Big Data in Education:
Developing Policy for Ethical Implementation in the US and Canada

Priscilla M. Regan
George Mason University

Jolene Jesse
National Science Foundation

Elsa Talat Khwaja
George Mason University

The growth of "big data" and the concomitant development of sophisticated analytical techniques for generating data, designing data sets, culling the data for patterns and trends, and identifying either individual or group prototypes of behavior raise the promise of a host of societal benefits – but also a number of more disquieting possibilities. In the education arena, the benefits include more sophisticated analyses of student learning and testing, more personalized learning, more effective delivery of educational materials, improved assessment, and more responsiveness to student needs. On the downside big data applications and products raise the possibility of discrimination as a result of profiling and tracking of students, as well as uses of student information for a wider range of purposes. With increased emphasis on the need to improve student learning, especially at the K-12 level, a number of actors are involved in marketing more sophisticated analytical products, approving the use of these products, and using them.

As all countries recognize the importance of competing in the global environment and as the world becomes more of a global village as a result of economic and social activities facilitated by the Internet, countries around the globe are directing attention and resources on improving educational achievement especially at the primary and secondary levels. With the concomitant increase in the costs of providing education and concerns about financial responsibility, heightened consideration of accountability and results, elevated awareness of the range of teacher skills and student learning styles and needs, more focus is being placed on the promises seemingly offered by online software and educational technology. Information technology companies recognize the huge market offered by K-12 education and are aggressively developing and marketing their products. Most of these companies are large international ones based in the United States, such as Google and Microsoft, but a range of new companies, still largely based in the United States, now populate the market.

This paper begins to explore the policy landscape in which the approval of "big data" educational tools is taking place in both the US and Canada. Are local schools making decisions? Or school boards? Or state and provincial education departments? Are legislative bodies or executives involved? The paper is particularly focused on exploring how the ethical issues with respect to educational technology and big data use are being framed and whether that differs by forum and/or by location.

Big Data in Education

Over the last twenty years, technology has become ubiquitous in classrooms at all levels, especially at the K-12 level. As noted above, this is explained by a number of factors including the focus on STEM education, the general social trend to a more technologically sophisticated society and the need to prepare upcoming generations to compete in that environment, and the fact that computer-assisted learning might entice students into engaging with material, which might normally have appeared less attractive. But other factors are also at work here, especially the pressure for student achievement, teaching effectiveness, controlling the school budget and also the interests of the technology companies in this seeming lucrative market. Additionally venture capitalists see the growth potential in the education market and are investing in ed tech start-ups – and finally, large wealthy foundations, such as the Bill and Melinda Gates Foundation, believe that technology offers many tools for improving the educational experience for children. The confluence of the importance of education achievement and effectiveness, the reality of the digital environment which students inhabit more generally, tighter educational budgets, and the profit interests of technology companies create an environment in which schools and departments of education are under pressure to adopt technology for a range of activities.

There is no question that technology-assisted education and the analytical possibilities that are presented by big data resulting in part from such education has great potential to improve student learning, teaching effectiveness, parent engagement, and accountability. At the same time, educational technology and particularly big data raise issues about the privacy and security of student data, the role of traditional educational actors – teachers, school administrators, school boards, state and provincial departments of education, and national departments of education – as well as the role of new educational actors, particularly online and software education technology firms.

Much of the discussion about Big Data in educational journal and newsletters reports on new initiatives conducted by educational firms, the promises of Big Data, and the positive effects on student learning and achievement. For example, Darrel West in a Brookings Report presents several potential benefits of Big Data including insights regarding student performance and approaches to learning, effectiveness of techniques, evaluation of student actions, and predictive and diagnostic assessments. He also notes several barriers complicating the achievement of these benefits including the need for data sharing networks, similar data formats, and balancing vital student privacy and confidentiality with access to data for research purposes but cautions that "Using privacy arguments to stop research that helps students is counter-productive."[1]

This paper will first examine the ethical policy concerns that have arisen in discussions about the use of Big Data in education and then examines how these issues were defined and discussed in the controversy surrounding InBloom, one of the first companies to establish a large footprint in this arena. Next, we identify the various stakeholders or actors in this environment including government institutions, technology companies, non-profits, and unions, and explore the language and positions they are adopting in discussing ethical issues. Finally we conclude with an analysis of the current status of the ethical discussions about Big Data in education.

Ethical Policy Concerns about use of Big Data in Education

Much of the discussion about Big Data in education has been framed in terms of "privacy." This is not particularly surprising both because privacy is viewed as a multi-faceted concept with several different components and because discussions about ethics and information technology in other sectors and over time have often been categorized under the value of privacy. We can identify at least six concerns traditionally associated with privacy that are challenged by big data generally and in the context of education.

