University Libraries Faculty and Staff Publications   University Libraries

2019

Practices, policies, and problems in the management of learning data: A survey of libraries' use of digital learning objects and the data they create

Graham Sherriff
University of Vermont, gsherrif@uvm.edu

Daisy Benson
University of Vermont, dsbenson@uvm.edu

Gary Atwood
University of Vermont, gatwood@uvm.edu

Follow this and additional works at: https://scholarworks.uvm.edu/libfacpub

Part of the Information Literacy Commons

Title

Practices, policies, and problems in the management of learning data: A survey of libraries' use of digital learning objects and the data they create

Abstract

This study analyzed libraries' management of the data generated by library digital learning objects (DLO's) such as forms, surveys, quizzes, and tutorials. A substantial proportion of respondents reported having a policy relevant to learning data, typically a campus-level policy, but most did not. Other problems included a lack of access to library learning data, concerns about student privacy, inadequate granularity or standardization, and a lack of knowledge about colleagues' practices. We propose more dialogue on learning data within libraries, between libraries and administrators, and across the library profession.

Keywords

library instruction; online learning; instructional technology; digital learning objects; student data; learning data; learning analytics

Introduction

The emergence of the web as a key platform in higher education has made possible innovations in technologies for delivering digital instruction and developments in the pedagogy of hybrid and distance education. One consequence of digital instruction has been the generation of data that describe online learning. Library instruction is no exception, with many libraries using digital

learning objects (DLO's) such as quizzes, tutorials, surveys and forms to facilitate, quantify and verify students' learning.

We are instructional librarians who make extensive use of these tools to provide instruction to several thousand students every year. These quizzes generate a significant volume of data, including personally identifying information (PII). For example, we manage a suite of Guide on the Side tutorials taken by approximately 3,000 first-year students each year. This volume of learning data is useful for assessment purposes, but it also poses important questions about the management of those data with potential implications for student privacy and data security. Following several years of using these tutorials and gathering data, we resolved that the collection, aggregation, retention and reporting of learning data should be guided by professional principles or best practices regarding the management of students' data.

Embracing this resolution, we wanted to develop robust in-house guidelines for the management of library learning data. We established that multiple governing or guiding documents were available to support such a project. These included campus-level guidelines for protecting student privacy (University of Vermont, 2012) and for compliance with the Family Educational Rights and Privacy Act (FERPA) (University of Vermont, 2016); our library's policy on user confidentiality (University of Vermont Libraries); state law (Vt. Stat. Ann. tit. 22, § 171-173); and profession-level guidelines such as the American Library Association (ALA)'s *Code of Ethics* (2008) and *Privacy Tool Kit* (2007).

However, we did not have any model of how to translate the principles, resolutions, and laws contained in these documents into a set of departmental procedures and best practices. We therefore designed a survey of librarians' data management practices and the policies that other libraries might be following. We hoped to learn how other libraries were approaching this matter, gather examples of policies or other governing documents, and hear from other libraries about the challenges they had encountered in applying these policies.

Our research questions:

- What types of learning data are libraries collecting?
- How are libraries managing the aggregation, storage, access, retention and reporting of learning data?
- What policies or best practices for the management of learning data are libraries following?
- What problems or uncertainties are libraries experiencing, related to their management of learning data?

"Learning data" is defined here as data representing students' engagement with DLO's: responses to questions, quiz scores, other submitted information (eg. name, email address, student ID, etc), or other data (eg. IP address, time to completion, etc).

Literature Review

Much has been written in recent years on learning data at the institutional and programmatic levels. This growth reflects the emergence of new technologies, making possible unprecedented opportunities and innovative techniques for analyzing data that may be related to students' academic progress and, in particular, the emergence of the field of learning analytics.

Learning analytics encompass the "measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs" (LAK '11, 2011). Learning analytics systems have the capacity to aggregate data from course activities and assignments, as well as multiple other data types that may have a relationship with learning performance. These may include "enrollment management, relationship management, business intelligence/reporting, learning management system activity/achievement monitoring, integrated planning and advising, early-alert warning, and degree mapping" (Oakleaf et al, 2017). Any correlations detected between different data types may then provide an empirical basis for 1) isolating and describing behaviors, characteristics or events that appear to drive students' academic success, difficulty or failure; 2) predicting which individual students are at risk of academic failure; and 3) identifying interventions by faculty or support services that may improve a student's risk profile (Oakleaf et al, 2017).

Integrating library-generated learning data into campus-level learning analytics systems has potential to enhance library instruction and its contribution to student success. "To move to the next logical step of library correlation research, librarians must prepare to engage and exploit the

'learning analytics' efforts that are proliferating throughout higher education institutions"
(Oakleaf 2018). Many libraries are now participating in library analytics initiatives - possibly as
many as 81% of Association of Research Libraries (ARL) institutions (Perry et al, 2018). In
reported case studies, librarians at the University of Huddersfield, the University of Wollongong,
the University of Minnesota, and Nottingham Trent University have analyzed students'
engagements with library services and resources, and reported correlations with key metrics such
as second-year retention and graduating GPA (Stone, 2015; Cox & Jantti, 2015; Nackerud et al,
2015; Sclater et al, 2016).

The collection and analysis of student data have clear ethical implications. For example, it is
possible to reason that higher education institutions and libraries should feel an obligation to
support the academic success of students who might otherwise fail or perform below their
abilities (Chowcat et al, 2015; Rubel & Jones, 2017; Oakleaf et al, 2017). There is a need for
more research into students' perspectives on learning analytics and data privacy, but recent
studies suggest that many students have a positive view of institutions' use of personal data "that
directly serve their matriculation interests", especially applications that support course
completion, effective learning, degree completion, or positive academic experiences (Dahlstrom
et al, 2015; also Arnold & Sclater, 2017).

