Code of practice for learning analytics

A literature review of the ethical and legal issues

Niall Sclater

Table of Contents

1: Introduction

Consultation by Jisc with representatives from the UK higher and further education sectors has identified a requirement for a code of practice for learning analytics. The complex ethical and legal issues around the collection and processing of student data to enhance educational processes are seen by universities and colleges as barriers to the development and adoption of learning analytics (Sclater 2014a). Consequently a literature review was commissioned by Jisc to document the main challenges likely to be faced by institutions and to provide the background for a sector-wide code of practice. This review incorporates many relevant issues raised in the literature and the legislation though it is not intended to provide definitive legal advice for institutions. It draws from 86 publications, more than a third of them published within the last year, from a wide range of sources including:

- The literature around learning analytics which makes explicit reference to legal and ethical issues
- Articles and blogs around the ethical and legal issues of big data
- A few papers which concentrate specifically on privacy
- Relevant legislation, in particular, the European Data Protection Directive 1995 and the UK Data Protection Act 1998
- Related codes of practice from education and industry

Expressing issues as questions can be a useful way of making some of the complexities more concrete. 93 questions have been extracted from the literature and are incorporated in the relevant sections of the review. They arise mainly in the areas of awareness, consent, ownership, control, the obligation to act, interventions, triage and the impacts on student behaviour. These headings, highlighted in the word cloud below, give an instant flavour of the main ethical, procedural and legal concerns around the implementation of learning analytics being raised by researchers and practitioners.

Triage
privacy
Stewardardship-of-data
transparency around algorithms and metrics
impacts-on-student-behaviour
interventions
outsourcing
consent
ownership-and-control
opting-out
publicly-available-data
student-responsibility
personal-circumstances
interpretation
awareness
obligation-to-act

In "*22. Principles for a code of practice*" sixteen codes of practice or lists of ethical principles from related fields have been summarised. A further word cloud created from the key principles points to some of the solutions others have come up with. Foremost among them are: transparency, clarity, respect, user control, consent, access and accountability.

This report presents a comprehensive review of the ethical and legal issues currently being reported on and likely to be encountered by institutions in their deployment of learning analytics. It is intended to form the groundwork for a consultation by Jisc with representatives from across further and higher education to develop a more concise code of practice. This will provide clear guidance to institutions and reassurance to students and staff that developments in learning analytics are being undertaken legally and ethically, primarily for the benefit of learners.

2: Rationale for a code of practice

Current legal and ethical guidelines have not caught up with innovations in the identification of patterns and new knowledge emerging from the vast datasets being accumulated by institutions. "We've driven off the existing legal and ethical maps", according to King & Richards (2014). A code of ethics is essential as the law alone cannot deal with the many different scenarios that will arise. We are in a critical window: whatever gets established today will be "sticky" and will affect public notions of what is acceptable for many years to come. If we fail to assert values such as privacy, transparency and free choice, society will abandon these in favour of technical innovation and commercial pressures.

Institutional power is increasing at the expense of individual identity, which can increasingly be shaped by applying analytics to our interactions. We need therefore to establish principles and best practices to guide the various stakeholders and encourage the ethical use of data rather than forcing users to share their personal data with little in return (Richards & King 2014). The challenges created to privacy by big data arise from the sheer amount of data that is collected and the efficient ways in which it can be analysed, enabling far more to be learnt about people than was ever anticipated. Technology has to be accompanied by policy in order for privacy to be protected (President's Council of Advisors on Science and Technology - PCAST 2014).

In a review of learning analytics literature Ferguson (2012) points out that there is a pressing need for a detailed ethical framework which helps institutions make decisions regarding the ownership and stewardship of learners' data. This should describe the rights and responsibilities stakeholders have in relation to the data, how researchers can obtain informed consent to use it, and how students can opt out of data collection or have their records removed. Ferguson also suggests that researchers in the field could include sections on ethics within their publications.

Pardo and Siemens (2014) find that institutions already struggle to define privacy policies in areas other than learning analytics which present new issues for ethics committees. The prominence of the Course Signals initiative at Purdue University and other innovations in analytics means that institutions need to consider the ethical issues arising from the data they hold and the possible uses they could put it to for enhancing retention and academic success (Willis, Campbell & Pistilli 2013). Berg (2014) argues that without a code of ethics and agreed practices institutions may act in an ad-hoc way thus reducing consistency and fairness for students. However, he believes that as learning analytics is still developing there is time to develop the code. It is likely that senior management will have a different perspective to teachers but which of these should be able to decide what is done with the analytics? Berg suggests that a code of practice would be the arbitrator in this case.

