Privacy In the Age of Analytics
Report of a CNI Executive Roundtable
Held April 13, 2015
Published August 2016

Background

Universities have always collected large amounts of information about their students, but in recent decades, with the rise of technology mediated teaching and learning, research, and even reading, and the slow development of the "smart campus," the amount of data being gathered, often of a highly personal and personalized nature, has become overwhelming. The move to digital systems has also allowed this data to become more visible, more centralized, and more persistent. Technological mediation has also brought a range of third parties into data collection and analysis processes. Often this data is being collected and retained without the knowledge, much less the informed consent, of the individuals whose activities it describes; not so much deliberately in secret but with little transparency or visibility.

At the heart of the idea of analytics is the application of various statistical and machine learning algorithms to detect patterns and correlations in these masses of data, and to use these patterns and correlations to predict future developments. In many cases organizations are then taking the next steps of trying to design interventions to alter these predictive outcomes before they occur. There are, of course, many problems in the accuracy of the predictions from analytics, and a tendency to confuse correlation with causation. The embrace of analytics in various areas of the educational process, frequently with the best of intentions, makes this amassing and retention of data much more attractive and much more dangerous because of its potential for reuse.

As a historically separate matter, libraries have had a long-standing need for various kinds of statistical and analytic data to guide collection, programmatic, and even facilities development. As demands for data to support decision-making have increased and opportunities to collect more data have become richer with the shift to digital resources, there are growing tensions with historical commitments to protect reader privacy (primarily through anonymity and non-collection of data). There are also disconcerting suggestions that these data, to the extent that they are collected, might become inputs to broader campus-wide learning analytics efforts; library views on privacy and data collection are often out of step with other parts of the institution. Note that library privacy is governed by specific state laws that may not apply to other university activities, and that there is also a very long-standing history of professional codes from organizations like the American Library Association (see especially their *Library Bill of Rights* and its position on reader privacy at ala.org/advocacy/intfreedom/librarybill/interpretations/privacy).

So we face multiple problems. The first is maintaining appropriate operational security and confidentiality controls on the masses of data now being collected and stored both by universities and by third parties. The second covers policy issues around what data is being collected, and how much of that is being collected without the knowledge or consent of the individuals and/or on a compulsory basis, as well as ongoing opaque policy and governance of this data. The third is reaching agreement about the appropriate design and use of analytics within academic contexts.

Note that our focus at this Roundtable was intended to be learning or student analytics, reading and other library related analytics, and the intersections between them; we did not explicitly consider the important emerging area of research analytics and metrics, though it did come up repeatedly in some of the institutional perspectives. It's also clear that there's a growing gap between the specific area of learning analytics and the much broader topic of student analytics, which include a lot of behavioral aspects. Obviously there is a connection in that behaviors influence academic outcomes, but how to best conceptualize and manage this broader area is a major source of tension at many institutions.

A discussion on this complex, dynamic and evolving landscape took place at the Coalition for Networked Information (CNI) Executive Roundtable in Seattle, Washington on the morning of April 13, 2015. Representatives from 17 higher education institutions, non-profit organizations, and one commercial publisher described their policies, strategies, concerns, and future plans.

Institutional Perspectives

Some key perspectives from institutional participants included these observations:

- Some university administrations are pressuring their departments and units to improve retention and time to degree, and this is prompting more attention to the development of analytics tools involving lots of data about students; quick results are taking precedence over other considerations and niceties. One individual commented that higher education is in a reactive mode in this arena.
- Universities employing online learning platforms at a large scale are (sometimes) addressing a variety of data privacy issues, and in some cases, the campus conversation surrounding this issue has created a lot of tension; one important dimension is *who* (faculty, staff, counselors, advisers) should be able to see this data, and to what extent, e.g. for a course, or for a student across courses. The move of textbooks to digital forms also creates a collection of issues, with the added complication of potential publisher interests and involvement in the capture and analysis of user data.
- Institutions are finding that they are gathering more information by card swipes (and therefore personally identifiable) than they had imagined. Logs related to mobile devices (both cellular and wireless) are another Pandora's box. Personal health trackers were mentioned as yet another emerging issue.
- It is not clear how much students care (or know) about the collection of personal data and whether that should matter in terms of developing policies on

notification, opting out, etc. One individual commented that students seem to care about privacy of personal medical data but not about privacy of personal academic data; other behavioral data might constitute a more ambiguous middle ground. Others believe that students care more about their privacy than the popular media would lead us to think. It's also important here to differentiate views on collection of data from policies surrounding subsequent retention and reuse.

