for surveillance transforms the intention of the data, which was initially to support how the system runs. Now it is also used to observe student movements in digital environments. Third, surveillance practices push the boundaries of acceptable student information practices at my institutions, where participants expressed a difference between data and information that students provide themselves and that which institutions observe from student movements. If students willfully disclosed information about themselves, participants argued, using that information was justifiable. However, using digital surveillance methods to obtain student information without students' knowledge was less defensible unless institutions made an effort to transparently explain and justify their data collection practices.

Second, an "all the data we can get" approach to learning analytics increases the risk that my institutions will develop personally identifiable digital dossiers. Not only will these dossiers include normal academic information, but if systems continue to develop more advanced surveillance technologies, they hold the potential to capture a comprehensive, time-stamped history of student movements, communications, and analyzable personal networks. While this may sound like hyperbole, findings indicate that there is an interest among powerful actors to record more and more student data and information to support learning analytics and inform institutional practices. Moreover, digital dossiers are especially concerning given an institution's inability to protect student records from third-party government actors who may force data exports to longitudinal databases.

Digital dossiers threaten contextual integrity by altering the attributes of educational records and changing the transmission principles that guide their use. Students expect their educational records to encompass that to which they refer: their education. But dossiers that capture information beyond what is generally considered to be educational in nature shatter well-established expectations regarding what an institution records about students. Furthermore, educational records-cum-dossiers that inform multiplying data-driven initiatives changes who sees student records and what they do with them due to learning analytics. FERPA may enable widespread use of data and information captured in student records under the umbrella of "institutional improvement," but it does not account for the fact that students generally expect their records to exist as static files that are used sparingly and purposefully. Instead, these new dossiers are ever-changing as newly connected systems increase data flows, fill data pools, and enable a greater swath of institutional actors to have access.

Finally, in relationship to emerging learning analytics practices and dossiers, FERPA is becoming feckless. When educational records existed as paper files in office drawers or minimally accessible data files in a registrar's computer, student privacy was easier to protect due to tighter flows of information. But interoperability between learning analytics systems, student information systems, and data warehousing enables student information flows once only imagined.

FERPA exists as a codified but now inadequate informational norm. Concerning student information, it governs how the information should flow, defines how it is characterized, and identifies relevant actors and their privileges all within the context of education. As such, it guides what institutions can and cannot do with educational information. To date, it has successfully protected students from privacy invasions stemming from bad actors outside a university's walls. But the greatest challenge to student privacy–and contextual integrity–stems not from outside actors: it comes from the institution itself. FERPA provides institutions too much latitude to do with student information what they wish. In the past, this concern has been less of an issue given technological limitations and the resources it would take for an institution to pry into its students' lives. Students continue to need protection from third parties, but they also need greater protections from the colleges and universities in whom they entrust their information.

An Assessment of Higher-Level Concerns

Contextual integrity is mostly at stake due to changes to established student information flows, but larger issues are also at play. Some of these issues were made explicit in the findings, while others lurk in the background. For instance, there are clear autonomy issues to consider, as evinced by concerns over self-fulfilling prophecies and the power of predictive analytics.

There are also less obvious dilemmas that need to be addressed. Findings also hinted at the fact that learning analytics at my institutions may impact the free will of students and their ability to experience higher education on their own terms. Additionally, learning analytics practices were propping up power structures that disenfranchise students. The following paragraphs tease apart the important aspects of each–autonomy, free will, and power–and their relationship to learning analytics.

As I discussed in section 7.11., participants identified concerns regarding how their learning analytics initiatives, especially those with predictive capabilities, may restrict student autonomy. Student privacy serves "a fundamental and ineliminable role" (Alfino & Mayes, 2003, p. 6) in autonomy by protecting students from undue intrusions into their life that could limit "individual conscience" (Richards, 2008, p. 404), such as developing intellectually, forming moral constructions, and assessing social values, all of which we value as educational goals. We also value student autonomy, because we wish not to influence a student's decisions to the point that they are not fully her own (Bloustein, 1964; Reiman, 1976).

There are two ways the learning analytics initiatives interact with autonomy and bring about student privacy concerns. First, as Rubel and Jones (forthcoming) outlined, privacy may support autonomy, insomuch as it protects against intrusions from others and diminishes outside influences on how one "acts or thinks for oneself" (p. 9). Second, restricting access to or hiding information from individuals denies them an opportunity to make accurate interpretations of their world, even if they do not use that information to support their actions. With regard to the first form of autonomy, students were unable to protect against institutional uses of learning analytics technologies that used their profile information, including private information they revealed on admissions applications, and data derived from their interactions with systems. Students were also unable to limit uses of predictive analytics, which informed instructors and advisors in ways that could be used to intervene in the lives of students. Moreover, students were often unaware that learning analytics systems made predictions in the first place. With regard to the second form of autonomy, it could be that students would prefer to see that information, even if they chose not act on it, but institutional actors did make the scores easily accessible to students.

Concerns about how institutions use predictive analytics to nudge and direct student behavior elicits free will concerns as well. Students have not and cannot expect to express *complete* free will during their time in college. There have always been limitations on students that dictate how they are to behave, act, and participate in a campus community. Some of these limitations are determined by class schedules and programmatic limitations, while others are more normative in nature. As such, students cannot always do what they want, whenever they want should they want to earn the credentials they seek. Even with these limitations, students have had free will to make choices for themselves and learn from their mistakes, but this may change as institutions adopt advanced predictive learning analytics technology.

Emerging assessment regimes built on learning analytics impact the ways by which institutional actors judge a student's academic success. The digital exhaust students leave as they interact with data-driven technologies enable learning analytics systems to track and analyze student activities and behaviors, providing fine-grained information to instructors and advisors. As a result, digital breadcrumbs and the analysis thereof supports new ways to judge a student's academic success, engagement, and potential. This is problematic for free will. When students intuit–or have been told–that their university is judging their decisions, movements, and behaviors based on digital traces, they may no longer act of their own accord.

Both of the learning analytics technologies at my case sites, Spotlight and Lighthouse, presented unique free will issues. Lighthouse captured when students were active in online courses; the system also captured what they wrote in forums and with whom they communicated, along with what parts of the system they interacted with. In part, it used this information to make predictions about their future success. Where Spotlight is concerned, comprehensive student profiles, which were embedded with retention and academic success predictions, enabled advisors to closely monitor and document a student's academic life, from grades to involvement in extracurriculars, and provided alerts about negative behaviors and academic progress. As the two institutions I researched continue to document, review, and intervene based on this data, students may feel coerced to act and behave in ways that align with their university's goals and views of assessment, not on their personal desires to pursue certain ends.

Some may argue that the ways learning analytics coerces students to behave does not differ in any notable way from current pedagogical and institutional practices. For instance, instructors use attendance, grades, reprimands, and university guidelines (e.g., student handbooks) to urge students to change their behaviors based on the values and normative expectations of the instructor and the institution. The difference between these coercions and those learning analytics brings about with data-driven nudges and predictions is a matter of detail. Learning analytics allows institutions to finely tune student behaviors based on detailed historical data, which cannot be achieved, for instance, just by taking attendance.

The outstanding autonomy and free will issues point to a larger problem regarding power imbalances among students and their respective colleges and universities. When information practices are hidden from a student's view yet intervene in her life in unknown and unexpected ways, colleges and universities hold the upper hand in the student-institution relationship; moreover, colleges and universities may be using the analytics insights for their own gain, not for the benefit of students. It could be, as was the case with Spotlight, that an advisor informs a student that she needs to become more active in extracurriculars. But the student may not know that this direction is based on data indicating low social engagement on campus; the student may infer that the advice is intended to improve her well-being and educational experience. In actuality, the institution's intent could be to use predictive scores of student retention to target specific students who are at risk of withdrawing or transferring to another institution, which could be detrimental to an institution's resource planning and funding levels.

Power imbalances between students and their institution are increased, as well, when FERPA cannot ensure their privacy rights. The law enables students to take action against their university when they feel that their privacy rights have been invaded or if their educational record is, in their opinion, incorrect. But the power for students to act on their privacy rights has been diluted by, as my findings revealed, the lack of a unified definition of what an educational record is and a technical inability to retrieve all personally identifiable data and information that is conceivably a part of a student's educational record.

Higher education institutions could develop programs or make better choices about learning analytics that would improve autonomy and increase a student's sense of free will, and in so doing decrease the power imbalance. Such programs, as my participants mentioned, may involve increasing transparency on campus regarding data-driven initiatives; it may also require adopting learning analytics technologies that may not be as feature-rich but promote these higher-level values, such as student autonomy. Until institutions pursue these initiatives and students are able to fully express their privacy rights, the power imbalances will continue to exist.

The Final Assessment of Contextual Integrity

The empirically-grounded assessments of contextual integrity ultimately led me to judge emerging learning analytics technologies and practices at my case sites as attacks against student privacy. When we consider the history of higher education, we can look upon learning analytics as an evolutionary use of student information in colleges and universities. That institutions want and will continue to seek student information to inform their practices, educational and otherwise, is a given. But, learning analytics is more than evolutionary or developmental; it is better characterized as a transformative use of student data and information. The practices are new, the technologies are new, and old policies and law can no longer comprehensively protect student privacy in light of data-driven educational technologies.

In some respects, this assessment of contextual integrity against learning analytics would not be a surprise to actors at Hammond University. Participants at Hammond University were especially aware of the fact that student privacy was a multifaceted problem they would have to address in order to push forward with the technology. As a result of their awareness, they were actively building data governance mechanisms.

With respect to Saint May State University, however, participants were less self-aware of the privacy problems attached to their learning analytics initiatives. Their concern was focused on self-fulfilling prophecies and ways to protect against this type of privacy harm. Findings indicate that participants at this case site believed they could manage privacy concerns by restricting access to learning analytics information using technical measures and ad hoc policies. At the time of the interviews, they expressed no desire to develop more comprehensive privacy protections like their peers at Hammond University.

While the differing levels of awareness among actors at both institutions is notable, it does not affect assessments of contextual integrity. What matters is how contextual integrity was impacted by learning analytics as a new socio-technical practice and what actors had done at the time to protect against negative impacts on student privacy. And the final assessment of contextual integrity indicates that student privacy was adversely influenced by learning analytics initiatives at both institutions. The emerging question, then, is what Hammond University and Saint May State University–and other institutions like them in their situation–need to do to regain contextual integrity. I take up this question in the coming sections by pointing out infrastructural weaknesses and providing recommendations that can shore up the information infrastructure in support of student privacy, but before I do so I offer up a critique of contextual integrity that argues that we need to pay special attention to micro-contextual issues.

8.3. A Critique of Contextual Integrity

The study has benefited from viewing the student privacy issues through Nissenbaum's (2010) framework of contextual integrity. A contextual approach to privacy provides researchers analytical focus and allows them to home in on particular issues threatening contextual values and expectations with regard to privacy. However, after using the framework as a lens to look at my empirical situation, I have discovered that it has an inherent weakness: it overlooks micro-level contexts.

Before addressing the weakness, we need to recall how the framework of contextual integrity defines contexts. According to Nissenbaum (2010), "contexts are structured social settings characterized by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)" (p. 132). Contexts exist as "abstract representations of social structures experienced in daily life" (Nissenbaum, 2010, p. 134), but they do not map onto specific definitions; contexts are fluid in that they are attached to specific times, places, and social situations. As such, contexts are often variable. For instance, public higher education remains relatively stable, but socio-technical systems (e.g., online learning technologies) and socio-political factors (e.g., neo-liberal movements) are reshaping this particular context.

Nissenbaum's approach to contextual integrity emphasizes macro- and meso-level contexts. For instance, she invokes education, healthcare, e-commerce, and governments (state and federal) as macro-level contexts in her examples. In doing so, she highlights the mélange of actor roles and activities and their effects on information flows as a result of a new or changed technological practice (e.g., implementing RFID systems for highway toll systems); the particular technological practice and the relevant characteristics that surround it (e.g., particular roles, values, and goals) create the meso-level context.

By emphasizing meso- and macro-level contexts in Nissenbaum's (2010) framework of contextual integrity, micro-level contexts are often lost in the background. But let me be clear here before moving onto a discussion of micro-level contexts: I am not saying that Nissenbaum does not account for micro-level contexts in her work. In fact, Nissenbaum built her framework on what we may call micro-level sociology, such as Bourdieu's (1984) field theory. However, the closest Nissenbaum (2010) comes to propping up the importance of micro-contexts is when she describes contextual nesting. Contextual nesting exists when a particular context exists within a more general context, such as my case sites existing within the larger context of higher education. However, when contexts are nested, conflicts may emerge. If one of my case sites was a for-profit, online institution, we could argue that the case site, then, would be in conflict with generally accepted values and ends of higher education. When conflicts emerge, Nissenbaum (2010) provides no resolutions, stating there are no "general solutions" and they are sometimes "simply intractable" (p. 137).

There are two opportunities we miss by not looking at micro-contextual levels. First, researchers can do important analytical work by looking at how micro-contexts relate to and conflict with the larger meso- and macro-level contexts in which they are nested; in shorthand, I call this phenomenon contextual interrelation. Second, by focusing our attention on meso- and macro-level contextual concerns, micro-level values, goals, and practices are overshadowed and downplayed; I call this particular effect contextual repression. In the sub-sections that follow, I use my findings to demonstrate the value of examining contextual interrelation and contextual repression as both relate to privacy concerns.

Contextual Interrelation

Contextual interrelation asks us to consider micro-level contexts and their relationship to each other and larger meso- and macro-level contexts. But what is a micro-level context? Pulling on Nissenbaum's (2010) own research, micro-level contexts are particular social spheres of actors who share the same goals, have similar values, and engage in compatible practices. In higher education institutions, micro-level contexts are usually defined by specific offices (e.g., the advising office), departments (e.g., the department of financial aid), and schools (e.g., the school of library and information studies). Actors within each of these micro-level contexts typically work towards the same ends.

Micro-level contexts naturally differ in the degree to which they are compatible with one another and overlap, much in the same way as meso- and macro-level contexts do, and conflicts across micro contexts may emerge. For example, academic departments and their institution's information technology division may not agree on what means should be used to improve online pedagogy, for example. And as discussed in the findings, institutional research offices do not always agree with other micro-contexts in how they analyze student data and information and derive analytic findings.

As I previously indicated, micro-level contexts may also be incompatible with the larger contexts in which they are embedded. For instance, an academic department in the humanities may not share its institution's goal of funneling students to STEM (science, technology, engineering, and math) departments because the job market needs employees with this type of educational background and the institution may benefit financially. Similarly, an academic department that values low class sizes will probably be at odds with recent trends throughout higher education to increase the student-to-faculty ratio in classes to defray other institutional costs.

Contextual interrelations and their incompatibilities implicate student information flows and, therefore, student privacy. Micro contexts may value student information differently, perceive student privacy in unique ways, and, therefore, could be at odds with one another. The clearest example of micro-contextual incompatibility in this study occurred with the advising offices at both of my case sites. As discussed previously, the advisors expressed significant concern about how student information should be used, noting self-fulfilling prophecies as a major issue. Other actors, like those in an institution's information technology department, sought freer flows of student information for analytic purposes. For advisors, then, invasions into a student's privacy may be autonomy reducing. But for actors who relied on student data for their work, student privacy and protections thereof needed to be questioned to optimize information flows.

Contextual Repression

When there are contextual incompatibilities, questions emerge about power and influence. Not every micro context can win out. Some values and goals related to using information, student information in this case, of one context will ultimately trump that of another. When one context does triumph over another, contextual repression takes place.

What contextual interrelations hint at is that student privacy, or privacy writ large, is not just a matter of informational norms and expected flows of information. Privacy is also a matter of political maneuvering and power plays among actors in micro contexts. Some micro-contextual goals, practices, and values will be repressed, while others will succeed.

For example, actors in information technology divisions may trump academic departments who wish to see student data used judiciously and to specific ends, and not just aggregated into data pools for later analysis. Here, the question is whether or not information technologists and information officers hold greater influence over actors in other micro contexts to do with the data what they want. Similarly, whoever becomes the ultimate student data gatekeeper may hold significant power over others in specific micro contexts. It is plausible that at some institutions the registrar's office holds dominion over student data in such a way that it gets to determine, based on this micro context's values, the attributes and transmission principles of all student information flows.

The issue at hand is that the framework of contextual integrity does not focus on the micro-contextual level and, therefore, we may lose some important findings. By looking at contextual interrelation and contextual repression, we may find out valuable information about how information flows are determined and by whom. This suggestion is not an attempt to negate the good contextual work that Nissenbaum has done with her framework of contextual integrity, instead it is meant to augment it by widening the focus of contextual privacy studies to include micro contexts with meso and macro contexts as well.

8.4. Student Control Over Information, Privacy Dashboards, and Identity Layers

Recent research suggests that students want control over their information, and my findings indicate that institutional actors feel that students should be given control. When students were asked about data practices in higher education, they made compelling statements in favor of control over their data and information; they also argued for fair and useful processes by which their institution would seek their consent and inform them of how it would use their data and information (Slade & Prinsloo, 2014). But, historically, higher education institutions have failed to promote informed consent practices within and outside of classrooms, using paternalistic justifications to warrant their information practices (Connelly, 2000). The discrepancy between what institutions think they can do with student data and what students expect is done with their data may “rupture the fragile balance of respect and trust upon which this relationship is founded” (Beattie, Woodley, & Souter, 2014, p. 424). This is a conflict that needs attention, and I provide recommendations to work towards a resolution in this section.

The sub-sections that follow provide a brief review of privacy-as-control as it relates to student privacy. I follow this conversation with an argument for why student privacy should be viewed in this way and how informed consent processes support privacy-as-control conceptualizations of student privacy. Next, I catalog the ways in which students have no control or are losing control over their information as learning analytics initiatives emerge. In order to support student controls over their information, I argue that higher education institutions should implement four informed consent principles, create student privacy dashboards, and design technical identity layers with privacy respecting technology.

Student Privacy as Control Over Data and Information

Privacy as a form of information control is a dominant theme in scholarly literature, serves as the basis for legal doctrine, and has even informed important Supreme Court decisions (Nissenbaum, 2010; Solove, 2008). According to Alan Westin’s (1967) seminal text, *Privacy and Freedom*, privacy is an individual’s “right to determine for themselves when, how, and to what extent information about them is communicated to others” (p. 7), which is in essence the right to control information about one’s self.

A control approach to student privacy does not assume institutional actors are not aware of student information, instead it supports actions that enable students to determine who can access information about themselves and limit to whom and under what conditions it is disclosed (Fried, 1968; Froomkin, 2000). Privacy-as-control is biased towards individual choice and treats information as a part of one's person that "flows naturally from self hood" (Solove, 2008, p. 26). If we were to define student privacy as a student's right to control data and information about herself, her right may supersede a university's claim to use that information. In effect, students could manage "the flow of personal information in all stages of processing–acquisition, disclosure, and use" (Kang, 1998, p. 1209) within the university's domain.

Learning analytics presents unique issues that dissolve student control over their information due to a number of technological conditions. First, data-driven practices make student lives transparent and but keep data-driven practices out of sight and, therefore, out of students' minds and without the wherewithal to argue for alternative practices. Second, students may wish to keep information private by expecting their university to deidentify their data, but the connected nature of databases and the power of personally identifiable analytic technologies makes deidentification efforts futile and reduces incentives for pursuing deidentification processes. Finally, higher education institutions continue to grow their privilege and power over students by exploiting their personal information, while students are left with few options to rein in flows of personal information. Even students were more aware of learning analytics practices, they increasingly face an institutional maze made up of bureaucracy, policy, and unclear legal interpretations that makes information control processes unapproachable, much less useful, if they exist at all (Solove, 2004; Tene & Polonetsky, 2013). Without some checks on student information flows that enable student control, the risk of digital dossiers becomes more significant, given that students will have little say in how their respective colleges and universities aggregate, store, and use their information for and against them.