Dealing with the ethical challenges relating to information technology is not simply a "good thing to do" but also helps organisations to develop an ethical environment where bad things are much less likely to occur. Negative publicity about the use of IT almost always results from failures to deal with ethical issues. Business-focussed staff may consider IT to be ethically neutral but they are not necessarily familiar with the many choices IT staff are required to make in the design and deployment of systems which have consequences for individuals.

This would seem to be particularly pertinent in learning analytics where the nature of the algorithms and how the results are presented and acted upon could have a significant impact on a student's academic success. It is certainly the case that when new and unfamiliar systems are introduced there may be pressures to cut corners and disregard ethical dilemmas which arise for which no advice has been provided. Users may also feel that if the data is held remotely in the cloud, ethical issues are more distant and can safely be ignored (Duquenoy, Dando & Harris 2010).

Without addressing the ethical issues there may be a backlash from users who feel their privacy is endangered and therefore the development of learning analytics may be held up (Greller & Draschler 2012; Siemens 2012). Two recent examples are worth noting. Firstly, an open-source product called "inBloom" was developed with $100m funding from the Gates and Carnegie Foundations. The aim was to store data in a common format which gave schools control over the data they collected and how it was used and shared. However, communications were mishandled and parents not properly consulted or informed. Families along with privacy advocates forced the closure of the programme, claiming that the system would have contained highly sensitive data such as disability status, and that there was no ability to opt out. They were also concerned that data on their children could fall into the wrong hands (K.N.C. 2014).

The second example is the “mood experiment” carried out by Facebook which placed positive and negative items and images in the timelines of 700,000 users to find out if they could manipulate users’ moods. The backlash from users and the media was huge and resulted in new guidelines to the company’s researchers. Mike Shroepfer, the Chief Technology Officer, admitted that they should have done things differently: they should have considered non-experimental ways to carry out the research, involved a wider and more senior group of people on the review panel, and communicated better why and how they were doing it. Review processes at Facebook for studies involving deeply personal matter such as emotions will therefore be enhanced (Shroepfer 2014). It is likely that the actions taken to *manipulate* emotions rather than merely analyse them is what sparked such outrage among users.

A survey of 144 data scientists at an event in Boston in 2014 found that 42% of them agreed that an industry standard ethical framework for collecting and using data should be available (Bertolucci 2014). 43% said that ethics played “a big part” in their research. 47% of the data scientists in the survey thought the Facebook study was unethical and 40% said they did not know. While research in science and health is already subject to strict ethical procedures there are few such guidelines in the technology industry. Analytics regarded as “inappropriate, creepy, intrusive or rude” are likely to adversely affect trust in an organisation and restrict their ability to develop their data processing capabilities (Schwartz 2010).

There are certainly lessons here for researchers and practitioners of learning analytics who wish to avoid accusations of unnecessary “big brother” type surveillance. Transparency and a recognition of potential unease from learners and educators may be helpful in preventing a backlash (Siemens 2012). Ellis (2013) believes that clearly communicated intentions such as improving learning rather than policing poor teachers are necessary. Certainly it is in the interests of students, staff and institutions that the uses to which learning analytics will be put is explained as clearly as possible. Slade & Prinsloo (2013) consider the management of student perceptions of learning analytics to be critical to its successful adoption. They would like to see learners collaborate with institutions to provide data and access to it so that they can be the primary beneficiaries of learning analytics, as well as helping the institutions.

Other industries have realised the critical importance of maintaining confidence in the organisation and its processes. ESOMAR (2011) suggests that market research depends for its success on public confidence and that researchers should avoid practices that potentially undermine it. Finally, the Information Commissioner’s Office

(ICO 2010) lists the benefits of following its code of practice for personal information online, all of which are of direct relevance for learning analytics:

- Greater trust and a better relationship with the people you collect information about
- Reduced reputational risk caused by the inappropriate or insecure processing of personal data
- Better take-up of online services, meaning economic savings and greater convenience for customers
- Minimised risk of breaches and consequent enforcement action by the Information Commissioner or other regulators
- Gaining a competitive advantage by reassuring the people you deal with that you take their privacy seriously
- Increasing people's confidence to provide more valuable information, because they are reassured that it will be used properly and kept securely, and
- Reduced risk of questions, complaints and disputes about your use of personal data

3: Ethical approaches

Pardo & Siemens (2014) define ethics in the digital context as "the systematization of correct and incorrect behaviour in virtual spaces according to all stakeholders". Surprisingly little of the literature around ethics for learning analytics however refers to the underlying philosophy and ethical theory. While useful in attempting to understand the issues, applied ethics is not a familiar area for most practitioners. Harris et al. (2008), in an attempt to help ICT staff assess ethical issues, suggest using four questions put by Mason et al. (1995), based on the two main ethical traditions of *teleology* and *deontology*:

- Who is the agent? (including their motives, interests and character)
- What action was taken or is being contemplated?
- What are the results or consequences of that action?
- Are those results fair or just?

