

Creepy Analytics and Learner Data Rights

Scott Beattie

Central Queensland University

Carolyn Woodley

RMIT

Kay Souter

Deakin University

Enthusiasm for the potential of learning analytics and big data technologies must be tempered with caution for fundamental learner rights to that data and concern for the ways in which these re-shape the learning environment and the learner-teacher-university relationship. This paper argues that there is a legitimate distrust of 'creepy' analytics that misuse surveillance technologies and that a Charter of Learner Data Rights would be a strong foundation on which to build analytic technologies that are open, personalised, portable, adaptive and engaging for learners.

Keywords: Learning analytics, learner data rights, privacy, intellectual property, charter of rights

Introduction

The day after tomorrow...

Welcome to Veridian Dynamics Learning Analytics, where we know what students are thinking – before they even think it. You will be pleased with the wide arrays of big data we can service you with, as well as the laser like precision with which we can drill down into an individual student's learning behaviour and biometric data as their own device cameras are harnessed to monitor attention, heart rate and changes in expression.

Our range of GPS linked options will allow you to see where students are learning and who they are learning with. We can tap into social networks to build extensive databases of who your students are and who they know. With predictive analytics we can even profile students on their likely performance, letting you know who is going to fail before they have even begun.

Our data aggregator can keep extensive records which, with a simple tick on the Terms & Conditions box, lets us cross reference learner data with library records, online searches, book purchases and a range of purchasing behaviour databases. This is also excellent for plagiarism detection and enforcement of intellectual property rights along with recovery of usage fees.

And our Veridian pledge to you is that our databases are almost never* hacked by identity thieves, we almost never* transfer data to government security services and we almost never* expose students to dangerous individuals working in our company or yours (although we cannot guarantee yours, conditions apply).

* The term 'almost never' is subject to terms and conditions. Please download appendices 9 through 18

The risks of analytics and big data

The techno-utopian dream of big data is in constant peril of succumbing to pervasive surveillance and consequently perpetrating privacy intrusion, stalking, criminal conduct and other forms of 'creepy' behaviour. While the term 'creepy' may not be scientific, it is a very apt and concise description for the feeling of unease that some new technologies provoke in learners and teachers. The 'creepy treehouse' syndrome describes inauthentic learning spaces that try to beckon students into online spaces by pretending to be genuine social networks (McBride, 2008). Creepy analytics, then, carry with them a sometimes hard-to-define, unsettling undercurrent below vendor claims of greater engagement and better retention outcomes.

Recently, Facebook and a group of academics conducted research into 'emotional contagion' by covertly

manipulating the data of 689,003 users (Kramer, Guillory & Hancock, 2014). When revealed, this clandestine research was met with extensive criticism and was variously described as “‘scandalous’, ‘spooky’ and ‘disturbing’” (Booth, 2014). Representations of creepy analytics provokes similar sorts of embodied reactions; frequently, ‘creepy’ is used to depict examples of organisations overstepping “the creepy line” (Connolly, 2014). While using analytics in ‘research’ may be legal and, no doubt, in the case of Facebook is encapsulated in evolving terms and conditions, such use of user data raises potent ethical questions for academic researchers.

Challenges for users of big data (Boyd & Crawford, 2012) include changes to how knowledge is defined, questions about claims to any objectivity, problems of decontextualised data, problems of localised data, doubts about the ethical use of data and the danger of creating new digital divides. Current debates within the big data field are particularly acute with the use of learning data. To earn the trust and encourage the engagement of students, learner data systems need to be open (Siemens, Gasevic, Haythornthwaite, Dawson, Buckingham Shum, et al. 2011) rather than based on proprietary technologies, transparent, personalised, networked (Dawson, Bakharia, Lockyer & Heathcote, 2011), transportable (Buckingham Shum & Ferguson, 2012), adaptive and interactive. The potential of various technologies comes with many risks. Certainly, more pervasive surveillance, including biometric, geo-positioning, predictive and network mapping technologies, can easily step over the line between good teaching practice and creepiness.