What is arguably more problematic about students losing control over their information is the effect learning analytics and other data-driven practices have on individual autonomy, as discussed previously. In his theory of intellectual privacy, Neil Richards (2015) argues that growing cognitive surveillance in digital spaces (e.g., learning management systems and eBooks) changes the conditions under which students can seek out knowledge, form opinions, and build their identity. Colleges and universities are increasingly turning to these digital environments and learning analytics to enable and augment instruction, but in doing so, they may limit a student's intellectual curiosity, negatively affect their ability to work through intimate ideas, and stifle heretical speech without concern that someone or something will record and disclose sensitive ideas before their time has come. As a result, students may "guard [their] words [and their] thoughts" (Richards, 2015, p. 101), which runs counter to higher education's longstanding support of and protection for intellectual freedom and inquiry.

Informed Consent

In order to promote information control and autonomy, higher education should look to informed consent practices. Students must be informed about and consent to data-driven practices at their universities if they are to gain control over their information. Informed consent, or "notice and choice" (Tene & Polonetsky, 2013, p. 260), is the process by which individuals are notified of how a secondary party, like a university, will use information about individuals, personally identifiable or otherwise. It also informs individuals of their rights to privacy, as well as the express rights the second party retains regarding the information. After being informed of rights and information practices, individuals can then choose whether or not to agree–to consent–to the terms in front of them and enter into a relationship with the second party or not.

Even though informed consent acts as "the gold standard for privacy protection" (Waldo, Lin, & Millett, 2007, p. 48), it is not a cure-all where privacy is concerned (Flaherty, 1999).

Informed consent can be problematic because individuals are rarely fully aware of what they are agreeing to. With the increasing risks of downstream data use, Moore (2010) argues that the benefits we gain from consenting to one use of our data are outnumbered by the harms that can accrue when the same information is used for previously unknown purposes later on. The problem is that we can rarely envision the downstream uses of our information, the unequal benefit to the second and third parties to whom we disclose information, and the potential consequences for our privacy (Hui & Png, 2005; Marx, 1999).

Informed consent procedures are usually biased towards those who seek out personal information, and they can be predatory. More often than not, individuals must choose to opt-out of inclusive information gathering practices, not opt-in, which produces the effect that more information is gathered than not. Even when individuals are required to opt-in and take stock of their privacy choices, they often choose the data-driven service over privacy because they desire immediate gratification and prefer not to do the work to limit information disclosures (Acquisti, 2004). Furthermore, simply providing a sense of control–even if this is not the case–motivates individuals to consent (Brandimarte, Acquisti, & Loewenstein, 2013). Organizations and institutions seeking personal information, then, can benefit from complicated informed consent processes that are easy to agree to; as such, informed consent is often a predatory structure that does promote an individual's ability "to make meaningful, uncoerced choices" (Goldman, 1999, p. 103) through negotiation of information disclosure terms.

Despite informed consent's faults, it still holds great promise and does not warrant a complete dismantling of the protections it offers. Furthermore, seeking consent to gather and use personal information is still normatively expected and legally required in many contexts. Where students are concerned, enabling their ability to take control over identifiable information and data flows will, arguably, support ownership rights, fight against technological conditions that favor higher education institutions, and promote autonomy. But the current student information infrastructure runs counter to these goals. Students are often required to disclose information about themselves, knowingly and unknowingly, at various times in their educational career. The following sub-sections outline points of information disclosure, such as at admission, while in class, and when students interact with ubiquitous campus systems.

Comprehensive Profiles

One driving motivation of those who advocate for learning analytics technologies is to understand how different populations of students learn. In order accomplish this, they must develop profiles about learners that capture the unique attributes and characteristics of a student. To do this, higher education institutions mine the trove of information within admissions applications.

The information students reveal about themselves on applications for admission is not trivial; in fact, it is often sensitive and comprehensive in nature. Commonly, admission applications include questions related to a student's academic achievement, including high school and other collegiate transcripts along with standardized test scores; chosen academic program and professional ambitions; demographic and socioeconomic information; family networks and their academic achievement level and association with the institution; and the student's financial standing. Some applications ask for descriptive essays related to the student's reading habits and cultural interests; others include sensitive questions about the prospective student's academic disciplinary and criminal history. Others may even solicit answers regarding the student's religion, sexual orientation, and gender identity (see Caldwell, 2012; Hoover, 2011; Steinberg, 2010). In total, this information serves to build comprehensive individual profiles that colleges and universities expend significant amounts of resources to achieve.

The problem is that higher education institutions rarely (if ever) fully inform their prospective students about how the details of their lives revealed on admissions applications will be used and by whom. Clearly, students expect that these applications will inform admissions decisions, but they fail to intuit downstream uses and institutions do not explicitly explain information practices reliant on this store of personal data. In fact, applications for admission, the point at which we may expect universities to establish informed consent, may not even express student privacy rights, especially with regard to information control; some institutions even claim a property right to prospective students' information. This practice is especially problematic considering that a student may feel that they have no option but to reveal all of the sensitive details about their lives, as there is always the chance that the student will be denied admission if she fails to provide the information a college or university asks for.

Classroom Disclosures

Besides the application for admission, students also reveal sensitive information about themselves by creating profiles on third-party applications–applications their instructors often require them to use in courses. Rarely does an institution, nor its instructors, fully inform students of the ways in which the companies responsible for these learning applications use and protect the information students disclose as users.

Consider the example of Piazza, a company that offers question and answer functionality as a stand-alone application or with direct integration into common learning management systems. Over 750,000 students at 1,000 institutions in 70 countries use Piazza (J. Gilmartin (Piazza representative), personal communication, March 26, 2015; Piazza, n.d.) to share information about themselves, access course materials, and communicate with their peers, instructors, and teaching assistants. To Piazza, data derived from students–including disclosures about their class history, internships, majors, and expected graduation year (Piazza Careers, n.d.)–has helped them to build a secondary service, Piazza Careers, which enables technology companies to court students for jobs if they fit a specific profile; of course, that is after the companies purchase access to Piazza Career's store of student data, analytics, and other services.

Higher education institutions often enter into contracts with third-party educational technology services–sometimes at no cost to the institution–in order to get access to compelling and very useful teaching and learning applications; in return, educational technology companies get access to valuable student data. With regard to Lighthouse, the vendor of the technology was able to use the student data it gleaned from both Hammond University and Saint May State University to improve its product. The evidence suggests that educational technology vendors enter into fruitful partnerships with institutions to build better technologies and scrape user profiles for information to build secondary, remunerative applications. These developments should be of no surprise to students, yet they often are.

While institutions often negotiate terms of service agreements on behalf of their students, students are rarely aware of the details of those agreements nor are they accessible. Simply because policies or memoranda of understanding exist that detail how student data should be used, we cannot assume that such agreements work to the benefit of students or protect their privacy. In fact, a lack of transparency regarding these agreements and a failure to fully inform students about how third-party companies use their data raises immediate concerns and questions. It may be that institutions are withholding information about data practices to keep student privacy concerns at bay, concerns that could potentially derail beneficial contracts with vendors or create friction in relationships between the student body and instructors, and by extension the institution.

Where in-class disclosures are concerned, colleges and universities may claim that hinderances to student information flow, like requiring informed consent, impede necessary institutional practices, like instruction. Recall that §99.31 of FERPA allows institutions to disclose private, identifiable student information–without informing students–to anyone within the institution who has a "legitimate educational interest" or to a third party who provides "institutional services or functions," like an educational technology company. But as we saw with Piazza and Lighthouse, third parties can use student data for their own benefit. It is true that under §99.33 of FERPA third parties cannot disclose that information downstream without proper consent from students, but that does not preclude companies from anonymizing data to build, improve, and sell services built on top of the trove of data without students knowing.

Continuously Tracked

If the Big Data ethos of "all the data we can get" becomes normative in higher education, it will encourage institutions to continually track a student's digital and physical movements and activities. By creating an infrastructure that supports that type of data gathering, students will unknowingly disclose information about themselves on a daily basis. What is most problematic about these types of data disclosures is that the technology that enables them seems benign, is embedded in normal day-to-day activities, or is sequestered from view completely. Students simply are not aware of the complex web of data capture technologies that store, aggregate, and analyze their information. Yet, there are particular types of data tracking that students need and, arguably, should be informed about in order for them to express some control.

Tracking technologies that capture geolocation, temporal data and, metadata raise serious concerns about student surveillance. Systems that can map in real time (or closely to) a student's physical and/or digital location and the time of her movements or activities disturbs our normative expectations and riles up our concerns regarding dataveillance. It is plausible that universities will use geolocation tracking to incentivize less social and more academically-oriented movements, like visiting the library, in order to improve learning outcomes. And special categories of students may come under higher scrutiny than others, such as student-athletes who are already under constant surveillance where their social media is concerned (see Reed, 2013). In both cases, students may more closely regulate their behaviors due to concerns about how their data trails could be used against them (Hier, 2003).

Students may be worried that insights mined from dataveillance will become a part of their permanent educational record, and they may wish to gain more control over this sensitive data and information. Anyone who has had the privilege of experiencing college would balk at data-based revelations that detail, for instance, a student's level of engagement with her courses, discover whether or not she's socially connected with peers, and reveal if she's experiencing emotional issues. College is a formative time for identity development and exploration, socially and intellectually, and data practices that record a student's life in detail worries students. Recent revelations at Stanford University (see Pérez-Peña, 2015), where students discovered that their institution logged when they used their identification cards to unlock doors, substantiates these concerns.

All of these points of information disclosure signal that new mechanisms need to be put into place to establish student privacy as a student's right to express control over their identifiable information. In the following sub-sections I recommend ways by which institutions, in combination with technology vendors, can reestablish informed consent and promote student privacy preferences with data dashboards and improved technical identity layers.

Recommendation One: Reestablish Informed Consent

Higher education institutions may rebuke claims that their information gathering practices require informed consent standards set by the Fair Information Practice Principles, as the principles are already embedded within FERPA (Ramirez, 2009). But FERPA enables institutional liberties with student information flows and consent practices, sometimes to the risk of student privacy. But as Rubel and Jones (forthcoming) point out, such freedoms also enable higher education institutions the ability to craft new privacy protections, to treat FERPA as the policy "floor" and not the "ceiling" (Family Policy Compliance Office, 2011, p. 5) of how institutions should regulate and safeguard student information flows. In this sub-section, I suggest that institutions should use these freedoms to develop a comprehensive informed consent scheme using data privacy dashboards built on top of a technical identity layer.

What is clearly lacking are procedures and policies that explicitly detail how student information will be acquired, who will use it, and for what purposes. New processes must be put in place to provide granular detail of information practices and provide students an opportunity to dictate, within limits, how their information should flow. To this end, I promote four principles to improve informed consent.

First, informed consent should be timely. Student information disclosures occur at discrete times during a student's tenure at an institution. Students reveal a significant amount of information about themselves on the application for admission, but they also provide a steady drip of data as they interact with information systems or reveal information about themselves when they take certain classes. To resolve these issues, informed consent processes should be instituted just-in-time when they disclose data.

The timeliness principle has its limitations, however. Certain disclosures, like geolocation data from RFID-enabled ID cards, happen on a daily basis. In cases like this, for example, students should be made aware of this type of disclosure when they apply for a student ID card. It may be easy for students to forget that they have consented to information disclosures, like these, that become part of daily life. Therefore, institutions should use annual notifications to remind students that they consented, what they consented to, and clearly detail any changes to the information disclosure practice that may require them to reconsent.

Second, informed consent should map actors involved with information flows. For informed consent to be useful, it must describe the actors involved with student information in as detailed a way as possible. For example, it is possible to map how student information will flow with regard to learning management system-based analytics. In this case, we can expect that flows will occur between the student, her instructor, and, potentially, instructional services professionals (e.g., instructional designers). If institutional researchers are also expected to access this data, that flow of information should also be described. Similarly, if the vendor of the learning management system will have access to student data in order to further develop its system, that should also be detailed along with any restrictions on the flow (e.g., if it will be minimized and kept confidential).

It may be that situations arise where institutional or third-party actors need access to student information that were unforeseeable when the flow map was constructed. This is to be expected, but institutions should shy away from using loopholes like FERPA's "legitimate educational interest," which enables any number of actors, within and outside the institution, to access student information. Institutions should make a good faith effort to clearly describe to students who will access and use their information and to what ends.

Third, informed consent should describe potential benefits and harms of information disclosures. Not all disclosures of private information are overly concerning (e.g., one's gender), but some are (e.g., one's sexual orientation). For informed consent procedures to be useful for students, they should clearly describe the primary ends to which the information will be put (e.g., improve learning outcomes for a department, enhance institutional efficiencies by studying student movements). As for harms, colleges and universities should be especially cognizant of third party interests in accessing student information for their own gain, so informed consent processes should demonstrate an awareness of downstream effects and the potential harm they can cause to a student in the future.

Finally, the first round of informed consent should not be the final say. As mentioned previously, students should be made aware of any substantive changes to information practices and flows. Upon review of these changes, students should be required to reconsent in order for the institution to continue to gather and use student data for the particular purpose under question. Some changes to information practices will not need students to reconsent, but students should still be made aware of minor amendments to what they consented to through timely and useful notifications.

Recommendation Two: Build Student Privacy Dashboards

Students need a way by which they can express their privacy preferences, and privacy dashboards present such an opportunity. As previously mentioned, learning analytics technology shares its statistical findings and predictions with institutional actors through visualizations (e.g., charts, trend lines, etc.). But, in order to promote self-awareness and encourage reflection among learners, some proponents of learning analytics advocate for creating data dashboards specifically for students (see Clow, 2012; Duval et al., 2012). Data dashboards enable self-management over learning, and they also serve as a model for how informed consent could be improved.

Student privacy dashboards would provide a central location where students would be informed about information practices that use their data and give them opportunities to opt out of personal data flows at a granular level. With such applications, students could dictate how they are informed (e.g., e-mail or text) and use toggle-like switches to determine what aspects of their information and data can be used for very specific purposes. Furthermore, privacy dashboards could archive and provide simple access to relevant information policies, as well as important communications from their institution regarding privacy concerns (e.g., data breaches).

While student privacy dashboards advantage student control over their information, they also benefit the university. Higher education institutions will require some data and information for business and other purposes, and they will need to set defaults that allow for these sorts of information flows. In order to achieve these ends, administrators of data dashboards should have the technical ability to disable some student data controls.

It may seem that institutional power over student data controls defeats the purpose of a privacy dashboard, but it actually presents opportunities for transparency. With many Big Data practices, "the purposes to which data is being put, who will have access to it, and how identities are being protected" remains opaque (Sclater, 2014, p. 20). Efforts–like student privacy dashboards–that inform students about how their institution uses their data and for what purposes hold potential to reduce opacity, lessen concerns about worrisome abuses brought about by analytics, and keep levels of trust high between students, faculty, and other institutional actors (Polonetsky & Tene, 2014; Richards & King, 2014; Slade & Prinsloo, 2014). If institutions need to set defaults on information flows, they can use privacy dashboards to communicate these needs and justify for whom that information is useful and beneficial.

Recommendation Three: Improve the Identity Layer

The success of student privacy dashboards depends entirely on the technical infrastructure on which they are built, specifically the identity layer of campus information technologies. In order to identify students, authenticate their credentials, and authorize access to campus services, higher education institutions use identity management technologies, such as Active Directory services and Single Sign-On protocols (Bruhn, Gettes, & West, 2003). These systems serve as the gatekeepers to student information systems, online learning applications, and to the campus's networked infrastructure, among any other application, system, or device that requires identity authentication. In combination with the web of campus technology, identity management systems establish an identity layer on which learning analytics relies and enables the technology to attribute data to identifiable students.

What identity layers in higher education do not do, at present, is act as a platform for privacy preferences (P3P). According to Lawrence Lessig (2006), P3P technology acts as a machine-readable protocol that enables technologies to communicate, assess, and respect individual privacy choices set in applications and digital tools, such as student privacy dashboards. Consider an application of P3P technology in an eAdvising-based learning analytics scenario. If a student chooses within her dashboard to deny others the right to use her geolocation information, the eAdvising system would have to respect that choice and would not pull the data from a data warehouse.

While P3P holds promise for protecting privacy, it has historically failed. P3P technology originally existed as a World Wide Web Consortium recommendation that emerged out of privacy working group, which involved academic, government, and private sector input (W3C Workshop on the Future of P3P, 2002). But a test of over 33,000 websites and their implementation of P3P found significant errors and limitations that limited the efficacy of the privacy protections they extended to users; moreover, researchers argued that there simply was not enough incentive (e.g., regulation or social pressure) for websites to respect privacy preferences (Leon, Cranor, McDonald, & McGuire, 2010). The Electronic Privacy Information Center (2000) argued that the protocol was too complicated, too limiting for users, and not legally enforceable.

One reason for P3P's demise was that it was attempting to work as a protocol between websites and browsers at a large scale. The odds were not in the protocol's favor given the diversity of actors and a lack of strict standards for designing websites. Furthermore, the Web had matured to a point where advanced privacy protections were not a part of the technical and social fabric of the technology. Simply put, there were just too many variables to enforce P3P and compel actors to design online environments with the protocol in mind. But what ultimately led to the demise of P3P on a large scale does not necessarily apply at a smaller scale.

An implementation of P3P on a college campus is, hypothetically, much easier to implement. Most importantly, the scale at which P3P would need to be enacted is much smaller. In essence, P3P settings would only need to be respected within the technical domain of the university and educational technology vendors. Additionally, the technical infrastructure of a university is in flux, which is a great advantage. The interest in and adoption of emerging learning analytics technologies signals that there are changes on the horizon for universities. More systems will need to be interoperable in order to maximize the purported gains of learning analytics. A key part of optimizing learning analytics is finding a way for data and information to flow freely between systems. The problem is that there are still very few data standards in existence that control the flows of student information on campuses; manual data exporting, cleaning, and importing is still the norm. An excellent opportunity exists to build P3P technology alongside new data standards, applications, and networks as the technical infrastructure develops to better support learning analytics initiatives.

While institutions could build student privacy dashboards, such efforts would be done in vain unless universities code P3P technology into their systems and expect the same of their vendors. For established technology systems, this may a difficult task, but the nascency of emerging educational technology systems and learning analytics provides an opening for institutions to create policy and effect change in ways that promote P3P and student privacy dashboards, before related conversations close off and technologies stabilize (Bijker, 1995). It may be difficult for individual institutions to influence vendors to design their technologies with P3P technology, but consortia, like the Washington Higher Education Technology Consortium (About WHETC, 2011) and Unizin (2014a), may have enough collective weight to support P3P adoption among leading technology vendors.

Chapter 9. A Conclusion to the Study

9.1. Introduction

The study has shown through richly described findings and a discussion of its important themes that student privacy is at stake in higher education institutions who pursue learning analytics technologies. With the study complete, this final chapter provides a summary of the study's findings and contributions as they relate to the research questions outlined in section 1.3. Other sections in the chapter also address implications of the study for practitioners and information policy, as well as opportunities for future research on learning analytics and student privacy. The final two sections in the chapter provide an overview of the study's limitations and leave the reader with some final remarks about the importance of studying student privacy in relationship to emerging data-driven educational technologies.

