Syracuse University

SURFACE

School of Information Studies - Faculty Scholarship

School of Information Studies (iSchool)

3-13-2017

An Information Policy Perspective on Learning Analytics

Caroline A. Haythornthwaite

Follow this and additional works at: https://surface.syr.edu/istpub

Part of the Library and Information Science Commons

Recommended Citation

Haythornthwaite, Caroline A., "An Information Policy Perspective on Learning Analytics" (2017). *School of Information Studies - Faculty Scholarship*. 181.
https://surface.syr.edu/istpub/181

An Information Policy Perspective on Learning Analytics

Caroline Haythornthwaite
Syracuse University
School of Information Studies
343 Hinds Hall
chaythor@syr.edu

ABSTRACT

Policy for learning analytics joins a stream of initiatives aimed at understanding the expanding world of information collection, storage, processing and dissemination that is being driven by computing technologies. This paper offers a information policy perspective on learning analytics, joining work by others on ethics and privacy in the management of learning analytics data [8], but extending to consider how issues play out across the information lifecycle and in the formation of policy. Drawing on principles from information policy both informs learning analytics and brings learning analytics into the information policy domain. The resulting combination can help inform policy development for educational institutions as they implement and manage learning analytics policy and practices. The paper begins with a brief summary of the information policy perspective, then addresses learning analytics with attention to various categories of consideration for policy development.

CCS Concepts

• **Social and professional topics~Computing / technology policy** [1]

Keywords

Information policy; learning analytics; educational policy

1. INFORMATION POLICY

Information policy is a relatively new area of research and practice. In its larger domain, information policy is concerned with public policy relating to information use and addresses legal, ethical and moral positions and practices pertaining to information from creation through use to disposition. Sandra Braman [3][4] most comprehensively defines the area, addressing the impact on government of the transformation to an information state, and providing this definition:

> "Information policy is comprised of laws, regulations, and doctrinal positions – and other decision making and practices with society-wide constitutive effects – involving information creation, processing, flows, access, and use." ([4], p. 3)

Of most interest here for application to learning analytics is the articulation of the information activities where policy adheres. The definition and distinction among these activities results "from a synthesis of the many, many conceptualizations of information production chains that are used implicitly and explicitly across disciplines" ([4], p.3). Braman argues that "[m]arking the boundaries of the domain with "information creation, processing, flows, access, and use" provides a synthetic and succinct heuristic that meets important evaluative criteria" that are valid, comprehensive, theoretically sound, and "*translatable* into the multiple languages in which the audiences of information policy speak" ([4], p. 3).

These information activities are easily translatable for systems analysis and learning analytics into the stages of the information lifecycle. Although 'activities' is no doubt a better term (given that processes of information creation, use, etc. will overlap in everyday practice), the lifecycle concept provides a procedural approach to understanding where policy is needed in the management of learning analytics data and information. And, indeed, the information lifecycle has been used by others as a framework for examining policy from an information perspective. Pasek [18], for example, synthesizes discussion from several papers in the information science area to tie information policy to the information lifecycle phases of creation, production, distribution, access, and use.

While intuitive and useful, an information lifecycle framework as outlined by Pasek does lack some of the nuance of the Braman categories. For example, by leaving out the concept of flows it is possible to miss the idea that 'creation' of an information resource can be achieved by selecting from an ongoing stream (flow) of information, e.g., by creating an archive of twitter posts, or by curating an information flow to create a resources of selected, topic specific postings. Similarly indexing is not explicitly addressed; for twitter this may be the hashtag, forming a crowd-based collaborative index to a stream of postings.

Both authors' short lists also seem to give little prominence to the disposition of information and data, where policies are needed to address both retention and deletion of records. Attention to this stage is well known in areas such as records management, and is rapidly emerging in internet policy relating to issues around the 'right to be forgotten' [7].

Since learning analytics is an area that itself includes "many, many conceptualizations of information production chains", it is an area that can benefit from engaging with information policy frameworks. At the same time, learning analytics may be facing issues not yet addressed in information policy, and examination of learning analytic activities may inform information policy development.

Writ broadly, information policy addresses issues relating to: intellectual property, such as copyright; privacy of personal information, particularly as related to government collection of personal data; security of personal privacy as well as national

security in relation to intrusions into government information systems; and access to information, including issues of literacy, digital divide, and the role of libraries, archives and museums.