At the same time, the collection of learning data raises concerns about data governance and data
privacy. The ability of library-generated learning data to provide reliable insights into student
performance is questionable. Any correlations with metrics of student success might not signify
causation, and it might not be feasible to isolate the effect of library-based instruction on big-

picture outcomes such as retention and graduation from the effects of other variables (Oakleaf, 2015). Also, the collection of library learning data could have adverse effects on students' academic performance. Students appear to become more reticent and cautious in an online learning environment when they are aware of being monitored by personnel other than the course instructor (Dawson, 2002). Other harmful effects that have been hypothesized included negative effects on students' autonomy (Rubel & Jones, 2016), their ability to learn from failure, and their motivation as a result of signaling to them that they are expected to fail (Arnold & Sclater, 2015).

There are also ethical concerns about the participation of libraries in campus-wide learning analytics systems - especially systems that also integrate data types that have traditionally been regarded as private. The ALA *Code of Ethics* asserts "each library user's right to privacy and confidentiality" (American Library Association, 2008) and ignoring or infringing that right may create a sense of being under surveillance (Knox, 2010) and result in students' "coded suspicion" of librarians (Jones & Salo, 2018). The *Code of Ethics* is also in tension with institutions that share or sell student data (Jones & Salo, 2018). The Family Educational Rights and Privacy Act (FERPA) prohibits educational institutions from releasing educational records without student's written consent (U.S. Department of Education, 2018), but higher education institutions are known to make use of statutory exceptions such as the "School Official Exception", which allows disclosure to third parties if they deem there to be a legitimate educational interest (Parks, 2017).

These opportunities, practical questions and ethical concerns underline the importance of purposeful data management practices for library data. In Perry et al's study of library practices and policies applied to all types of library data (2018), a majority of respondents reported that they were not collecting PII, while concerns for student privacy and confidentiality were a key factor for many respondents in their decisions about whether to share data outside the library. Many reported specific data protection measures, but some were not widespread - for example, fewer than half reported having a schedule or policy for retaining/deleting data - and only 25% reported having library-level guidelines or documentation for handling or protecting data.

Several efforts have been undertaken to develop, codify, and implement codes of practice specific to learning data (for example Sclater & Bailey, 2015, Williamson & Phillips, 2017). For the library profession, the ALA's *Code of Ethics* articulates core values with respect to user rights, notably (but not limited to) item 4: "We protect each library user's right to privacy and confidentiality…" (2008). Subsequent ALA documents on intellectual freedom have reaffirmed the values stated in the *Code of Ethics* and collectively may be considered a set of best practices for the profession (Magi, 2010). These include having a library policy on user confidentiality, the avoidance of unnecessary data collection, and disposal of PII unless it is need for a specific purpose.

The ALA has also published documents supporting operational practices with immediate relevance to learning data, such as "Library Privacy Guidelines for Data Exchange Between Networked Devices and Services" (2016), "Questions and Answers on Privacy and Confidentiality" (2014), and "Library Privacy Guidelines for Students in K-12 Schools" (2016).

In contrast with many of the ALA's publications on intellectual freedom that are well-known and widely used, such as the *Code of Ethics* and the *Privacy Tool Kit* (2007), these newer documents with more immediate relevance to learning data appear to have received less attention across the profession. Our review of the literature found no reference to them.

There has been negligible research into the learning data generated by library DLO's. For example, it was not one of the 17 library data types presented in Perry et al's broadly-scoped survey (2018). This shortage of research comes despite the fact that management of these data is a key aspect of using and administering DLO's, and different formats and platforms may have different technical features and different levels of confidentiality and security (Sherriff, 2017). Little has been published regarding the collection, storage, retention, access, or dissemination of DLO data, or on the design of policies that might guide those practices. Rempel & Slebodnik (2015) contend that data-gathering is an integral element of evaluation and assessment, though they caution that the integration of DLO data into program-level evaluation can be problematic "because the library is part of a much larger environment and is very rarely the sole factor influencing these big-picture outcomes [such as retention, graduation and skills for lifelong learning]". A 2011 survey found that librarians' criteria for evaluating DLO's comprised nine major elements, none of which were directly related to data collection and management (Mestre et al, 2011).

Methods

The scope of this study was intended to be the learning data generated by the question and quizzing features of "third-party" DLO's and other digital technologies. Examples include

interactive tutorials and quizzes created in platforms such as Google Forms, Captivate, Guide on the Side, LibWizard Tutorials or Qualtrics. They might also include other data-gathering technologies, like clickers and online polling.

Our survey was designed to exclude tutorials and quizzes created with the native features of learning management systems (LMS's), based on the assumptions that most LMS's have a robust level of data security; are governed by established practices and policies; and are administered by an institutional unit other than the library. However, several respondents provided insights into their management of learning data within their institution's LMS because that is where their core instruction takes place. For example, some respondents described how they were embedding third-party DLO's into their LMS, which made it difficult to make a clear distinction between the two.

We distributed the survey in March 2018 following approval by the University of Vermont's Institutional Review Board (IRB). We posted invitations to participate to multiple librarianship listservs. These included several listservs supporting generalist interests and several oriented towards instruction, assessment, and technology. We also posted invitations to participate in online forums maintained by the developers of some DLO platforms. We sent direct invitations to librarians whom we knew to have a specialized interest in online instruction, assessment or technology.