The *first* is that collection of information about an individual should take place with the knowledge of the individual and that the amount of information should be minimized to that which is required for the particular purpose for which it was collected. This is the classic information privacy concern that from a policy perspective has been addressed by the Fair Information Practice Principles (FIPPs) often summarized by notice, consent, choice and transparency. These principles are the basis of much privacy and data protection legislation around the world including in the United States in the Family Educational Rights and Privacy Act (FERPA) and the Children's Online Privacy Protection Act (COPPA) and in Canada in the Privacy Act, governing the public sector, and the Personal Information Protection and Electronic Documents Act (PIPEDA).

Although many have questioned the effectiveness of the FIPPs approach more generally, there is almost universal agreement among privacy scholars and experts that the FIPPs approach is appropriate in the big data environment. With big data there is more collection of information, by more parties, about more aspects of an individual's life, and with more granularity about that life. But the issue is not merely "more" or even the qualitative changes that quantity does not convey. The

[1] Darrell M. West, "Big Data for Education: Data mining, Data Analytics, and Web Dashboards," *Governance Studies at Brookings* (September 2012). http://www.brookings.edu/~/media/research/files/papers/2012/9/04-education-technology-west/04-education-technology-west.pdf

issue is also how much of big data collection takes place without the individual's awareness. As the President's Council of Advisors on Science and Technolosy (PCAST) noted in 2014 individuals "constantly *emit* into the environment information whose use or misuse may be a source of privacy concerns."[2]

Moreover, enhancements in digital storage capacity combined with improvements in computational power and developments of more sophisticated algorithms for analyzing data have enabled organizations to probe and dissect datasets in ways unimagined even twenty years ago. As Rubinstein similarly points out big data make possible the extraction of new, potentially useful information from data – this "newly discovered information is not only unintuitive and unpredictable, but also results from a fairly opaque process."[3] The entire enterprise of big data challenges all previous ideas about how to limit data collection about individuals and how to involve the individual in the process of data collection and subsequent uses so that the individual could exercise some meaningful control.

With respect to education and big data, this issue of notice, consent and transparency becomes even more complicated than it does in other contexts both because records of children and hence the concerns of parents come into play and also because the educational relationship is mandatory, not voluntary. Educational technology firms usually do not have a direct contractual relationship with the students and parents but with the schools, school boards or teachers. Thus providing information and controls about the uses of big data are at least one step removed from the data subject.

A *second* concern long associated with privacy is that individuals should be able to remain anonymous or obscure if they so choose to do so. But with an ever-increasing number of social relationships and practices becoming data points, it becomes more difficult for individuals to remain unidentified or unfindable. Algorithmic searches of datasets now can rather quickly eradicate what had been high transaction costs on finding meaningful information.[4] Most privacy and data protection laws cover "personal information" or "personally identifiable information" meaning that the information was directly associated with a particular individual. With big data, such distinctions are obviated as more and more bits of

[2] President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective* (May 2014), p. 38. Available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf, x

[3] Ira S. Rubinstein, "Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law* (2013) 3(2): 74-87. Available at: http://idpl.oxfordjournals.org/content/early/2013/01/24/idpl.ips036.full.pdf+html (pp.1-14), p.3

[4] Woodrow Hartzog and Evan Selinger, "Big Data in Small Hands," *Stanford Law Review Online* (Sept. 3, 2013) 66:81-88 and Woodrow Hartzog and Evan Selinger, "Obscurity: A Better Way to Think about Your Data than Privacy," *Atlantic* (Jan. 17, 2013). Available at: http://www.theatlantic.com/technology/archive/2013/01/obscurity-a-better-way-to-think-about-your-data-than-privacy/267283/

unidentified information can in effect be attached to a particular individual with just a bit of searching and analysis.

With big data, anonymization of information about individuals becomes more difficult, if not impossible, as big data makes reidentifying data rather easy. The debate about anonymity and reidentification began in 2000 with Latanya Sweeney's study of the 1990 US Census data in which she found that one's 5-digit ZIP code, date of birth, and gender provided unique identification for 87 percent of the population or 216 of 248 million people.[5] In effect few characteristics are actually needed to uniquely identify an individual, making it very difficult to anonymize databases by removing some characteristics, because the bundle of characteristics remaining will likely prove sufficient to identify individuals once a database is merged with other databases and searched using sophisticated algorithms. More recently, Sweeney and colleagues identified the names of volunteer participants in the de-identified public, Personal Genome Project by linking the Project's profiles to public records and data mining the results.[6]

Educational data are often stored in large, longitudinal data sets from which personally identifiable variables have been removed. These data sets are used for reporting purposes from the school to district to state or province and finally to the federal government. They are also used for research purposes to identify trends over time and to analyze factors that affect student performance. They have traditionally been referred to as aggregate, anonymized data – but this tradition is being challenged in the era of big data.