Along with privacy fears, there are core questions in the analytics terrain about who owns individual learners' data: who has rights of access to it and how is conduct in relation to that data governed? This paper explores some of the risks of unbridled collection, access and interpretation of learner analytics and argues that a charter of learner data rights, agreed to by both public educators and private edutech firms, would provide a foundation of a relationship for future learning analytics to be designed for respectful and ethical learning environments.

There are dual dangers in proposing a learner data rights: naïve learners may agree to 'clickwrap' conditions without reading or understanding them and more security-savvy learners will avoid engaging in the learning environment altogether for fear of compromise.

Further, the portable nature of technologies means that learning analytics might extend far beyond the typical patterns of Learning Management System (LMS) use. As learners progress through life, they will generate an array of different data sets and preferences that could be of use to themselves as an evolving learner profile. This data has additional value: if learners choose to share their data with others, there is the potential for pattern recognition and learning recommendations drawn from that big data to create adaptive, interactive learning technologies (Office of Educational Technology, 2013).

The problem with creepiness

Many educators have an instinctive, gut reaction to new technologies, which may, in part, come from the shock of the new but, also, from the challenge technologies present to deeply felt, embodied pedagogical practice. For techno-enthusiasts, any concerns about new technologies can be easily dismissed as Ludditism. But there are real pedagogical concerns about technologies transforming relations between learner, teacher and institution. Particular concerns arise from intrusion into the private domain, from the use of biometric technologies, from the aggregation and sharing of data where consent is not first obtained or where it is done under 'click wrap' agreements that are usually technically legal but in reality seldom amount to full informed consent.

Creepy analytics create real pedagogical concerns with the potential to impact on student learning:

- The ease with which data can be collected encourages collection and correlation of more and more data regardless of need but rather on the chance that it might be useful. This makes individuals more exposed, for example, to identity theft from hackers or to stalking and malfeasance within an organisation. It is negligent to disregard these real risks.
- Where analytic systems are closed, learners (and institutions where the systems are corporate) do not know what is being collected, how it is being aggregated and who it is being shared with. Potentially, undisclosed marketing money is being made from sharing aggregated private data as well.
- Data can be reductive and can expose individuals to mass-profiling that puts them at risk. For example, library records on seditious or ‘dangerous’ books, especially when correlated with other data such as political party affiliations, family connections, racial and religious profiling, could produce risky decontextualised data.
- In the 21st century, data is a commodity and should be the property of those who generate it. Data should not be at risk of being sold off without owner’s permission or even knowledge. If we are able to generate

profiles of successful learners based on their research strategies, who should own this?

- Academic freedom is at risk. The learner is not able to autonomously negotiate their learning relationship with their educational institution. Rather, the learning relationship is dictated behind closed doors and under locked code by decision makers who are often corporate outsiders and located in a different jurisdiction.
- Analytics represents changes in the relationship of trust and respect with learners that have not been agreed upon. These unknown intruders in the relationship challenge learner and teacher identity in a way that is unsettling and “creepy”.

These pedagogical concerns tap into fundamental legal and ethical dangers, including the following:

- lack of openness and transparency in use of learning data shapes the learning contract, the institutional duty of care and statutory freedom of information rights;
- privacy rights of students and staff in the use of learning analytic data;
- the difficulty of securing cross-institutional agreement on the sharing and use of data across universities, including protocols for sharing data can effect individual rights without full consent and even with consent which may reflect a tangled net of different agreements;
- intellectual property rights and the fair dealing conditions under which data can be re-used and distributed;
- aggregation and disidentification issues;
- fairness and due process in the management of data systems including validation and deletion of erroneous data;
- learning analytics data ownership and stewardship, duty of care in the management and storage of analytic data;
- conditions of access by security and law enforcement bodies, including foreign powers especially where vendor servers are located in another jurisdiction;
- confidentiality, duty of care and other aspects governing the learner/institutional relationship; and
- ethical requirements for the use, analysis and reporting of student analytic data, including research ethics in publishing data derived from learning analytics.