Well-known US government policies relating to information policy include the Digital Millenium Copyright Act (DMCA), Family Educational Rights and Protection Act (FERPA), Health Insurance Portability and Accountability Act (HIPPA), Freedom of Information Act (FOIA), with similar acts found in many other countries. Also known are government requests for information, e.g., in the US Patriot Act (now Freedom Act), and undisclosed domestic information gathering (e.g., by the US National Security Agency)[14], with public opinion divided between acceptance and rejection of such activities [19].

Organizations such as the American Library Association, and movements such as 'Freedom from Surveillance' [1][2] and 'Freedom to Read' [3] work in the information policy domain to effect change and education relating to government monitoring, censorship and privacy relating to information access (e.g., non-disclosure of library or video borrower records). The ubiquity of records and personal information on the Internet has driven efforts for the Right To Be Forgotten (RTBF), with European initiatives in this area leading the way [7][11][21]. Open data, open government data, and sunlight initiatives advocate for open access to government information to increase access to resources and to support government transparency [20].

These areas of attention provide a framework for examining learning analytics policy. It is possible then to frame the discussion based on the information activities outlined by Braman and/or the information lifecycle. We can ask, for example, how issues of information privacy are found in the activities of creation, processing, flows, access, and use, or the stages of creation, production, distribution, access, use, and disposition in relation to learning data.

While it is beyond the scope of this short paper to address how each information policy area relates to each information activity or stage in the information lifecycle, what follows provides some examples of how these information policy areas draw attention to activities that relate to learning analytics practice and policy. Important discussion in the information policy area address:

- Protection of individual records, including legal frameworks for privacy protection such as FERPA and HIPPA; attention to the ramifications of the failure of anonymization techniques to keep records anonymous; and invasions of privacy whether for law enforcement efforts or by breaches of secruity

- Monitoring and dataveillance [5], including collection of data without participant knowledge or consent; general surveillance of individual actions in public and/or online [15]; information discovery through connecting databases and/or data mining techniques [12][22]

- Access to information, including literacy relating to reading, computing and technologies, and the digital divide; transparency in government or other data collection, use and disposition activities

- Redress mechanisms for correction, removal or amendment of records

- Ownership of information, including copyright, patents, trademarking, etc.

The following sections discuss these areas in relation to learning analytics.

2. PROTECTION OF DATA RECORDS AND INDIVIDUAL PRIVACY

Protection of learner privacy opens up a wide array of questions about data across the information lifecycle, from what data should be collected, to how it is stored, used, and disposed of. Policies on record retention are undoubtedly present in educational institutions, relating to legal requirements and social practices of record keeping. For learning analytics, the question of what data is collected has to be revisited as analytics can – conceivably – collect more than just in-class records, and can connect to more than educational data. Policies in this area may be more about what data are not to be collected than about collection of data, and about keeping databases separate rather than connecting them. For example, what policies should apply to connecting student aid data to student performance data? Or connecting foreign versus domestic status to student data?

The issue of connecting databases—or joining them in the database management sense—is becoming a major issue in information policy areas because the 'safe harbor' of anonymization as a means of protecting individual identity is no longer providing the protection it should. This big data problem arises from widespread data collection, and a general trend to information discovery by combining data across collections. Recent studies reveal how little data it takes to reidentify individuals. Following the release of anonymized data on movie recommendations made in Netflix, researchers Arvind Narayanan and Vitaly Shmatikov were able to reidentify 68 percent of the users by knowing only *when* a user rated a movie and the ratings on *two movies*; with data on *when* and six ratings 99 percent of users could be reidentified (cited in [17]); similarly, researchers at MIT found that the dates and locations of four credit card purchases allowed identification of 90 percent of users in a dataset of 1.1 million records [12].

These are issues that change the direction of both policy and law, and they are appearing through analytics first. Writing in 2010, Ohm [17] notes that

> "Nearly every information privacy law or regulation grants a get-out-of-jail-free card to those who anonymize their data … the public policy debate … centers almost entirely on squabbles over magical phrases like "personally identifiable information" (PII) or "personal data. … Prior to these [reidentification/ deanonymization] studies, nobody would have classified ZIP code, birth date, sex, or movie ratings as PII." ([17], p. 1705)

Thus, while most of the discussion here is of using information policy as a roadmap to learning analytics policy, each area can inform the other.