9.2. A Summary of the Study's Findings and Contributions

The study began with a review of the literature, which revealed a number of significant student privacy issues related to learning analytics. Additionally, the review of the extant literature showed that little work had been completed to date to understand how institutional actors perceive and address student privacy issues while building capacity for learning analytics. This study worked to fill that important gap.

The research questions that framed this study were as follows:

- How do institutional actors who are building capacity for learning analytics perceive related student privacy issues?
- How do those perceptions influence their practice with the technology?

- How do actors resolve privacy problems as they encounter them?
- What institutionally contextual factors influence student privacy practices related to learning analytics?

The exploratory design of this study allowed my participants to address these questions from their own perspective and based on their own experiences. As such, the data led to a number of emergent themes that addressed the research questions.

Institutional actors perceived that student privacy was a concern and it impacted their work with learning analytics. Most notably, the findings indicated that actors were unsure of their obligations to protect student privacy and sought advice from registrars, members of institutional legal counsel, and their institutional review boards to determine what, if any policies, needed to be respected. There were no specific institutional policies in place to guide their use of learning analytics, the learning analytics initiatives did not fall under the purview of institutional review boards, and FERPA provided actors significant leverage to do with student data and information what they wanted.

Institutional actors raised concerns about the surveillance aspects of learning analytics, and in doing so evoked the specter of Big Brother. However, the data also indicated that institutional actors were reconsidering their obligations to protect student privacy, since they felt norms were changing among their students in light of how students shared information about themselves using social media technologies. Most importantly, there were conflicts among actors regarding the degree to which they should aggregate and analyze all the data they can get about students. While more data could possibly improve the predictive models learning analytics technologies use, the data practice also had significant implications for student privacy.

The privacy issues impacted how actors employed learning analytics. For instance, the data showed that privacy concerns required technological changes to one learning analytics system to protect student privacy. Furthermore, the privacy issues were strong enough to motivate one of the institutions to hire a new information officer to build a comprehensive data governance strategy. Moreover, particularly strong concerns about the potentially negative effects of predictive analytics impacted how and what advisors told students about the information borne from one learning analytics system.

The data also indicated relevant contextual factors that influenced student privacy practices. For instance, actors at Saint May State University use ad hoc policy to determine student privacy protections with respect to learning analytics; whereas actors at Hammond University did their due diligence to determine their institutional and federal privacy obligations, as well start to develop more comprehensive privacy protections as they built of data governance mechanisms.

This study has contributed to a more fully developed understanding of why student privacy is at stake at institutions who employ learning analytics technology. Moreover, the study has provided further depth to the scholarly community's understanding of Nissenbaum's (2010) theory of contextual integrity by putting focus on the value of interrogating micro-contextual elements as they relate to questions of privacy. Also, the study has presented a new approach to further enhance student privacy by suggesting that higher education institutions and educational technology vendors could develop data dashboards and technical identity layers to enable students granular control over their data and information.

9.3. Implications and Future Research Opportunities

The study's findings and the discussion that resulted from it points to a number of implications. In the sub-sections that follow I suggest implications for practitioners, propose information policy institutions should consider, and I close with recommendations for researchers to consider.

Implications for Practitioners

The findings highlight a number of important implications for practitioners, such as those I interviewed. For instance, there are important lessons herein that can apply to other contexts related to how institutional actors and their respective institutions should be transparent about learning analytics. For instance, instructors should consider writing privacy statements in their syllabi if they use learning analytics technologies in order to make students aware of how their instructor is using personally identifiable data and information about them. The same can be said for advisors. If advisors use predictive analytics to understand a student's aptitudes and probability of success on a particular academic path, advisors should share with students that they are doing so and how analytic systems establish predictive scores. At an institutional level, colleges and universities should make it obvious and clear that they are using analytic technologies and to what ends.

More importantly, practitioners need to consider the harms of aggregating all the data they can get for learning analytics, especially given that amassing significant amounts of data is the dominant approach when pursuing Big Data's insights. Practitioners need to step back from this persuasive way of analyzing personally identifiable data and reflect on how the information they obtain could negatively affect important relationships (e.g., between instructors and students), identity formation (e.g., how students perceive their academic potential), and institutional values (e.g., developing introspective, self-motivated students).

Implications for Policy

The findings suggest that there are significant implications for institutional policy. Colleges and universities who seek the benefits of learning analytics must consider the privacy issues involved with their projects. The problem is that institutions have long relied on FERPA for guidance where student privacy is concerned. Since the data indicates that FERPA can no longer do a comprehensive job of protecting student privacy, institutions are tasked with developing specific policies to cover the variety of privacy problems inherent to learning analytics.

Specifically, this study indicates that institutions need to consider creating explicit information policy in three areas. First, institutions need to define what data and information is included in an educational record and make that definition transparent to students. Establishing a detailed definition will create clear boundaries around student data and information, enable students to express their privacy rights, and guide contract negotiation between institutions and technology vendors in ways that clearly explain that identifiable data needs to be accessible, that is if it is defined as part of an educational record.

Second, new policies should make it explicitly clear which actor roles should have access to particular types of student data and information, instead of relying on ambiguous and broad interpretations of FERPA, such as the "legitimate educational interest" clause. Moreover, these policies should require mechanisms by which actors request access to student data and provide clear justifications for why they need access in the first place (see The Open University, 2014 for an example).

Finally, institutions should establish policies that define what types of student data can be used for the purpose of learning analytics, especially where sensitive data is concerned. For instance, such a policy might categorize a student's academic history as low in sensitivity. But if builders of predictive models want to use a student's gender, race, or sexual orientation, which all represent sensitive data types, the policy should require the builders to justify to data gatekeepers or governance boards why this sensitive data is necessary for their model and how they will protect against profiling and discrimination.

Implications for Researchers

The developing nature of learning analytics has created a number of veins for research, and the findings and discussion in this study point to some potentially fruitful new areas of inquiry. First, the discussion signals that new work needs to be done in micro-contextual studies of student privacy. The findings indicate that there may be frictions between particular actor roles and their institution with regard to how the university protects (or fails to protect) student privacy; this incongruity may impact how these actor roles use or rebel against learning analytics.

Researchers also need to consider how institutional policies, or a lack thereof, guide uses of student information. The findings indicate that, at least among my two case sites, there is little institutional policy and, in fact, institutional actors rely on ad hoc decisions to determine how actors should use student information. Researchers should study how ad hoc policies are made *in situ* and the effects thereof on student privacy. As more institutions develop capacity for learning analytics, the body of research in this area would benefit from studies that work to understand how colleges and universities develop specific privacy policies and justify how they open or restrict information flows.

Students were lacking a voice in this study, and the body of research related to learning analytics would greatly benefit from studies that incorporate student perspectives on privacy. Very little is known about this area of the research. It could be that students, as some of my participants explained, do not care about how institutions use their data and information; my anecdotal evidence suggests that students do, in fact, care about how institutions use their information and employ learning analytics technologies. Practitioners would be well-served by research that explores student privacy from a student's perspective, and theoretical work is especially needed given the current dearth of conceptual research in this area.

The move to aggregate all the data institutions can get by developing interconnected information systems brings about interesting questions about what is necessary or appropriate for developing learning analytics. Researchers should consider employing "epistemological Luddism" (Winner, 1977, p. 330) alongside practitioners to investigate what technical systems and data flows institutions can live without. Epistemological Luddism questions if technological arrangements run counter to human values, norms, and goals. As a method, epistemological Luddism asks those who employ it to "dismantle or unplug" (Winner, 1977, p. 331) parts of technological assemblages. Researchers should study dismantling effects, make an assessment of whether or not actors who rely on the system can live without the part unplugged from the system, and determine if the unplugged part aligns with human values; if it does not, the part should be redeveloped to be compatible with human expectations and needs or left dismantled.

Finally, researchers, especially those in higher education policy, philosophy, and history, need to take stock of the values driving higher education and determine if learning analytics is compatible with them. As I discussed in earlier work with Alan Rubel (Rubel & Jones, forthcoming), there is an open question as to whether or not learning analytics is compatible with some core values of higher education, namely producing autonomous citizens who can think critically, act according to their values, and participate in our liberal democracy. Proponents of learning analytics have not proven that the technology is compatible with these values; and if the technology runs counter to them, then we should question its legitimacy. However, it may also be the case that higher education's longstanding values are no longer applicable given immense social, political, and economic pressures on colleges and universities to adapt. It could, in fact, be the case that learning analytics is compatible with what society wants and needs out of its colleges and universities today and in the future.

9.4. Limitations

All research has limitations, and my study was not without its own weak points. In the following sub-sections, I address limitations of this study with respect to problem framing, researcher skill and bias, sampling problems, the transferability of the study's findings, and claims of goods and harms.

Problem Framing

I framed this study about learning analytics and student privacy problems thereof in a particular way in order to scope the project. However, in so doing I set out research questions that privileged some assumptions to the disservice of others. For instance, by framing the problems and research questions in the way that I have, I did not account for the fact that there may, in fact, be very beneficial uses of learning analytics. My focus on student privacy problems has not considered the good research in progress that is advancing our understanding of how students learn, for instance.

Moreover, my work does not address ways how higher education institutions value student data differently from the students themselves. Instead, I make a prima facie assumption that the interests of the student–her privacy interests–should be accounted above the interests of the institution; there are defensible positions where institutions should not respect a student's privacy wishes (e.g., in the case of an active shooter on campus where student safety may depend on accessing private student information). Some will argue student privacy should be respected regardless of institutional interests or even variation in student interests (e.g., some students may not care that their instructors are employing predictive analytics). This is an argument I did not make, because I sought institutional perspectives on student privacy; forcing such a position would have been antithetical to my methodological approach. Nonetheless, such an argument is an interesting one worth considering. These problem framing limitations are all worthy of their own separate research, but they were not the subject of this dissertation and outside the scope of my data analysis.

Researcher Skill and Bias

Qualitative research, and grounded theory in particular, produces large amounts of data that can be daunting to the untrained novice researcher. And grounded theory researchers need to be carefully trained in order to carefully sift through the deluge of data and come out on the other side with a conceptually-driven story. If the researcher lacks the necessary skills, she may not be able to craft an analytic story out of the mass of data.

A dissertation is a starting point in a researcher's career, and I admit that one of the implications of this study is that I am still honing my qualitative craft. I believe that I employed grounded theory methods sufficiently in this study, and I was able to successfully build a data-based story out of 600-plus pages of transcripts. However, I willingly admit that more experience with the methods may have enabled me to develop more insightful findings, work more efficiently, and develop conceptually-advanced themes from the data.

Researcher bias is also related to researcher skill as a limitation. New grounded theory researchers may not understand what mechanisms need to be put in place to reduce bias due to a lack of experience. I cannot deny the role of my personal values and ontological perspective in this study. However, I attempted to reduce researcher bias by employing methods that induced reflexivity, such as transparently expressing my values herein and writing journal entries while in the field (see section 6.5. for more on these methods and how I employed them). Using these methods, I believe I reduced researcher bias, but such biases will always exist in qualitative research.

Sampling

I used two sampling methods for this study: criterion and nominated sampling. Criterion sampling proved effective for scoping relevant participants, but nominated sampling created limitations in two ways. First, some of the nominated individuals for the project were unresponsive when I asked them to participate in this study. According to my interviewees, these nominated individuals were relevant to my study and could have added greater insight in my findings and subsequent discussion. Second, since some of the nominated participants did not participate, my theoretical sampling strategies were limited as well. As a result, it is possible that my data saturation was less than ideal. Both of these sampling limitations are less of an issue in other studies when participant criteria are less limiting and participants are greater in number and more widely accessible. However, I must recognize that, for this study where there were only limited numbers of individuals who fit the criteria and were willing to be interviewed, sampling limitations existed.

Transferability

Transferability of findings and discussion themes are always a concern in qualitative research. Qualitative researchers are usually interested in maximizing the transferability of their research since transferability increases its usefulness across contexts. I believe that I successfully described the emergent themes using thick description and abstracted useful conceptual ideas in the discussion in such a way that other researchers and practitioners can make use of this study. However, one limitation of this study related to transferability exists. Many of the privacy problems, and the lessons learned about them, were contextually dependent, meaning that they occurred in relation to specific applications of learning analytics technologies and initiatives at particular institutions. As such, readers should be aware of the fact that the degree to which the findings are transferable will vary from institution to institution.

Claims of Goods and Harms

This study can make no empirical claims regarding the goods and harms of learning analytics. I can state neither that learning analytics improves, in any way, the welfare of individuals or higher education institutions, nor that it negatively impacts the lives of those I interviewed or the students they serve. The work I have done herein is based on the perceptions of institutional actors about student privacy issues as they related to their learning analytics initiatives. Therefore, the data I gathered and the conclusions I made cannot definitively state that any individuals experienced any kind of positive (e.g., improved student learning outcomes) or negative (e.g., real invasions of student privacy) effects that resulted from using the technology. Such claims would require different data and different research questions.

9.5. Final Remarks

Public and scholarly interest in student privacy has been low since related conversations waned in the late-1970s. But with the development of Big Data practices, such as learning analytics to direct student behaviors using data-driven insights, both the public and the scholarly community has taken interest in student privacy once again, and with good reason. The years students spend in higher education institutions are some of the most formative years of their life; they represent a time of intellectual and personal exploration. Student privacy provides a protective sphere against those who wish to manipulate student lives during this sensitive time. While learning analytics may reap significant benefits for students and the colleges and universities in which they are enrolled, advocates have yet to prove the technology's efficacy nor have they shown they can adequately protect student privacy. Until they do so, the public needs to continue to push for student privacy protections and the scholarly community needs to investigate the technology's privacy weaknesses.

Finally, proponents and critics of Big Data practices should continue to keep a watchful eye on how higher education institutions navigate this socio-technical space. The history suggests that colleges and universities have and continue to see student data and information as a valuable resource, and to maximize this resource they are willing to expend significant resources. Most importantly, higher education institutions have significant control over their information technology infrastructure and, therefore, are in a privileged position; they are be able to optimize it to aggregate and analyze data in ways the commercial sector and even government institutions may only dream of. The outstanding questions are thus: Will colleges and universities ethically build information infrastructures in support of data analytics with benevolence and careful forethought regarding the privacy issues? Or will they build socio-technical systems in ways that disenfranchise their data subjects, permit hidden data practices, and eschew information control mechanisms, like others who have built capacity for data analytics programs? We should be concerned if higher education, as a value-conscience and morally sensitive institution, chooses the latter.

References

About IPEDS. (n.d.). *The National Center for Education Statistics*. Retrieved from http://nces.ed.gov/ipeds/about/

About WHETC. (2011). *Washington Higher Education Technology Consortium.* Retrieved from http://councilofpresidents.org/whetc/main.html

Acquisti, A. (2004). Privacy in electronic commerce and economics of immediate gratification. *Proceedings of the ACM Electronic Commerce Conference, USA*, 21–29. doi: 10.1145/988772.988777

Alexander, F. K. (2000). The changing face of accountability: Monitoring and assessing institutional performance in higher education. *Journal of Higher Education, 71*(4), 411–431. Retrieved from http://www.jstor.org/stable/2649146

Alfino, M., & Mayes, G. R. (2003). Reconstructing the right to privacy. *Social Theory and Practice, 29*(1), 1–18. Retrieved from http://www.jstor.org/stable/23559211

American Council on Education. (1927). Report of the committee on uniform records and reports of the department of superintendence of the National Education Association. *Research Bulletin of the National Education Association, 5*(5), 225–346.

American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115 (2009).

Anderson, R. (2015, April 10). EdData privacy update 4/10/2015 [Web log post]. Retrieved from http://www.dataqualitycampaign.org/blog/2015/4/eddata-privacy-update-4-10-2015

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance & Society, 12*(2), 185–196. Retrieved from http://queens.scholarsportal.info/ojs/index.php/surveillance-and-society/article/viewFile/bds_ed/bds_editorial

Arkin, H. (1935). Development and principles of the punched card method (Hollerith). In G. W. Baehne (Ed.), *Practical applications of the punched card method in colleges and universities* (pp. 1–20). New York, NY: Columbia University Press.

Arnold, K. E., Lynch, G., Huston, D., Wong, L., Jorn, L., & Olsen, C. W. (2014). Building institutional capacities and competencies for systemic learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, USA*, 257–260. doi: 10.1145/2567574.2567593

Arnsdorf, H. G. (1935). New York University. In G. W. Baehne (Ed.), *Practical applications of the punched card method in colleges and universities* (pp. 23–30). New York, NY: Columbia University Press.

Asay, M. (2013, May 11). Who's getting rich in the big data gold rush? *ReadWrite*. Retrieved from http://readwrite.com/2013/11/11/whos-getting-rich-in-the-big-data-gold-rush

Ash, K. (2010, January 29). Student ID cards sport new digital features. *Education Week*. Retrieved from http://www.edweek.org/dd/articles/2010/02/03/02id.h03.html

Asilomar Convention. (2014). *The Asilomar convention for learning research in higher education*. Retrieved from http://asilomar-highered.info/index.html

Atkinson, R. C., & Blanpied, W. A. (2008). Research universities: Core of the US science and technology system. *Technology in Society, 30*(2008), 30–48. doi: 10.1016/j.techsoc. 2007.10.004

Babey, E. R. (2006). Costs of enterprise resource planning system implementation–and then some. *New Directions for Higher Education, 2006*(136), 21–33.

Baer, L. & Campbell, J. (2012). From metrics to analytics, reporting to action: Analytics' role in changing the learning environment. In D. G. Oblinger (Ed.), *Game changers: Education and information technologies* (pp. 54–66). Louisville, CO: EDUCAUSE.

Bagley, C. H. (1967). Institutional research and information control. Paper presented at the meeting of The Association for Educational Data Systems, USA. Retrieved from http://eric.ed.gov/?id=ED014794

Balderston, F. E. (1974). The design of uses of information systems. *New Directions for Institutional Research, 1974*(1), 47–66. doi: 10.1002/ir.37019740106

Barbaro, M., & Zeller Jr., T. (2006, August 9). A face is exposed for AOL searcher no. 4417749. *The New York Times*. Retrieved from http://www.nytimes.com/2006/08/09/technology/09aol.html

Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. *Proceedings of the Second International Conference on Learning Analytics and Knowledge, USA*, 259–262. doi: 10.1145/2330601.2330664

Basken, P. (2010, January 3). States embrace student-data tracking, with prodding from White House. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/States-Embrace-Student-Data/63376/

Bauman, Z., & Lyon, D. (2013). *Liquid surveillance: A conversation*. Cambridge, United Kingdom: Polity Press.

Beattie, S., Woodley, C., & Souter, K. (2014). Creepy analytics and learner data rights. *Proceedings of Rhetoric and Reality: Critical Perspectives on Educational Technology, NZ*, 421–425. Retrieved from http://ascilite2014.otago.ac.nz/files/concisepapers/69-Beattie.pdf

Becker, H., & Geer, B. (1960). Participant observation: The analysis of qualitative field data. In R. N. Adams & J. J. Preiss (Eds.), *Human organization research: Field relations and techniques* (pp. 267–289). Homewood, IL: The Dorsey Press.

Beitin, B. K. (2012). Interviewing and sampling: How many and whom. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (pp. 243–253). Los Angeles, CA: SAGE Publications.

Benn, S. I. (1984). Privacy, freedom, and respect for persons. In F. D. Schoeman (Ed.), *Philosophical dimensions of privacy: An anthology* (pp. 223–244). Cambridge, UK: Cambridge University Press.