Computer scientists and privacy policy experts and advocates continue to press for better techniques for anonymizing data, for example by using only 3 digits of one's ZIP code or redacting year of birth or day of month. However, as databases become larger and more integrated these attempts increasingly prove to be ineffective. After reviewing the computer science and legal literatures on anonymity and reidentification, Paul Ohm concludes that: "Data can be either useful or perfectly anonymous but never both."[7] This is a conclusion that is becoming more widely shared as various big data projects by companies such as Netflix, AOL and Google reveal that individuals can indeed be identified in studies that were using supposedly anonymous data. And there is increasing recognition that data can either be useful or protective of privacy, but not both. As a biomedical researcher

[5] Latanya Sweeney, *Uniqueness of Simple Demographics in the US Population* (Laboratory for International Data Privacy, Working Paper LIDAP-WP4, 2000). Available at: http://dataprivacylab.org/projects/identifiability/index.html

[6] Latanya Sweeney, Akua Abu, and Julia Winn, "Identifying Paticipants in the Personal Genome Project by Name," *Harvard University Data privacy Lab,* White Paper 1021-1 (April 24, 2013). Available at: http://dataprivacylab.org/projects/pgp/1021-1.pdf

[7] Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," 57 *UCLA Law Review* 1701-1777, 1704 (2010).

notes: "I can't anonymize your genome without wiping out the information that I need to analyze."[8] Much the same holds true in the educational context.

A *third* concern involves the surveillance or tracking that provides more and more detailed information for big data analytics – and that big data require to be even more powerful. A key element of this surveillance is what is now being referred to as the "internet of things," where all our smart devices pick up and transmit detailed information. Big data not only entails more monitoring of activities and extraction of data about those activities, but also involve analysis of those activities to determine likely future activities. This more sophisticated prediction that is built into many big data analytics transforms surveillance into a more omniscient phenomenon.

In the area of big data and education, online testing and teaching programs monitor how long it takes students to answer a question or read a page – and often also capture key strokes or patterns of reading or responding that might indicate the thought processes of the student. The programs also track where (home, school, computer lab) the student is working and what time of day – and often also record what other students are working on the same programs at that time. The results of all this tracking are cross-matched with more traditional information about the student as well as new information from various devices (such as how much a student moves throughout the day or how much time a student spends on social networking sites) – and all of this is fed into predictive analytics programs to determine student learning patterns, strengths and weaknesses, and advice about how best to personalize the learning environment for that student – and raises a *fourth* ethical concern regarding autonomy.

Big data, especially the analytics powered by big data, challenge individual autonomy, the individual's ability to govern his or her life as that individual thinks best. Big data algorithms jeopardize autonomy by leading people in certain directions – to buy certain items, try certain routes or restaurants – and in a certain way challenge the self as defined throughout much of Western philosophy. Some have expressed this concern as about social fragmentation into "filter bubbles," where individuals are subject to feedback loops that limit individuals' sense of their options.[9] Ian Kerr and Jessica Earle distinguish among three types of predictions that affect autonomy: consequential predictions that allow individuals to act more in their self-interest and avoid unfavorable outcomes; preferential predictions that lead one to act in a way expected from the data; and preemptive predictions that are not based on the preferences of the actor but reduce the range of options available to the actor.[10] Tene and Polonetsky point to the dangers of predictive analysis

[8] John Quackenbush quoted in Jonathan Shaw, "Why 'Big Data' is a Big Deal," *Harvard Magazine* (March/April 2014), 30-35, 74-75, p. 34. Available at: http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal

[9] Eli Pariser, *The Filter Bubble: How the new Personalized Web is Changing What We Read and How We Think* New York: Penguin Books, 2011.

[10] Ian Kerr and Jessica Earle, "Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy, *Stanford Law Review Online* (Sept. 3, 2013) 66: 65-72.

including the perpetuation of old prejudices and the accentuation of social stratification.[11]

Autonomy is thus related to a *fifth* privacy concern associated with big data, which involves traditional due process for individuals, the principle that individuals are treated fairly and equally and not discriminated against based on race, gender, age or other personal attributes – or based on factors of which they are not aware. Big data's use of mathematical algorithms and artificial intelligence to make predictions about individuals based on conglomerates of their information and that of others raises questions about treating individuals as individuals fairly, accurately, and in ways they can understand.[12] This concern involves issues of profiling and discrimination.