Joining databases is also an issue in relation to ideas of freedom to read and keeping borrower records private [10]. Learning analytics implementations float the idea of connecting individual records about library use to the outcomes on courses. Freedom to read would suggest that these databases should not be connected, and that general library borrowing and reading practices should be kept private, including in relation to the course in question. The ethical issue that arises is whether students borrowing from a university library are to be granted the same protections or considerations as individuals borrowing from a public library.

3. DISCLOSURE OF INFORMATION GATHERING AND USE

Perhaps the biggest challenge will be transparency, in identifying and making evident the kinds of information collected and their use – perhaps even the ability for students to opt out of their use. Educational institutions, many of which are government supported, may need to set policy that addresses openness in a way that discloses what data are collected, how they are used, etc., and in a way that is accessible – technically, intellectually – to the relevant constituencies. Thus, educational settings that adopt learning analytics may also find themselves setting policies about how and when to convey this information to relevant constituencies.

Approaching learning analytics from an information policy perspective calls for greater emphasis on disclosure, and this in turn may change the emphasis of systems development. For instance, one of the relevant constituencies is the student population that supplies this data. Considering information use in the context of openness can influence learning analytics development by establishing a higher value to providing learning data back to students than aggregating such data for institutional use. Where this value enters design, it can set a different policy and practice perspective for learning analytics, one that differentiates if from academic analytic initiatives. (For more on values in design, see [13].)

Disclosure also has other ramifications. In particular it can place a bureaucratic burden on educational institutions to consider the impact of each new data stream on existing policy. This could have a chilling effect on innovation and spontaneity in learning analytics systems development; and it is also likely to tip the balance to testing of data collection options to identify those with repeatable and useful outcomes for educational practice. Policy about research and test environments are then needed to balance the needs of development and production, including policy to safeguard conditions for innovation, e.g., by designating and maintaining separate exploratory data repositories, and by establishing short rather than long-term retention policies about test data.

Along with disclosure about data collection and use, communication in clear language, and opportunities for amendment of records, freedom from surveillance draws attention to the need to address whether and for what data individuals may withhold consent for data collection. Moreover, while the emphasis is normally on surveillance, i.e., top down, organizational monitoring, not all surveillance comes from above. As monitoring technologies have become more prevalent, they are now found in the hands of students, and thus policies about sousveillance [16] need to be considered. There are already issues arising around student filming of lectures, raising questions about intellectual property: Who owns the lecture? Who owns its distribution rights? Ubiquitous recording technologies may even be co-opted into educational practice, e.g., requiring students to record interviews, events, etc. as part of the class discussion. This should immediately open up instituituional discussion and policy around ownership, privacy, ethics and academic behavior.

With every case of information gathering comes the issue of ownership. Copyright protects original works of authorship (not ideas, systems, or methods), but within that realm transcripts of conversations may be included [9]. While university policies may already lay out the parameters of ownership of materials generated in the pursuit of education, new ground may need to be broken to understand the copyright of online conversations in an educational context, and perhaps even the patenting process associated with ideas generated in online forums.

4. ACCESS TO RECORDS

Information policy developments, and Freedom of Information/ Right to Information legislation, have generated an expectation that stored data records that are not routinely disclosed can be made available by request. These generates a need to manage requests for student access to their records. Learning analytics systems may need to implement means of record production in anticipation of such requests, and indeed such systems may be the means for production of such records. Thus, policies about implementation must grapple early with the eventuality of a freedom of information request.

Similarly, requests for data may come from outside the institution, for example in compliance with the Patriot Act/Freedom Act. Learning analytics systems may hold data on individuals that is not in the student record. Such data may be requested, and become part of a record that by policy or law must be shared with agencies outside the university. Thus, policy needs to consider the way new data streams from learning analytics can potentially end up in student records.

5. AMENDMENT AND REMOVAL OF RECORDS

Mistakes, data corruption, ineffective data practices, and individual requests can all lead to the need to amend and/or remove data records. Where learning analytics data and systems are involved in creating and generating these records, policies need to address not only whether changes can be made and under what circumstances, but also how data changes will be communicated to record holders.