Berg, A. (2013, September 13). Towards a uniform code of ethics and practices for learning analytics [Web log post]. Retrieved from: http://ict-innovatie.uva.nl/2013/09/13/towards-a-uniform-code-of-ethics-and-practices-for-learning-analytics/

Bernard, H. R. (2000). *Social research methods*. Thousand Oaks, CA: SAGE Publications.

Bertaux, D. (1981). From the life-history approach to the transformation of sociological practice. In D. Bertaux (Ed.), *Biography and society: The life history approach in the social sciences* (pp. 29–45). London, UK: SAGE Publications.

BeVier, L. R. (1995). Information about individuals in the hands of government: Some reflections on mechanisms for privacy protection. *William and Mary Bill of Rights Journal, 4*(2), 455–506. Retrieved from http://scholarship.law.wm.edu/wmborj/vol4/iss2/3/

Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress and recommendations* (Report No. ERS1207). Retrieved from the EDUCAUSE Center for Applied Research website: http://net.educause.edu/ir/library/pdf/ers1207/ers1207.pdf

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief* (U.S. Department of Educatoin No. ED-04-CO-0040). Retrieved from http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

Bijker, W. E. (1995). *Of bicycles, Bakelites, and bulbs: Toward a theory of sociotechnical change*. Cambridge, MA: The MIT Press.

Bloustein, E. J. (1964). Privacy as an aspect of human dignity: An answer to Dean Prosser. *New York Law Review, 39*. 962–1007.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method.* Englewood Cliffs, NJ: Prentice Hall.

Boehner, J. (2005, June 17). A monster database is not the answer. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/A-Monster-Database-Is-Not-the/5808/

Bollier, D. (2010). *The promise and peril of big data* (Research Report). Retrieved from the Aspen Institute website: http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

Bok, D. C. (2006). *Our underachieving colleges: A candid look at how much students learn and why they should be learning more*. Princeton, N.J.: Princeton University Press.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* (R. Nice, Trans.). Cambridge, MA: Harvard University Press.

Bowker, G. C. (2005). *Memory practices in the sciences.* Cambridge, MA: The MIT Press.

Bowker, G. C. (2013). Data flakes: An afterword to "raw data" is an oxymoron. In L. Gitelman (Ed.), *"Raw data" is an oxymoron* (pp. 167–171). Cambridge, MA: The MIT Press.

boyd, d., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication, and Society, 15*(5), 662–679. doi: 10.1080/1369118X.2012.678878

Brandeis, L. D. (1913, December 20). What publicity can do. *Harper's Weekly*. Retrieved from http://louisville.edu/law/library/special-collections/the-louis-d.-brandeis-collection/other-peoples-money-chapter-v

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2013). Misplaced confidences: Privacy and the control paradox. *Social Psychology & Personality Science, 4*(3), 340–347. doi: 10.1177/1948550612455931

Brazy, D. (2010, May 4). Ariz. college to position sensors to check class attendance. *The Badger Herald*. Retrieved from http://badgerherald.com/news/2010/05/04/ariz-college-to-posi/

Brown, M. (2011). *Learning analytics: The coming third wave* (ELI No. ELIB1101). Retrieved from the EDUCAUSE Learning Initiative website: http://net.educause.edu/ir/library/pdf/ELIB1101.pdf

Bruhn, M., Gettes, M., & West, A. (2003). Identity and access management and security in higher education. *EDUCAUSE Quarterly, 4*, 12–16. Retrieved from https://net.educause.edu/ir/library/pdf/eqm0342.pdf

Bryant, A. (2002). Re-grounding grounded theory. *Journal of Information Technology Theory and Application, 4*(1), 25–42. Retrieved from http://aisel.aisnet.org/jitta/vol4/iss1/7/

Bryant, A. (2009). Grounded theory and pragmatism: The curious case of Anselm Strauss. *Forum: Qualitative Social Research, 10*(3). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1358/2851

Buglear, J. (2009). Logging in and dropping out: Exploring student non-completion in higher education using electronic footprint analysis. *Journal of Further and Higher Education, 33*(4), 381–393. doi: 10.1080/03098770903272479

Burd, S. (2005, May 6). Plan to track students steps into political quicksand. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Plan-to-Track-Students-Steps/4053/

Burke, J. C. (1998). Performance funding: Present status and future prospects. In J.C. Burke & A. Serban (Eds.), *Performance funding for public higher education: Fad or trend* (5–14). San Francisco, CA: Jossey-Bass.

Burns, P., Hilton, J., & Patterson, L. (2014). *RUCC webinar* [PowerPoint slides]. Retrieved from http://unizin.org/wp-content/uploads/2014/07/Unizin-RUCC-Overview.pdf

Caison, A. L. (2007). Analysis of institutionally specific retention research: A comparison between survey and institutional database methods. *Research in Higher Education, 48*(4), 435-449. doi: 10.1007/sl1162-006-9032-5

Caldwell, T. (2012, March 14). More college students may be asked to declare sexual orientation. *The New York Times*. Retrieved from http://thechoice.blogs.nytimes.com/2012/03/14/sexual-orientation-university-of-california/

Campbell, J. P. (2007). *Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Global: http://search.proquest.com/docview/304837810?accountid=465

Campbell, J. P., DeBlois, P. B., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 40–57. Retrieved from http://www.educause.edu/ero/article/academic-analytics-new-tool-new-era

Campbell, J. P., & Oblinger, D. A. (2007). Academic analytics (Report No. PUB6101). Retrieved from the EDUCAUSE website: https://net.educause.edu/ir/library/pdf/pub6101.pdf

Campus Labs. (2014a). *Case study: Northern Arizona University–Using data to improve student retention*. Retrieved from Campus Labs website: http://www.campuslabs.com/pdf/caseStudy-NAU.pdf

Campus Labs. (2014b). *Shining a light on student success*. Retrieved from Campus Labs website: http://www.campuslabs.com/pdf/product-sheets/BeaconProductSheet.pdf

Caruso, L. R. (1971). Privacy of students and confidentiality of student records. *Case Western Reserve Law Review, 22*(3), 379–389.

Ceruzzi, P. E. (2012). *Computing: A concise history*. Cambridge, MA: The MIT Press.

Charmaz, K. (2000). Constructivist and objectivist grounded theory. In N. K. Denzin and Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., 509–535). Thousand Oaks, CA: SAGE Publications.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Los Angeles, CA: SAGE Publications.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Los Angeles, CA: SAGE Publications.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*(5/6), 318–331. doi: 10.1504/IJTEL.2012.051815

Clancy, F., Hoke, S., & T. Mullan. (1975). An approach to a multifaceted information system in large medical school. *Proceedings of the International Conference on APL, USA, 7*, 84–94. doi: 10.1145/800117.803787

Clark, R. (1987). Information technology and dataveillance. Retrieved from http://rogerclarke.com/DV/CACM88.html

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the Second International Conference on Learning Analytics and Knowledge, USA*, 134–138. doi: 10.1145/2330601.2330636

Cobb, W. H., & Bray, D. W. (1935). The State University of Iowa. In G. W. Baehne (Ed.), *Practical applications of the punched card method in colleges and universities* (pp. 93–104). New York, NY: Columbia University Press.

Cohen, A. M., & Kisker, C. B. (2010). *The shaping of American higher education: Emergence and growth of the contemporary system*. San Francisco, CA: Jossey-Bass.

College Access and Opportunity Act of 2006, H.R. 609, 109th Cong. (2006).

Common Education Data Standards. (n.d.). Retrieved from https://ceds.ed.gov/

Confidentiality laws. (n.d.). *The National Center for Education Statistics*. Retrieved from http://nces.ed.gov/statprog/conflaws.asp

Conley, A., Datta, A., Nissenbaum, H., & Sharma, D. (2012). Sustaining privacy and open justice in the transition to online court records: A multidisciplinary inquiry. *Maryland Law Review, 71*(3), 773–847. Retrieved from http://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=3504&context=mlr

Connelly, R. J. (2000). Intentional learning: The need for explicit informed consent in higher education. *The Journal of General Education, 49*(3), 211–230. Retrieved from http://www.jstor.org/stable/27797469

Constance, C. L. (1935). University of Oregon. In G. W. Baehne (Ed.), *Practical applications of the punched card method in colleges and universities* (pp. 59–65). New York, NY: Columbia University Press.

Cooke, M. L. (1910). *Academic and industrial efficiency* (Bulletin No. 5). The Carnegie Foundation for the Advancement of Teaching and Learning. Boston, MA: Merrymount Press.

Cooley, T. C. (1888). *The law of torts* (2nd ed.). Chicago, IL: Callaghan.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, CA: SAGE Publications.

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review, 55*(1), 93–128. Retrieved from http://heinonline.org/HOL/Page?handle=hein.journals/bclr55&id=93

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Los Angeles, CA: SAGE Publications.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: SAGE Publications.

Crossing, G., Heagney, M., & Thomas, L. (2009). Improving student retention in higher education: Improving teaching and learning. *The Australian Universities' Review, 51*(2), 9–18.

Cunningham, A. F., and Milam, J. (2005). *Feasibility of a student unit record system within the Integrated Postsecondary Education Data System* (NCES Report No. 2005–160). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Retrieved from http://nces.ed.gov/pubs2005/2005160.pdf

Cutcliffe, J. R. (2003). Reconsidering reflexivity: Introducing the case for intellectual entrepreneurship. *Qualitative Health Research, 13*(1), 136–148. doi: 10.1177/1049732302239416

Darke, P., Shanks, G., & Broadbent, M. (1998). Successfully completing case study research: Combining rigour, relevance and pragmatism. *Information Systems Journal, 8*(4), 273–289. doi: 10.1046/j.1365-2575.1998.00040.x

Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Boston, MA: Harvard Business Review Press.

Davie v. Board of Regents, University of California, 227 P. 243 (Cal. Dist. Ct. App. 1924).

De Filippi, P. (2014). Big data, big responsibilities. *Internet Policy Review, 3*(1). doi: 10.14763/2014.1.227

DeCew, J. (1997). *The pursuit of privacy: Law, ethics, and the rise of technology*. Ithaca, NY: Cornell University Press.

Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention, 13*(1), 17–35. doi: 10.2190/CS.13.1.b

DeLillo, D. (1985). *White noise*. New York, NY: Viking Penguin.

Denzin, N. K. (2007). Grounded theory and the politics of interpretation. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 454–471). London, UK: SAGE Publications.

Denzin, N. K., & Lincoln, Y. S. (2003). *The landscape of qualitative research: Theories and issues*. Thousand Oaks, CA: SAGE Publications.

Dexter, F. B. (1896). *Biographical sketches of the graduates of Yale College with annals of the college history, 1745–1763* (vol. 2). New York, NY: Henry Holt.

Dey, I. (1999). *Grounding grounded theory*. San Diego, CA: Academic Press.

Diaz, V. & Brown, M. (2012). *Learning analytics: A report on the ELI focus session* (ELI No. ELI3027). Retrieved from the EDUCAUSE Learning Initiative website: http://net.educause.edu/ir/library/PDF/ELI3027.pdf

Dixon v. Alabama State Board of Education, 294 F.2d 150 (1961).

Draschler, H., & Greller, W. (2012). The pulse of learning analytics: Understandings and expectations from the stakeholders. *Proceedings of the Second International Conference on Learning Analytics and Knowledge, Canada*, 120–129. doi: 10.1145/2330601.2330634

Draucker, C. B., Martsolf, D. S., Ross, R., & Rusk, T. B. (2007). Theoretical sampling and category development in grounded theory. *Qualitative Health Research, 17*(8), 1137–1148. doi: 10.1177/1049732307308450

Duhigg, C. (2012a, February 6). How companies learn your secrets. *The New York Times*. Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Duhigg, C. (2012b). *The power of habit: Why we do what we do in life and business*. New York, NY: Random House.

Duval, E., Klerkx, J., Verbert, K., Nagel, T., Govaerts, S., Parra, G., …Vandeputte, B. (2012). Learning dashboards and learnscapes. *Proceedings of CHI 2012, USA*, 1–5. Retrieved from https://lirias.kuleuven.be/bitstream/123456789/344525/1/eist2012_submission_6-2.pdf

Duval, E., & Koskinen, T. (2014). Learning analytics and assessment. *eLearning Papers, 2014*(36), 2. Retrieved from http://www.openeducationeuropa.eu/sites/default/files/old/Issue36_0.pdf

Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review, 14*(4), 532–550. doi: 10.2307/258557

Electronic Privacy Information Center. (2000). *Pretty poor privacy: An assessment of P3P and Internet Privacy*. Retrieved from https://epic.org/reports/prettypoorprivacy.html

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662–664. doi: 10.1111/bjet.12028

Evans, D. (2011). The Internet of things [Web log post]. Retrieved from http://blogs.cisco.com/news/the-internet-of-things-infographic/

Fain, P. (2012, February 1). Big data's arrival. *Inside Higher Ed*. Retrieved from http://www.insidehighered.com/news/2012/02/01/using-big-data-predict-online- student-success

Family Policy Compliance Office. (2011). *The Family Educational Rights Privacy Act: Guidance for reasonable methods and written agreements*. Retrieved from Department of Education Family Policy Compliance Office website: https://www2.ed.gov/policy/gen/guid/fpco/pdf/reasonablemtd_agreement.pdf

Fefferman, N. H., O'Neil, E. A., & Naumova, E. N. (2005). Confidentiality and confidence: Is data aggregation a means to achieve both? *Journal of Public Health Policy, 26*, 430–449. doi: 10.1057/palgrave.jphp.3200029

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317. doi: 10.1504/IJTEL.2012.051816

Fichtenbaum, M. (1935). The University of Texas. In G. W. Baehne (Ed.), *Practical applications of the punched card method in colleges and universities* (pp. 47–58). New York, NY: Columbia University Press.

Fichtenbaum, M., & Shipp, W. B. (1947). Grade records and tabulating machines. *Journal of the American Association of Collegiate Registrars, 22*(3), 293–301.

Fincher, C. I. (1977). Institutional practices and threats to individual privacy. *New Directions for Institutional Research, 1977*(14), 17–31. doi: 10.1002/ir.37019771405

Fischer, K. (2005, March 18). Accountability panel says government should collect more data on students. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Accountability-Panel-Says/19725/

Flaherty, D. H. (1999). Visions of privacy: Past, present, and future. In C. J. Bennett & R. Grant (Eds.), *Visions of privacy: Policy choices for the digital age* (pp. 19–38). Toronto, Canada: University of Toronto Press.

Fleck, J. (1994). Learning by trying: The implementation of configurational technology. *Research Policy, 23*(6), 637–652. doi: 10.1016/0048-7333(94)90014-0

Flick, U. (2007). *Managing quality in qualitative research*. Los Angeles, CA: SAGE Publications.

Foucault, M. (1995). *Discipline and punish: The birth of the prison*. New York, NY: Vintage Books.

Foucault, M. (1998). *The history of sexuality: The will to knowledge*. London, UK: Penguin.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & Health, 25*(10), 1229–1245. doi: 10.1080/08870440903194015

Frankfort, J., Salim, K., & Carmean, C. (2012). Analytics, nudges, and learner persistence. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/analytics-nudges-and-learner-persistence

Freund, P. A. (1971). Privacy: One concept or many? In J. R. Penncock & J. W. Chapman (Eds.), *Privacy: Nomos XIII* (pp. 182–198). New York, NY: Atherton Press.

Freund, P. A. (1975). Address to the American Law Institute. *Proceedings of the 52nd Annual Meeting of the American Law Institute*, 42–43. Retrieved from http://heinonline.org/HOL/Page?handle=hein.ali/alimetsp1975&div=1

Fried, C. (1968). Privacy: A moral analysis. *Yale Law Journal, 77*(3), 475–493. Retrieved from http:// www.jstor.org/stable/794941

Friedman, B. (1999). *Value-sensitive design: A research agenda for information technology* (Contract No: SBR-9729633). Arlington, VA: National Science Foundation.

Fritz, J. (2013). *Using analytics at UMBC: Encouraging student responsibility and identifying effective course designs* (Report No. ERB1304). Retrieved from the EDUCAUSE Center for Applied Research website: https://net.educause.edu/ir/library/pdf/ERB1304.pdf

Froomkin, A. M. (2000). The death of privacy? *Stanford Law Review, 52*(5), 1461–1543. Retrieved from http://www.jstor.org/stable/1229519

Garcia-Kaplan, O. (2015, April 10). A FERPA rewrite [Web log post]? Retrieved from http://blog.ferpasherpa.org/?p=194

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. doi: 10.1007/s11528-014-0822-x

Gates, K. F. (2014, January 2). Classroom attendance via ID card [Msg 6]. Message posted to http://www.educause.edu/discuss/constituent-groups-about-information-technology-management-and-leadership/cio-constituent-group/classroom-attendance-

Gillham, B. (2000). *The research interview*. London, UK: Continuum.

Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley, CA: The Sociology Press.

Glaser, B. G. (2001). *The grounded theory perspective: Conceptualization contrasted with description*. Mill Valley, CA: The Sociology Press.

Glaser, B. G., Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: Aldine Transaction.

Godkin, E. L. (1890). The rights of the citizen, IV–To his own reputation. *Scribner's Magazine, 8*(1), 58–67. Retrieved from http://digital.library.cornell.edu

Goff, J. W., & Shaffer, C. M. (2014). Big data's impact on college admission practices and recruitment strategies. In J. E. Lane (Ed.), *Building a smarter university: Big data, innovation, and analytics* (pp. 93–120). Albany, NY: SUNY Press.

Goldin, C., & Katz, L. F. (1999). The shaping of higher education: The formative years in the United States, 1890 to 1940. *Journal of Economic Perspectives, 13*(1), 37–62. Retrieved from http://www.jstor.org/stable/2647136

Goldman, J. (1999). Privacy and individual empowerment in the interactive age. In C. J. Bennett & R. Grant (Eds.), *Visions of privacy: Policy choices for the digital age* (pp. 96–115). Toronto, Canada: University of Toronto Press.

Goldstein, P. J., & Katz, R. N. (2005). *Academic analytics: The uses of management information and technology in higher education* (Report No. ERS0508). Retrieved from the EDUCAUSE Center for Applied Research website: https://net.educause.edu/ir/library/pdf/ers0508/rs/ers0508w.pdf

Gorr, W., & Hossler, D. (2006). Why all the fuss about student information systems? Or information systems as golden anchors in higher education. *New Directions for Higher Education, 2006*(136), 7–20.

Gott v. Berea College, 156 Ky. 376, 161 S.W. 204 (1913).

Goulielmos, M. (2004) Systems development approach: Transcending methodology. *Information Systems Journal, 14*, 363–386. doi: 10.1111/j.1365-2575.2004.00175.x/full

Greely, H. T. (2007). The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annual Review of Genomics and Human Genetics, 8*, 343–364. doi: 10.1146/ annurev.genom.7.080505.115721

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society, 15*(3), 42–57. Retrieved from http://www.ifets.info/download_pdf.php?j_id=56&a_id=1256

Gruber, C. S. (2007). Backdrop. In H. S. Wechsler, L. F. Goodchild, & L. Eisenmann (Eds.), *The history of higher education* (3 ed.) (pp. 260–277). Boston, MA: Pearson Custom Publishing.