In the education environment, with its recognition of the importance of education to equal opportunity, there is a longstanding concern for not discriminating and for watching closely for subtle, as well as obvious, signs of discrimination. But with big data such subtle signs may be difficult to discern. For example, Ohm points out that "big data helps companies find a reasonable proxy for race."[13] But perhaps more troubling in education is that big data facilitates the creation of more refined, intersectional categories that discriminate among students in more insidious and harder to read ways. As Jonas Lerman points out: "The big data revolution may create new forms of inequality and subordination, and thus raise broad democracy concerns."[14] At a Data and Civil Rights Conference in 2014, these issues were explicitly addressed in one paper in which the authors pointed out that: "the complexity of algorithmic analysis makes identification of bias and discrimination difficult;" the difficulty of reversing or avoiding "flawed algorithmic assessments;" the danger of self-fulfilling prophecies or prejudging students; and the risk of increasing stratification.[15]

A *sixth* issue that has long been part of the debate about privacy, especially information privacy, is the question of the ownership of data about an individual. Does the individual "own" the information or does the third party holding the information in a database? Although many privacy scholars question whether the

[11] Omar Tene and Jules Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics," *Northwestern Journal of Technology and Intellectual Property* (2013) 11(5): 239-273 ,253. Available at: http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip

[12] Danielle Keats Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washington Law Review* (2014) 89: 101-133.

[13] Paul Ohm, General Principles for Data Use and Analysis," in Julia Lane Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum (eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement.* New York: Cambridge University Press, 2014, pp. 96-111.

[14] Jonas Lerman, "Big Data and Its Exclusions," *Stanford Law Review Online* (Sept. 3, 2013) 66: 55-63, 60.

[15] Andrea Alarcon, Elana Zeide, Alex Rosenblat, Kate Wikelius, danah boyd, Seeta Pena Gangadharan, and Corrine Yu, "Data & Civil Rights: Education Primer," produced for Data & Civil Rights Conerence (October 30, 2014) available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2542268

property model provides a workable framework for talking about privacy,[16] the property rhetoric and rationales have become part of the policy discussion about big data, as they had been in earlier iterations of debates about privacy policy. As one moves further from either submitting personal information to one organization or clicks "I agree" on a website, any ownership in that information arguably fades. And if that information becomes part of a dataset that is then reused or reconfigured or combined with another or sold to another organization, the claim of personal ownership in that information diminishes even more.

In the education arena, student records are generally "owned" by the school or school district. The involvement of ed tech companies has somewhat muddied the question of ownership – depending on how contracts with these firms are written.

At this point in policy discussions about big data in education, the focus is on:

- the security of the data
- deidentification of student data for analytical purposes
- prohibitions on targeted advertising using student data
- ownership of information – trend seems to be that ownership should remain with the local school district
- transparency re online practices

The issue of profiling of students and the potential discriminatory effects has not yet been incorporated directly into these evolving policy discussions.

In order to provide a concrete context for understanding how big data innovations raise ethical concerns, the following section provides an overview of the controversy surrounding InBloom in New York State.

InBloom: Controversy leads to legislation and bankruptcy

In the fall of 2013 twelve parents filed a lawsuit to stop an agreement between the State of New York and InBloom, a nonprofit corporation started by the Council of Chief State School Officers and underwritten by a $100 million grant from the Bill and Melinda Gates Foundation and the Carnegie Corporation of New York. At the time of the lawsuit, InBloom had commitments from nine states to adopt its cloud service, although only New York, Louisiana and Colorado had actually signed contracts and were undertaking pilot efforts to upload data with the non-profit. By October 2013, New York State had already uploaded 90 percent of the data from 2.7

[16] Paul M. Schwartz, "Property, Privacy, and Personal Data," *Harvard Law Review* 117(7):2055-2128 (May 2004) and Julie E. Cohen, "Examined Lives: Informational Privacy and the Subject as Object," *Stanford Law Review* 52: 1373-1438 (2000).

million public and charter school students into the system.[17] Education technology vendors also liked InBloom and were signing on to the service.

InBloom was supposed to be a data aggregator, meant to serve as a repository for the streams of data being generated by multiple education technology sources. InBloom would allow the data gathered from disparate educational software programs and apps to be uploaded into a cloud repository, translated into a common language, and made accessible through a dashboard by teachers, school administrators, school boards, and state departments of education, along with other "third parties". Users could then track individual students' progress through various educational stages, and teachers and others could intervene or "personalize" the learning experiences of individual students as they either struggled with or needed more challenge from the curriculum.[18]

In February 2014, the parents' lawsuit was dismissed, but by that point the New State Legislature had put provisions in the state budget restricting the State Department of Education from undertaking any contracts with third party data aggregators. InBloom closed its doors in April 2014 after school districts in Louisiana and Colorado followed New York State's lead and pulled out of pilots involving the data repository.[19] What had ultimately led to InBloom's demise was a cacophony of voices from many sides concerned about privacy, parental consent and access to the aggregated data. InBloom's software had included some 400 "optional fields" that schools could choose to fill in and that included some fairly sensitive information such as disability status, social security numbers, family relationships, reasons for enrollment changes, and disciplinary actions.

Parents and privacy advocates balked at what they saw as intrusive data gathering that seemed like surveillance. Questions were raised about who could and would access the data, especially data like disciplinary actions, with subjective terms like "'perpetrator,' 'victim,' and 'principal watch list,'" as well as the potential for such data to be used to "stratify or channel children."[20] Parents were particularly incensed that InBloom would not allow any opting out of the data collection. Teachers and other education professionals were concerned about state-level officials having access to student-level data, and about the potential use of

[17] Sanger, Natasha, October 5, 2013, "Deciding Who Sees Students' Data," The New York Times, http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html; accessed February 16, 2016.

