



STUDENT PERSPECTIVES ON THE USE OF THEIR DATA: BETWEEN INTRUSION, SURVEILLANCE AND CARE

*Sharon Slade, The Open University, United Kingdom,
Paul Prinsloo, University of South Africa, South Africa*

Introduction

The Open University (OU) is a large, open distance learning institution with more than 200,000 students. In common with many other higher education institutions (HEIs), the University is looking more closely at its use of learning analytics. Learning analytics has been defined as the collection and analysis of data generated during the learning process in order to improve the quality of learning and teaching (Siemens, Dawson, & Lynch, 2013). In the context of the Open University, learning analytics is the use of raw and analysed student data to, inter alia, proactively identify interventions which aim to support students in completing their study goals. Such interventions may be designed to support students as individuals as well as at a cohort level.

The use of a learning analytics approach to inform and provide direction to student support within the Open University is relatively new and, as such, existing policies relating and referring to potential uses of student data have required fresh scrutiny to ensure their continued relevance and completeness (Prinsloo & Slade, 2013). In response, The Open University made the decision to address a range of ethical issues relating to the University's approach to learning analytics via the implementation of new policy. In order to formulate a clear policy which reflected the University's mission and key principles, it was considered essential to consult with a wide range of stakeholders, including students.

Literature review

Amidst the hype surrounding the potential of Big Data and more specifically the use of student data in learning analytics (Booth, 2012; Wagner & Ice, 2012), there are increasing concerns regarding the ethical implications of the harvesting, analysis, use and storage of student data (Prinsloo & Slade, 2013). Central to the general concerns regarding the protecting of privacy and informed consent, is the notion of "privacy self-management" which has its origins in the Fair Information Practice Principles (1973) which covers, amidst other issues, "individuals' rights to be notified of the collection and use of personal data; the right to prevent personal data from being used for new purposes without consent; the right to correct

or amend one's records, and the responsibilities of the holders of data to prevent its misuse" (Solove, 2013, p.1882).

Most of the current strategies regarding the harvesting, analysis, use and storage of student data focuses on issues surrounding informing students of the harvesting and use of their data, but as Solove (2013) and others indicate, most of these initiatives to inform individuals don't work because of the fact that

"(1) people do not read privacy policies; (2) if people read them, they do not understand them; (3) if people read and understand them, they often lack enough background knowledge to make an informed choice; and (4) if people read them, understand them, and can make an informed choice, their choice might be skewed by various decision-making difficulties" (Solove, 2013, p.1888).

Although many HEIs have adopted learning analytics to some extent, there is little formal evidence that students are aware or explicitly consulted on the broader uses of their data beyond research. Despite various claims regarding the success of learning analytics to improve student success and retention (e.g., Arnold, 2010; Clow 2013), Watters (2013) warns that "the claims about big data and education are incredibly bold, and as of yet, mostly unproven" (par. 17).

At Purdue University, students can access a video which explains how their data is used to produce predictions of their success and alerts them to potential progress problems using the Course Signals tool (Pistilli et al., 2012). The University of Maryland (UMBC) introduced a Check My Activity tool which gives students more control of their own data by allowing them to compare their online activity and grades to those of other students. UMBC has promoted awareness of this tool and its purpose. It plans to track which students engage with this tool (but provide an opt out facility for students who don't want their usage to be tracked) (Fritz, 2010).

More generally, there is broad discussion on the issue of transparency and the possibility – or even the desirability – of an opt out option, but little to suggest that HEIs consult directly with, or inform students explicitly regarding, the ways in which their data is used. In the light of the literature that indicates that we need to move "beyond privacy self-management", we should perhaps rethink issues such as consent and the unequal power-relationship between the institution and students, the advantages of opting in rather than opting out, addressing privacy's timing and focus and the codification of privacy norms and developing substantive rules for data collection (Solove, 2013).

