Changing Policies Concerning Student Privacy and Ethics in Online Education

Anthony E. Kelly and Mika Seppälä

***Abstract*—Whereas the growth in global distance and online education has blossomed, especially with the arrival of massive open online courses (MOOCs), the same technological infrastructure permits unprecedented access to knowledge about students and their behaviors. This knowledge extends far beyond scores on tests to include the measurement of noncognitive factors such as persistence, and intrusive metadata such as geolocation information. Moreover, the growth in the internet of things (e.g., via smart phones and RFID chips) is rapidly complexifying the problem of intrusive data collection. In this paper, we review some of the policy challenges facing student privacy in online learning.**

***Index Terms*—Privacy, online, learning, policy.**

I. Introduction

Changes in technology are radically changing how people learn in cyber infrastructure learning environment such as massive open online courses (MOOCs). These data ecosystems, which may span formal education, informal education, and out-of-school settings, allow and track activities, locally, using the internet of things (e.g., smart phones, smart sensors and other cyber physical devices), and globally, via the internet. New data interoperability protocols allow tracking of behavior across an ecosystem of devices and platforms. A wide range of learner behaviors (many implicit or non-obvious, such as those collected via metadata emitted by smart phones) generate rich and vast data-streams, which may be stored on servers controlled or not controlled by the online learning platform.

Big data applications pose significant opportunities and challenges for researchers. For example, a partnership between Facebook and Wolfram Alpha (http://www.wolframalpha.com/facebook/) generates a profile of how one's friends cluster in groups, where in the world they are located, the global reach of one's network, the popularity of one's friends, what one talks about on Facebook, when one uses Facebook by hour, and what activities one engages in during those hours, the types of relationships one's friends are engaged in, how one's friends connect to you by age, gender, relationship status, how one's friends are interconnected and who the "gateway" friends are across networks, and an analysis of the photos one likes. These same data may be generated for each friend in one's network. Thus, for research purpose, potential human subjects may already be members of numerous digital and social networks, they may have access to data sources on other people, and may already have privacy and data security agreements in place with third party providers.

Online behaviors may generate data directly related to learning from membership in social networks and also from many non-learning indicators, including "quantified self" data from wearable devices that can reveal extraordinarily detailed insight into research subjects, and their lives. The implications of available correlational data mean that the accepted boundaries of research studies are evaporating. Thus, policies on human subjects' protections in research are key challenge facing designers of learning systems, especially in the US (e.g., under IRB, FERPA, COPPA and, perhaps HIPPA regulations, which are described in more detail, below).

II. The Policy Framework in the US

Human subjects' protection principles in the United States emanate from the Belmont Report [1], which established general guidelines for the treatment of people in research in response to notorious treatment of human subjects, including the Tuskegee syphilis studies [2]. The Belmont principles may be summarized as: a) respect for persons; b) beneficence; and c) justice. The respect criterion is concerned with subjects' autonomy, informed consent, courteous treatment and (unless sanctioned) absence of deception during research. Beneficence is a criterion emphasizing minimizing risks to subjects while maximizing the benefits for the subjects and the project. Beneficence is a "do no harm" admonition. Justice requires non-exploitation of subjects and their fair treatment. Justice also concerns a fair balance of costs and benefits for current and future participants.

Over time, the Belmont Report has become expressed in the Common Rule, which now directs most US federally funded research considerations:

- Minimize participant risks through sound research methodology [46.111a(1)]
- Risks appropriate to benefits [46.111a(2)]
- Equitable subject recruitment [46.111a(3)]
- Informed consent [46.111a(4) and (5)]
- Monitor data for participant safety [46.111a(6)]
- Appropriately protect privacy and confidentiality of

Manuscript received November 28, 2014; revised February 27, 2015. This report is based upon work supported by the National Science Foundation under Grant No. 1419055.Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Anthony Kelly is with the College of Education and Human Development at George Mason University, USA (e-mail: akelly1@gmu.edu).