[18] Sanger, Natasha, October 5, 2013, "Deciding Who Sees Students' Data," The New York Times, http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html; accessed February 16, 2016.

[19] Sanger, Natasha, April 21, 2014, "InBloom Student Data Repository to Close," The New York Times Bit Blog, http://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close/?_r=0, accessed March 8, 2016.

[20] Sanger, Natasha, October 5, 2013, "Deciding Who Sees Students' Data," The New York Times, http://www.nytimes.com/2013/10/06/business/deciding-who-sees-students-data.html; accessed February 16, 2016.

sometimes dubious measures to assess the effectiveness of teachers in the classroom.

But to really understand the morass that was InBloom, it is necessary to take another step back and describe the controversy surrounding the Common Core State Standards (CCSS) and the US Department of Education's Race to the Top initiative. Launched in 2009, the Common Core State Standards are meant to harmonize "learning expectations" across states for students at different grade levels leading to "college- and career-ready" outcomes.[21] The initiative was organized by the National Governors Association and the Council of Chief State School Officers and the standards were written by working groups of academics, education advocacy groups, and experts from testing companies. Only after teachers' unions protested were K-12 educators added to the working groups. The CCSS guidelines outline the skills students should have mastered at each grade level; they are not a curriculum. So school districts and teachers could craft their own teaching plans, although testing would eventually align across districts and states to harmonize expectations of student achievement. As states implement CCSS, there has been a major push for education technology software solutions to provide data to assist districts with the transition to CCSS expectations and that could track student progress.

While the US Department of Education's $4 billion Race to the Top Initiative did not require states to adopt CCSS, applicants were encouraged to adopt "college- and career-ready standards." The call for proposals also strongly suggested that states adopt technology solutions to ensure that data would drive decision making in terms of curriculum and testing for achievement. Between CCSS and Race to the Top, schools and school districts were increasingly under pressure to use education software that would improve outcomes in the college- and career-ready space, and to improve student achievement on both national and international tests.
By 2013, even as most states had embraced CCSS, grass roots protests were in full swing as parents and teachers protested the educational changes suggested by CCSS, and local officials began to object to the erosion of local control of education. Some advocacy groups also chaffed at what they perceived as federal encroachment on state's rights, despite the founding of CCSS in the Governors' and States' associations. Federal government funding of new assessments for CCSS was largely seen as a takeover of curriculum and instruction by the federal government.

InBloom has insisted its efforts were misunderstood. As a data repository, InBloom officials insist they were not controlling or using data, simply storing it for schools and school districts to have easier access across the large number of data platforms, software, and apps. In other words, they were to be a middleman between software

[21] This section adapted from Catherine Gewertz, The Common Core Explained, Education Week, accessed at http://www.edweek.org/ew/issues/common-core-state-standards/index.html?r=877434580&preview=1 on March 7, 2016.

vendors and school districts, with the districts controlling their own data.[22] InBloom was not alone in the data aggregation space; there are a number of data aggregators who are currently doing exactly what InBloom had promised to do, including Pearson (PowerSchool student information system) and Clever, based in San Francisco. Pearson and Clever both house data on 13 million school children and 15,000 school districts respectively.

However, InBloom got caught in the middle of the national debate about the future of education, and privacy became the issue that would unite the opposition and prove convincing to legislators that a limit had been reached. It didn't help that InBloom fought all efforts to allow parents to opt out of the service, and that the New York State Department of Education refused to listen to public concerns over access to the data. The controversy ballooned into a large scale lack of trust in InBloom and widespread perceptions that InBloom and the State were arrogant and insensitive.[23]

The confounding of CCSS and InBloom, however, did not mean that privacy protests about InBloom were not legitimately about privacy. Critics justifiably pointed out that InBloom and the NY State Department of Education hadn't fully assessed risks and liabilities surrounding both privacy and data security. Parent and teacher groups began to coalesce around the privacy issues and new organizations began to form that objected to privacy violations within the InBloom context. The Parent Coalition for Student Privacy, an advocacy group started by Leonie Haimson, a parent advocate in New York City, and Rachael Strickland of Colorado, was particularly effective at articulating the objections of InBloom adversaries, including the threats to student privacy through the weakening of FERPA, data sharing practices among school districts and states, the development of longitudinal data tracking systems, and the push for continuously quantifying students for personalize learning. Of particular concern was the sharing of data with for-profit data-mining vendors and other third party commercial concerns who might then market products directly to students, or the theft of student data by hackers.[24]

Policy Discussions – Responses to Ed Tech and Big Data

[22] Herold, Benjamin, April 21, 2014, InBloom to Shut Down Amid Growing Data-Privacy Concerns, Education week, http://blogs.edweek.org/edweek/DigitalEducation/2014/04/inbloom_to_shut_down_amid_growing_data_privacy_concerns.html; accessed March 8, 2016.