Research design and methodology

This study takes a directed content analysis approach. The “goal of a directed approach to content analysis is to validate or extend conceptually a theoretical framework or theory” (Hsieh & Shannon, 2005, p. 1281). Directed content analysis is therefore much more structured than conventional approaches to content analysis. Using a deductive, directed content analysis approach entails identifying key concepts of variables as initial coding categories, defined by theoretical frameworks and published research (Hsieh & Shannon, 2005).

The purpose of the study was to inform the development of new policy relating to how student data is used to shape student support by sharing an early draft and providing a number of structured questions for consideration. Data was collected over a period of 3 weeks in 2014 from the University Students’ Consultative Forum. The role of the forum is to enable students to consider and discuss matters affecting study at the OU and potential changes to University strategy and policy. It comprises 50 volunteered members who each participate for a minimum period of a year, appointed from the following categories:

- A core of representatives from the Open University Student Association, to include three central representatives, one from each of the four UK Nations, and two international students;
- One student from each of the curriculum-based consultation forums (17 in total) and one student from the University postgraduate research students consultative forum; and
- Student representatives on the various committees that make up the University governance structure.

Student representatives are asked to constructively contribute their personal views (as individual students rather than as formal representatives of other groups of students) to separate online discussions on matters affecting study and the student experience. The topic forums are limited to forum members for discussion purposes, but are open to all staff and students to read.

Details of the discussion questions

In preparation for the discussion, all forum members were given access to the draft policy which outlined the context and background to the University’s historical data collection, potential use of learning analytics, definitions of activity and data types which would be in or out of scope, and the set of eight overarching principles which would form the basis for the University’s approach to making more effective use of student data. Nine questions aimed at exploring both student awareness of the issues associated with a learning analytics approach and their reactions to those issues were posted online in a phased manner to the University Students Consultative Forum. One author had access to the forum as an invited moderator to

support the discussion and to provide clarification where needed. In the review of the posts, moderator posts were not considered.

Students were asked to explore issues around keeping their information up to date, transparency issues (why the University collects data and how it is used), to discuss their experiences of receiving student support messages during their studies, and to consider concerns regarding data collection and storage. The questions posted over the period are given below:

1. Do you regularly keep your StudentHome profile and other information such as study goals up to date? Is there anything the OU could do to make it easier to do this?
2. How often should the OU give you opportunities to check and update your data, and give consent to its use? What would be the most effective way of doing this?
3. Do you think the OU makes clear enough how and why it collects and analyses data? How do you think the OU should communicate its approach to students in the future?
4. Can you think of occasions when the OU has actively used data it has built up about you to offer you support tailored to your needs? Have these initiatives effectively used the information the OU holds about you?
5. Have you ever been offered support that you felt has not been based on relevant, up-to-date and accurate information the University holds about you?
6. Have you any other concerns about data collection, storage, updating and that you think the OU should address?
7. Do you think there is any information that the OU doesn't collect or use at present which it should do in the future?
8. Is there any information which the OU currently collects that you think is not relevant to supporting you as a learner?
9. Is there anything else you would like us to consider?

Limitations of this study

The forum is a recognised means of allowing students views to be reflected, and, as such, consults on matters relating to policy affecting students. It would be difficult to argue that the views of forum members can be accepted as representative of the whole of the OU student body. Their views can, at most, represent an initial basis for further research.

Analysis and Findings

Over the 3 week period, there were over 300 posts. 35 forum members made at least one post and six students made 16 or more posts. The questions and resulting discussions fell into four broad categories: keeping student information up to date; transparency issues, discussion of student support experiences; and data collection, storage and analysis storage. A summary of the discussion is captured below.