One author who has examined the application of ethics to learning analytics is Willis (2014) who believes that learning analytics has a basis in *utilitarianism*, which argues that action should be based on what does the most good for the most people. As a learner, allowing your data to be combined with that of others, potentially to help them, would fit in with this stance. *Moral utopianism* similarly suggests that people act in a way that betters others. Applying this framework to learning analytics, the technologies would ensure meaningful interventions to help students learn and develop themselves. Data would be kept securely and predictions would be accurate.

In another paper, Willis, Campbell & Pistilli (2013) outline some ethical principles which can be applied to learning analytics, including:

- **Aristotle's *Golden Mean*: "The moral virtue is the appropriate location between two extremes"**. Analytics can be used to identify extremes of student behaviours e.g. under-confidence and over-confidence, and thus help move towards a more moderate position
- **Immanuel Kant's *Categorical Imperative*: "Act on the maxim that you wish to have become a universal law"**. Institutions have a duty to act if predictive algorithms are demonstrated to be effective in producing actionable insights
- **John Stuart Mill's *Principle of Utility*: "Seek the greatest happiness for the greatest number"**. Analytics can be used to improve the success of large numbers of students

Ethical decision making brings together people's world views (including their epistemology and values), their individual positions on methodology, the academic and political environment, and the assumptions of individual disciplines (AoIR 2012). However, ethical systems can be lofty and vague; they become more meaningful when given a set of principles (Willis, Campbell & Pistilli 2013). Many of the fundamental principles behind codes of ethics come from documents such as the UN Declaration of Human Rights and the Nuremberg Code. These include human dignity, autonomy, protection, safety, and the minimisation of harm (AoIR 2012). The avoidance of harm is manifested in the well-known, if slightly unconvincing, "Don't be evil" slogan at the start of Google's (2012) Code of Conduct. This principle is particularly relevant to learning analytics which has the potential to adversely affect students' academic success if used improperly.

Slade & Prinsloo (2013) point out that the origins of learning analytics in various research areas bring a number of different but overlapping ethical perspectives in relation to areas such as data ownership and privacy. They themselves take a *socio-critical* standpoint, discussing the role of power relations between students, their institutions and other stakeholders such as funding bodies. They position learning analytics as a "transparent moral practice", viewing students as participants in the process. They classify the ethical issues of learning analytics in three overlapping categories:

- The location and interpretation of data
- Informed consent, privacy and the de-identification of data, and
- The management, classification and storage of data

The imbalance of power between the institution, staff and students has important ethical implications for learning analytics. Other professions are aware of this inequality and how it can affect individuals: the larger the differential in power between the professional and the subject, the heavier is the responsibility of the professional (European Federation of Psychologists' Association, EFPA 2005). Relevant principles from the literature and many guidelines from related fields which might be appropriate for a code of practice for learning analytics are included in "*22 Principles for a code of practice*".

4. Legal context

Developments in the law are slow and cannot match the speed of innovation. Willis (2014) argues that this is a good thing as otherwise technological development would be impeded. Principled reflection however should attempt to match the speed of innovation. He calls for companies which are developing new learning analytics systems and institutions evaluating the products to put discussions on ethics at the forefront. Despite a lack of case law to provide the context for all aspects of learning analytics (Kay, Korn & Oppenheim 2012) there are many existing legal restrictions on the collection and processing of data in particular which need to be considered. Learning analytics is likely to be carried out alongside an institution's other data processing activities so many of the existing policies and processes for legal compliance and managing risk should already be in place (ICO 2014).

Rules about the collection and processing of personal information in different jurisdictions reflect varying legal, social and cultural values (Schwartz 2010). In Europe, the right to privacy is recognised in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms. This is one of the principles behind the Data Protection Directive (European Commission 1995) (DPD), which is enacted in the UK through the Data Protection Act (UK Government 1998) (DPA). Together these form the key current legislation of relevance to the implementation of learning analytics. The DPD and the DPA provide the bare minimal requirements; guidance on how to implement the DPA in practice is provided in ICO (2010).