[23] Bogle, Ariel, April 24, 2014, "What the Failure of InBloom Means for the Student-Data Industry," Slate Future Tense Blog, http://www.slate.com/blogs/future_tense/2014/04/24/what_the_failure_of_inbloom_means_for_the_student_data_industry.html; accessed March 8, 2016.

[24] Kharif, Olga, May 1, 2014, "Privacy Fears over Student Data Tracking Lead to InBloom's Shutdown," Bloomberg Business, http://www.bloomberg.com/bw/articles/2014-05-01/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking; accessed March 8, 2016.

In this section we identify three large actors that have participated in policy discussions thus far and that are likely to continue to be key actors – government and public sector institutions at the federal, state/provincial, and local school level; the education technology companies; and the advocacy community, particularly nonprofits and unions. Our goal here is to identify the themes and discourse that is emerging at this stage of the debate about these ethical issues, to examine how they vary by actors, and how they vary between the US and Canada.

Federal Level

United States: In the United States, four federal statutes affect student data privacy and the student data gathering activities of other entities. First is the Family Educational Rights and Privacy Act of 1974 (FERPA), which requires schools in principle to acquire consent before disclosing student information but allows a number of exceptions including to "organizations conducting certain studies for or on behalf of the school."[25] Additionally, collection and dissemination on students may be subject to the Children's Online Privacy Protection Act of 1998, and amended in 2013, affecting primarily private sector activities and enforced by the Federal Trade Commission. Third the Protection of Pupil Rights Amendment of 1978 governs the administration to students of a survey, analysis, or evaluation that concerns one or more of the following eight protected areas:

- political affiliations or beliefs of the student or the student's parent;
- mental or psychological problems of the student or the student's family;
- sex behavior or attitudes;
- illegal, anti-social, self-incriminating, or demeaning behavior;
- critical appraisals of other individuals with whom respondents have close family relationships;
- legally recognized privileged or analogous relationships, such as those of lawyers, physicians, and ministers;
- religious practices, affiliations, or beliefs of the student or student's parent; or,
- income (other than that required by law to determine eligibility for participation in a program or for receiving financial assistance under such program).[26]

Finally the Education Sciences Reform Act of 2002 strengthens confidentiality requirements for student records especially with respect to the activities of the National Center for Education Statistics (NCES).

The application of these laws to big data in education is still unclear. At the US Department of Education, the initial emphasis in 2012 was on Big Data as a source of information for making teachers more effective: "this is where "big data" comes in. As technology is used to support instruction and assessment, we'll have more and more information available about what works, for what type of learners, under what conditions."[27] The continuing focus is primarily on big data's availability, ease of collection, low cost, and potential for providing more and more granular

[25] http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

[26] http://familypolicy.ed.gov/ppra

[27] Remarks from Joanne Weiss for the 2012 NCES STATS-DC Data Conference, July 11, 2012. http://nces.ed.gov/whatsnew/conferences/Statsdc/2012/STATSDC2012keynote.pdf

information about how and why students are learning, as well as possibilities for predicting student learning and enabling early interventions to improve the learning process.

In 2012, the Department of Education's Office of Educational Technology contracted with SRI International to do a report on "Enhancing Teaching and Learning Through Educational Analytics and Data Mining."[28] The authors identified three challenges – technical, institutional capacity, and privacy and ethics. Privacy and/or confidentiality does not appear to be a particular concern until 2014, following the OSTP and PCAST reports on Big Data and Secretary Duncan's interest in the topic. The Chief Privacy Officer gave an update on privacy, including Big Data, at a National Forum on Education Statistics.[29] She noted concern about marketing student data and the need for "self-policing" by commercial entities, and the OSTP Big Data May 2014 Report's recommendation that "The federal government should ensure that data collected in schools is used for educational purposes and continue to support investment and innovation ...it should explore how to modernize the privacy regulatory framework under FERPA and COPPA ...." She emphasized the trend that schools are contracting out school functions, that schools have more and new types of data, that data are not collected using the traditional 2-party contract model, and that more data are commercialized. But she also noted that most of the activity, and responsibility, is at the state level. The Department of Education's primary role would then be support and training with an emphasis on providing transparency.