The survey comprised 13 multiple-choice and short open-response questions. Each question was optional. This allowed respondents to decline to answer questions about their institution and

position title that - in combination - had the potential to be personally identifying. It also allowed respondents to decline to answer questions on sensitive issues, such as evaluations of colleagues' compliance with institutional policies or legal statutes.

Findings

Participation

A total of 169 respondents participated in the survey. This aligned with our expectations for the likely size of the response set. In order to provide informed responses, respondents needed to have specific knowledge about the use of DLO's across their library and about institutional policies. Another limiting factor was the invitation to participate, which encouraged librarians to forward it to the person at their institution who would be best able to provide these responses.

Findings are reported here only from the 71 full responses that were submitted. Partial submissions (n=98) are omitted. Most partial submissions contained few or zero question responses. On account of the relatively low number of full submissions, we are reporting response rates as percentages rounded to the nearest integer.

The survey asked participants to identify their role in the administration of DLO's and the management of any learning data generated by them. The most common role stated was designing and developing DLO's (58%), whether for an instruction program or for the respondent's individual use. Every respondent identified themselves as having responsibilities in one or more of these areas:

- design and development of DLO's

- instruction

- online learning

- administration of a DLO platform

- library assessment

- leadership of an instruction program

- library leadership

Two respondents self-identified as having responsibility for the management of learning data and the implementation of workflows, policies, and best practices. The identification of roles met our expectations for likely participation and confirmed that all respondents had responsibilities relevant to the study.

What types of learning data are libraries collecting?

There is a wide variety of third-party programs and platforms that are commonly used by libraries that generate learning data. The most common, as reported in this survey, were Google Forms (53%), Qualtrics (32%), and Springshare's LibWizard Tutorials (32%). There was also a long tail of other programs reported by a much smaller percentage of respondents. These include polling platforms, such as PollEverywhere and Mentimeter; survey platforms such as SurveyMonkey, Typeform and REDCap; quizzing platforms, such as Kahoot and ProProfs; tutorial platforms such as Articulate Storyline and Credo; and data visualization tools such as

Tableau. There were also responses beyond the intended scope of this survey, specifically native

LMS features and paper handouts.

These programs and platforms were being used to capture and generate multiple data types (see

Table 1). Most respondents were gathering students' overall scores or grades for quizzes and

tutorials (66%), scores for individual responses to quiz questions (68%), and the actual responses

to those questions (79%). A significant minority of respondents were gathering PII, such as

students' names (42%) and email addresses (38%). This tracks with Perry et al's finding that for

the collection 16 of 17 named library data types, a minority of libraries include PII (2018)

Single-count responses in our results included "Student ID number", "Satisfaction", "time spent

on tutorial … number of attempts", "reflections", and geographical coordinates.

Type	Count	Percentage
Name	30	42%
Email address	27	38%
Responses to open questions	56	79%
Scores or grades for individual questions	48	68%
Overall scores or grades for DLO's	47	66%
Other	14	20%

Table 1. Types of learning data collected via DLO's. Respondents were asked to select all

options that apply.

Results indicated that a substantial proportion of respondents' libraries were taking measures to address the sensitive nature of PII. Forty-five percent reported redacting these types of information before reporting or sharing learning data. This figure may be even higher: another 24% of respondents reported not knowing what their practices for redacting reported data might have been.

The volumes of learning data gathered by libraries are typically small (see Table 2). Thirty-eight percent of respondents reported that their libraries were generating fewer than 500 separate data records (data representing the performance of an individual learner on an individual tutorial) each year. Seventy percent reported generating fewer than 10,000 each year and only 3% reported generating more than 10,000.

Type	Count	Percentage
Fewer than 500	27	38%
500 - 1,000	12	17%
1,001 - 10,000	18	25%
More than 10,000	3	4%
Don't know or not possible to estimate	11	15%

Table 2. Number of unique data records generated annually.

The other respondents were unable to estimate their volume of learning data. Comments indicated several possible reasons for this. Some respondents stated that they had no way of knowing the quantities of learning data generated across their library. In some cases, this was due to the autonomy of instruction librarians to select, implement and manage teaching tools independently from their colleagues. In others, this was because the library's DLO's were positioned within the institution's LMS, effectively blocking the library from accessing the data generated.

How are libraries managing the aggregation, storage, access, retention and reporting of learning data?

Access to the data generated by DLO's was a complicated matter for many respondents. Twenty-four percent reported a situation where the instruction librarian did not have direct access to the raw data (as opposed to reports of processed data). In cases where the program had an administrator in the library, only 42% had direct access to the data. In cases where the library had an assessment librarian, only 35% had direct access. Among course instructors, just 14% had direct access to the data. It may be possible that each responding library had at least one person in one of these groups and so had access to the data in some way. However, as noted, several respondents observed that direct access was not available to their library, often because the data were gathered and stored within the LMS or other places to which they did not have access.

A very low number of libraries reported having a specific period of time for retaining learning data before deletion (8%). This is significantly lower than Perry et al's finding that 38% of ARL libraries had a records retention schedule or policy covering student data (2018). For those that

did report a specific retention period in our study, the duration of this period was two years (n=3), three years (n=1) or five years (n=2). One respondent noted that there might be a divergence between policy and practice: "Two years according to policy. In actuality, I'm not sure." Most respondents (55%) reported retaining these data indefinitely. Thirty-four per cent reported not knowing how long their learning data were being retained. We hypothesize that this indicates a lack of access to data housed within an LMS or a situation where librarians are creating DLO's individually and managing the data independently.