Gubrium, E., & Koro-Ljungberg, M. (2005). Contending with border making in the social constructionist interview. *Qualitative Inquiry, 11*(5), 689–715. doi: 10.1177/1077800405278776

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation. *Field Methods, 18*, 58–82. doi: 10.1177/1525822X05279903

Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology, 51*(4), 605–622. doi: 10.1080/00071310020015280

Hallberg, L. R-M. (2006). The “core category” of grounded theory: Making constant comparisons. *International Journal of Qualitative Studies on Health and Well-being, 1*(3), 141–148. doi: 10.1080/17482620600858399

Halverson, R., & Shapiro, R. B. (2013). Technologies for education and technologies for learners: How information technologies are (and should be) changing schools. In D. Anagnostopoulos, S. A. Rutledge, & R. Jacobsen (Eds.), *The infrastructure of accountability:*

Data use and the transformation of American education (pp. 163–179). Cambridge, MA: Harvard Education Press.

Hamburger v. Cornell University, 148 N.E. 539 (NY 1925).

Hardy, Q. (2012, June 4). Rethinking privacy in an era of big data. *The New York Times*. Retrieved from bits.blogs.nytimes.com//2012/06/04/rethinking-privacy-in-an-era-of-big-data/

Hearn, J. C., McKlendon, M. K., & Mokher, C. G. (2008). Accounting for student success: An empirical analysis of the origins and spread of state unit-record systems. *Research in Higher Education, 49*(8), 665–683. doi: 10.1007/s11162-008-9101-z

Hermanowicz, J. C. (2003). *College attrition at American research universities: Comparative case studies*. New York: Agathon Press.

Herold, B. (2015, April 7). Major FERPA overhaul under consideration in U.S. House. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/DigitalEducation/2015/04/ferpa_overhaul_US_House.html

Herzog, H. (2012). Interview location and its social meaning. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (pp. 207–217). Los Angeles, CA: SAGE Publications.

Hier, S. P. (2003). Probing the surveillant assemblage: On the dialectics of surveillance practices as processes of social control. *Surveillance & Society, 1*(3), 399–411. Retrieved from http://www.surveillance-and-society.org/articles1(3)/probing.pdf

Higher Education Act, 20 U.S.C. § 1094 (1965).

Higher Education Opportunity Act, Pub. L. No. 110-315, 122 Stat. 3078 (2008).

Holden, H. P. (1976). Student records: The Harvard experience. *American Archivist, 39*(4), 461–467. Retrieved from http://archivists.metapress.com/content/b82315m3q3944545/fulltext.pdf

Holton, J. A. (2007). The coding process and its challenges. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 265–289). Los Angeles, CA: SAGE Publications.

Holton, J. A. (2008). Grounded theory as a general research methodology. *Grounded Theory Review, 7*(2). Retrieved from http://groundedtheoryreview.com/2008/06/30/grounded-theory-as-a-general-research-methodology/

Hoover, E. (2011). Elmhurst College will ask applicants about sexual orientation. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/headcount/elmhurst-college-will-ask-applicants-about-sexual-orientation/28553

Hoover, E. (2012). Facebook meets predictive analytics. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/headcount/facebook-meets-predictive-analytics/32770

Hopkins, B., & Evelson, B. (2011). *Expand your digital horizon with big data* (Report No. RES60751). Retrieved from Forrester website: http://www.forrester.com/Expand+Your+Digital+Horizon+With+Big+Data/fulltext/- /E-RES60751?objectid=RES60751

Hossler, D. (2006). The impact of implementing new information systems on the priorities, management, and allocation of resources of colleges and universities. *New Directions for Higher Education, 2006*(136), 69–76.

Hossler, D., & Gorr, W. P. (2006). Enterprise systems. In D. Priest & E. St. John (Eds.), *Privatization and public universities* (pp. 203–227). Bloomington, IN: Indiana University Press.

Hossler, D., & Pape, S. (2006). Editors' notes. *New Directions for Higher Education, 2006*(136), 1–6.

Hughes, T. P. (1969). Technological momentum in history: Hydrogenation in Germany 1898-1933. *Past & Present, 44*(Aug., 1969), 106–132. Retrieved from http://www.jstor.org/stable/649734

Hui, K., & Png, I. P. L. (2006). The economics of privacy. In T. Hendershott (Ed.), *Handbook of information systems and economics* (pp. 471–497). North Holland, Netherlands: Elsevier Science.

Huisman, J., & Currie, J. (2004). Accountability in higher education: Bridge over troubled water? *Higher Education, 48*(4), 529–551.

IBM. (2013). IBM and Georgia's largest school system bring personalized learning to life [Press release]. Retrieved from http://www-03.ibm.com/press/us/en/pressrelease/42759.wss

IBM Research. (n.d.). Smarter education group. Retrieved from http://researcher.watson.ibm.com/researcher/view_group.php?id=4977

Ingham, J. (2000). Data warehousing: A tool for the outcomes assessment process. *IEEE Transactions on Education, 43*(2), 132–136. doi: 10.1109/13.848064

Inness, J. C. (1992). *Privacy, intimacy, and isolation*. New York, NY: Oxford University Press.

Jisc. (n.d.). *Effective learning analytics*. Retrieved from http://www.jisc.ac.uk/rd/projects/effective-learning-analytics

Johnson, J. A. (2014). The ethics of big data in higher education. *International Review of Information Ethics, 21*(2014), 3-10. Retrieved from http://www.i-r-i-e.net/inhalt/021/IRIE-021-Johnson.pdf

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). *The NMC horizon report: 2014 higher education edition*. Austin, TX: The New Media Consortium.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). *The NMC horizon report: 2015 higher education edition*. Austin, TX: The New Media Consortium.

Johnson, L., Levine, A., Smith, R., & Stone, S. (2010). *The 2010 horizon report*. Austin, TX: The New Media Consortium.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 horizon report*. Austin, TX: The New Media Consortium.

Johnson, M. J., & Rowlands, T. (2012). The interpersonal dynamics of in-depth interviewing. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (2nd ed.) (pp. 99–113). Thousand Oaks, CA: SAGE Publications.

Jones, K. M. L. (2015, January 27). The Student Digital Privacy Act: Where's the protection for college students [Web log post]. Retrieved from http://thecorkboard.org/blog/2015/01/27/the-student-digital-privacy-act-wheres-the-protection-for-college-students/

Jones, K. M. L., Thomson, J., & Arnold, K. (2014). Questions of data ownership on campus. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/questions-data-ownership-campus

Kafka, F. (1968). *The trial*. New York, NY: Schocken Books.

Kang, J. (1998). Information privacy in cyberspace transactions. *Stanford Law Review, 50*(4), 1193–1294. Retrieved from http://www.jstor.org/stable/1229286

Kay, D., Korn, N., & Oppenheim, C. (2012). *CETIS analytics series: Legal, risk, and ethical aspects of analytics in higher education* (Volume 1, Number 6). Retrieved from CETIS website: http://publications.cetis.ac.uk/wp-content/uploads/2012/11/Legal-Risk-and-Ethical-Aspects-of-Analytics-in-Higher-Education-Vol1-No6.pdf

King, J. H., & Richards, N. M. (2014, March 28). What's up with big data ethics? *Forbes*. Retrieved from http://www.forbes.com/sites/oreillymedia/2014/03/28/whats-up-with-big-data-ethics/

Kirkpatrick, F. H. (1941). Records and counseling in the small college. *Journal of the American Association of Collegiate Registrars, 16*(3), 316–318.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly, 23*(1), 67–94. doi: 10.2307/249410

Kling, R., Rosenbaum, H., & Sawyer, S. (2005). *Understanding and communicating social informatics: A framework for studying and teaching the human contexts of information and communication technologies*. Medford, NJ: Information Today.

Kolowich, S. (2013, January 25). The new intelligence. *Insider Higher Ed*. Retrieved from http://www.insidehighered.com/news/2013/01/25/arizona-st-and-knewtons-grand-experiment-adaptive-learning

Kolowich, S. (2015, January 13). Obama proposes bill to protect student data, but not in higher education. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/wiredcampus/obama-proposes-bill-to-protect-student-data-but-not-in-higher-education

Kruse, A., & Pongsajapan, R. (2012). *Student-centered learning analytics* (Thought paper). Retrieved from Georgetown's Center for New Designs in Learning and Scholarship website: https://cndls.georgetown.edu/m/documents/thoughtpaper-krusepongsajapan.pdf

Kvale, S. (2007). *Doing interviews*. Los Angeles, CA: SAGE Publications.

Lake, P. F. (2013). *The rights and responsibilities of the modern university: The rise of the facilitator university*. Durham, NC: Carolina University Press.

Land, R., & Bayne, S. (2004). Screen or monitor? Surveillance and disciplinary power in online learning environments. In R. Land & S. Bayne (Eds.), *Education in cyberspace* (pp. 165–178). London, UK: RoutledgeFalmer.

Lane, E. (2014, October 13). Moneyball: How businesses are using data to outsmart their rivals. *CNN*. Retrieved from http://edition.cnn.com/2014/10/13/business/moneyball-businesses-outsmarting-rivals/

Lane, J. E., & Finsel, B. A. (2014). Fostering smarter colleges and universities. In J. E. Lane (Ed.), *Building a smarter university: Big data, innovation, and analytics* (pp. 3–26). Albany, NY: State University of New York Press.

Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety* (META Group No. 949). Retrieved from Gartner website: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Learning analytics initiative. (n.d.). *Apereo*. Retrieved from https://www.apereo.org/content/learning-analytics-initiative

Leber, J. (2013, August 16). How big data could replace your credit score. *Fast Company*. Retrieved from http://www.fastcoexist.com/3015894/how-big-data-could-replace-your-credit-score

Lee, P. (2011). The curious life of *in loco parentis* in American universities. *Higher Education in Review, 8*, 65-90. Retrieved from http://www.higheredinreview.org/articles/vol8Lee.pdf

Leon, G. P., Cranor, L. F., McDonald, A. M., & McGuire, R. (2010). *Token attempt: The misrepresentation of website privacy policies through the misuse of P3P compact policy tokens*. Retrieved from Carnegie Mellon University CyLab website: https://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab10014.pdf

Lessig, L. (2006). *Code: Version 2.0*. New York, NY: Basic Books.

Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. New York, NY: W.W. Norton.

Levine, A. H., Cary, E., & Divoky, D. (1973). *The rights of students: The basic ACLU guide to a student's rights*. New York, NY: The American Civil Liberties Union, Inc.

Lianos, M. (2003). Social control after Foucault. *Surveillance & Society, 1*(3), 412–430. Retrieved from http://www.surveillance-and-society.org/articles1(3)/AfterFoucault.pdf

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: SAGE Publications.

Lipset, S. M. (1993). *Rebellion in the university*. New Brunswick, NJ: Transaction Publishers.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31–40. Retrieved from https://net.educause.edu/ir/library/pdf/ERM1151.pdf

Los, M. (2006). Looking into the future: Surveillance, globalization and the totalitarian potential. In D. Lyon (Ed.), *Theorizing surveillance: The panopticon and beyond* (pp. 69–94). Cullompton, United Kingdom: Willan Publishing.

Lowendahl, J-M. (2013). *Hype cycle for education, 2013* (Report No. G00251104). Retrieved from Gartner website: https://www.gartner.com/doc/2559615/hype-cycle-education-

Lyon, D. (2007). *Surveillance studies: An overview*. Cambridge, United Kingdom: Polity Press.

Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society, 2014, 1*(2), 1–13. doi: 10.1177/2053951714541861

MacCarthy, M. (2014). Student privacy: Harm and context. *International Review of Information Ethics, 21*, 11–24. Retrieved from: http://i-r-i-e.net/inhalt/021/IRIE-021-MacCarthy.pdf

Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough: Why e–learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society, 15*(3), 149–163. Retrieved from http://www.ifets.info/journals/15_3/11.pdf

Macfadyen, L. P., Dawson, S., Pardo, A., & Gašević, D. (2014). Embracing big data in complete educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment, 9*(Winter 2014), 17–28. Retrieved from http://www.rpajournal.com/dev/wp-content/uploads/2014/10/A2.pdf

Mackenzie, D., & Wajcman, J. (1985). *The social shaping of technology: How the refrigerator got its hum.* Milton Keynes, UK: Open University Press.

Mandge, O. L. (2013). A data mining tool for prediction of suicides among students. *Proceedings of the National Conference on New Horizons in IT, Italy,* 178–181. Retrieved from http://www.met.edu/Institutes/ICS/NCNHIT/papers/41.pdf

Manovich, L. (2011). Trending: The promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 460–475). Minneapolis, MN: The University of Minnesota Press. Retrieved from http://dhdebates.gc.cuny.edu/debates/text/15

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition and productivity* (Report). Retrieved from McKinsey Global Institute website: http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

Marx, G. T. (1999). Ethics for the new surveillance. In C. J. Bennett & R. Grant (Eds.), *Visions of privacy: Policy choices for the digital age* (pp. 39–67). Toronto, Canada: University of Toronto Press.

Mayer-Schönberger, V. (2009). *Delete: The virtue of forgetting in the digital age*. Princeton, NJ: Princeton University Press.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt.

Mayer-Schönberger, V., & Cukier, K. (2014). *Learning with big data: The future of education*. New York, NY: Houghton Mifflin Harcourt.

McCann, C., & Laitinen, A. (2014). *College blackout: How the higher education lobby fought to keep students in the dark*. Washington, DC: New America. Retrieved from http://newamerica.net/sites/newamerica.net/files/policydocs/CollegeBlackoutFINAL.pdf

McCarthy, J. T. (2015). *The rights of publicity and privacy* . Eagan, MN: Clark Boardman Callaghan.

McClure, W. E. (1936). An inclusive record. *The Journal of Higher Education, 7*(1), 12–15. Retrieved from http://www.jstor.org/stable/197430

McCosh, J. (1878). Discipline in American colleges. *The North American Review, 126*, 428–441. Retrieved from http://www.jstor.org/stable/25110205

McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society, 4*(3), 540–565.

McKechnie, L. E. F. (2008). Reactivity. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (Vol. 2, pp. 729–730). Thousand Oaks, CA: Sage Publications. Retrieved from http://search.credoreference.com/content/entry/sagequalrm/reactivity/0

Miller, E. C. (1938). Records and reports compiled by the registrar. *Journal of the American Association of Collegiate Registrars, 13*(2), 225–230.

Miller, K., & Chapin, K. (2013, February 15). How big data changes lives. *WGBH News*. Retrieved from http://wgbhnews.org/post/how-big-data-changes-lives

Moore, A. D. (2010). *Privacy rights: Moral and legal foundations*. University Park, PA: The Pennsylvania State University Press.

Morse, J. M. (1994). Designing funded qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook for qualitative research* (pp. 220–235). Thousand Oaks, CA: SAGE Publications.

Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research, 5*, 147–149. doi: 10.1177/104973239500500201

Morse, J. M. (2007). Sampling in grounded theory. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 229–244). Los Angeles, CA: SAGE Publications.

Moss, S. (2014, October 23). Big data: New oil or snake oil? *Wired*. Retrieved from http://www.wired.com/2014/10/big-data-new-oil-or-snake-oil/

Mumford, E. (2000). Socio-technical design: An unfulfilled promise or a future opportunity? In R. Baskerville, J. Stage, & J. DeGross (Eds.), *Organizational and social perspectives on information technology* (pp. 33–46). Norwell, MA: Kluwer Academic Publishers.

Myers, M. D. (1997) Qualitative research in information systems. *MIS Quarterly, 21*(2), 241–242.

Myers, M. D. (2009). *Qualitative research in business & management*. Thousand Oak, CA: SAGE Publications.

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *Proceedings of the IEEE Symposium on Security and Privacy, USA*, 111–125. doi: 10.1109/SP.2008.33

National Center for Education Statistics. (n.d.). About the SLDS grant program. Retrieved from http://nces.ed.gov/programs/slds/

Nelson, K., & Creagh, T. (2013). *A good practice guide: Safeguarding student learning engagement* (Research Report). Retrieved from Safeguarding Student Learning website: http://

safeguardingstudentlearning.net/wp-content/uploads/2012/04/LTU_Good-practice-guide_eBook_20130320.pdf

Newfield, C. (2007). The rise of university management. In H. S. Wechsler, L. F. Goodchild, & L. Eisenmann (Eds.), *The history of higher education* (3 ed.) (pp. 346–358). Boston, MA: Pearson Custom Publishing.

Ng, K. and Hase, S. (2008). Grounded suggestions for doing a grounded theory business research. *The Electronic Journal of Business Research Methods, 6*(2),155–170. Retrieved from http://www.ejbrm.com/volume6/issue2/p183

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*(1), 119–158.

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford Law Books.

Norris, D. M., & Baer, L. L. (2013). *Building organizational capacity for analytics* (Research Report No. PUB9012). Retrieved from EDUCAUSE website: https://net.educause.edu/ir/library/pdf/PUB9012.pdf

Norris, D., Baer, L., Leonard, J., Pugliese, L., & Lefrere, P. (2008). Action analytics: Measuring and improving performance that matters in higher education. *EDUCAUSE Review, 43*(1), 42–67. Retrieved from https://net.educause.edu/ir/library/pdf/ERM0813.pdf

O'Brien, D. M. (1979). *Privacy, law, and public policy*. New York, NY: Praeger Publishers.

O'Connor, M. C. (2010, May 24). Northern Arizona University to use RFID student cards for attendance tracking. *RFID Journal*. Retrieved from http://www.rfidjournal.com/articles/view?7628

O'Neil, M. (2014a, February 25). Data breach at Indiana U. exposes information on 146,000 students. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/wiredcampus/data-breach-at-indiana-u-exposes-information-on-146000-students/50729

O'Neil, M. (2014b, March 5). QuickWire: Lawsuit is filed in Arizona data-breach case. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/wiredcampus/quickwire-lawsuit-is-filed-in-arizona-data-breach-case/52381

O'Neil, M. (2014c, March 17). Data breaches put a dent in colleges' finances as well as reputations. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Data-Breaches-Put-a-Dent-in/145341

OECD. (1980). Annex to the recommendation of the Council of 23 September 1980: Guidelines governing the protection of privacy and transborder flows of personal data (Report no. C(80)58/FINAL). Retrieved from the OECD website: http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=114&InstrumentPID=312&Lang=en&Book=False

OECD. (2013). *The OECD privacy framework* (Report). Retrieved from the OECD website: http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57*(6), 1701–1777. Retrieved from http://www.uclalawreview.org/pdf/57-6-3.pdf

Opp, K-D. (2001). Ho do norms emerge? An outline of a theory. *Mind & Society, 2*(3), 101–128. doi: 10.1007/BF02512077

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*(3), 398–427. doi: 10.1287/orsc.3.3.398

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research, 2*(1), 1–28. doi: 10.1287/isre.2.1.1

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438–450. doi: 10.1111/bjet.12152

Parks v. Northwestern University, 75 N.E. 991 (Ill. 1905).

Parry, M. (2012, July 18). Big data on campus. *The New York Times*. Retrieved from http://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html

Peltier, G. L., Laden, R., & Matranga, M. (1999). Student persistence in college: A review of research. *Journal of College Student Retention, 1*(4), 357–375.

Pérez-Peña, R. (2015, January 16). Students gain access to files on admission to Stanford. *The New York Times*. Retrieved from http://www.nytimes.com/2015/01/17/us/students-gain-access-to-files-on-admission-to-stanford.html

Peters, B. (2012, June 21). The big data gold rush. *Forbes*. Retrieved from http://www.forbes.com/sites/bradpeters/2012/06/21/the-big-data-gold-rush/

Peterson, G. E. (1964). *The New England college in the age of the university*. Amherst, MA: Amherst College Press.