Mika Seppala was with the Department of Mathematics at Florida State University, Tallahassee, FL 32306 USA (e-mail: mika.seppala@fsu.edu).

participants [46.111a(7)]

The Common Rule in the US guides the activity of institutional review boards (IRB) that oversee human protection issues. The issues facing IRB boards are never static, but remain under active review, especially in the light of internet-based research on human subjects (e.g., [3]). In the US, IRB oversight is linked to related regulations that impact how human subjects are treated.

FERPA. In addition to IRB concerns are those that relate to human subjects' protections and data privacy under the Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99), which protects the privacy of student education records. According to the FERPA websitehttp://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html:

"Generally, schools must have written permission from the parent or eligible student in order to release any information from a student's education record. However, FERPA allows schools to disclose those records, without consent, to the following parties or under the following conditions (34 CFR § 99.31):

- School officials with legitimate educational interest;
- Other schools to which a student is transferring;
- Specified officials for audit or evaluation purposes;
- Appropriate parties in connection with financial aid to a student;
- Organizations conducting certain studies for or on behalf of the school;
- Accrediting organizations;
- To comply with a judicial order or lawfully issued subpoena;
- Appropriate officials in cases of health and safety emergencies; and
- State and local authorities, within a juvenile justice system, pursuant to specific State law.
- Schools may disclose, without consent, "directory" information such as a student's name, address, telephone number, date and place of birth, honors and awards, and dates of attendance. However, schools must tell parents and eligible students about directory information and allow parents and eligible students a reasonable amount of time to request that the school not disclose directory information about them. Schools must notify parents and eligible students annually of their rights under FERPA. The actual means of notification (special letter, inclusion in a PTA bulletin, student handbook, or newspaper article) is left to the discretion of each school.

COPPA. Additionally, in the US, anyone conducting research on young children must follow the Children's Online Privacy Protection Act (COPPA). According to the Federal Trade Commission website (http://www.ftc.gov):

- …the Rule covers a child-directed site or service that integrates outside services, such as plug-ins or advertising networks, that collect personal information from its visitors…
- The definition of a *website or online service directed to children* is expanded to include plug-ins or ad networks that have actual knowledge that they are collecting personal information through a child-directed website or online service. In addition, in contrast to sites and services whose primary target audience is children, and who must presume all users are children, sites and services that target children only as a secondary audience or to a lesser degree may differentiate among users, and will be required to provide notice and obtain parental consent only for those users who identify themselves as being younger than 13.
- The definition of *personal information* now also includes geolocation information, as well as photos, videos, and audio files that contain a child's image or voice.
- The definition of *personal information* requiring parental notice and consent before collection now includes "persistent identifiers" that can be used to recognize users over time and across different websites or online services. However, no parental notice and consent is required when an operator collects a persistent identifier for the sole purpose of supporting the website or online service's internal operations, such as contextual advertising, frequency capping, legal compliance, site analysis, and network communications. Without parental consent, such information may never be used or disclosed to contact a specific individual, including through behavioral advertising, to amass a profile on a specific individual, or for any other purpose.
- The amended Final Rule revises the parental notice provisions to help ensure that operators' privacy policies, and the direct notices they must give parents before collecting children's personal information, are concise and timely.
- The amended Final Rule requires operators to take reasonable steps to make sure that children's personal information is released only to service providers and third parties that are capable of maintaining the confidentiality, security, and integrity of such information, and who assure that they will do so. The Rule also requires operators to retain children's personal information for only as long as is reasonably necessary, and to protect against unauthorized access or use while the information is being disposed of.

HIPPA. Traditionally, education research has not concerned itself with privacy issues related to the Health Insurance Portability and Accountability Act of 1996 (HIPAA).It is unclear if the collection of "quantified self" (biometric) data (e.g., the collection and use of "FitBit" data) for a nursing or health-related course may fall under this Act.