What problems or uncertainties are libraries experiencing relating to current practices?

The survey invited respondents to comment on the questions or challenges that their library had encountered relating to the management of learning data. Forty-five comments were submitted. The most common (n=14) was frustration that the library was collecting learning data that were not detailed enough to be meaningful or actionable. Possible actions identified by respondents ranged from small-scale, in-library actions like improving tutorials to higher-level actions like "relating the library learning outcomes to causal (as compared to correlational) retention and graduation data".

Regarding the reasons for what respondents described as non-actionable data, several respondents cited strategic decisions by their libraries due to concerns about student privacy or the capacity to manage the data:

- "We opted not to collect any individually-identifiable data, even though doing so would have provided us with information that might have been useful in assessment of student learning."

Other responses related dissatisfaction with the quality of data to the placement of DLO's within an LMS. As noted, the survey explicitly stated we considered learning data within an LMS to be beyond the scope of our study, but several respondents were motivated to describe problems and frustrations caused by exactly that situation. As seen above, this kind of placement may have the advantage of taking responsibility for learning data management out of the library's hands, and may subject it to rigorous campus-level data management practices. On the other hand, in several cases at least it diminishes the quality of the data, as well as the overall availability:

- "It stinks that I cannot directly access the results of the information literacy test that I design. It is deployed through our Learning Management System (Blackboard)... I rely on [instructors] sending me the data with student identifying information redacted. This gives me overall scores, but not information about individual questions."
- "All of our tutorials are embedded in courses (LMS), and have an assessment, results of which are collected through the LMS, and accessed by faculty only. This creates a discrepancy as we (librarians) don't have access to this data, and cannot therefore identify the [sic] whether our instructional intervention has had any educational impact."
- "The gap between individual librarians teaching and creating learning objects and those that oversee, manage and access data is huge to the point where most front line librarians don't have access, don't know their options or just don't go there"

Another obstacle to gathering meaningful data is the technical limitations of DLO platforms. Reported limitations include inadequate question options and inadequate compatibility of data generated by different platforms. One comment about "disparate products" suggested a lack of standardization of data across DLO platforms. One set of standards for DLO's is available: Sharable Content Object Reference Model (SCORM) standards, which two respondents' libraries were applying to their tutorials (Advanced Distributed Learning Initiative, n.d.). The availability of this set of standards indicates the problem may be a lack of awareness of SCORM or a lack of expertise or resources for keeping DLO's in conformance.

- "Our digital learning object, mainly a quiz, does not have an option for students to enter any personal information so we don't know which students got the questions right and which ones didn't. We also can't follow up with the students about their answers."
- "The data that is easiest to get yields the least actionable information. Better to ask open-ended questions, scored and compiled by humans, than a simple multiple choice computer scored assessment."
- "It has been challenging to use data from disparate products to give us insight into more than superficial trends in student performance."

A library's organizational structure can be another factor complicating the availability of learning data for analysis and reporting. Several responses described situations where members of instruction teams or departments were creating and using DLO's individually, and results

were not being shared or aggregated. This created silos of learning data and a lack of awareness about how DLO's were being used. (n=7).

- "Any digital learning objects are created individually and I don't necessarily know what the [other instruction librarians] are doing"
- "We don't approach assessment seriously or systematically. Unless someone is teaching a for credit class, everything is up to individual discretion."

In contrast with the concerns about inadequate availability of data and low-quality data, numerous respondents described situations where they had learning data at their disposal, but not the resources needed to analyze them (n=9). In some situations, there was a perceived lack of human resources. In others, there was a perceived lack of time. In these cases, there were expressions of anxiety that the library was not able to perform the analysis needed to assess and communicate the impact of library instruction.

- "We have only one person to collect the data… We can't really spend a lot of time analyzing it like we should. We also do not do a great job of communicating our story in an accessible way to demonstrate what we do well/what we need to improve for the administration. It is all we can do to react to the student needs that come to our attention through this data."
- "We don't have an assessment librarian at our small liberal arts college... [T]here is tons of data and lots of ways to analyze it and there just isn't enough time."

- "It's easy to collect a huge amount of assessment data and then feel overwhelmed trying to do any meaningful analysis… We're currently challenged with how to present our assessments to university administration."

Questions relating to student privacy were a significant issue for multiple respondents (n=9). Respondents expressed discomfort with the collection of PII and the collection of student learning data without a stated plan for analysis or assessment. There was also an expressed desire that the library should manage learning data in alignment with established library values regarding users' privacy:

- "many librarians are uncomfortable with platforms that would allow the collection of the data"
- "All of the tutorial data we collect is filtered through our CMS. As far as I know there is not [sic] data management plan for this data or any long term plan for privacy and access for any information entered into the CMS. I am working with other units to try and establish something in the future, especially following [the institution]'s release of their approach to learning analytics and privacy:"
- "The campus has greater interest in collecting personally identifiable information which the library has been able to avoid at this point in time."

The library profession's position on the ethics of privacy and user confidentiality, as codified in the ALA *Code of Ethics* (Jones & Salo, 2018), appears to have created points of tension on some respondents' campuses - or, at least, the perception of such tensions. The last quotation bulleted

above suggests an administration seeking to gather types of data that go beyond library ethics - in contrast with the previous bullet, which refers to one institution's initiative to implement a policy of responsible and user-centered data collection. Tensions over the storage and management of learning data were also reported:

- "Some faculty and university administrators would like librarians to take on the role of storing and reporting quiz scores, but we have declined these requests partially because of staffing/workload and partially because we don't have systems in place to protect identifiable data and believe that we have a responsibility to protect our students' data. "

Several libraries reported having no problems or concerns. In some cases (n=9), this was because the library was using DLO's within the LMS, whether this was DLO's native to the LMS or embedded objects that had been created with third-party platforms. Respondents reported that using the LMS meant their DLO's were positioned in an environment with robust security measures and satisfactory data management practices.