Other Acts which may need to be considered for the application of learning analytics in the UK are the:

- Equality Act 2010 and the Disability Discrimination Act 2004 (which still applies in Northern Ireland) - ensuring that analytics does not disadvantage any particular group e.g. tools and dashboards should be fully accessible to students with disabilities
- Freedom of Information Act 2000 (and equivalent Scottish legislation in 2002) - where the confidentiality of students should be preserved in any request for information

The DPD and the DPA specify the responsibilities of *data controllers* and *data processors*, and the rights of *data subjects*. In the case of learning analytics the data controller is likely to be the educational institution, a data processor may for example be an organisation which is hosting the virtual learning environment or learning analytics system, and students will be the primary data subjects. The DPA is built around eight key principles from the Directive. In summary, personal data must be:

- processed fairly and lawfully
- obtained only for specific lawful purposes
- adequate, relevant and not excessive for those purposes
- kept accurate and up to date
- kept for no longer than is necessary for those purposes
- processed in accordance with the data subject's rights
- kept safe from unauthorised or unlawful processing, accidental loss, destruction or damage to the data

- not transferred outside the European Economic Area unless that country has equivalent levels of protection for processing personal data

Consent is an important part of the Directive which has flowed into the DPA, and is highly relevant to the collection of data for learning analytics. In summary, personal data can only be processed if one or more of the following circumstances apply:

- the data subject has unambiguously given their consent
- processing is necessary for the performance of a contract (to which the data subject is party)
- processing is necessary for legal compliance
- processing is necessary to protect the vital interests of the data subject
- processing is necessary to protect the legitimate interests of the controller except where these are overridden by the interests or fundamental rights and freedoms of the data subject

The latter point regarding the protection of the data controller's *legitimate interests* may be taken by institutions as justification for not obtaining proper consent from students. However, it may be difficult to argue that the individual's privacy is less important than the institution's right to carry out learning analytics without consent. The processing must be *necessary* and not just of potential interest – there may be another way of achieving the legitimate interests of the controller which is less invasive of individual privacy (ICO 2014). Claiming legitimate interests is however potentially as valid as any of the other above grounds for processing an individual's data; institutions may consider that learning analytics are in the best interests of the individual and the wider student body. They should first carry out a "balancing test" which weighs up the legitimate interests of the controller against the rights and interests of the subject (European Commission 2014b).

There are also particular provisions regarding the processing of personal data revealing race or ethnicity, political opinions, religious or philosophical beliefs, trade-union membership and information on an individual's health or sex life. Students must give their consent for such "sensitive" data to be used.

Another relevant part of the legislation is that data controllers must on request provide data subjects with their personal data and details of the purposes of the processing for which the data are intended "without constraint at reasonable intervals and without excessive delay or expense". Data subjects must also be allowed to rectify any errors. The controller must be prepared to provide information regarding other recipients of the data and "knowledge of the logic involved in any automatic processing of data concerning him at least in the case of … automated decisions". This implies that institutions should prepare to be completely clear and transparent around the algorithms and metrics they are developing for learning analytics.

Data subjects have the right to object to the processing of data about them "on compelling legitimate grounds" – so enabling students to opt out of data collection may be necessary. Also, critically for learning analytics, the data subject has the right "not to be subject to a decision which … significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.", including presumably his or her studies.

A further key aspect of the legislation is the prohibition of the transfer of personal data outside the European Union except where adequate safeguards (e.g. the US Safe Harbor agreement) have been put in place. This potentially restricts the use by institutions of cloud-based learning analytics services hosted in the US or elsewhere.

In January 2012 the European Commission proposed a comprehensive reform of the Directive "to strengthen privacy rights and boost Europe's digital economy". The aim is to create a single law for the 27 member states of the EU, each of whom have enacted the DPD in different ways, boosting consumer confidence and saving an estimated €2.3 billion per year (European Commission 2012). The following principles, all of which are of relevance to learning analytics, are included in the proposal, which has progressed though EU bureaucracy and is likely to be made into law imminently (European Commission 2014a):

- **A right to be forgotten**: When individuals no longer want their data to be processed and there are no legitimate grounds for retaining it the data will be deleted. This is intended to empower people rather than to erase past events or restrict freedom of the press
- **Easier access to your own data**: A right to data portability will make it easier for users to transfer their personal data between service providers
- **Putting you in control**: When consent is required to process an individual's data, they must be asked to give it explicitly. It cannot be assumed. Saying nothing is not the same thing as saying yes. Businesses and organisations will also need to inform users without undue delay about data breaches that could adversely affect them
- **Data protection first, not an afterthought**: 'Privacy by design' and 'privacy by default' will also become essential principles in EU data protection rules – this means that data protection safeguards should be built into products and services from the earliest stage of development, and that privacy-friendly default settings should be the norm – for example on social networks

Rubinstein (2013), a commentator based in the US, claims that European law fails to acknowledge the "impending big data tsunami" which she believes will overwhelm the core principles of informed choice and data minimisation i.e. restricting the collection and holding of data to the minimum necessary in order to carry out the specified purpose. Organisations which use data mining may not be able to provide adequate notice to individuals as they cannot know in advance what they might discover. Since they lack the knowledge of possible correlations users cannot consent to the use of their data either. Moreover, the distinction between personal and non-personal data is not necessarily sustainable in the world of big data. The principle of data minimisation conflicts with the philosophy of big data which applies analytics to massive data sets, attempting to do so without any restrictions. New rights such as the right to be forgotten, she believes, may be impractical and conflict with rights of free expression.