The interpretation and applicability of policies and regulations such as these will vary by country and jurisdiction. On the other hand, many countries have established the equivalent of research ethics boards not unlike the Intuitional Review Boards (IRB)in the US. Despite many years of applying ethical principles for research, IRB principles are not decided or clear in all cases [1], [3], [4]. Indeed, there is some variation in the review of ethical concerns by IRB panels [5], [6].

Importantly, social science researchers have raised questions about the applicability of *existing* clinical medical studies on non-clinical research [7]. The path forward for attending to student privacy and ethical treatment is less clear given the rapidly changing data infrastructure, and the regulatory environment.

III. Some Ethical Issues for Online Learning Research

A. Recruitment of Participants

Recruiting participants assumes the ability to uniquely identify each participant [8], [9]. What rules apply to participants in massive open online courses (MOOCs) or other learning platforms who may be using guest IDs?

Does recruitment of participants from one network imply recruitment from their other social networks? May a researcher explore data from un-recruited networks or generalize findings to multiple networks [10]?

What is the status of un-recruited (and perhaps unknown) students who become tied to the recruited participant as part of learning activities on the platform? Can these students' data be collected and used?

Can participants be recruited via avatars, screen names or other pseudonyms? If so, how would age and other descriptors be verified in these cases?

Since many participants will be members of social networks, are "Facebook friends" or linked active members in social networks recruited "by default"? What are the rights of existing, and newly-added linked network members?

If participants are members of known networks (e.g., around some theme controversial or otherwise), how should this knowledge impact recruitment solicitations?

B. Informed Consent

The centerpiece of ethical research practice is obtaining informed consent from a participant. For online learning, may informed consent be obtained from avatar, screen-name or pseudonyms? Is written consent required in all cases, and how should this be handled, digitally? For example, can an avatar sign a consent waiver?

Should informed consent be also obtained for associated, linked or networked people, especially if potential data may come from protected populations?

Is it sufficient for the purposes of informed consent if participants "click through" a link to a research activity after reading a disclaimer since this is common practice on many commercial websites?

C. Privacy

If commercial tools are being used by the researcher or by the participant, what are the implications for data collection for third party "terms and conditions" agreements already agreed to by the participant? [11]. For example, if a student has selected certain privacy settings for Facebook do these settings apply, downstream, to the researcher [12]? According to [12], one third of surveyed IRB panels ignored the privacy and security policies of commercial companies when reviewing subject protections.

What is the privacy status of avatars, pseudonyms, screen names, or other user IDs, especially if these may be changed by the participant or be adopted by other people? (see [13], [14]).

Even with the consent by a known participant, what protections can reasonably be provided by the researcher regarding de-anonymization of consented data, now or in the future? (see [15], [16]). Moreover, research activity on social networks (e.g., Facebook, Twitter feeds) may become part of the "scrapable data" collected by third parties via data-aggregation tools, including data on associates, friends, etc.

What rules guide research use these and of mobile data including geolocation, GIS identifiers, and IP addresses? Should these data and biometric and "quantified self" data included in protected personally identifiable information (PII)? [17]

D. Jurisdictional Issues

The European Parliament voted in 2013 to approve the *EU General Data Protection Regulation*.Once operative, this European Union regulation on privacy aims to establish a "right to erasure," and wishes to prohibit "profiling" of users. Profiling could include inferring characteristics of users including health conditions, or socioeconomic status. As e-learning becomes increasingly global, it is increasingly likely that human subjects from many countries will be included in research studies. What are jurisdictional issues for researchers, and the implications for research data, and the conclusions drawn from them if certain behaviors legal in one country are proscribed in another? For example, if US researchers have collected data on German students, and have used these data as part of central hypothesis testing in a study, what actions must they take or can they take if a sizable portion of the German sample assert rights "to be forgotten" or for data erasure? Must ongoing research be halted? Must published papers be withdrawn?