- Universities have concerns about what (student) data they have that is subject to subpoena or other demands from law enforcement and realize that they may have unintentionally stored data without having made an intentional decision about retention. Risk and liability lurk everywhere here.
- One campus noted a shift to centralizing information at the university level; as an example, data on student performance in courses had been the purview of individual faculty and is now being tracked on an institutional basis.
- One university has standard contract language to use with external providers to address data privacy issues and works to influence faculty not to accept licenses as individuals that might give away control of data.
- The question about access to student information by members of athletics departments, or indeed by compliance groups dealing with college athletics, has come up recently.
- Faculty or researcher profiling and output tracking systems raise many privacy issues and are raising concerns at several institutions. One institution discussed the use of a faculty research profiling system (VIVO) and the concerns they have about including graduate student profiles without having specific opt-in mechanisms.
- Data quality deserves greater focus; quality data, and data repurposed or recombined in accurate ways, are essential prerequisites to meaningful analytics. Poor quality data will lead likely to bad decision making, as will data use that is inconsistent with the parameters under which the data was gathered. To the extent that inaccurate data leads to bad outcomes, it is increasingly a potential source of serious liability.
- Data privacy is increasingly a topic of discussion within some units of campus, such as the library, and in some campus-wide, cross-unit venues.
- Campus police, other law enforcement, Student Life, and other groups are important stakeholders in the discussion, as issues arise about life safety, suicide prevention and similar matters.
- Within the library community, staff see the need for better education about privacy trends and wish to raise the profile of this issue within libraries and their communities. While protecting user privacy, libraries as a community would benefit from addressing what might be a baseline of data that could be collected and used as shared analytics.
- Libraries are debating the extent to which, or even whether, they can leverage usage data and share it publicly to assist with the improvement of library discovery systems.
- Some participants expressed concerns about web-based library services, many of which are passing non-encrypted data about users all the time.

- Libraries collecting data using Google Analytics are realizing they may be violating the ALA *Library Bill of Rights* and possibly state laws and are starting to look at this issue; this is but one example of how easily convenient web-based service offerings can come with unexpected consequences. Libraries in consortia that bring together institutions in a region that encompasses several states need to help their members reconcile the variety of state laws addressing privacy.
- Some libraries may have relatively restrictive data privacy policies in place that are in conflict with (more restrictive than) campus policies.
- It should be recognized (and we were briefly reminded, though we did not focus on this issue) that publishers and content platform providers also face a complex set of issues about reader privacy, with conflicting pressures from authors, readers, and libraries about data collection, analysis, reuse and sharing. Note that many universities are themselves publishers in one form or another.

A representative from Jisc in the UK described some of their excellent, broadly based, cross-institutional work on student privacy issues. They are working with the National Union of Students and developing approaches to make things related to privacy as transparent as possible. They are developing a consent platform that will communicate policy changes and also give students a degree of control over how their data would be used. The Jisc representative reported on the development of a code of practice for learning analytics (now available at jisc.ac.uk/guides/code-of-practice-for-learning-analytics along with a wealth of other related material) that addresses responsibility, transparency and consent, privacy, validity, access, enabling positive interventions, minimizing adverse impacts, and stewardship of data.

When the Roundtable participants were asked about whether their institutions had a clear policy on access to and retention of data related to specific classes, most did not. They noted that most administrators or researchers interested in student analytics were more interested in aggregating student data from a variety of systems and that the learning management system was only one source of data. Others noted that on their campuses, a variety of learning management systems were in use and that some faculty employed third party solutions on their own, often with little awareness of privacy issues. One participant noted that some consortial arrangements for learning management systems are promoting their potential use for collective analytics but have not addressed how they will do this in the face of multiple campus policies.

Concluding Thoughts

If the attendees at this Roundtable reflected general trends in the community, then privacy related to student analytics is clearly an area in flux in most higher education institutions. Participants realize that much is at stake and that they should be addressing these issues in a more systematic way within their institutions. There are questions about whether analytics are truly as useful as the hype would indicate, about what policies should address and who should formulate them, and about what parameters concerning student privacy should be acceptable.