- "We use the campus LMS to host SCORM packaged library tutorials, so our students' data is protected according to campus standards."
- "If we care about the results/grade of a tutorial, we create the learning object (module, quiz, discussion, etc) in our LMS (Canvas) which grades it and keeps the data confidential."

In other cases, reports of no data management issues attributed this to purposefully not collecting (n=7). They contended that this meant any data being collected were not sensitive.

- "[No issues around learning data] because for the most part, we don't gather it."
- "We do not ask for private information, so that's not an issue."

One comment cited "FERPA protections addressed for student data", but it was unclear whether this was an assertion that the respondent saw no problems relating to library learning data because of FERPA, or a questioning of whether FERPA offers sufficient protection for these data. As discussed below, compliance with FERPA may provide a lower level of protection for student data than might be assumed.

What policies or best practices are libraries following?

Survey responses suggested that a majority of libraries have a campus-level (institutional) policy applicable to learning data (63%). Some cited institutional policies, while others referred to the policies and requirements of their IRB.
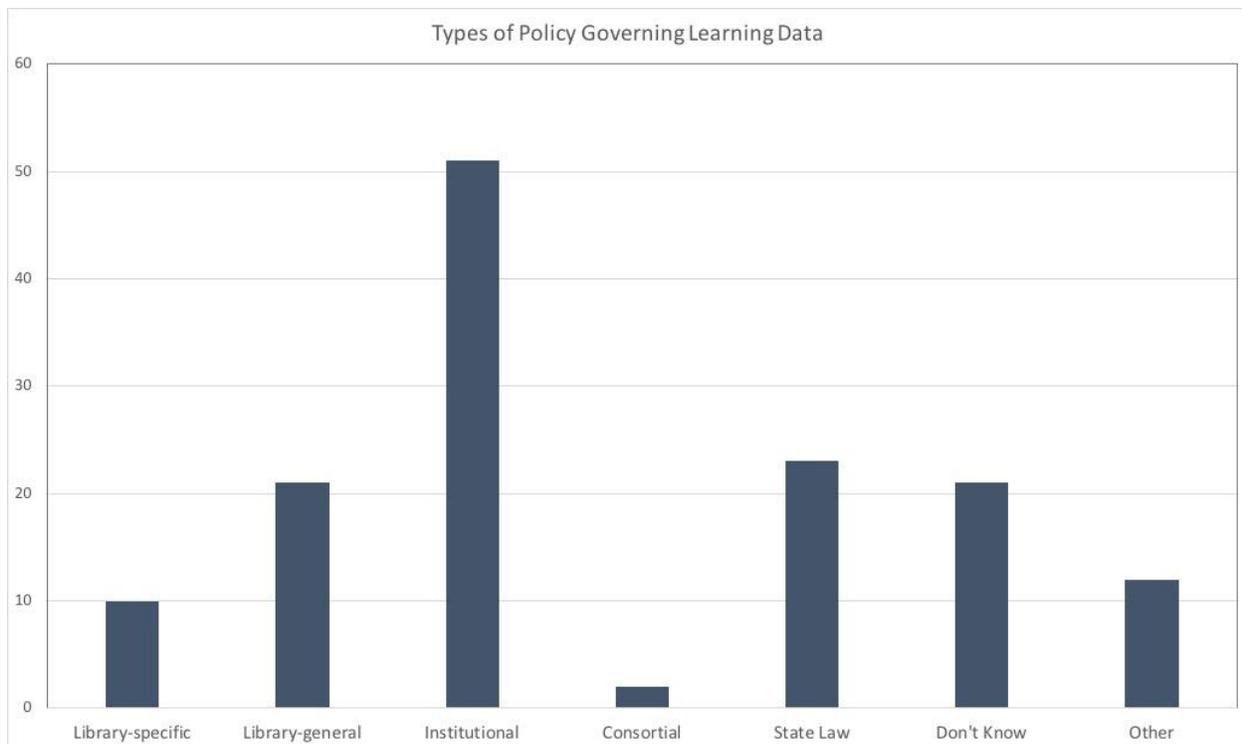
Figure 1. Types of policy governing library learning data. Respondents were asked to select all options that apply.

A significant proportion of respondents reported having an institutional policy that governed library learning data (37%). Few reported having library-level policies in place. Thirteen percent of responses (n=9) identified a library-level policy specific to learning data and it is this set of responses that indicated the highest level of adherence to best practices regarding student learning data. Twenty-five per cent of responses (potentially overlapping with the 13%) identified a library-level policy with general applicability to user data, though it was unclear how adequately these policies were addressing learning data issues. These statistics align with Perry et al's finding that a minority (25%) of ARL respondents reported having library staff documents or guidelines that informed any participation in library analytics initiatives (2018).

Further analysis found significant correlations between the existence of a campus-level policy (as opposed to other types of policy) with attention to the following issues: collection of demographic or personally identifying data; responses or scores; management of receipts; data storage; and data retention.