In February 2014, the Department of Education's Privacy Technical Assistance Center issued a brief addressing requirements and best practices in protecting privacy while using online educational practices. Its response to the questions of whether or how FERPA and PPRA covered online educational services and protected student records in this environment was basically – "it depends. Because of the diversity and variety of online educational services, there is no universal answer…"[30]

ED has also been offering guidance to states on statistical methods to protect student privacy and confidentiality when aggregate data is released, which may unintendedly disclose personally identifiable information. The possibility of such unintended disclosures increases with Big Data. As can be gleaned from the above, there appear to be two arenas within the Department of Education in which discussions of privacy and Big Data are or might occur – the

[28] Marie Bienkowski, Mingyu Feng, and Barbara Means, "Enhancing Teaching and Learning Through Educational Analytics and Data Mining." Center for Technology and Learning, SRI International (October 2012). https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

[29] Kathleen M. Styles, "Change is in the Air: An Update on Student Privacy" (July 28, 2014) National Forum on Education Statistics. http://nces.ed.gov/forum/pdf/S2014Styles.pdf

[30] Department of Education, Privacy Technical Assistance Center, "Protecting Student Privacy While Using Online Educational Services: Requirements and Best Practices," PTAC-FAQ-3 (February 2014), available at: https://tech.ed.gov/wp-content/uploads/2014/09/Student-Privacy-and-Online-Educational-Services-February-2014.pdf

privacy/legal/regulatory arena and the technology/private vendor arena. In the first arena, the discussion is taking place in the context of the current legal environment – without yet recognizing how dramatically and fundamentally big data undercuts this legal environment. At this point, it would appear that more resources and attention are being devoted to the latter arena.

Canada: In Canada, there is no ministry or department of education at the federal level as the Canadian constitution gives the provincial governments exclusive responsibility for all levels of education. The two federal privacy laws, however, affect student privacy and the activities of educational technology firms – with the Privacy Act regulating public educational institutions and the Personal Information Protection and Electronic Documents Act (PIPEDA) covering private sector entities such as ed tech companies. There is no federal law pertaining particularly to student privacy or addressing student data practices.

State/Provincial Level

United States: In the United States, the issue of privacy and student data emerged as a topic of state lawmaking due to the Snowden revelations and concerns about data surveillance generally, publicity surrounding data breaches at retailers such as Target, and the increased use of educational technology for a number of functions – administrative systems, classroom instruction, homework, student collaborations, and incorporation of social networking.[31] In 2013, Joel Reidenberg directed a study on cloud computing in public schools, which examined found that school districts were not addressing privacy concerns in a uniform or informed manner when they transfer student information to cloud computing service providers.[32] Based on their detailed investigation into a sample of schools, they concluded that "cloud services are poorly understood, non-transparent and weakly governed"[33] and "an overwhelming majority of cloud services do not address parental notice, consent, or access to student information."[34]

In 2014, 110 student data privacy bills were introduced in 36 states with 21 states passing 24 such bills into law. The latter half of 2014 saw a shift in policy discussions from concern with data in state systems to the privacy implications of

[31] Material regarding state laws is derived from two publications from the Data Quality Campaign : "State Student Data Privacy Legislation: What Happened in 2014, and What is Next?" (August 2014) available at: http://dataqualitycampaign.org/wp-content/uploads/2014/09/DQC-Data-Privacy-whats-next-Sept22.pdf
and "State Student Data Privacy Legislation: What Happened in 2015, and What is Next?" (September 2015) available at: http://dataqualitycampaign.org/wp-content/uploads/2015/09/DQC-Student-Data-Laws-2015-Sept23.pdf

[32] Joel Reidenberg, N. Cameron Russell, Jordan Kovnot, Thomas B. Norton, Ryan Cloutier, and Daniela Alvardado, "Privacy and Cloud Computing in Public Schools," Center on Law and Information Policy, Fordham Law School (Dec. 13, 2013) available at: http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1001&context=clip

[33] Ibid, p. 6

[34] Ibid, p. 7.

student data collected, held and analyzed by third party service providers following the controversies and press attention from InBloom's activities in New York and Colorado. California passed the first law explicitly targeting online providers in its Student Online Personal Information Protection Act (SOPIPA). State legislative interest increased in 2015 with 182 student privacy bills introduced in 46 states and 15 states passed 28 new student privacy laws. Bills in 25 states were modeled after SOPIPA and in 31 states bills articulated requirements for service providers. Many states also addressed concerns about the capacity and resource needs of school districts in managing the issues around student privacy, especially with respect to staff training and explicit policies such as those for contracts with service providers.