The need for accurate information and a shared responsibility

The term ‘student profile’ caused some initial confusion. The discussion was intended to explore how students felt about updating information initially collected at the point of registration - mostly demographic data around ethnicity, disability, academic history etc, but including other information aimed to help provide relevant support, for example, study goals, reasons for study, etc. This information can occasionally become outdated, for example, if a student’s financial circumstances or geographic location change, but several students first interpreted this as referring to their visible forum profile (their photo or other picture and released contact details). Once clarified, it was agreed that the process of updating held demographic information was straightforward, although some felt that there needed to be clear, persuasive reasons given to keep the profile up to date. The purpose of collecting some data was not always understood. For example, study goals were felt to be liable to change, but if students were unaware of the use that is made of that information, there would be less of a driver to keep it up date.

“I have not yet seen any persuasive argument for reviewing my profile on any basis, regular or otherwise.”

Many students felt that the collection of this ‘core’ information was both irrelevant and intrusive, and resented being prompted – often several times – to update it and fill any gaps in University records.

“I do not approve of the OU holding personal details on file about me other than those very specifically related to my studies. Some questions I view as impertinent and intrusive. Is the University hell-bent on gaining a reputation as an intrusive busy body?”

This view was expressed by other students, suggesting that most were unaware that the Open University, like other HEIs, routinely collects a wide range of data for regulatory and reporting purposes.

Generally, there was no single consensus on any of the issues flagged here – rather than developing an understanding that students have a responsibility to ensure that their personal information is accurate and relevant, the discussion appeared to provoke further concerns and bring to the surface unease and irritation. In developing the idea of how often students might be prompted to review their personal information, many felt that students could do this at any time, others felt that an annual reminder to check (or at module registration) would be sufficient. The discussion began to unearth concerns about the implications of registration itself

“and any registration such not depend on giving consent to be part of – choose your words here [tailored support; a research project; marketing data; alumni communications].”

One contributor felt more comfortable with broader uses of his data, posting

“it should be just an advisory indicating that it would be useful to update your profile both from the student’s own perspective and to aid the OU in its analytical exercises for the benefit of the university as an institution and to future students. Perhaps some sort of opt out clause for any data that some might have objections to, although I can’t imagine there would be anything that controversial that might produce flag waving student protests of epic proportions.”

Transparency of purpose

The issue of purpose perhaps raised the most emotive discussion over the period, generating more posts than any other topic. Students were clearly concerned that any data collection would be a new activity (*“didn’t know they used the information for stuff, so yes, I would appreciate knowing what it’s being used for”*), and that as a result, actions might be applied to them (*“some students have been appalled that an educational establishment has been collecting data that they didn’t expect. For example, monitoring their use of websites and how far they have progressed through a module.”*) or they might be prevented from making particular study choices. The general view was that more could be done to make clear what data is being collected, how it is being collected, where it is being collected from, the uses for which it being collected and who will have access.

Although contributors did recognise the positive intentions associated with a learning analytics approach, there were some murmurings of disquiet, perhaps best summarised by this student:

“There’s a huge difference IMO between anonymised data to observe/monitor large scale trends and the “snooping” variety of data collection tracking the individual. I’m happy for any of my data to be used in the former; with the latter I would be uncomfortable about the prospect that it would be used to label and categorise students in an unhelpful or intrusive way”.

There were a number of suggestions for communicating the approach to students in the future, for example

- stating exactly how information is used, with links to the detail;
- providing a basic summary of the key points on the student’s home page,
- communicating the approach at the point that a student is about to supply any data that is to be used;
- providing a fairly inclusive set of examples of what information is gathered and how it may be used.

Experiences of student support

At this stage, contributors were more aware of the background to the discussion (the wish to make greater use of student data to tailor student support) and many had begun to voice concerns around how such an approach might lead to assumptions and generalisations. Against this background then it was a little surprising to have a largely consensus view that their experiences of student support to date did not appear to have been based on relevant (to them as individuals), up to date and accurate information. Indeed, there was a clear view that, as a result of generalisation, the volume of emails received from the University was excessive, with the result that potentially important (to them as individuals) messages could be overlooked, for example,

“I get strange emails from time to time that are just not targeted at me (I tend to hit delete fast now). I concluded that the OU doesn’t make best use of data on what modules people have done/ what quals they already have.” and “The problem with an apparent blunderbuss approach is that it devalues the credibility of OU postings, so that any useful information is quite likely to end up under the delete key.” One student did spot the conflict here with other discussions relating to intrusiveness by posting: “Difficult for the University though to flag issues like this to students without holding data about what we do/how well we do/whether we use the forums/need advice...”