Some cited policies or governing documents that are not proximate or specific to learning data, including state laws (30%) - for example Washington state law on Student Education Records (Washington Administrative Code, 2018) - and FERPA (10%). "When I think about data management best practices, I just think we follow our university's FERPA guidelines" was a typical sentiment. However, reliance on FERPA protections may be inadequate for libraries that aspire to uphold high levels of student privacy. FERPA explicitly prohibits educational institutions from releasing a student's records without their written consent. However, as noted above, exceptions allow institutions to share records with third-party "school officials" without student consent when there is an educational reason to do so; and higher education institutions have expanded both the scope of an academic record, sometimes including geolocational and social media data, and expanded the definition of a "school official" to include commercial entities (Parks, 2017). No response cited the ALA's or ACRL's guidance, apparently confirming a lack of awareness or application of these documents.

Many respondents noted a lack of information or clarity around relevant policies and guidelines. In the words of one respondent, "I don't think we have anything formally written." A significant number (25%) stated they did not know what policies, regulations, laws or guidelines governed their library's management of learning data.

Even in cases where a policy was in place, there appeared to be inadequate guidance regarding many aspects of learning data management that might be considered critical (see Table 3). For each of the following aspects, a minority of policies had relevant content: sharing and transfer of data (46%), collection of demographic data (45%), location of storage (38%), collection of response/scoring data (34%), retention (32%), receipt and ownership of data (24%), and third-party ownership (23%).

Type	Count	Percentage
Collection of demographic data	32	45%
Collection of responses / scoring data	24	34%
Individual receipt / ownership	17	24%
Location of storage	27	38%
Sharing / transfer of data	33	46%
Retention	23	32%
Third-party ownership / storage	16	23%
Other	4	6%

Table 3. Aspects of data management covered by governing policies or best practices. Respondents were asked to select all options that apply.

We also hypothesized that having a policy in place would not necessarily equate to practices that meet the requirements of that policy. Survey responses confirmed that some libraries were not conforming to a learning data policy in some or all aspects. Six percent assessed their library as "mostly" or "moderately" conforming with policy and 3% as "not at all". This compares with 6% that assessed their library as "fully" conforming. However, by far the most common response was "don't know" (68%), adding to the evidence that many librarians working in online learning have limited knowledge about their library's current practices in this area.

What problems or uncertainties are libraries experiencing relating to policies?

Our survey asked participants to describe any issues they might have encountered in developing, or aligning with, policies, regulations, laws or guidelines relating to learning data. This elicited significantly fewer comments (n=20) than the number received regarding issues stemming from current practices (n=45). It seems reasonable to interpret this difference as a reflection of the lower number of libraries with policies in place, rather than an indication that policies are less problematic than practices.

Indeed, the most-frequently cited issue was the absence of a learning data policy (n=7). Respondents noted that the lack of a policy had created issues relating to the standardization (or not) of data and intensified concerns about the library's role in preserving student privacy.

- "We don't have a policy on learning performance data, which is desperately needed as assessment becomes an increasingly big component of library responsibilities."
- "Any student generated digital data - there is no present policy."

- "In the absence of administrative support to develop these kinds of policies, I (head of instruction) have elected not to collect identifying data whenever possible and to purge names and emails from our data on a semester-by-semester basis."

For respondents whose libraries did have a policy in place, the issues were varied. Several observed (n=3) that their library's learning data were governed by general data policies at the institutional level, or by FERPA - in other words, by frameworks not specific to learning data and likely to offer weakened protections for student privacy.

- "There is no standard at our university except that standards that apply to all student data"
- "We had to complete an online course regarding privacy and Federal law and obtain a certificate showing that we passed."
- "[W]e have institutional rights to the student data we collect via library learning objects… [A]ll data use is covered under normal policies for instructors."

Some respondents also identified their institution's IRB as a source of learning data governance. Applying for authorization of DLO's in library instruction would be an onerous and atypical process, so it was unclear to us whether librarians were really expected to obtain IRB approval, or maybe instead trying to apply the underlying principles of IRB review. IRB principles and the processes for obtaining authorization were in one case seen as appropriately rigorous - "[No issues…] our survey complies with IRB policies." - and in another as prohibitively rigorous:

- "The IRB at my institution is so strict that it is a huge disincentive to collecting this type of data or conducing [sic] studies on this type of data."

Non-compliance (or possibly partial compliance) was seen by some respondents as diminishing the role of any policy. For example, "So far, we don't have a mechanism to really enforce compliance here." We can speculate that compliance may be less likely to be enforced when the policy is issued at the institutional level and is generally applicable across all campus units, as opposed to being developed and administered at the library level. One respondent described additional factors that limit librarians' ability to adjust non-compliant practices, specifically the positioning of DLO's within the LMS and resistance from faculty:

- "In the past, we have modified our learning objects (specifically our e-courses) so that we can be compliant with FERPA and our university's policy on confidential data. We are limited in what we can do in our course management system, and we've had pushback from faculty as we have changed the assessments/outputs (i.e., we switched from a certificate that students could print to an electronic badge). This has been an issue, but we have had the support of our e-Learning department and the issues have diminished somewhat."

In some cases, a lack of engagement with issues concerning libraries' learning data has been a problem. Inadequate attention has impaired some libraries' compliance with existing policies or other work to address learning data issues. Inadequate attention was ascribed to lack of interest, lack of expertise, lack of information, and lack of communication:

- "[P]olicy compliance is influenced heavily by our librarians' interest"

- "I struggle to get my colleagues to engage in conversations about or care about student data at all."