The official US government approach also differs markedly from the EU philosophy. This is important because learning analytics products may be developed primarily in the US and will be potentially unusable in Europe unless appropriate privacy safeguards are built in. PCAST (2014) argues that policies which concentrate on the collection and storage of data are unlikely to improve privacy and could only be enforced at the expense of "severe and economically damaging measures". The overall environment in the US generally relies on utilitarian approaches which weigh up the possible benefits against the risks and costs. Meanwhile in Europe basic human rights are so fundamental that it is difficult to justify breaches of them despite any benefits to society or industry that might accrue (Ess and AoIR 2012).

Schwartz (2010) puts it simply: the American approach tends to permit the use of personal information unless a law prohibits it, partly because of the protections for freedom of expression in the First Amendment. The EU however has laws which cover all processing of personal data. The UK's Information Commissioner is clear: "The complexity of big data analytics is not an excuse for failing to obtain consent where it is required" (ICO 2014).

Prinsloo and Slade (2013) suggest that despite most universities having policy frameworks to safeguard data privacy and regulate access, the frameworks are not necessarily adequate to address the specific ethical challenges of learning analytics. What leaps out from their evaluation of policy frameworks in their institutions is the extraordinary number of documents relating to legal and ethical issues that the student may be required to navigate through. It is difficult to consider this to be *informed* consent. Surely the systems themselves must provide prompts rather than the institution holding such data in a range of policy documents which few students will have the time or inclination to read.

5: Institutional approach

Ethical stances vary across domains and geography. Practices in market research and social networks may not be applicable to education for example (Kay, Korn & Oppenheim 2012). Meanwhile within a single institution there will be competing objectives. Prinsloo, Slade & Galpin (2012) describe how institutional goals such as widening participation may conflict with the findings of learning analytics. Organisations may concentrate on increasing the number of graduates, improving the completion rates of disadvantaged students or maximising profits. Learning analytics will be applied differently depending on the key drivers of the institution (Slade & Prinsloo 2013).

A number of publications refer to the underlying motivations of an institution regarding how they approach struggling students. Contact North (2012) uses the example of a student who is struggling with Calculus – do the analytics which discover this trigger additional help from a tutor or advice to withdraw from the subject? Willis & Pistilli (2014) ask if it is "unethical for an institution not to readily offer support when it can identify students who might benefit from various resources?" Campbell, DeBlois & Oblinger (2007) extend this potential "obligation to act" to the students themselves. Kay, Korn & Oppenheim (2012) describe how institutions have an increasing obligation to use data to support progression and to maximise learners' employability.

Systems and methodologies may be developed based on an institutional philosophy which values high grades and rapid progress; this may not always be what the individual wants nor in their best interests (Johnson 2014). Slade and Prinsloo (2013) go further in arguing that it is inevitable that institutional algorithms will "reflect and perpetuate current biases and prejudices". They believe that students should be enabled to act outside such imposed models.

While there will be considerable common ground a code of practice for learning analytics may therefore need to take account of the potentially quite different ethical approaches of institutions.

6: Awareness and consent

Awareness - questions	
1. Does the administration let the students/staff know their academic behaviours are being tracked?	Hoel et al. 2014
2. How much information does the institution give the teachers?	Hoel et al. 2014
3. Do students appreciate that information is being gathered about them?	Slade & Galpin 2012
4. Are we explicit about what we might do with that information?	Slade & Galpin 2012
5. Is it appropriate for students to have an awareness of the labels attached to them?	Slade & Prinsloo 2013
6. What and how much information should be provided to the student?	Willis, Campbell & Pistilli 2013
7. Should the results [of predictive analytics] be shared with the student, faculty or other staff?	Campbell et al. 2010
8. Should the information collected in one course be made available to teaching staff of another course?	Pardo & Siemens 2014
9. To what extent should students have access to the content of their digital dossiers, who has access to these dossiers?	Prinsloo 2013
10. Will adult learners expect unmediated access to internal analysis of their performance?	Reilly 2013
11. Who can see the data collected?	Slade & Galpin 2012