Peterson, K. (1944). A revision of the student record forms in the college of education. *Educational Research Bulletin, 23*(7), 191–196. Retrieved from http://www.jstor.org/stable/1472719

Piazza. (n.d.). *To infinity and beyond*. Retrieved from https://recruiting.piazza.com/insights

Piazza Careers. (n.d.). *Transforming college recruiting*. Retrieved from https://recruiting.piazza.com

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9–20. Retrieved from http://onlinelearningconsortium.org/jaln/v16n3/evolution-big-data-and-learning-analytics-american-higher-education

Polonetsky, J., & Tene, O. (2014). The ethics of student privacy: Building trust for ed tech. *International Review of Information Ethics, 21*, 25–34. Retrieved from: http://i-r-i-e.net/inhalt/021/IRIE-021-Polonetsky-Tene.pdf

Porter, N. (1870). *The American colleges and the American public*. New York, NY: Arno Press.

Poster, M. (1996). Databases as discourse; or, electronic interpellations. In D. Lyon & E. Zureik (Eds.), *Computers, surveillance, and privacy* (pp. 175–192). Minneapolis, MN: University of Minnesota Press.

Powers, E. (2006, July 6). Wrangling over unit records. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2006/07/07/unitrecord

Pratt v. Wheaton College, 40 Ill. 186 (1866).

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge, Belgium*, 240–244. doi: 10.1145/2460296.2460344

Privacy Protection Study Commission. (1977). *Personal privacy in an information society: The report of the Privacy Protection Study Commission* (Report No. 052-003-00395-3). Retrieved from National Criminal Justice Reference Service website: https://www.ncjrs.gov/pdffiles1/Digitization/49602NCJRS.pdf

Protection of Human Subjects, 45 C.F.R. § 46 (2009).

Public-private split on possible student privacy threat. (2004). *National On-Campus Report, 32*(24), 1–4.

Ragin, C. C., & Becker, H. S. (Eds.). (1992). *What is a case? Exploring the foundations of social inquiry*. Cambridge, UK: Cambridge University Press.

Ramirez, C. A. (2009). *FERPA clear and simple: The college professional's guide to compliance*. San Francisco, CA: Jossey-Bass.

Reed, S. (2013). Four areas of collegiate student-athlete privacy invasion. *Communication & Sport*. doi: 10.1177/2167479513510910

Regan, P. M. (1995). *Legislating privacy: Technology, social values, and public policy*. Chapel Hill, NC: The University of North Carolina Press.

Reilly, M. (2013). *Further education learning technology: A horizon scan for the UK Government Foresight Horizon Scanning Centre* (Report). Ariel Research Services. Retrieved from http://arielresearchservices.com/wp-content/uploads/2014/03/Further-Education-and-Learning-Technology-Final-Draft.pdf

Reiman, J. H. (1976). Privacy, intimacy, and personhood. *Philosophy & Public Affairs, 6*(1), 26–44. Retrieved from http://www.jstor.org/stable/2265060

Richards, N. M. (2008). Intellectual privacy. *Texas Law Review, 87*(2), 387–446.

Richards, N. M. (2013). The dangers of surveillance. *Harvard Law Review, 126*, 1934–1965. Retrieved from http://harvardlawreview.org/2013/05/the-dangers-of-surveillance/

Richards, N. (2015). *Intellectual privacy*. New York, NY: Oxford University Press.

Richards, N. M., & King, J. H. (2013). Three paradoxes of big data. *Stanford Law Review Online, 66*. Retrieved from http://www.stanfordlawreview.org/online/privacy-and-big-data/three-paradoxes-big-data

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review, 49*(2), 393–432. Retrieved from http://heinonline.org./HOL/Page?handle=hein.journals/wflr49&id=405

Rotella, P. (2012, April 2). Is data the new oil? *Forbes*. Retrieved from http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/

Rothbauer, P. (2008). Triangulation. In *The SAGE encyclopedia of qualitative research methods*. Retrieved from http://search.credoreference.com/content/entry/sagequalrm/triangulation/0

Rourke, F. E., & Brooks, G. E. (1966). *The managerial revolution in higher education.* Baltimore, MD: The Johns Hopkins Press.

Rubel, A., & Jones, K. M. L. (forthcoming). Student privacy in learning analytics: An information ethics perspective. *The Information Society*.

Russell, J. D., & Reeves, F. W. (1936). *The evaluation of higher education institutions* (vol. 6). Chicago, IL: The University of Chicago Press.

Samarati, P., & Sweeney, L. (1998). *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression* (Technical Report No. SRI-CSL-98-04). SRI Computer Science Laboratory, Palo Alto, CA. Retrieved from https://epic.org/privacy/reidentification/Samarati_Sweeney_paper.pdf

Sawyer, S., Crowston, K., Wigand, R. T., & Allbritton, M. (2003). The social embeddedness of transactions: Evidence from the residential real estate industry. *The Information Society, 19*(2), 135–154. doi: 10.1080/01972240309460

Sawyer, S., & Eschenfelder, K. (2002). Social informatics: Perspectives, examples, and trends. *Annual Review of Information Science and Technology, 36*(1), 427–465. doi: 10.1002/aris.1440360111

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Journal of Educational Technology & Society, 17*(4), 117–132. Retrieved from http://www.ifets.info/journals/17_4/8.pdf

Schrage, M. (2014). Big data's dangerous new era of discrimination. *Harvard Business Review.* Retrieved from https://hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination/

Schwartz, P. M. (2010). *Data protection law and the ethical use of analytics* (Report). Prepared for The Hunton and Williams Centre for Information Policy Leadership. Retrieved from the Privacy Association website: http://privacyassociation.org/media/pdf/knowledge_center/Ethical_Underpinnings_of_Analytics.pdf

Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review, 86*(6), 1814–1894. Retrieved from http://www.nyulawreview.org/sites/default/files/pdf/NYULawReview-86-6-Schwartz-Solove.pdf

Sclater, N. (2014). *Code of practice for learning analytics: A literature review of the ethical and legal issues* (Jisc Report). Retrieved from Jisc website: http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

Sclater, N. (2015, April 21). Code of practice for learning analytics – public consultation (Web log post). Retrieved from http://analytics.jiscinvolve.org/wp/2015/04/21/code-of-practice-for-learning-analytics-public-consultation/

Scott, W. R. (2008). *Institutions and organizations: Ideas and interests* (3rd ed.). Los Angeles, CA: SAGE Publications.

Seidman, I. E. (1991). *Interviewing as qualitative research*. New York, NY: Teachers College Press.

Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. *Proceedings of the Second International Conference on Learning Analytics and Knowledge, USA,* 4–8. doi: 10.1145/2330601.2330605

Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., … Baker, R. S. J. D. (2011). *Open learning analytics: An integrated and modularized platform* (Report). Retrieved from the Society for Learning Analytics Research website: http://solaresearch.org/OpenLearningAnalytics.pdf

Singel, R. (2009, December 17). Netflix spilled your *Brokeback Mountain* secret, lawsuit claims. *Wired*. Retrieved from http://www.wired.com/2009/12/netflix-privacy-lawsuit/

Singer, N. (2015, March 5). Digital learning companies falling short of student privacy pledge. *The New York Times*. Retrieved from http://bits.blogs.nytimes.com/2015/03/05/digital-learning-companies-falling-short-of-student-privacy-pledge/

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. doi: 10.1177/0002764213479366

Slade, S. & Prinsloo, P. (2014). Student perspectives on the use of their data: Between intrusion, surveillance and care. *Proceedings of the European Distance and E-Learning Network, UK*, 291–300. Retrieved from http://oro.open.ac.uk/41229/

Society for Learning Analytics Research. (2014a, August 5). *LASI14 overview of the discipline George Siemens and Ryan Baker June 30, 2014* [Video file]. Retrieved from https://www.youtube.com/watch?v=oR4prUnEm2g&list=UUpqoUFmYIt33x9rzFzixtKw

Society for Learning Analytics Research. (2014b). Leaders in learning analytics and open source software hold open learning analytics summit [Press release]. Retrieved from http://www.solaresearch.org/wp-content/uploads/2014/04/OLAPressReleaseFINALv2.pdf

Solove, D. J. (2004). *The digital person: Technology and privacy in the information age*. New York, NY: New York University Press.

Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review, 154*(3), 477–560. Retrieved from https://www.law.upenn.edu/journals/lawreview/articles/volume154/issue3/Solove154U.Pa.L.Rev.477(2006).pdf

Solove, D. J. (2008). *Understanding privacy*. Cambridge, MA: Harvard University Press.

Stanley, J. (2012, April 30). The potential chilling effects of big data. *American Civil Liberties Union.* Retrieved from https://www.aclu.org/blog/technology-and-liberty/potential-chilling-effects-big-data

Steinberg, D. (2013, April 25). Unlocking the big data goldmine for SMBs. Retrieved from http://www.wired.com/2013/04/unlocking-the-big-data-goldmine-for-smbs/

Steinberg, J. (2010, February 26). University of Pennsylvania tries outreach based on sexual orientation. *The New York Times*. Retrieved from http://thechoice.blogs.nytimes.com/2010/02/26/penn/

Stetson University v. Hunt, 88 Fla. 510, 102 So. 637 (1915).

Stevens, G. E. (1980). Invasion of student privacy. *Journal of Law & Education, 9*(3), 343–351.

Stevens, M. L., & Silbey, S. S. (2014). Ethical framework for learning research in higher education [Press release]. Retrieved from http://asilomar-highered.info/pressrelease.html

Stiles, R. J. (2012). *Understanding and managing the risks of analytics in higher education: A guide* (Research Report No. EPUB1201). Retrieved from EDUCAUSE website: https://net.educause.edu/ir/library/pdf/EPUB1201.pdf

Straumsheim, C. (2013, October 18). Before the fact. *Insider Higher Ed*. Retrieved from https://www.insidehighered.com/news/2013/10/18/u-kentucky-hopes-boost-student-retention-prescriptive-analytics

Stross, R. (2012, November 24). So you're a good driver? Let's go to the monitor. *The New York Times*. Retrieved from http://www.nytimes.com/2012/11/25/business/seeking-cheaper-insurance-drivers-accept-monitoring-devices.html?_r=0

Student Privacy Pledge. (2015). Privacy pledge. Retrieved from http://studentprivacypledge.org/?page_id=45

Sun, J. C. (2014). Legal issues associated with big data in higher education: Ethical considerations and cautionary tales. In J. E. Lane (Ed.), *Building a smarter university: Big data, innovation, and analytics* (pp. 27–56). Albany, NY: State University of New York Press.

Svitek, P., & Anderson, N. (2014, February, 19). University of Maryland computer security breach exposes 300,000 records. *The Washington Post*. Retrieved from http://www.washingtonpost.com/local/college-park-shady-grove-campuses-affected-by-university-of-maryland-security-breach/2014/02/19/ce438108-99bd-11e3-80ac-63a8ba7f7942_story.html

Sweeney, L. (2000). *Simple demographics often identify people uniquely* (Data Privacy Working Paper No. 3). Carnegie Mellon University, Pittsburgh, PA. Retrieved from http://dataprivacylab.org/projects/identifiability/paper1.pdf

Swenson, J. (2014). Establishing an ethical literacy for learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, USA*, 246–250. doi: 10.1145/2567574.2567613

Tansil, R. C. (1941). Development and appraisal of cumulative records at the State Teachers College, Towson, Maryland. Journal of the American Association of Collegiate Registrars, 16(2), 170–186.

Taylor, C. (2012, November 7). Triumph of the nerds: Nate Silver wins in 50 states. *Mashable*. Retrieved from http://mashable.com/2012/11/07/nate-silver-wins/

Tene, O., & Polonetsky, J. (2012). Privacy in the age of big data: A time for big decisions. *Stanford Law Review Online, 64*. 63–69. Retrieved from https://www.stanfordlawreview.org/online/privacy-paradox/big-data

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Privacy, 11*(5), 239–273. Retrieved from http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1/

The confidentiality of university student records: A common law analysis. (1976). *Brigham Young University Law Review, 1976*(2), 477–498. Retrieved from http://digitalcommons.law.byu.edu/lawreview/vol1976/iss2/5/

The Family Educational Rights and Privacy Act of 1974, 20 U.S.C. §1232g.

The Institute for Higher Education Policy. (1998). *Reaping the benefits: Defining the public and private value of going to college*. The Institute for Higher Education Policy, Washington, DC. Retrieved from http://www.ihep.org/research/publications/reaping-benefits-defining-public-and-private-value-going-college

The Open University. (2014). *Policy on ethical use of student data for learning analytics* (Institutional policy). Retrieved from The Open University website: http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf

The Student Right-to-Know and Campus Security Act of 1990, 20 U.S.C. § 1092.

The Student Right to Know Before You Go Act of 2012, H.R. 4061, 112th Cong. (2012).

The Student Right to Know Before You Go Act of 2012, S. 2098, 112th Cong. (2012).

The Student Right to Know Before You Go Act of 2013, H.R. 1937, 113th Cong. (2013).

The Student Right to Know Before You Go Act of 2013, S. 915, 113th Cong. (2013).

The White House. (2015). FACT SHEET: Safeguarding American consumers & families [Press release]. Retrieved from https://www.whitehouse.gov/the-press-office/2015/01/12/fact-sheet-safeguarding-american-consumers-families

Thelin, J. R. (2007). Colleges in the Colonial era. In H. S. Wechsler, L. F. Goodchild, & L. Eisenmann (Eds.), *The history of higher education* (3 ed.) (pp. 54–74). Boston, MA: Pearson Custom Publishing.

Thomas, D. A. (1978). Legal issues in the use and abuse of student records. *The Midwestern Archivist, 3*(1), 3–12. Retrieved from http://www.jstor.org/stable/41101393

Thomas, M. K. (2011). The utility and efficacy of qualitative research software in Grounded Theory research. In V. B. Martin & A. Gynnild (Eds.), *Grounded Theory: The philosophy, method, and work of Barney Glaser*. Boca Raton, FL: BrownWalker Press.

Thomson, J. J. (1975). The right to privacy. *Philosophy & Public Affairs, 4*(4), 295–314. Retrieved from http://www.jstor.org/stable/2265075

Thwing, C. F. (1900). *College administration*. New York, NY: The Century Co.

Tinto, V. (1997). Classroom as communities: Exploring the educational character of student persistence. *Journal of Higher Education, 68*(6), 599–623. Retrieved from http://www.jstor.org/stable/2959965

Trow, M. (1996). Trust, markets and accountability in higher education: A comparative perspective. *Higher Education Policy, 9*(4), 309–324. doi: 10.1057/palgrave.hep.8380051

U.S. Department of Education. (n.d.). Individualized, personalized, and differentiated instruction. Retrieved from http://www.ed.gov/technology/draft-netp-2010/individualized-personalized-differentiated-instruction

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. Higher education* (Contract No. ED-06-C0-0013). Retrieved from http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf

U.S. Department of Education. (2014). *The condition of education 2014* (NCES 2014-083). Retrieved from http://nces.ed.gov/programs/coe/indicator_cva.asp

U.S. Department of Health, Education, and Welfare. (1973). *Records, computers and the rights of citizens* (DHEW Publication No. (OS)73-94). Retrieved from The United States Department of Justice website: http://www.justice.gov/sites/default/files/opcl/docs/rec-com-rights.pdf

U.S. Government Accountability Office. (2014). *State funding trends and policies on affordability* (GAO Report No. GAO-15-151). Retrieved from http://www.gao.gov/assets/670/667557.pdf

Unizin. (2014a, June 11). Why Unizin [Web log post]. Retrieved from http://unizin.org/2014/06/why-unizin/

Unizin. (2014b, September 27). Unizin pilots open content repositories [Web log post]. Retrieved from http://unizin.org/2014/09/unizin-pilots-open-content-repositories/

Urquhart, C., Lehmann, H., & Myers M. D. (2010). Putting the 'theory' back into grounded theory: Guidelines for grounded theory studies in information systems. *Information Systems Journal, 20*, 357–381. doi: 10.1111/j.1365-2575.2009.00328.x

Van Allen v. McCleary, 211 N.Y.S. 2d 501 (1961).

Van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). *Analytics in higher education: establishing a common language* (Report No. ELI3026). Retrieved from the EDUCAUSE Learning Initiative website: http://net.educause.edu/ir/library/pdf/ELI3026.pdf

Van Dijk, J. (2014). Datafication, dataism, and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197–208. Retrieved from http://library.queensu.ca/ojs/index.php/surveillance-and-society/article/view/datafication/datafic

Veysey, L. R. (1965). *The emergence of the American university*. Chicago, IL: University of Chicago Press.

Virunurm, V., & Gaunt, R. N. (1977). Safeguards for the protection of individual records in computerized data banks. *New Directions for Institutional Research, 1977*(14), 55–70. doi: 10.1002/ir.37019771407

W3C Workshop on the Future of P3P. (2002). *World Wide Web Consoritium*. Retrieved from http://www.w3.org/2002/p3p-ws/Overview.html

Waldo, J., Lin, H. S., & Millett, L. I. (Eds.). (2007) *Engaging privacy and information technology in a digital age*. Washington, D.C.: The National Academies Press.

Walsham, G. (1993). *Interpreting information systems in organizations*. Chichester, England: Wiley.

Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems, 4*(2), 74–81. doi: 10.1057/ejis.1995.9

Wang, J., & Yan, Y. (2012). The interview question. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (pp. 231–242). Los Angeles, CA: SAGE Publications.

Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review, 4*(5), 193–220. Retrieved from http://www.jstor.org/stable/1321160

Wellman, B. (2001). Computer networks as social networks. *Science, 293*(14), 2031–2034. doi: 10.1126/science.1065547

Westin, A. F. (1967). *Privacy and freedom*. New York, NY: Atheneum.

Westin, A. F., & Baker, M. A. (1972). *Databanks in a free society: Computers, record-keeping, and privacy*. New York, NY: Quadrangle.

Will, K. H. (2005, March 29). Alma mater as big brother. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/articles/A8331-2005Mar28.html

Will, K. H. (2006, July 23). Big brother on campus. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2006/07/21/AR2006072101396.html

Williams, R. L. (2007). The origins of federal support for higher education. In H. S. Wechsler, L. F. Goodchild, & L. Eisenmann (Eds.), *The history of higher education* (3 ed.) (pp. 310–314). Boston, MA: Pearson Custom Publishing.

Willis, J. E., Campbell, J. P., & Pistilli, M. D. (2013). Ethics, big data, and analytics: A model for application. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application

Willis, J. E., & Pistilli, M. D. (2014). Ethical discourse: Guiding the future of learning analytics. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/ethical-discourse-guiding-future-learning-analytics

Winner, L. (1977). *Autonomous technology: Technics-out-ot-control as a theme in political thought*. Cambridge, MA: The MIT Press.

Winner, L. (1980). Do artifacts have politics? *Daedalus, 109*(9), 121–136. Retrieved from http://www.jstor.org/stable/20024652

Wood, D. M., Ball, K., Lyon, D., Norris, C. & Raab, C. (2006). *A report on the surveillance society: For the Information Commissioner by the surveillance studies network* (Research Report). Retrieved from the Information Commissioner's Office website: https://ico.org.uk/media/about-the-ico/documents/1042390/surveillance-society-full-report-2006.pdf

Wyden, R., & Rubio, M. (2012, September 19). Learning blind. *USA Today*. Retrieved from http://usatoday.com/news/opinion/forum/story/2012-09-17/wyden-%20rubio-student-loans-college/57806404/1

Wyden, R., & Rubio, M. (2014, February 6). Reform starts with good data. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/views/2014/02/06/higher-ed-needs-better-data-spur-reform-essay

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Yorke, M., & Longden, B. (2004). *Retention and student success in higher education*. Maidenhead, United Kingdom: Open University Press.