This small post generated lots of useful discussion about how data could and perhaps should be used to provide a more personalised and relevant support service, with students suggesting that a learning analytics approach applied in conjunction with support delivered by a personal tutor might ameliorate the risks of labelling students incorrectly. Others felt that the involvement of tutors could effectively prejudice the tutor:student relationship by impacting on the tutor’s expectations of that student. Another set of students felt that if the analysis of their data resulted in a ‘false positive’ identification, the risk of mislabelling could be managed if the consequence were limited to the offer of a service (which could be declined) rather than the removal of study options.

Data collection, storage and analysis

Views around the issues of data collection and storage were fairly non-contentious. Generally contributors expressed similar views which may be best summarised below:

- It is important to have a clear purpose for data collection and to communicate that purpose effectively ; to explain what data will/won’t be used for, and who can see it (e.g. on each student, in aggregate).
- A set of frequently asked questions developed for staff dealing with declaring personal information around diversity could usefully be replicated for students.
- There should be transparent policies about how long data can be held for and what the process is for handling requests for deletion of data.

- Data should only be shared on a ‘need to know basis’ – particularly where it is personal/sensitive.
- There should be strong and transparent governance in this area with a focus on ethics.
- Data handling protocols are important and should be enforced effectively.
- There should be periodic data audits.
- There should be an up-to-date data dictionary.
- It is important to address any concerns about the sharing of information with other organisations or the processing of information by other organisations.

The issue of analysis of that data caused more interesting discussion though with students flagging the differences between raw data and ‘derived information’. This theme cropped up in many of the separate discussions with concerns flagged about the reliability of the models used (“*people simply cannot draw the conclusions that they want to on the basis of a data pattern*”) and the ways in which model results might be employed (“*I have a concern that increased data-richness resulting in over-reliance on data and ‘computer says no’ responses. Catering for the individual is what’s needed. If data collection is used to help appropriate questions to be asked, fine - if it’s providing answers, very much not so.*”). Several students also flagged the need for staff involved in data analysis and in the delivery of intervention based on that analysis to be well-informed and appropriately trained.

Discussion

The range of issues flagged in direct response to the questions posted has provided useful additional understanding of the student perspective. In addition, the discussions have occasionally touched on aspects of the application of learning analytics within higher education that were not explicitly sought. The two most prominent topics of debate centred around third party data sources and the issue of informed consent/opt out.

Students were quick to flag the dangers of data protection and privacy in relation to having their data passed on – e.g., where a third party undertakes a service on behalf of the University. These issues were assumed to be neatly dealt with by existing policy. However, there was also a view expressed that the University should not attempt to draw in information from third party sites for its *own* purposes. One post stated

“I don’t object to somebody at the OU who I have seen fit to add to a Facebook group commenting. that’s why they have added to a group. But I don’t add The OU as an entity to a group. And I don’t expect it to go wandering about the web picking up snippets about me and feeding that in as data to be used in an analytical programme.”

This student felt that such information could be easily misconstrued and would overstep the boundaries of acceptable permissions.

The most dominant issue raised across all of the questions posted though concerned the need for consideration of informed consent and/or opt out. This was flagged several times with students stating that

“I think an opt-out option is essential for students who do not want to share data for whatever reason. No one should feel compelled to provide data if they don’t want to and they should be able to keep their reasons for this, which may be very personal, private.”

and

“Basically informed consent should be required. A right to refuse without compromising study ought to be built in.”