The Data Quality Campaign identified two overlapping approaches in these state bills. First, the *prohibitive approach*, which restricted or prevented the collection of certain types of data (e.g., biometric) or certain uses of data (e.g., predictive analytics), was adopted in 79 of 110 bills in 2014 and 125 of 182 bills in 2015. Second, the *governance approach*, which established procedures (e.g., audits and inventories), roles and responsibilities to ensure appropriate student data practices, was found in 52 of 110 bills in 2014 and 122 of 182 bills in 2015.[35] States are still sorting out the appropriate roles of state boards of education, school districts, and school boards. In 2014, 32 bills charged state boards of education with student privacy responsibilities and 7 of these became law; 28 bills gave this responsibility to school districts. In 2015, 35 bills charged state boards of education with student privacy roles and 5 of these became law; 63 bills gave this responsibility to school districts and 9 became law; and 23 bills tasked local school boards with the responsibility and 7 of these became law.[36]

Canada:

School Districts and Schools

Big Data Companies

Educational technology is shaped by the changes in the education space itself; everything from financial constraints on schools and school boards, new demands for accountability and outcome measures, innovations in teaching and learning, and new laws and requirements (both state and federal) have driven the market for new technologies and the data generated by them to inform decision making around teaching, learning and policy. The education technology sector is booming, with more than $1.8 billion in venture capital currently being invested in the estimated $8 billion market for education technology software. This has meant a surge in new startups, increasing attention paid by large and established technology companies, and increased competition for domain space. During and in the wake of the InBloom saga, major companies like McGraw Hill Education, Pearson, and even News Corp

[35] Data Quality Campaign, 2015

[36] Data Quality Campaign, 2014 and 2015.

had developed data tracking software for education.[37] Facebook, Google, and Microsoft have also acquired or are developing their own education software subsidiaries. Rather than kill the education software sector, the demise of InBloom simply provided more space for other companies to come in and fill the void.[38]

Nonprofits and Unions

While the education technology sector explodes, the student privacy advocacy space has also welcomed new actors, both for and against the use of data in education. Those advocating for the development of education technologies tend to focus on the benefits of using technology in offering teachers, schools and education policy makers the kind of evidence that would lead to more success in the classroom. These groups, including the Data Quality Campaign (DQC), the Future of Privacy Forum (FPF), the Consortium for School Networking (CoSN), the Student Privacy Resource Center (FERPASherpa), and the Software and Information Industry Association (SIIA), among others, are largely funded by large technology corporations and their foundation arms. Their websites and informational brochures tend to focus on the benefits of using technology and data in the classroom, along with information about the various privacy laws and current student privacy protections.

Among the activities of these organizations are the creation of "pledges" and "certifications" that educational technology companies and education leaders could sign on to by promising to adopt prescribed privacy practices. The Student Privacy Pledge, for example, was developed by FPF and SIAA as a way for educational technology companies to pledge to more open communication about their products and privacy safeguards and to encourage the adoption of practices that "meet or go beyond" federal regulations. The website claims 243 current signatories.[39] CoSN is also developing a "Trusted Learning Environment Seal" targeting "school system leaders" who have undergone the organization's certification programs to become "certified education technology leaders."[40] Finally, DQC also targets school leaders with information about communicating about the benefits of using data on student achievement, and on applicable privacy laws and protections through online training modules and awards for state and local officials who "have embraced a culture of data in service of students."[41]

[37] Ibid.

[38] Bogle, Ariel, April 24, 2014, "What the Failure of InBloom Means for the Student-Data Industry," Slate Future Tense Blog, http://www.slate.com/blogs/future_tense/2014/04/24/what_the_failure_of_inbloom_means_for_the_student_data_industry.html; accessed March 8, 2016.

[39] More information about the Student Privacy Pledge may be found at https://studentprivacypledge.org/ accessed March 8, 2016.

[40] More information about the Trusted Learning Environment Seal may be found at http://www.cosn.org/about/news/national-education-organizations-launch-effort-build-%E2%80%98trusted-learning-environment%E2%80%99-us-1 accessed March 8, 2016.

[41] More information about the Data Quality Campaign and their Flashlight Awards may be found at http://dataqualitycampaign.org/success-stories/data-flashlight-awards/ accessed March 8, 2016.

Some question the efficacy of privacy pledges and certification, even as the President and others have embraced the movement. Natasha Sanger, reporting in the New York Times Bit Blog in February 2015, noted that a Student Privacy Pledge signature does not guarantee that companies have adopted the best encryption practices to protect student data on unsecured networks. Additionally, the education technology companies that sign the pledge, while promising to protect student data, do not commit to protecting teacher and/or parent data collected.[42] Others have raised issues of data privacy equity as well. While well-funded school districts might be able to afford well-designed education software and apps with top-of-the-line privacy and security protections, poorer school districts may find they rely more on free software from non-profits or fledgling startups that might not be able to afford the best data encryption measures, regardless of whether they have signed a pledge to do so.[43]

[42] Sanger, Natasha, February 11, 2015, "Data Security Gaps in an Industry Student Privacy Pledge," New York Times Bit Blog available at http://bits.blog.ntimes.come/2015/02/11/data-security-gaps-in-an-industry-student-privacy-pledge/ accessed February 16, 2016.

[43] Sanger outlines instances of poor data encryption, and issues of equity are brought up in "From Mining to Minding Student Data," EdSurge, accessed March 8, 2016 at https://www.edsurge.com/research/special-reports/state-of-edtech-2016/k12_edtech_trends/data_privacy