Zeide, E. (2014). The proverbial 'permanent record.' Retrieved from http://ssrn.com/abstract=2507326

Appendices

A. Initial Interview Request E-mail

Dear [INSERT NAME]-

Introduction:

My name is Kyle Jones and I am a doctoral candidate at the School of Library and Information Studies at the University of Wisconsin-Madison. I'm e-mailing you to request an interview about learning analytics and your institution's development of the technology. Specifically, I'm interested in some of the student privacy issues surrounding learning analytics and how your institution is addressing them. This project is directly related to my dissertation. I am contacting you because others on campus have identified you as someone who can speak about learning analytics, the use of student information for learning analytics, and/or student privacy concerns.

Potential Benefit:

Your participation in this research project will help others to further understand issues of privacy related to learning analytics in higher education institutions.

Interview Process and Request:

Interviews can be done in either a face-to-face or web-based setting and will last about an hour. I am willing to drive to a location of your choosing to speak with you.

If you are willing to participate, please e-mail me at kmjones8@wisc.edu.

If you are unwilling or unable to participate, please feel free to forward this message to your colleagues if you feel they may be able to assist in this research. Also, please let me know that you're not interested in participating by e-mailing me.

Questions:

If you have any questions about this project, myself, or the research team, please feel free to contact me.

Thank you for considering this request,

~Kyle M. L. Jones~

Doctoral Candidate

School of Library and Information Studies, UW-Madison

B. Interview Setup E-mail

Dear [INSERT NAME]-

Thank you for agree to participate in an interview. Please send me a list of days and times for us to meet that work well with your schedule.

When choosing a location for us to meet, please take into consideration the following:

- choose a location that is comfortable for you;
- choose a location that allows you to speak freely without concern for others who may overhear our conversation;
- choose a location that does not have a lot of ambient noise so as to limit potential interference with the audio recording of the interview.

Many people I speak with choose to meet in their office, and this is usually a great spot. Before we meet, I request that you review and sign the virtual consent form available at: [INSERT LINK TO CONSENT FORM] I will bring a print copy to our meeting in order to review the form and answer any questions. At a later time, I will e-mail you a copy of the signed consent form for your records.

Thanks again for participating, [INSERT NAME]. I am looking forward to hearing your perspective on learning analytics and student privacy.

Sincerely,

~Kyle M. L. Jones~

Doctoral Candidate

School of Library and Information Studies, UW-Madison

C. Consent Form

UNIVERSITY OF WISCONSIN-MADISON

Research Participant Information and Consent Form

Title of the Study: Building capacity for learning analytics technologies and related student privacy issues

Principal Investigator: Kristin Eschenfelder, PhD (phone: 608-263-2105) (email: eschenfelder@wisc.edu)

Student Researcher: Kyle M. L. Jones, MLIS (phone: 608-263-2900) (email: kmjones8@wisc.edu)

DESCRIPTION OF THE RESEARCH

You are invited to participate in a research study about student privacy issues related to learning analytics technologies in higher education institutions.

You have been asked to participate because you have been identified as an actor at a higher education institution who is directly or indirectly involved in your institution's capacity building for learning analytics technology.

The purpose of the research is to:

1. Identify what issues concerning student privacy are relevant to actors involved in designing, adopting, and using the technology;
1. address how student privacy issues are managed through system design and institutional practices;
1. and establish larger contextual factors that influence student privacy as it relates to the use of learning analytics technologies.

This study will include potential subjects who fill these roles:

1. Higher education administrators and policy makers (e.g., presidents, CIOs, library directors, registrars, legal services, etc.)
1. Technologists (e.g., instructional technologists, instructional designers, campus technology administrators, etc.)
1. Faculty (e.g., who are on technology committees, who use learning analytics in their teaching and learning practices)
1. Developers of learning analytics technologies as suggested by interviewees (e.g., commercial vendors, members of the open source community related to learning analytics)

This research will either be conducted in a private place of your choosing or in Adobe Connect, a web- conferencing application for audio and video communication. Audio recordings will be made of your participation. Only the research team will hear the audio recordings made of the interview. Transcripts will be made of the recording to be analyzed by the research team. Both audio recordings and transcripts will be stored securely on UW-Madison servers.

WHAT WILL MY PARTICIPATION INVOLVE?

If you decide to participate in this research you will be asked to participate in one face-to-face or web- based, open-ended interview with a member of the research team.

Your participation will last approximately 30-60 minutes per session and will require 1 session which will require 1 hour in total. Follow-up interviews may be requested of you, but you are not required to participate then if you do not wish to.

ARE THERE ANY RISKS TO ME?

There are minimal risks to you. You could reveal information that may negatively impact the reputation of yourself and/or your institution. To reduce this risk, you and your institution will be given pseudonyms to protect your identity and your institution's identity; furthermore, any characteristics that clearly identify you or your institution will be removed or significantly altered in project disseminations. The pseudonyms will be used when the findings of the research project are disseminated. If you decide not to participate or to withdraw from the study it will have no effect on any services or treatment you are currently receiving. You may at any time opt-out of answering questions and/or stop your participation completely. Digital audio recordings will be used as part of the data collection process. These audio recordings and the transcripts made from them will be stored using encryption technologies to protect your identity. All audio recordings will be destroyed after the project's completion.

ARE THERE ANY BENEFITS TO ME?

There are no direct benefits for you for participating in this study. However, learning analytics technologies have the potential to enhance assessment processes in higher education, create improved learning outcomes, and increase institutional transparency and efficiency. But, the literature recognizes that there are privacy issues that need to be addressed in order to reduce risk. Your participation in this study will help to address those privacy issues and make recommendations regarding policies and procedures about student privacy, which may improve higher education.

HOW WILL MY CONFIDENTIALITY BE PROTECTED?

This study is confidential. Except where indicated on this consent form and when captured by the digital audio recording, neither your name or any other identifiable information will be recorded.

WHOM SHOULD I CONTACT IF I HAVE QUESTIONS?

You may ask any questions about the research at any time. If you have questions about the research after you leave today you should contact the Principal Investigator Kristin Eschenfelder, PhD at 608-263-2105. You may also call the student researcher, Kyle M. L. Jones, MLIS at 608-263-2900.

If you are not satisfied with response of research team, have more questions, or want to talk with someone about your rights as a research participant, you should contact the Education Research and Social & Behavioral Science IRB Office at 608-263-2320.

Your participation is completely voluntary. If you decide not to participate or to withdraw from the study it will have no effect on any services or treatment you are currently receiving.

Your signature indicates that you have read this consent form, had an opportunity to ask any questions about your participation in this research and voluntarily consent to participate.

You will receive a copy of this form for your records.

SIGNATURE OF CONSENT

Name of Participant (please print):

____

E-Mail Address of Participant (please print):

____

Signature: ____

Date: ____

IRB Approval Date: 10/9/2014

Date IRB Approval Expires: 10/8/2015

FWA00005399 ED/SBS IRB University of Wisconsin – Madison

D. Interview Script

Interview Protocol - Interviewer Script

Date:____

Time:____

Institution/Organization:________________________________

Participant:____

Pre-Interview Protocol

Introduction

Thank you for agreeing to meet with me today and participate in this study. The purpose of this study is to discuss some of the student privacy issues that are related to learning analytics technologies. I'm interested in your relationship to the learning analytics project at [participant's institution], if your institution has encountered privacy problems, and what they've done–if anything–to resolve them.

Today, we'll be having an open-ended conversation. While I have questions I ask all of the research participants, most of this conversation will be focused on issues, concerns, and ideas important to you. As such, there is no right or wrong answer, and I am not making any particular type of judgement about you or your answers.

Think of our conversation today as a way for you to educate me, a stranger to your institution and a person who is trying to make sense of how you are experiencing some of the issues in your position. I may ask some simple questions, and I may ask some seemingly simple follow-up questions, but this is because I do not know much about your experiences and role at this institution.

Informed Consent & Participant Rights

If you haven't yet filled out the informed consent form online, I'd like to take this time to for you to review it, ask any questions of it, and sign it.

[Provide consent form, either a physical copy or a link to the online consent form]

I'd also like to remind you that you may opt-out from answering any questions or withdraw from the interview completely if at any time you feel uncomfortable or do not want to continue. Let's review the consent form. Take some time to read through it. At this point, do you have any questions?

Audio Record the Interview

At this time, I will begin to record our conversation. No one besides the research team will have access to the audio recording. After our time today, I will transcribe the recording. This process allows me to accurately capture what you're saying and helps me build a more complex and rich understanding of our conversation. Remember, all personally identifiable characteristics will be anonymized, and any findings from my interviews will hide your identity and that of your institution/organization. Do you have any concerns about the audio recording? Would you prefer not to be audio recorded?

[IF NO, Start recording]

Interview Protocol

Information about the interviewee

I'd like to begin by getting a better understanding of your role at the institution and your relationship to the learning analytics project here on campus. Let's begin with some general questions.

- To help me understand your role at this institution, please briefly and generally describe your job and some of your responsibilities.
- Please tell me what role you have in the learning analytics project on your campus. I encourage you to talk about specific aspects in the project you've worked on.
- As a follow up, please explain why you think you are involved in the project. For example, some interviewees have detailed that they have a particular set of skills to offer to the project or leadership acumen to assist the project's progress.

Motivations for using learning analytics

Quickly, I want to ask you about what you think is driving learning analytics.

- Why do you think learning analytics is emerging now at colleges and universities?
- Is there anything in particular about the motivations behind learning analytics that you wonder about or question?
- Why do you think your institution has decided to pursue learning analytics?

Information privacy and learning analytics

Let's move forward, getting into specifics now about student privacy and learning analytics. In the literature about learning analytics, researchers and practitioners alike have mentioned that student privacy is a concern that needs to be addressed in learning analytics projects. For this set of questions, I'll ask about student privacy and learning analytics generally.

- First, how do you define student privacy?

- As a follow up, what specific privacy rights should students be afforded? Some have said that students should be allowed to control their student data; others state that students should expect a general right of privacy, but that some exceptions would have to be made for the institution to use their information.
- From your point of view, would you agree with those who say that learning analytics presents student privacy problems, and please explain your opinion to help me understand why or why not.
- I have heard some people say that student privacy is an issue because of the types and amounts of data used; others say that it's not the data that is the problem, but the learning analytics technology itself that makes the private information visible; elsewhere I've heard that the privacy problems exist because colleges and universities don't have the policies and procedures in place to handle the use of the sensitive data. From your vantage point, where do you think privacy problems are most evident: data use, technologies that make the data visible, or in inadequate or missing policies? Do you think the privacy problems occur because of other reasons not listed?
- In the United States, the Family Educational Rights and Privacy Act, or FERPA, is the guiding law where student privacy is concerned. Are you familiar with FERPA, and how do you think the law helps people like yourself and those on your learning analytics projects address and resolve student privacy issues?
- Some interviewees have said that third parties, such as future employers, should be given access to raw student data and/or the information learning analytics creates as a byproduct. What are your thoughts on this?

Privacy issues relevant to the interviewee's project

Ok, now I'd like to ask you specifically about your institution's learning analytics project. Other interviewees have reported that they've had to address student privacy issues in number of ways as they build up capacity for learning analytics on their respective campuses.

- First, to your knowledge, are you aware of any student privacy issues or concerns that you or anyone on your campus has had to address in order to build capacity for learning analytics?
- If yes: Please detail what the issue or issues were, who was involved, and what was done to address the problem.
- If no: Since you're unaware of any issues, what do you think your institution will have to do to make sure they are able to address any privacy issues that emerge as a result of learning analytics?
- Does your institution have in place or is it actively creating policy to address some of the unique privacy issues related to learning analytics?
- If yes: Please describe some of the aspects of the policy, as well as how and why the policy was developed.
- If no: You've said that you're unaware of any policies related to learning analytics. In your opinion, does your institution need to develop policies and what do you think those policies would address?
- Some institutions I've spoken with are developing data governance committees, due in part to some of the privacy issues presented by learning analytics. Do you think such a committee needs to be created or bolstered at your institution, and how do you think it would provide guidance in relation to learning analytics technologies and practices?

- Now, let's talk for a moment about learning analytics technology. What kind of student privacy issues does learning analytics technology present to your institution? For example, some past interviewees have said that it affects network security, others state that data aggregation and quality have become concerns, while others have said that some of the interfaces students and teachers use needed to be redesigned to hide some sensitive information.

Values

For the final set of questions, I want us to talk about learning analytics a little more abstractly.

- Do you think that the use of a student's personally identifiable information for learning analytics is justifiable? When might it not be justifiable?
- Who do you think benefits from learning analytics? Some say students do, others think the benefits go to the institution.
- If your institution used personally identifiable data for learning analytics but didn't tell the students they were doing so, do you think that would be fair to students? When is the use of the information fair and when isn't it?

Post-Interview Protocol

Participant Questions

Are there any questions about this topic that are interesting to you that you feel we did not cover in our interview today?

Participant Comfort/Researcher Effect

During our conversation today, did anything prompt you to respond differently than if you were speaking with a close, trustworthy friend?

Nominated Sample

Is there anyone else at your institution/organization or involved in the learning analytics project that you feel I should contact to inquire about an interview? Why these individuals?

Follow-Up

If you have any questions about the progress of this research or our interview, or you would like to follow-up with more information, please feel free to contact me.

[Provide business card]

E. Codes

CODE	PAGE NUMBERS	INTERVIEWEE
"A tool or a bludgeon"	29	LR
"At the table"	9	JP
"How you slice the data" determines la's success, usefulness, invasiveness, etc.	5	JR
"Make the case for data"	34	SM
"Normal range of academic information"	49	LR
"On the cutting edge"	19	SM
"Prioritizing" vs. "Targeting"	12,13	SebMuel.
"School official"	11	JS
"Self-Fulfilling Prophecies"	11, 23, 24, 25, 30	SN
"Tail wagging the dog" with FERPA and LA	27	JR
"The black box"	17, 18, 26, 28, 31	JP
Access and use concerns	40, 45, 47, 49, 52	SebMuel.
Access levels and roles	60, 61	MP
Accessing more data	14, 15, 16	SW & LG
Accountability over privacy	21, 22	RM
Acting on the data	13, 17	SJ
Actionable information	6,8,12	SebMuel.
Ad Hoc Policies	11,23	SN
Aggregate vs. targeted data	11,12	AS
All privacy is "the same thing"	59	JP
All the data they can get	14	RM
Analytic frame	38, 39, 45, 50	LR

Asking critical questions of data	16	SM
At the table	69	LR
Balance of benefits	77, 78	JP
Balancing privacy with educational opportunities	4,30,31	RM
Balancing privacy with job/ institutional responsibilities	13, 17	SW & LG
BI interest in LA	27	MW
Bias	42, 48	SebMuel.
Big brother	15, 16	SW & LG
Bleeding edge	6	SW & LG
Bridging roles	5	LR
Brightline: what is necessary and appropriate (e.g., race identification is questionable)	7	JR
Building relationships	2,21	SW & LG
Campus decision to not show past performance	5	JR
Capacity problems	9,10,11	SJ
Capacity, efficiency, and effectiveness	19, 25, 26, 27, 28, 45	LR
Card swipe as data source	37, 38, 40	SN
Case management	10	LR
Case site differences	9	RM
Centralizing control	10	JS
Chain of data choices, conservative choice	19, 26, 29	JR
Changes to FERPA	21	RM
Changing policy norms	66, 67	SN

Changing roles with regard to data	5,7	RM
Changing technological landscape	8	SJ
Choosing Models	15, 16, 17, 18, 20, 21	SJ
Classifying data	29	RM
Co-opting LA	33	SebMuel.
Conflicts and limits with record review rights	21, 22	JS
Connecting data	7,8	CJ
Contextual factors	5	SM
Contextual privacy expectations	22	RM
Creepiness of LA	37, 49	RP
Critical informants	1,2	JR
Cultural incompatibilities	57	MP
Culture of innovation in higher education / technology enhanced learning / Big Data "elsewhere"	8,9	JR
Data access concerns	23	JP
Data and information inaccuracies	15	SM
Data choices should be driven by institutional needs	33	SM
Data Dashboards as a form of control	22, 23	AS
Data dictionary	3, 7, 9, 10, 12	JP
Data gatekeepers	4, 13, 14, 15, 16, 17, 7	AS
Data governance	3,4	SW & LG

Data grabs and awareness thereof	36, 37	AS
Data inclusion criteria	27, 28, 29, 31	SebMuel.
Data must be used with real life experience	11, 25	JR
Data ownership	13, 29, 30, 48	SJ
Data policies	20	CJ
Data problems	11,12,13	SJ
Data silos	9	JP
Data steward vs. data gatekeeper	4,5	RM
Data stewardship	11,12	SJ
Data warehousing	4	SW & LG
De-identification concerns	7	JS
Dealing with the "whole person"	11	LR
Defining educational records	19, 24, 25	RM
Defining LA	17, 18, 25, 26	CJ
Defining student privacy as a systems process	23	JR
Denying student choice	13, 14	JS
Desensitized to privacy issues of LA	21	SM
Difference between gathering and using data	8,17,19	JR
Different institutional approaches to privacy due to culture	30	SM
Digital dossiers	40	SN
Directory vs. confidential data / information	2	RM

Disclosing data to build relationships	29	RM
Distribution of knowledge about LA across roles		JR
Early Alert Systems	8,9	SN
Educational records	18, 19, 20, 21, 22	JS
Effecting positive interventions	21, 22	LR
Ethical conflict	18	JS
Ethical data use	14, 15, 20, 23, 24	CJ
Expectation of de-identification	59	MP
Expectation that institutions will protect data	57	RP
Expert vs. community approach	7	SW & LG
Faculty analytics	20	JP
FERPA as the floor	67	MP
FERPA compliance problems	20	SW & LG
FERPA flag	31	CJ
FERPA influences perspectives about privacy	18	AS
FERPA loopholes	31, 33	CJ
Fine tuning of LA	42	RP
FoCI flag	25, 26	JS
Focusing on the algorithm	10	AS
Frustrations with LA	5,6	SM
Gaining institutional efficiency, increasing effectiveness	12, 13, 14, 17, 18	SM
Gatekeepers	6,10	JS

Gatekeeping activities	3, 4, 24, 25, 34	RM
Gathering vs. using the data	23	CJ
Geolocation data	46	CJ
Getting vendor technology to work together	16	SM
Goals of higher ed. in conflict with LA	47	LR
Going beyond privacy baselines	67	SebMuel.
Good alignment	7,9,8	LR
Have to be very careful with predictions	7	JR
Having conversations	8,9,10	SW & LG
Hidden data	9, 10, 14, 22, 29, 31, 34	CJ
Hiding data/information	6,7	RM
Higher ed lacks interest in student privacy issues	30	MW
Higher student/faculty awareness of privacy issues	21, 29	SM
Historical data about students	11,14	LR
Historical interest in analytics	5	LR
Holistic view of students	54	SN
Hype / on the cutting edge	27	MW
Hypothesizing	29, 2, 8, 46	CJ
Identifying at-risk students	10	MP
Identifying trends	3,31	MP
Improving data access	2, 5, 36, 38, 57, 63, 86	JP
Improving predictions	7	SebMuel.
Inclusivity mitigates issues	9	RM