- "Such a poor area of communication, understanding and practice"

Conclusions

Limitations

The study was designed as a survey, not a census, and the results describe the perspectives of a small proportion of the academic libraries in the United States. The number of responses to this survey was sufficient to collect meaningful data and identify patterns and commonalities across different librarians' experiences. However, a larger number of responses would provide a larger dataset, probably a richer set of comments and observations, and a stronger basis for our analysis and interpretations.

As we anticipated, some respondents tried to describe the practices of their library or instruction department as a whole, with imperfect information about their colleagues' work. This was sometimes true even when the respondent was the head of the instruction department or when the institution was small and the number of instruction librarians was low.

Further directions

This study was motivated by a desire to develop robust practices and policies that would guide our management of the learning data generated by library DLO's. Our findings appear to confirm

that learning data are a significant aspect of library user data, and that many librarians share our concerns about following appropriate practices and policies. To some extent, this validates our premise that libraries should be strategic and intentional in their management of this type of data.

In practice, the current situation is that many libraries have insufficient guidance or governing documentation. For a majority of respondents, either no applicable policy is in place or there is inadequate awareness of such a policy. Approximately one-third reported having an institutional policy and significantly fewer reported having a library-level policy. For our purpose of developing a policy, this means that currently there are few library-level policies available that we might take as a direct model.

Where institutional policies are in place, our study shows these may be too general or too broadly applicable to be effective instruments for guiding the management of learning data. Most do not address key aspects of data management such as data collection, storage, sharing, and retention. Some documents cited by our respondents, such as FERPA, have little specificity to learning data and suggest some confusion about what constitutes data governance. We concur with Hinchliffe & Asher (2015) that libraries need to write their own policies when campus-level policies are not available or do not address the needs of libraries or learning data. Of course, any library-level policy needs to align with other governing documents, whether they are institutional policies, state laws, academic best practices, or the recommendations adopted by librarianship's professional bodies. And as we note in our Introduction, one challenge for libraries is translating widely-scoped policies on privacy and data management into library-specific data management practices.

What should library-specific data management practices encompass? This study assumed that libraries need to address core practices such as the types of data to be collected, the location of data storage, access to stored data, the aggregation and reporting of data, and timelines for retention or deletion. In addition, our respondents highlighted a range of other aspects of data management that could be made less problematic by clearly-defined practices and procedures:

- Balance security measures and convenience of access, possibly by defining positions with access privileges and procedures for submitting requests for data.
- Establish processes for requesting or obtaining library data generated within an LMS.
- Consider standardizing learning data generated by different platforms, collected by different librarians, or collected at different points in time. In particular, consider defining standards for the granularity of learner data. What level of granularity is required to make the data meaningful? What level is required to make it compatible with other campus metrics?
- Define circumstances in which it is appropriate (or not) to collect data tied to PII.
- Identify resources (time and expertise) for analysis and assessment of learning data, for example by formalizing responsibilities.
- Define the library's position on the disclosure of learning data and participation in learning analytics.
- Consider the possible distinctions between "small data" that stay within the library's control and can be used for in-library assessment; and data compatible institution-wide "big data" analytics projects.

- Encourage dialog within the library and across campus on learning data and learning analytics, with the goals of raising awareness and developing relationships with other interested parties.
- Consider how to communicate the library's learning data principles and practices to students.

Libraries also need to participate in conversations around learning data, including dialog across the profession, discussions around data ethics, and campus data initiatives such as learning analytics. This could include serving on the campus IRB or learning analytics steering groups (Robertshaw, 2018). In all of these forums, libraries need to articulate the value and limitations of library data. For example, to what extent can library data support metrics for student retention and graduation?

Libraries also need to articulate the ethical positions held by the profession. Are those positions clear and established? Our survey findings indicated many librarians have concerns that collecting and sharing learning data have negative implications for student privacy. In other words, there is continued adherence to values codified by the ALA *Code of Ethics* and other documents. At the same time, our literature review showed that the profession's longstanding consensus on user privacy is questioned by librarians who attach greater priority to supporting student success, by contributing personally identifying student data to learning analytics projects (Oakleaf et al, 2017, Oakleaf, 2018). This is a debate for our profession that needs more dialogue, more leadership, more engagement, and continued research into how exactly learning data can help - not harm - our students.

References

Advanced Distributed Learning Initiative. (n.d.). SCORM overview. Retrieved September 26,

    2018, from https://adlnet.gov/research/scorm/

American Library Association. (2008, January 22). Professional ethics [Text]. Retrieved April

    12, 2018, from http://www.ala.org/tools/ethics

American Library Association. (2014). Questions and answers on privacy and confidentiality

    [Text]. Retrieved May 31, 2018, from http://www.ala.org/advocacy/privacy/FAQ

American Library Association. (2016, May 5). Library privacy guidelines for students in K-12

    schools [Text]. Retrieved May 31, 2018, from

    http://www.ala.org/advocacy/privacy/guidelines/students

American Library Association. (2007, May 29). Privacy Tool Kit [Text]. Retrieved September

    25, 2018, from http://www.ala.org/advocacy/privacy/toolkit

Arnold, K. E., & Sclater, N. (2017). Student perceptions of their privacy in learning analytics

    applications. In *Proceedings of the Seventh International Learning Analytics &*

    Knowledge Conference (pp. 66–69). New York, NY, USA: ACM.

    https://doi.org/10.1145/3027385.3027392

Chowcat, I., Kay, D., & Korn, N. (2015). The legal, risk and ethical aspects of analytics. In

    Library Analytics and Metrics: Using Data to Drive Decisions and Services (pp. 157–

    165). London, UK: Facet Publishing.