People have very little idea of what data is being collected about them or shared with third parties (Richards & King 2014). It is not clear to what extent students are aware of how much personal data is being recorded by their institutions, though they are increasingly used to being monitored in other aspects of their lives (Slade & Prinsloo 2013). The Open University (2014a) in a recent policy document makes clear to staff and students the categories of data potentially available for learning analytics: personal information provided by the student, the student's study record, sensitive information such as ethnic origin and disability, details of contacts between the student and the University, interactive content generated by the student, system-generated data such as accesses to the Virtual Learning Environment (VLE), data derived from other data and data generated internally e.g. student use of a library subscription service. In addition provisions are made for the use of anonymised data both internally, e.g. forum posts, and from external datasets such as social networking sites – but only to generate information on a cohort rather than individuals. Also listed are data types out of scope for learning analytics such as data on student complaints.

Information Commissioners Office (ICO) (2014) reports on an Organisation for Economic Co-operation and Development (OECD) meeting which attempted to classify personal data in a new way, based on its origins:

- **Provided data** – consciously given by individuals e.g. when filling in an online form
- **Observed data** – recorded automatically e.g. by cookies, sensors or facial recognition from CCTV pictures
- **Derived data** – produced from other data e.g. calculating customer profitability from the number of items purchased in a store and the number of visits
- **Inferred data** – produced using analytics to find correlations between datasets in order to categorise or profile people e.g. predicting future health outcomes

Big data generally uses observed, derived or inferred rather than provided data – this has implications for privacy as individuals may be unaware that the data is being collected and processed.

Two important ethical and legal principles for personal data are *notice*, the disclosure of what data controllers are doing with information, and *choice*, the ability for people to opt out of particular uses of their data (Richards & King 2014). *Informed consent* is recognised as key to the analysis of learner data by many commentators (e.g. Esposito 2012; Slade & Prinsloo 2013) and is a basic principle of scientific research on humans (ASA 1999). This can demonstrate credibility and accountability on the part of researchers and, by extension, the institution.

Fairness is also a key principle in the Data Protection Directive and relates to how data is obtained. Processing is unlikely to be regarded as fair if individuals are misled about how their data will be used when they provide their consent. Meanwhile if individuals do not have a real choice and cannot withhold their consent, then data controllers may be in danger of breaching the requirements of the Data Protection Act (DPA) (ICO 2014).

Consent - questions	
12. Does an individual need to provide formal consent before data can be collected and/or analysed	Campbell et al. 2010
13. To what extent do we provide students the option to update their digital dossiers and provide extra (possibly qualitative) data?	Prinsloo 2013
14. Do students have the right to request that their digital dossiers be deleted on graduation?	Prinsloo 2013
15. Can students opt to disguise themselves online?	Slade & Galpin 2012
16. Should students have input regarding what data is stored and how it is used?	Willis & Pistilli 2014

Traxler & Bridges (2005) describe informed consent as referring to students' "understanding of the nature, extent, duration and significance" of their participation. However, in order to provide their consent students need to know what data are collected, and when and how the data is being manipulated (Pardo & Siemens 2014). When requesting informed consent it is important to use language that is respectful and easy to understand, and allow participants to ask questions about the research at any stage. In conventional research the subjects should be told that their participation is voluntary and that withdrawal does not incur any penalty. They should also be informed about the possible consequences of non-participation or withdrawal (ASA 1999).

Researchers should be aware that they can subtly influence participants because of the researchers' expertise or authority, and they should take this into account when developing informed consent procedures (American Educational Research Association - AERA 2011). However declining to participate in learning analytics could have a negative impact on academic success so a student may not feel they have much choice in agreeing to be monitored. The concept of *voluntary* informed consent, as used by the British Educational Research Association (BERA, 2011), is laudable but not necessarily achievable by institutions which wish to make extensive use of learning analytics.

So how can consent best be granted by students? Clearly the ticking of a box at the end of a lengthy agreement in a small font in complex legal language which few users have the time or inclination to read cannot be regarded as *informed*. Even if they do read the policies they can find them vague regarding what is actually done with their personal information. The Data Protection Act requires that users give proper informed consent but the Act does not refer to the logistics of how this is obtained. A checkbox for opting into data collection, which requires a positive action, is more likely to be considered valid than one for opting out (Kay, Korn & Oppenheim 2012). However, users may fail to check the box thus denying them the benefits of learning analytics and posing further ethical and logistical problems for the institution.