Many of these concerns stem from the fundamental issue of learner data ownership, where learners must have rights to access and to make sense of data (Siemens & Smith, 2011) as the first step in understanding the risks and negotiating how that data is used. There is a marked difference of opinion between those who believe that learners own their learning data and those who believe that the learning institution, as owner of the technical infrastructure, also owns the data generated. Also worth considering is that, in this brave new world of big data, teachers also generate learner data in their own research and teaching activities. This data may be subject to the same legal and ethical risks which students face and which could have implications for industrial relations.

Towards Ethical Learning Analytics

For learning analytics to move beyond the hype and to earn acceptance from all relevant parties, there must be a convergence of interests focussed on *learner-centred* learning analytics. Learner-centred learning analytics provide a clear benefit to learners, are owned and co-produced by learners and are transparently and openly used in an environment of trust where data will not be used to disadvantage or negatively stereotype anyone. This convergence depends on all parties being able to participate in the debate and being informed of key legal and ethical issues so that they can critically evaluate various proposed technologies and frameworks.

The Office of Learning and Teaching (OLT) good practice guide, *Safeguarding Student Learning Engagement* (Nelson & Creagh, 2013), sets out a social justice framework for student engagement procedures which emphasises the centrality of the concepts of self-determination, rights, access, equity and participation. These concepts are central to a learning analytics framework that proceeds in good faith and with the support of all parties. With the complexity and sheer number of the legal and ethical issues, a Charter of Learner Data Rights would seem prudent. A Charter would acknowledge individual rights and to demonstrate a willingness to self-regulate before legal risks result in litigation. Self regulation does not, of course, replace other legal and ethical frameworks but works alongside and sets a higher standard for good practice than can be compelled through external regulation.

Existing privacy laws and the Information Privacy Principles (s14 of the Privacy Act 1988 (Cth) and also embodied in the state legislation) would form a strong backbone of a Charter of Learner Data Rights. Even so, a broad range of rights would have to be recognised. The Charter would need to establish core principles and make specific undertakings about the various uses of permissible and agreed learner analytics. Other codes of

practice could be integrated; for instance, private edutech companies are beginning to see the importance of this kind of self regulation. Learning technologies company Knewton has emphasised the importance of protecting student data (Ferreira, 2014) in ten principles which focus on use and management of data:

1. Student data belongs to the student.
2. Student data should never be sold or shared without explicit permission.
3. Student data should only be used to improve learning outcomes.
4. Student data should be easy to manage.
5. Student data should be very carefully protected.
6. Student data should be clear and comprehensible.
7. Students should be able to consolidate their data.
8. Student data should be portable.
9. Student data analysis should be completely stoppable — and recoverable.
10. Institutional IP should be protected (Ferreira, 2014).

Corporate self-governance such as this is a good idea, but a common public standard would be better, particularly as learners move between different contexts. Recently, the *Asilomar Convention for Learning Research in Higher Education* (2014) released a set of ethical guidelines for online learning, including the use of learning analytics. Most relevant to this discussion are the six principles to 'inform the collection, storage, distribution and analysis of data derived from human engagement with learning resources', based on United States legal and ethical frameworks including the 1973 Code of Fair Information Practices (Department of Health, Education & Welfare, 1973) and the Belmont Report of 1979 (Department of Health, Education & Welfare, 1979). These principles include:

- Respect for the rights and dignity of learners (responsible governance including transparent collection and sharing practices, consideration of appropriate consent)
- Beneficence (maximising benefits and minimising possible harms)
- Justice (benefits for all learners, specifically aimed at reducing inequalities between learners)
- Openness (research as a democratic public good, access to data and analytic techniques)
- The Humanity of Learning (digital technologies to enhance insight, judgment and discretion)
- Continuous Consideration (ethical practice as an ongoing, evolving practice and standards).

The *Asilomar Convention* forms an excellent foundation of ethical considerations but a charter should indicate legal aspects as well. The charter must also integrate the legal standards of multiple jurisdictions as far as this is possible. Education is regulated under Australia's federal structure, which structure renders it difficult to form a legally accurate summary of the various rights and duties under different sets of legislation and common law. However, a charter would be a code of practice and so would aim for concise expression of the common ground, and promise of adherence to a common set of values. As a voluntary public document, a charter of learner data rights would not necessarily have enforcement teeth of its own but it would certainly have a shaming effect should the code be breached. A charter would support a complaint to the ombuds office or to other areas of oversight and accountability. Educational institutions could not be *compelled* to be signatories to such a charter; however, refusal to do so would reflect poorly on their values and internal governance.