Perhaps the view was expressed best by one post which observed the apparent correlation between certain study behaviours (the behaviour cited was lack of online engagement in the early stages of study) and success in a module. The author also noted the argument for a duty of care to advise people against making a potentially costly mistake by continuing on a course they might not complete. S/he concluded this by stating *“But it is ultimately their choice.”*

Interestingly, the Open University has approved the policy which will establish its position on the ethical use of learning analytics, but has not implemented the provision of an opt out clause. The background to this is complex reflecting the need to fully explore both the practical issues associated with enabling full (or partial) opt out, but most importantly to establish where the duty of care primarily lies. It is this latter issue that has led to a further consultation to establish what lies at the heart of supporting its students: to assume a moral responsibility for employing information which aims to provide more effective and relevant support for all students, or to recognize students as informed individuals with the right to choose not to receive targeted intervention and support based on their own information.

Conclusion

The use of a forum to gather representative student views to feed into the development of policy covering a learning analytics approach to student support proved to be hugely useful. The range and complexity of many issues flagged has helped to inform and more clearly define the policy document and will feed into the ways in which communication of both the policy and the implementation of learning analytics will be rolled out across the Open University.

The direct involvement of the student voice in shaping a policy dealing with the ethics of learning analytics has offered a unique insight into the ways in which students regard their data – as a valuable entity to be carefully protected and even more carefully applied. In progressing the development of learning analytics in higher education, it is crucial to explicitly address the benefits and potential pitfalls of some an approach from the perspectives of all key stakeholders. This study has offered an opportunity to explore how students might react to

increasing uses of their personal and study data, and to facilitate a more considered and informed response.

References

1. Arnold, K. (2010). Signals: Applying academic analytics. In *EDUCAUSEreview*. Retrieved from <http://www.educause.edu/ero/article/signals-applying-academic-analytics>
2. Booth, M. (2012). Learning analytics: the new black. In *EDUCAUSEreview*, [online]. Retrieved from <http://www.educause.edu/ero/article/learning-analytics-new-black>
3. Clow, D. (2013). Looking harder at Course Signals. In *Doug Clow's Imaginatively-Titled Blog: New Technology in Higher Education*. November 13, 2013. <http://dougclow.org/2013/11/>
4. Fritz, J. (2010). Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. In *The Internet and Higher Education*, 14(2), (pp. 89-97). <http://www.sciencedirect.com/science/article/pii/S109675161000062X>, DOI: 10.1016/j.iheduc.2010.07.007
5. Hsieh, H-F. and Shannon, S.E. (2005). Three approaches to qualitative content analysis. In *Qualitative Health Research*, 15(9), (pp.1277 – 1288) 15(9), DOI: 10.1177/1049732305276687.
6. Pistilli, M.D.; Arnold, K.; Bethune, M. (2012). Signals: Using Academic Analytics to Promote Student Success. In *EDUCAUSE Review Online*. <http://www.educause.edu/ero/article/signals-using-academic-analytics-promote-student-success>
7. Prinsloo, P. And Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. In *Learning Analytics and Knowledge 2013 – Leuven, Belgium*, 8-12 April 2013.
8. Siemens, G.; Dawson, S.; Lynch, G. (2013). Improving the Quality and Productivity of the Higher Education Sector: policy and Strategy for Systems-Level Deployment of Learning Analytics. In *Society for Learning Analytics Research* (p. X), December 2013.
9. Solove, D. (2013). Introduction: Privacy self-management and the consent dilemma. In *126 Harvard Law Review*, 1880, (p. 1882), <http://ssrn.com/abstract=2171018>
10. Wagner, E. and Ice, P. (2012). Data changes everything: delivering on the promise of learning analytics in higher education In *EDUCAUSEreview*.
11. Watters, A. (2013) Student data is the new oil: MOOCs, metaphor and money. In *Hack Education blog*, 17 Oct. 2013. <http://hackeducation.com/2013/10/17/student-data-is-the-new-oil/>