Info. in models is the same "any faculty member" would be thinking about	6,10	JR
Informed consent	32	SJ
Informing students	39, 40	JP
Institution benefits from LA	77	MP
Instructional incompatibilities	4,5,6	MP
Interpretations of FERPA	6,2	RM
IRB tipping point	6	CJ
IRB: research vs. business purpose	4	JR
Is identifying students appropriate	5	JR
Iterative consent	53, 54	LR
Justifying data collection	15, 16, 17	JS
Justifying information requests	38, 42	CJ
Keystroke tracking transforms normal measurements of student tracking	6	JR
LA affecting student autonomy	29, 44	LR
LA as "underwriting" student success	15, 27	JP
LA as a "tempting target"	42	MW
LA as a leveraging technology	18, 20, 27	LR
LA as Big Data	3, 10, 34, 43	AS
LA as identifiable data	59, 71	MP
LA as processed data / not educational record	75, 76	MP

LA data and educational records	68	SN
LA reveals new information, trips privacy flag	4	JR
LA will force FERPA changes	40, 41	CJ
Large-scale LA	3, 10, 11, 34, 39	SJ
Layers of access	31, 33, 41, 42, 49	SN
Learning analytics data and appeals/hearings	20,21	SN
Legal policies and powerful actors determine privacy choices (FERPA, legal, IRB)	6	JR
Less awareness of higher ed privacy problems	30, 31, 41	MW
Levels of LA	6,37	LR
Library data	41	SJ
Limit profile data	72, 73	MP
Limited opt-out opportunities	65	MP
Limited predictive ability	60	SebMuel.
Limits on access to LA data	34	JP
Limits on building predictive models	11	RM
Limits on control	19, 20, 25	AS
Limits on info flows are situational	23, 24	SM
Location of data	5	SW & LG
Longitudinal databases	23	JS
Losing control of data to feds	23	JS
Making a better student	19	JP

Making informed decisions	44	MP
Making the system reflect the campus	5	AS
Managing the data lifecycle	2	RM
Maximizing data accuracy	55	MP
metadata vs. content	15	LR
Missing a data baseline with LA	54	MP
Missing roles in LA	8,9	RM
Missing voices	41, 2, 4, 5	AS
Mitigating downstream effects	13	JS
Moral imperative to help at-risk students	78	SebMuel.
More data the better	3	AS
Motivations for LA	23, 24, 25	LR
Need for centralization of data efforts	14	JP
Need for training	30	SN
Need to de-identify for reporting	60, 62	MP
Need to justify LA practices	48. 50, 54	LR
Need to know	9, 22, 29, 31	CJ
Needing an infrastructure	7	SW & LG
Needing consent	6	JS
Needing new policies	39	SM
Needing the "right data"	25, 35	SM
Needing to define student data	48, 49	JP
Negotiating contracts not new	11	CJ

Negotiating data access	11	CJ
New normal and opt-in/out	16	JR
New privacy problems	3	CJ
New role responsibilities	13	SW & LG
New roles	2,3,13	JP
New tools	1	LR
Not knowledgeable of LA privacy issues	68	MP
NOT needing all the data	38, 40	LR
Not needing consent	13, 17	JS
Off Limits Data	26, 27, 29, 30	SebMuel.
Offloading privacy conversations (related to data gatekeepers)	17	AS
On the cutting edge	36	RP
Open questions	2, 13, 21, 23, 24	CJ
Optimizing access	34	SN
Opting In/Out	29, 30, 31, 33	SJ
Overwhelmed with, noise of data	25, 27, 35	SM
P3P and "protected areas"	36	MW
Patching FERPA with institutional policies	58, 59	LR
Paternalism	3,4,38	CJ
Permissions	9,10	JS
Personal and professional conflict	29	MW
Personalized learning	12	RM
PII: limitations and opportunities	6,11	JS

Policy: missing, needing, building	4,6	SW & LG
Politics of LA	22	RP
Politics of student privacy	26, 27	SJ
Predictions limited by data sources	18, 27, 28, 31	JP
Prioritizing resources	8	SebMuel.
Privacy and campus vs. public contexts	71	SebMuel.
Privacy as limited access	16	SW & LG
Privacy dashboard	38	MW
Privacy flags	37	LR
Privacy flags and tracking	15, 16, 30	RM
Privacy inverse in online ed.	4	CJ
Privacy literacy	18	SW & LG
Privacy-as-control	29, 30, 31, 32, 35, 47	SJ
Privacy-as-trust	29, 30, 31, 32, 36, 41	SM
Problems with decentralized LA	33, 34	LR
Problems with LA project	5,6	SM
Professional / personal conflicts	47, 48	SJ
Profiling	42, 16, 23	SJ
Protected data	14, 15	SJ
Protecting against bias	7	RM
Protecting against misuse	63	SN
Questioning perception of changing privacy norms	33	MW
Re-identifying data	17	SW & LG

Redefining educational record	13	SW & LG
Registrar and enrollment services are key determiners of FERPA	4	JR
Removing data weakens the model (due to privacy)	5	JR
Respecting privacy expectations	38, 39	CJ
Retention and academic predictions	44, 45	SebMuel.
Revealing surveillance practices	26	SN
Review and inspect rights and expectations	26, 27, 28	RM
ROI of privacy projects	36, 37	RM
Role of predictive analytics	16	CJ
Roles	1	JR
Scaling LA	10,11	SW & LG
Self-fulfilling prophecy	40, 41, 42, 48	SebMuel.
Self-governing	20	SW & LG
Signing away control rights	61	RP
Single source data / data warehousing	66	LR
Smorgasboard approach	24, 25, 26	JP
Solving problems	6	CJ
Some info/data required to start an institutional relationship	33	SM
Some statistical error allowed	6	JR
Specific data and systems as official educational records	12,15	LR

Squishy data	10	LR
Strong predictions require lots of data	13	JR
Student autonomy concerns	39	SebMuel.
Student awareness and wherewithal	11	AS
Student centered culture limits privacy issues	21, 22, 28	SM
Student control of data	12	SW & LG
Student participation in data use	33	SM
Student privacy as limited access	59	MP
Student privacy is self-defined	74	SN
Student privacy rights	62, 63, 74	SebMuel.
Student response will be "interesting"	14	JR
Surveillance	26, 27, 41	CJ
Swallowing the rule	7,23,24	JS
Tail wagging the dog	19, 20	SW & LG
Technically difficult to provide privacy controls	65	MP
Technological enthusiasm lacking balance	29	MW
Telling stories	27, 31	AS
Telling students	18, 20	SW & LG
The "ultimate" justification	32, 21	JP
The data gatekeeper	12	CJ
The important questions	13	JS
The right data	2,3	CJ

Third party sharing	11,17	JS
Third-Party data	38, 40	SJ
Tracking already required	18	JR
Tracking faculty activity	50	SN
Training	102, 103	JP
Transparency	55, 58, 60	LR
Trusting the code	28	AS
Turf wars	7	AS
Unanswered questions	3	SW & LG
Updates will "erode" FERPA	45	MW
Using data	3,15	JS
Using LA information carefully	54	RP
Using LA to assess faculty	33, 41	MP
Using LA to make informed decisions	32	MP
Using LA with real-life experience	41	LR
Vendor lock-in	7, 8, 29, 40	AS
Where the data resides	12	JR
Who benefits from LA	73	RP
Who owns student data	49, 50	MW

F. Memo Summaries

MEMO TITLE	SUMMARY
"Same Thing We've Always Done"/Data Use with Experience	Memo discusses an opinion that using learning analytics is nothing different from what teachers due with already existing student data.
A Holistic Understanding	Memo discusses how learning analytics runs counter to institutional values that encourage actors to get a complete understanding of a student, not just a data-driven snapshot.
A Need to Know	Memo discusses an interpretation of FERPA that allows specific institutional actors to get access to protected student data.
Acting on Data	Memo discusses a perspective that instructors need access to raw data as well as the information borne from learning analytics systems.
Aggregated Versus Targeted Data	Memo discusses how privacy issues are reduced when student data is used in aggregate instead of to target interventions with identifiable students.
Analytic Frames	Memo discusses how different actor roles have unique perspectives of student privacy.
Balancing Privacy	Memo discusses a perspective that student privacy needs to be balanced with the benefits of learning analytics.
Changing Privacy Norms	Memo discusses a perspective that students no longer care about privacy.
Data Gatekeepers and Czars	Memo discusses how certain actor roles are perceived to be in control of student data.
Data Gatekeepers and Turf Wars	Memo discusses who controls student data and politics surrounding those controls

Data Governance	Memo discusses the need for data governance mechanisms.
Data Grabs	Memo discusses concerns about third-party access to student data.
Data Inclusion Criteria	Memo discusses a perspective that institutions should use clear criteria for selecting what data should be included in learning analytics technologies.
Data Oversight	Memo discusses issues related to data governance.
De-Identified Data	Memo discusses a perspective that student data should be de-identified whenever possible.
Defining Educational Records	Memo discusses how educational records are defined and by whom.
Educational Records, Rights, and Limitations	Memo discusses how educational records are defined and limitations to how students can express their legal privacy rights.
Effects of Campus Decisions	Memo discusses how campus decisions with regard to information policy and technological design impacts the usefulness of learning analytics technology.
Ethical Obligations	Memo discusses how model builders have an ethical obligation to build statistically powerful models.
Ethics	Memo discusses codes of ethics and limitations thereof.
Finding Efficiency and Cost Reductions with Learning Analytics	Memo discusses a perspective that learning analytics is being used to increase institutional efficiency due to institutional budget cuts.
Gatekeeping Activities	Memo discusses specific activities gatekeepers, such as registrars, undertake with regard to student data.
Gathering Versus Using the Data	Memo discusses how privacy issues are impacted by how the data is used, not aggregated.

General Reflections for Alan Simons	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Connie Johanson	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Conor Vergne	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Johanna Stevens	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for John Roberts	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Josef Prost	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Lewis Rosberg	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Lisa Goldsmith	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Lynn Unser	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Marco Power	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Matthew Walker	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Richard May	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.

General Reflections for Rick Penske	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Sabine Nice	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Sebastian Mueller	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Shannon Jones	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Shea Montoya	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Sierra Jones	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
General Reflections for Stanley Williams	Memo discusses the interviewee's final thoughts and reflections about the interview with the participant.
Goal Alignment	Memo discusses how the goals of learning analytics may or may not align with the goals of particular actor roles.
Greater Awareness	Memo discusses how learning analytics is raising awareness of student privacy among institutional actors.
Impacting Student Autonomy and Academic Success	Memo discusses how predictive analytics may reduce students autonomy with regard to information that could affect their decision-making capability, especially as it relates to the student-advisor relationship.
Importance of Training	Memo discusses how institutional actors should be trained to interpret learning analytics information in order to reduce privacy issues.

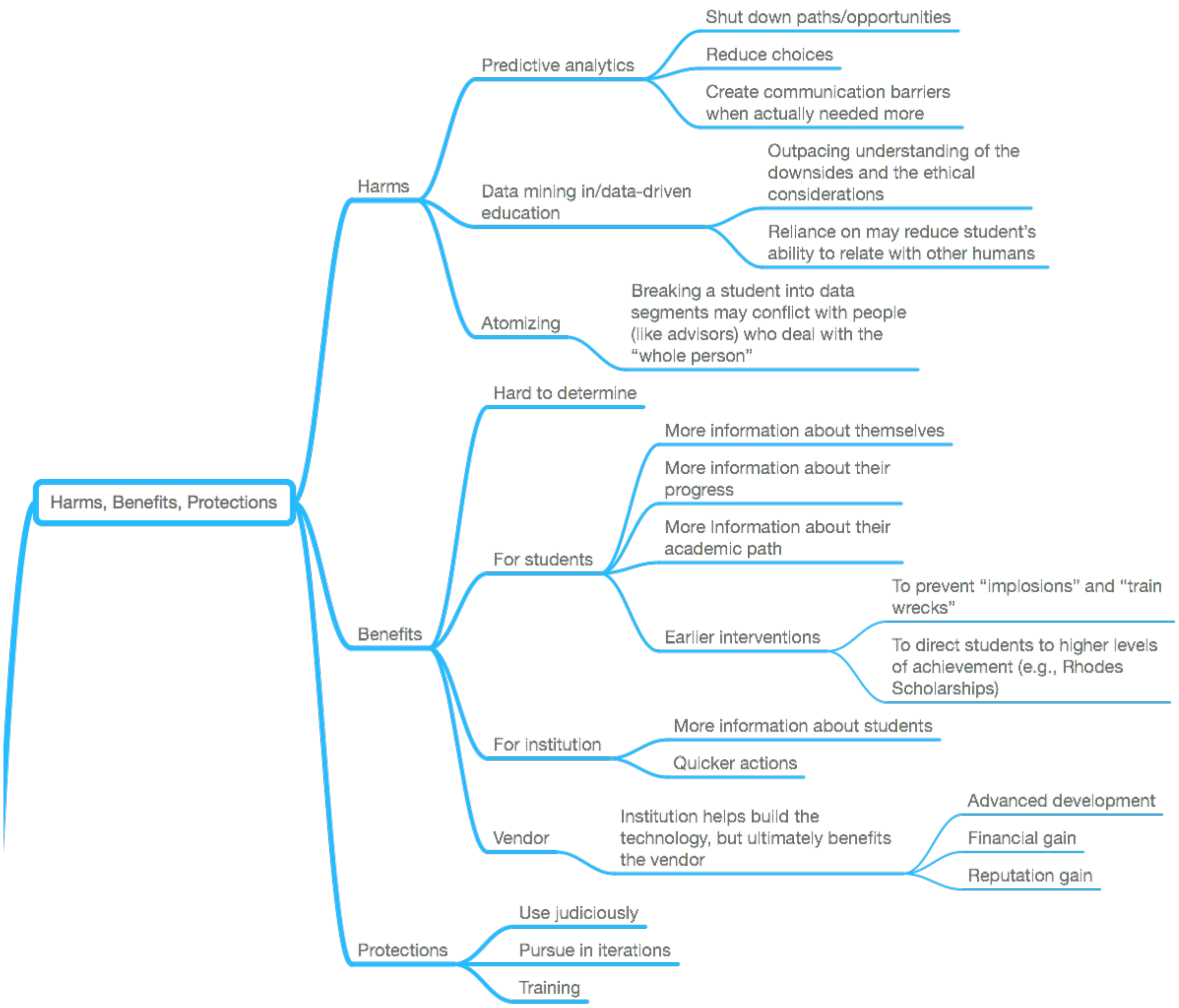
Incompatibilities	Memo discusses instructional, institutional, and cultural incompatibilities with learning analytics technologies.
Informing Students (1)	Memo discusses a perspective that the institution needs to inform students of learning analytics practices.
Informing Students (2)	Memo discusses how students should be told how learning analytics is being used and by whom on campus; emphasizes the role of transparency.
Inspecting Learning Analytics Data	Memo discusses why students would want to inspect data derived from learning analytics technologies.
Interpreting FERPA	Memo discusses how FERPA is interpreted differently by institutional actors and the effects thereof.
Learning Analytics as a Big Data Practice	Memo discusses how learning analytics is comparable to other Big Data practices.
Learning Analytics: New or Old Practice?	Memo discusses a perspective learning analytics is not a new practice with regard to how institutions use and analyze student data.
New Information from Learning Analytics is a Norm Changer	Memo discusses whether or not learning analytics affects information norms.
Paternalism	Memo discusses how institutions have a paternalistic responsibility to protect students and their data.
Politics of Learning Analytics	Memo discusses how some institutional actors politicize their involvement in learning analytics initiatives.
Predictive Analytics: Autonomy Reducing	Memo discusses how predictive analytics may reduce students autonomy with regard to information that could affect their decision-making capability.
Privacy Flags	Memo discusses specific privacy concerns related to learning analytics practices.

Privacy Hinges on Data Access	Memo discusses how who gets access to student data and how ultimately determines student privacy protections.
Privacy Norms and the Law	Memo discusses how changing privacy norms have influence amendments to FERPA.
Privacy-as-Control (1)	Memo discusses student privacy as a student's right to control their information and how FERPA supports that definition.
Privacy-as-Control (2)	Memo discusses a growing argument among participants that student privacy is the right to control data about themselves.
Privacy-as-Control and Limits Thereof	Memo discusses existing controls students have to protect their privacy and institutional limitations on controls.
Problems with Data Ownership	Memo discusses how providing students an ownership right over their data is problematic.
Relationship Building and Control Definitions of Privacy	Memo discusses a perspective that students need to disclose information about themselves to establish a relationship with the institution.
Required to Track/Automated Tracking	Memo discusses how higher education institutions are required to track student data for reporting measures.
Roles	Memo discusses new and emerging roles.
Scaling Learning Analytics (1)	Memo discusses the value of scaling learning analytics throughout an entire institution.
Scaling Learning Analytics (2)	Memo discusses how scaling learning analytics is problematic and challenging.
Self-Fulfilling Prophecies	Memo discusses how predictive analytics may limit how students perceive their abilities and potential for success.
Signal and the Noise	Memo discusses how aggregating as much student data as possible may actually limit learning analytics.

Slicing the Data	Memo discusses how privacy issues are impacted by how the data is analyzed.
Student Response	Memo discusses perspectives among institutional actors regarding the student response to learning analytics.
Swallowing the Rule	Memo discusses how particular interpretations of FERPA reduce student privacy protections.
Tail Wagging the Dog	Memo discusses the perspective that policy and law is play catching up with emerging technologies.
Telling Stories	Memo discusses a strategy for reducing privacy concerns among students.
Telling Students (1)	Memo discusses how students should be told how learning analytics is being used and by whom on campus.
Telling the Students (2)	Memo discusses a perspective that instructors should tell students how they are using learning analytics in a class.
The Black Box	Memo discusses how actors do not understand how learning analytics systems come to predictive conclusions about students.
The Holistic View	Memo discusses a perspective that an advisor's job is to get a complete view of a student.
The More Data the Better	Memo discusses an argument that aggregating as much data possible for learning analytics will improve the insights borne from the technology.
The New Normal	Memo discusses opt in/opt out processes if learning analytics becomes normalized in the future.
The Right Data	Memo discusses an opinion that institutions need the right data, not all the data.

Tipping Point for the IRB	Memo discusses the role of the institutional review board in determining human subjects protections
Trust and Privacy	Memo discusses a perspective that student privacy involves trust between students and their institution.
Trusting the Code (1)	Memo discusses a perspective that institutional actors should trust the code technology vendors write.
Trusting the Code (2)	Memo discusses a perspective that trusting the code technology vendors write is problematic.
Turf Issues	Memo discusses how some actors are politicizing learning analytics for their own gain and due to a lack of trust.
Vendor Lock-In (1)	Memo discusses the effects of being contractually attached to one educational technology vendor.
Vendor Lock-In (2)	Memo discusses the effects of being contractually attached to one educational technology vendor.

G. MindNode Sample


Harms, Benefits, Protections
Harms
Predictive analytics
Shut down paths/opportunities
Reduce choices
Create communication barriers when actually needed more
Data mining in/data-driven education
Outpacing understanding of the downsides and the ethical considerations
Reliance on may reduce student's ability to relate with other humans
Atomizing
Breaking a student into data segments may conflict with people (like advisors) who deal with the "whole person"
Benefits
Hard to determine
For students
More information about themselves
More information about their progress
More information about their academic path
Earlier interventions
To prevent "implosions" and "train wrecks"
To direct students to higher levels of achievement (e.g., Rhodes Scholarships)
For institution
More information about students
Quicker actions
Vendor
Institution helps build the technology, but ultimately benefits the vendor
Advanced development
Financial gain
Reputation gain
Protections
Use judiciously
Pursue in iterations
Training