Sometimes various options are provided but ICO (2010) suggests that people generally do not use the privacy choices available to them. They may not look for them in the first place, may not find them, and might not understand or realise their significance. They may also expect basic privacy protections without having to make any active choices themselves. It is good practice to set privacy defaults to reflect what most users are likely to expect. If users then alter their privacy settings from the defaults this may be an indication that they should be reset. PCAST (2014) suggests that people could sign up to one of a series of "privacy preference profiles" which might put particular emphasis on individual rights or value to the consumer. In the UK the DPA requires institutions to provide a "fair processing notice" or privacy notice, informing users as to when their data will be collected and how it will be used. It needs to provide details of the organisation collecting the data, the purposes for which it will be processed and other information required to ensure the processing is fair (ICO 2014).

Complex terms of contract and long-winded privacy notices may send out the wrong message to students – something to beware of in light of the inBloom fiasco. Instead institutions must develop a trusting environment and a culture of ethical data use (Polenetsky & Tene 2014). One study suggests that people will not read privacy policies if they perceive that the cost of reading them is greater than the benefit of doing so, and that users tend to scan the documents for particular information rather than read them comprehensively. The authors estimate that if people were to read the privacy policies for all the websites they visit it would take them 201 hours per year, a cost of $3,534 for the average Internet user in the US (McDonald & Cranor 1998).

Facebook's privacy policy itself is reported to contain more words than the US Constitution and while users are able to change their privacy settings they are presented with over 50 toggles which result in over 170 privacy options (Rayport 2011). Adding to the confusion providers regularly change their privacy policies and may or may not make that clear to the user (Pardo and Siemens 2014). There is a fundamental imbalance in the relationship between the provider, which offers a lengthy set of terms on a take-it-or-leave-it basis, and the user who has little time to read the document (PCAST 2014). However, ICO (2014) challenges data controllers to be "as innovative in [the provision of privacy notices] as they are in their analytics, and to find new ways of conveying information concisely."

Few people are aware of the many players operating in the big data world, particularly data brokers and analysis companies, so it becomes very difficult for an individual to request access to their data. (IWGDPT 2014). Some approaches from industry are worth noting however. Amazon provides information on its recommendation system to customers who can manage their own browsing history, deleting irrelevant items from it (such as gifts to others) or turning off Amazon's ability to track them. This transparent approach provides useful customer feedback to the company and heightens trust. Errors that arise from poor data, mistaken assumptions or faulty algorithms can also be reduced by allowing users to steward their own data (Schwartz 2011). Clearly only some types of data could be available for correction by students – the dates and times they are connected to a university system for example might not be appropriate for modification (Pardo and Siemens 2014).

Slade and Prinsloo (2013) speculate as to whether there are ever special circumstances which would trump the need for informed consent. They consider that there are few or no reasons in higher education to avoid full transparency regarding the uses to which student data is being put. They add that the methods for obtaining informed consent should be renewed regularly as the technologies and applications are evolving so rapidly.

As stated earlier, a key benefit of big data is the potential for finding patterns or drawing conclusions which could not have been predicted in advance. In the same way it is not possible to predict all the potential impacts on privacy when asking users for their consent. Big data systems do not necessarily produce personally identifiable information when the data is initially gathered (Crawford & Schultz 2013) although it can be assumed that most data used for learning analytics will be attributable to an individual. Crawford & Schultz also argue that the greater the implications of a particular decision based on big data, the greater right a person should have to question how that decision was arrived at. While the Data Protection Directive may require all such decisions to be made transparent to the user, it can be assumed that learning analytics systems which have a direct impact on the individual (e.g. helping to produce a grade) need to be thoroughly explainable.

As noted earlier the DPD and DPA require that students are informed when data about them is being collected or processed. Not only must they be told about the process but they should also be able to access any of their data. Finally they should be informed about how any automated decisions are made:

There is though a provision in the law for institutions to process data where it is not viable to obtain the data subject's consent, based on the "legitimate interest" of the data controller. These must not however override the interests of the individual. The data controller must balance their legitimate interests and the individual's against each other before deciding on a course of action. The way forward then is to ensure that students are informed about the data collected about them, where it comes from, what is done with it and whether it will be passed to third parties. Individuals should be given access to their profile and all the information held about them – in a user-friendly, portable format. They should be able to correct it and to opt out where possible (International Working Group on Data Protection in Telecommunications - IWGDPT 2014).

A potential problem if students have the opportunity to adapt the data about them is that they may choose to actively misrepresent themselves to avoid being labelled (Slade and Prinsloo 2013) or to portray themselves in a more favourable light.

The logistics of giving access to individuals about the data held about them are highly complex, given the number of systems in use by universities and colleges. The Open University (2014b) aims "in the near future … to provide basic aggregated results to students who request this information". The institution states that there are still technical and organisational impediments to giving the students access to their data securely and transparently.