E. Ownership of Data

Who owns research data on subjects? This is not a trivial question if the data are used to generate profits for Universities or other research institutions, or researchers. Here we engage issues such as the data "flow" from one data service to another. For example, if a researcher uses a social network platform to observe user behavior, what is the relationship between the restrictions on research assumed in the agreement between the research and the IRB board, and the relationship between the user and the commercial company that owns the social network platform? Are the data and inferences drawn from the data owned by the user, or by the university, or by the commercial entity? What if the social network platform stores the data in the cloud in a different jurisdiction? Under what conditions may directly-collected research data, and aggregated "scrapable" data, on a social network site be used and sold to third parties?(see [18]).

IV. Conclusion

It is becoming increasingly clear that while massive data collection and data-mining inferences can add to the science of learning, the same advances can be used to track students' behaviors, identify them, and characterize not just their behaviors, but infer health and other conditions. Parents, students and policymakers are becoming aware of the possible violations of privacy. In a series of articles in a special issue, *Science* magazine opined on the *end of privacy* [19]. For the US, the regulatory picture is complex and in flux. This picture is made more complex by the rapidly changing

sources of data on research subjects, both online and via many sources of metadata. The picture is rendered even more complex by the fact that e-leaning platforms are becoming increasingly global so that participants may be subjected to an array of ethical protection boards, commercial law, and privacy regulations. Moreover, the data collected on learners and human subjects may be stored on servers that are in locations far from where the research is conducted and approved, and where the research subjects live and learn.

For those interested in cutting-edge research on the computer science of privacy, many leading scholars are supported by the US National Science Foundation. Of particular interest may be the National Science Foundation Secure and Trustworthy Cyberspace (SaTC) program. This program had its 2015 principal investigators' meeting, and discussed a number of issues related to privacy and security of data. A link to presentations may be found here: https://www.usenix.org/conference/satcpi15. Newly funded research at the National Science Foundation may be found by using the search engine at nsf.gov.

As developers of online learning opportunities build even more impressive opportunities for learning, the privacy risks grow apace. At this point, the social mores, guidelines, directives, regulations, and legislation have failed to keep pace with the explosion of activities that have become digital, and that leave a myriad of digital trails.

The ethics of researching online learning is still a matter of debate. If the promise of online and e-learning [20] is to be realized, the various associations that represent the interests of researchers must stay abreast of the developments in a variety of areas [21], [22]. What is most important is that those who develop the various technologies in education, those who employ them for learning, and those who use these platforms for research must, daily, earn the trust of the public that supports them [23].

REFERENCES

[1] J. Sims, "A brief review of the Belmont report," *Dimensions of Critical Care Nursing*, vol. 29, no. 4, pp. 173–174, 2010.

[2] R. V. Katz, S. S. Kegeles, N. R. Kressin *et al*., "The Tuskegee legacy project: Willingness of minorities to participate in biomedical research," *Journal of Health Care for the Poor and Underserved*, vol. 17, no. 4, pp. 698–715, 2006.

[3] P. Y. Mahon, "Internet research and ethics: Transformative issues in nursing education research," *Journal of Professional Nursing*, 2013.

[4] W. Banks and M. Eble, "Digital spaces, online environments, and human participant research: Interfacing with institutional review boards," in *Digital Writing Research: Technologies, Methodologies, and Ethical Issues*, H. McKee and D. DeVoss (eds.), Cresskill, NJ: Hampton Press, pp. 27–47, 2007.

[5] E. A. Buchanan and C. M. Ess, "Internet research ethics and the institutional review board: Current practices and issues," *SIGCAS Computers and Society*, vol. 39, pp. 43–49, 2009.

[6] L. Stark, *Behind Closed Doors: IRBs and the Making of Ethical Research*, Chicago, IL: University of Chicago Press, 2012.

[7] S. W. Glickman, S. Galhenage, L. McNair, Z. Barber, K. Patel, K. A. Schulman, and J. McHutchison, "The potential influence of internet-based social networking on the conduct of clinical research studies," *Journal of Empirical Research on Human Research Ethics*, vol. 7, no. 1, pp. 71–80, 2012.