This discussion does leave open the difficult question of who should lead the way in the creation of a charter of learner data rights. The ethical and legal risks of learning analytics are present and imminent and should not wait for prolonged bureaucratic processes and inter-institutional negotiation. Two alternatives suggest themselves: either one institution leads the way to create a charter that others will follow; or a charter will be created by an independent body such as the OLT, a telecommunications ombuds office or a privacy commissioner.

The tripartite relationship between learner, teacher and educational institution exists in a rapidly evolving social, legal and technological environment. Any technological changes can (and do) rupture the fragile balance of respect and trust upon which this relationship is founded. Open, transparent and respectful use of learning analytics, operating within a framework such as a charter of learning data rights, presents an opportunity to reinforce this relationship while creepy analytics will only undermine its foundations.

References

Asilomar Convention for Learning Research in Higher Education (2014). *The Asilomar Convention for Learning Research in Higher Education*, <http://asilomar-highered.info/> [viewed 22 May 2014].

- Booth, R. (2014). Facebook admits manipulating users' emotions by modifying news feeds. *The Guardian*. 30 June.
- Boyd, D. & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication and Society*, (15)5: 662-679.
- Buckingham Shum, S., & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology and Society*, 15(3): 3-26.
- Connolly, B. (2014). Don't cross the 'creepy line' of data analytics. *CIO Australia*, http://www.cio.com.au/article/538947/don_t_cross_creepy_line_data_analytics/ [viewed 13 Mar 2014].
- Dawson, S., Bakharia, A., Lockyer, L. & Heathcote, E. (2011). *'Seeing' networks: visualising and evaluating student learning network*. Canberra: Australian Learning and Teaching Council.
- Department of Health, Education & Welfare. (1973) *Records, Computers and the Rights of Citizens*, <http://epic.org/privacy/hew1973report/> [viewed 20 March 2014].
- Department of Health, Education & Welfare. (1979) *The Belmont Report*. <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html> [viewed 24 March 2014].
- Ferreira, J. (2014). Protecting Student Data: 10 Guiding Principles, *The Knewton Blog*, <http://www.knewton.com/blog/ceo-jose-ferreira/data-principles/> [viewed 13 March 2014].
- Information Privacy Principles from the *Privacy Act 1988 (Cth)*, Retrieved from <http://www.oaic.gov.au/privacy/privacy-act/information-privacy-principles>
- Kramer, A. D. I., Guillory, J. E. & Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24): 8788-8790.
- McBride, M. (2008). Classroom 2.0: Avoiding the 'Creepy Treehouse', <http://melaniemcbride.net/2008/04/26/creepy-treehouse-v-digital-literacies/> [viewed 6 April 2014].
- Nelson, K. & Creagh, T. (2013). *A good practice guide: Safeguarding student learning engagement*. Brisbane, Australia: Queensland University of Technology.
- Office of Educational Technology (2013). *Expanding Evidence Approaches for Learning in a Digital World*. US Department of Education. Washington, D.C: U.S. Department of Education, <http://www.ed.gov/technology>
- Siemens, G. & Smith, V. (2011). Transforming Learning Through Analytics: A Conversation with George Siemens and Vernon Smith, video, *Educause Annual Conference*, October 21 2011, <http://educease.mediasite.com/Mediasite/Play/bdac04349c7246a6b6f949ded44397501d>
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. J. d. (2011). *Open Learning Analytics: an integrated and modularised platform*. United States Society for Learning Analytics Research, <http://solaresearch.org/OpenLearningAnalytics.pdf>

Please cite as: Beattie, S., Woodley, C., & Souter, K. (2014). Creepy Analytics and Learner Data Rights. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Rhetoric and Reality: Critical perspectives on educational technology. Proceedings ascilite Dunedin 2014* (pp. 421-425).

Note: All published papers are refereed, having undergone a double-blind peer-review process.



The author(s) assign a Creative Commons by attribution licence enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.