New analytic techniques may also generate new data, which may be retroactively attached to existing records. For example, later analyses may generate predictive statistics on a student's likelihood of success, perhaps with the benevolent intention of identifiying ways to increase success. Where does such data belong in the student record? If these data are generated based on records of graduated students, should such a score be entered retroactively into student records? Where amendment might include such new data points, policy needs to address whether such data is retroactively attached to records or are added only from an official start date; and where such data may appear prejudicial in future review, policy may be needed to limi the lifespan of the data, e.g., removing data at graduation or at another well-considered point in time or student progress.

6. CONCLUSION

Information policy has emerged as a field of inquiry over the last 20 years, responding to the increasing presence, availability, use and misuse of data and information streams. As a field, it aims to affect national level policies relevant to the information activities identified in many disciplines. While most examples here have been of US initiatives and acts, information policies are in place in most countries, and often stem from initiatives in intergovernmental agencies. This paper advocates for bringing the information policy knowledge into learning analytics policy as well as exploring how learning analytics can add to information policy.

7. REFERENCES

[1] American Library Association (nd). *Vanishing liberties: The rise of state surveillance in the Digital Age* (video 27 minutes). Retrieved October 12, 2016 from: https://chooseprivacyweek.org/

[2] American Library Association (nd). Choose Privacy Week. chooseprivacyweek.org/.

[3] Braman, S. (2006). *Change of state: Information, policy, and power*. Cambridge, MA: MIT Press.

[4] Braman, S. (2011). Defining information policy. *Journal of Information Policy, 1*, 1-5.

[5] Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM, 31*(5), 498-512. http://www.rogerclarke.com/DV/CACM88.html.

[6] De Montjoye, Y-A, Radaelli, L., Singh, V.K. & Pentland, A. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, *347*(6221), 536-539. DOI: 10.1126/science.1256297

[7] European Commission (nd). Factsheet on the "Right to be forgotten" ruling (C-131/12). http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

[8] Ferguson, R. Hoel, T., Scheffel, M., & Drachsler, H. (2016). Special section on ethics and privacy in learning analytics. *Journal of Learning Analytics*, *3*(1). http://learning-analytics.info/journals/index.php/JLA/issue/view/373/showToc

[9] Fowler, M. (Friday, January 7, 2011). Who "owns" an Interview. Retrieved Oct. 14, 2016 from http://www.rightsofwriters.com/2011/01/who-owns-interview.html.

[10] Freedom to Read Foundation (nd). Retrieved from http://www.ftrf.org/

[11] Garcia-Murillo, M. & MacInnes, I. (Nov.22, 2014). *The right to be forgotten: Its weaknesses and alternatives*. Paper presented at the Institute for Information Policy Workshop, Washington, DC. http://dx.doi.org/10.2139/ssrn.2529396

[12] Hardesty, L. (January 29, 2015). Privacy challenges: Analysis: It's surprisingly easy to identify individuals from credit-card metadata. Retrieved Oct. 14, 2016 from http://news.mit.edu/2015/identify-from-credit-card-metadata-0129.

[13] Knobel, C., & Bowker, G. C. (2011). Values in design. *Communications of the ACM,* 54(7), 26-28.

[14] Lepore, J. (June 24, 2013). The prism: Privacy in the age of publicity. *The New Yorker.* http://www.newyorker.com/magazine/2013/06/24/the-prism

[15] Lyon, D. (2002). *Surveillance society: Monitoring everyday life*. Buckingham, England: Open University Press.

[16] Mann, S., Nolan, J., & Wellman, B. (2002). Sousveillance: Inventing and using wearable computing devices for data collection in surveillance environments. *Surveillance & Society, 1*, 331-355. http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/3344/3306

[17] Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57*, 1701-1777.

[18] Pasek, J. E. (2015). Defining information policy: Relating issues to the information cycle. *New Review of Academic Librarianship, 21*(3), 286-303, DOI: 10.1080/13614533.2015.1009126

[19] Rainie, L. (Sept. 21, 2016). The state of privacy in post-Snowden America. http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/

[20] The Annotated 8 Principles of Open Government Data (nd). Retrieved from https://opengovdata.org/.

[21] Toobin, J. (Sept. 29, 2014). The solace of oblivion: In Europe, the right to be forgotten trumps the Internet. *The New Yorker*.

[22] Wall St. Journal (January 30, 2015) Your credit cards say a lot more about you than you think? Retrieved October 12, 2016 from http://blogs.wsj.com/moneybeat/2015/01/30/your-credit-cards-say-a-lot-more-about-you-than-you-think/