Cox, B., & Jantti, M. (2015). The Library Cube: Revealing the impact of library use on student performance (University of Wollongong). In B. Showers (Ed.), *Library Analytics and Metrics: Using Data to Drive Decisions and Services*. London, UK: Facet Publishing.

Dawson, S. (2002). The impact of institutional surveillance technologies on student behaviour. *Surveillance & Society*, *4*(1/2). Retrieved from https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/3455

Hinchliffe, L., & Asher, A. (2015, June). *All the data: Privacy, service quality, and analytics*. Presented at the ALA Annual Conference & Exhibition, San Francisco, CA. Retrieved from http://alaac15.ala.org/node/28724

Jones, K., & Salo, D. (2018). Learning analytics and the academic library: Professional ethics commitments at a crossroads. *College & Research Libraries*, *79*(3), 304–323. https://doi.org/10.5860/crl.79.3.304

Knox, D. (2010). A good horse runs at the shadow of the whip: Surveillance and organizational trust in online learning environments. *Canadian Journal of Media Studies*, *7*(35).

LAK '11. (n.d.). 1st International Conference on Learning Analytics and Knowledge 2011. Retrieved May 31, 2018, from https://tekri.athabascau.ca/analytics/

Magi, T. (2010). A content analysis of library vendor privacy policies: Do they meet our standards? *College & Research Libraries*, *71*(3), 254–272.

Mestre, L. S., Baures, L., Niedbala, M., Bishop, C., Cantrell, S., Perez, A., & Silfen, K. (2011). Learning objects as tools for teaching information literacy online: A survey of librarian usage. *College & Research Libraries*, *72*(3). https://doi.org/10.5860/crl-130rl

Nackerud, S., Fransen, J., Peterson, Kate, & Mastel, K. (2015). Retention, student success and academic engagement at Minnesota (University of Minnesota). In B. Showers (Ed.),

Library Analytics and Metrics: Using Data to Drive Decisions and Services. London, UK: Facet Publishing.

Oakleaf, M. (2015). The library's contribution to student learning: Inspirations and aspirations. *College & Research Libraries*, *76*(3), 353–358. https://doi.org/10.5860/crl.76.3.353

Oakleaf, M. (2018). The problems and promise of learning analytics for increasing and demonstrating library value and impact. *Information and Learning Science*, *119*(1/2), 16–24. https://doi.org/10.1108/ILS-08-2017-0080

Oakleaf, M., Whyte, A., Lynema, E., & Brown, M. (2017). Academic libraries & institutional learning analytics: One path to integration. *The Journal of Academic Librarianship*, *43*(5), 454–461. https://doi.org/10.1016/j.acalib.2017.08.008

Parks, C. (2017). Beyond compliance: Students and FERPA in the age of big data. *Journal of Intellectual Freedom and Privacy*, *2*(2), 23–33. https://doi.org/10.5860/jifp.v2i2.6253

Perry, M. R., Briney, K. A., Goben, A., Asher, A., Jones, K. M. L., Robertshaw, M. B., & Salo, D. (2018). *SPEC Kit 360: Learning Analytics*. Washington, DC: Association of Research Libraries. Retrieved from https://publications.arl.org/Learning-Analytics-SPEC-Kit-360/

Rempel, H. G., & Slebodnik, M. (2015). *Creating online tutorials: A practical guide for Librarians*. Lanham, MD: Rowman & Littlefield.

Robertshaw, M. B. (2018, August). *The lock to the safe has been tampered with: Why FERPA & IRB aren't enough to protect student data in higher education*. Education presented at the NISO Virtual Conference: The Computer Campus: Integrating Information Systems and Services, Online. Retrieved from https://www.niso.org/events/2018/08/computer-campus-integrating-information-systems-and-services

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, *32*(2), 143–159. https://doi.org/10.1080/01972243.2016.1130502

Sclater, N., & Bailey, P. (2015, June 4). Code of practice for learning analytics. Retrieved July 9, 2018, from https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning analytics in higher education: A review of UK and international practice*. Jisc. Retrieved from https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education

Sherriff, G. (2017). Guide on the Side and LibWizard Tutorials side-by-side: How do the two platforms for split-screen online tutorials compare? *Journal of Web Librarianship*, *11*(2), 124–142. https://doi.org/10.1080/19322909.2017.1295821

University of Vermont. (2016, May 5). FERPA Rights Disclosure. Retrieved from https://www.uvm.edu/policies/student/ferpa.pdf

University of Vermont. (2012, August 12). Privacy Procedures for Protected Personal Data. Retrieved from https://www.uvm.edu/policies/general_html/privacyprocedures.pdf

University of Vermont. (2009). Confidentiality Policy Statement and Procedures. Retrieved September 25, 2018, from https://www.uvm.edu/about/library_policies/privacy_policy_and_procedures

U.S. Department of Education. (2018, March 1). Family Educational Rights and Privacy Act (FERPA) [Guides]. Retrieved June 22, 2018, from https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html?

Vermont State Legislature. Vermont Statutes Annotated Title 22, § 171-173 Library Patron Records. Retrieved from https://legislature.vermont.gov/statutes/fullchapter/22/004

Washington State Legislature. Washington Administrative Code (WAC) Title 172 Chapter 172-

    191 Student Education Records (2018). Retrieved from

    http://apps.leg.wa.gov/wac/default.aspx?cite=172-191

Williamson, J., & Phillips, J. (2017, January 30). Consenting adults? Privacy in an age of

    liberated learning data. Retrieved April 2, 2018, from

    https://er.educause.edu/articles/2017/1/consenting-adults-privacy-in-an-age-of-liberated-

    learning-data