[8] J. Aycock, E. Buchanan, S. Dexter, and D. Dittrich, "Human subjects, agents, or bots" in *Current Issues in Ethics and Computer Security Research*, G. Danezis, S. Dietrich, and K. Sako, (eds.), *FC 2011 Workshops, LNCS*, Springer, Heidelberg, pp. 138–145, 2012.

[9] C. Mendelson, "Recruiting participants for research from online communities," *Computers, Informatics, Nursing*, vol. 25, no. 6, pp. 317–323, 2007.

[10] J. Bonneau and S. Preibusch, "The privacy jungle: On the market for data protection in social networks," presented at the the Eighth Workshop on the Economics of Information Security, 2009.

[11] B. Gilbert, "Getting to conscionable: Negotiating virtual world's end user license agreements without getting externally regulated," *Journal of International Commercial Law and Technology*, vol. 4, no. 4, pp. 238–251, 2009.

[12] M. Taddicken, "Privacy, surveillance and self-disclosure in the social web: Exploring the user's perspective via focus groups," in F. Christian, B. Kees, A. Anders, and S. Marisol, *Internet and Surveillance. The Challenges of Web 2.0 and Social Media*, New York, pp. 255-272, 2012.

[13] E. A. Buchanan and E. E. Hvizdak, "Online survey tools: Ethical and methodological concerns of human research ethics committees," *Journal of Empirical Research on Human Research Ethics*, vol. 37, 2009.

[14] M. Madejski, M. Johnson, and S. Bellovin, "The failure of online social network privacy settings," *Columbia Research Report* (CUCS-010-11), pp. 1–20, 2011.

[15] A. Rosenberg, "Virtual world research ethics and the private/public distinction," *International Journal of Internet Research Ethics*, vol. 3, no. 1, pp. 23–37, 2010.

[16] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proc. the 29th IEEE Symposium on Security and Privacy*, Oakland, CA, May 2008, pp. 111–125.

[17] P. Schwartz and D. Solove, "The PII problem: Privacy and a new concept of personally identifiable information," *New York University Law Review*, vol. 86, p. 1814, 2011.

[18] P. Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization," *UCLA Law Review*, vol. 57, pp. 1701–1777, 2010.

[19] M. Enserink and G. Chin, "The end of privacy." *Science*, vol. 347, no. 6221, pp. 490-491.

[20] J. L Moore, C. Dickson-Deane, and K. Galyen, "E-learning, online learning, and distance learning environments: Are they the same?" *The Internet and Higher Education*, vol. 14, no. 2, pp. 129–135, 2011.

[21] P. Elias, "A European perspective on research and big data analysis," in J. Lane, V. Stodden, H. Nissenbaum, and S. Bender, eds., *Privacy, Big Data and the Public Good: Frameworks for Engagement*, New York: Cambridge University Press, pp. 173-191., 2014.

[22] J. Lane, V. Stodden, S. Bender, and H. Nissenbaum, *Privacy, Big Data, and the Public Good*, Cambridge: Cambridge University Press, 2014.

[23] H. Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Palo Alto, CA: Stanford University Press, 2010.

Anthony E. Kelly was born in the Republic of Ireland. He completed a bachelor's degree at St. Patrick's College in Dublin, 1979. He pursued a doctoral degree at Stanford University, Stanford, CA, US. He graduated with a doctorate in psychological studies in education with a doctoral minor in psychology. He is currently on leave from George Mason University in Fairfax VA, USA, and he is serving as a senior advisor at the US National Science Foundation in Arlington, VA. His research interests include research methodology, and public policy.

Mika Seppälä was born in Finland. He graduated from the University of Helsinki with a doctorate in 1978. He was a professor of mathematics at Florida State University in Tallahassee, Florida, USA. Dr. Seppälä's original mathematical interest was in the study of riemann surfaces, algebraic curves, and their moduli spaces. More recently Dr. Seppälä had developed learning analytics for massive open online courses like the ones offered by the myweps.com service.