One participant expressed concern about the increasing use of analytics to understand student learning, questioning whether the things that can be measured are the best

ways to conceptualize and describe what students have actually learned at university. Another expressed concerns that students who might not score well in a foundational course may do very well in future courses because of the rigorous education they had in the prior course but this is not addressed by current analytics practice. Analytics are also not generally used to identify students with exceptional abilities and the current systems in use are rarely set up to make suggestions for what might be done to enrich the educational experience of talented students.

It is worth noting that almost all discussion focused on relatively high-level analytics, rather than the micro data that's collected in adaptive learning systems or from interactive platforms (including massive open online courses, or MOOCs).

What is mature data governance policy and practice for an institution? In universities and colleges, how will students be involved in the creation of the policies? One participant remarked that they have had difficulty engaging students in these discussions. Social media policy was another area touched on by participants. Is the institution keeping the chats and other interactions of students when they occur within a learning management system environment, for example, or in intermediate spaces like class-related Google Hangout sessions? To what extent, if at all, is it considering student interactions on public social media platforms (Facebook, Twitter, etc.) as potential inputs to analytics (or other activities)?

With the collection and retention of data comes institutional responsibility to protect it effectively. In policy development, institutions need to keep in mind how they would feel about the data they are retaining being breached and exposed publically, given that there have been regular reports of data breaches on campuses in the past few years. It may be that the risks exceed the value of collecting or retaining the information; if nothing else, it underscores the importance of good data inventories and retention schedules, and the destruction of data that isn't deliberately being retained.

At present, some institutions have what one participant described as a patchwork of policies, and others have virtually no institutional policies addressing privacy, other than those mandated by state and federal government. One participant noted that institutions need people who have the background and expertise to lead informed conversations about privacy, and several expressed the importance of investing in educating staff and faculty about these matters.

When the group was asked to identify the locus of responsibility for data governance in their institution, there was no common answer. Replies included a data stewards council (or similar group), the provost, the CIO, IT governance group, the Office of Compliance, the Vice President for Student Life, the Cabinet, and a security/privacy officer reporting directly to the president. The role and relevance of institutional review boards (IRBs) came up several times; there is anything but clear consensus on this.

More and more data is being held externally by vendors and other partners; institutions are struggling to find the right contractual provisions and business arrangements to deal with this, and to achieve any reasonable degree of consistency across the institution. Higher education institutions need to be aware that when corporate entities are holding some of their confidential data, they may face unexpected difficult

situations even when they believe agreements are in place. As one cautionary tale: a company recently in the news had a strong customer privacy policy in place, but when it entered into bankruptcy, this amassed data was viewed as a corporate asset that could be sold off to meet obligations.

One participant encouraged the colleagues present to continue to push hard on privacy issues as a community since we trust each other and have common interests.

We left the roundtable with a very strong sense that both policy development and implementation, and the effective communication of personal data collection and reuse practices to institutional communities, are running far behind actual implementations of systems that use this data at many institutions. This situation suggests that, in an arena like student privacy, institutions are at risk of sudden and embarrassing public controversies that force them into a reactive mode of retrospectively justifying practices and making policy quickly, and without sufficient deliberation and consensus.

This is an area that continues to be important to the CNI agenda, and we will be tracking developments, which are happening quickly, through posts about important documents, events and trends on the CNI-ANNOUNCE mailing list and through sessions at our membership meetings and in other venues.

CNI Executive Roundtables, held at CNI's semi-annual membership meetings, bring together a group of campus partners, usually senior library and information technology leaders, to discuss a key digital information topic and its strategic implications. The roundtables build on the theme of collaboration that is at the foundation of the Coalition; they serve as a forum for frank, unattributed intra and inter-institutional dialogue on digital information issues and their organizational and strategic implications. In addition, CNI uses roundtable discussions to inform our ongoing program planning process.

The Coalition for Networked Information (CNI) is a joint program of the Association of Research Libraries (ARL) and EDUCAUSE that promotes the use of information technology to advance scholarship and education. Some 230 institutions representing higher education, publishing, information technology, scholarly and professional organizations, foundations, and libraries and library organizations, make up CNI's members. Learn more at cni.org.