Providing students' information to them may at least be more feasible in a web-based environment than where learning is taking place through mobile devices. Traxler and Bridges (2005) suggest that it may be impossible to explain the scope of the activity in a way which is succinct enough to be appropriate for presentation on a mobile device.

Finally, a key issue for big data and learning analytics is that the DPD requires that consent must be given if personally identifiable data is to be used for a different purpose to which it was originally intended. Institutions should assess the compatibility between the original and proposed purpose for every case (IWGDPT 2014).

7: Transparency around algorithms and metrics

Transparency around algorithms and metrics - questions	
17. How transparent are the algorithms that transform the data into analytics?	Reilly 2013
18. Who can see the models?	Slade & Galpin 2012

Users of online services may have control over what information they share but are highly unlikely to understand the complexities of how their data is being processed subsequently. Arguably the granting of consent is meaningless if learners have no conception of the way their data is being used – or could potentially be misused. The paradox is that any document which properly explains the complexities is unlikely to be understood or read while summaries are likely to be over-simplistic. Meanwhile some organisations attempt to give an impression of transparency which can hide what is really going on (Office of the Privacy Commissioner of Canada - OPCC 2012).

Secret processes and opaque and unaccountable algorithms can hide arbitrary or unfair decision making (MacCarthy 2014). Transparency regarding the purposes to which data is being put, who will have access to it, and how identities are being protected is a responsibility of the institution. However, it may not be possible to make the software tools and their predictive models public if they are proprietary (Pardo & Siemens 2014). Another possible tension is between academics' desire to publicise research findings about the algorithms they have developed and the interests of their university in protecting that information in order to maintain competitive advantage or to exploit the products commercially (Sun 2014).

The use of big data must be transparent according to Richards & King (2014) who argue that transparency helps to prevent abuses of institutional power, helps individuals to feel safer in sharing their data and results in better predictions. Slade & Prinsloo (2013) go further in arguing that universities should inform students of the risks when learning takes place in the wider Internet, outside the confines of institutional systems. They also warn that full transparency regarding methodologies brings the potential that students will abuse the system, providing false or incomplete information to obtain extra support.

8: Ownership and control of data

Ownership and control - questions	
19. Who determines which data is collected and shared?	Campbell et al. 2010
20. How are use decisions made [regarding the data warehouse]?	Campbell et al. 2010
21. Can anyone use the data warehouse for any purpose?	Campbell et al. 2010
22. Who really owns the [analytics] information?	Kay et al. 2012
23. Who is ultimately responsible for maintaining [the information]?	Kay et al. 2012
24. Are the goals of the instructor, department chair, dean and provost aligned?	Contact North 2012
25. Who owns a student's data?	Prinsloo 2013
26. Who can influence the models?	Slade & Galpin 2012
27. Who can mine our data for other purposes?	Slade & Galpin 2012
28. Who gets to decide what happens next?	Slade & Galpin 2012
29. Who can choose which students get more/less support?	Slade & Galpin 2012
30. Do teachers, learners and administrators have the same authority/rights to determine what support is provided?	Slade & Galpin 2012
31. Is there a shared responsibility to ensure that information is accurate?	Slade & Galpin 2012
32. Who has the power to make decisions about the learning analytics model and data?	Swenson 2014
33. Who has the power to legitimize some student knowledge or data and not others?	Swenson 2014
34. Who has the power to focus on potential intervention strategies and not others?	Swenson 2014
35. Who has the power to give voice to certain students and not others?	Swenson 2014
36. Who has the power to validate some student stories and not others?	Swenson 2014
37. Who decides what feedback is valid and how often it should be delivered?	Willis & Pistilli 2014
38. Who determines what constitutes a successful outcome in a student career?	Willis & Pistilli 2014

Ownership and control - questions	
39. Who is responsible when a predictive analytic is incorrect?	Willis, Campbell & Pistilli 2013

Bollier (2010) refers to a "Declaration of Health Data Rights" by an organisation called HealthDataRights.org, which now seems to be defunct. It asserts that we should:

- Have the right to our own health data
- Have the right to know the source of each health data element
- Have the right to take possession of a complete copy of our individual health data, without delay, at minimal or no cost; if data exist in computable form, they must be made available in that form
- Have the right to share our health data with others as we see fit

This ethical stance would seem to be appropriate for learners' data and to comply with the Directive. Another analogy is with credit rating agencies which are obliged to give consumers copies of their credit records on request (Bollier 2010). One issue here for institutions is that the vast amount of data assembled on every electronically captured interaction the institution has about an individual would be overwhelming and meaningless. Do we therefore ensure that we give students copies of the summarised data with our interpretations and visualisations instead of or as well as the raw data?

Issues of ownership are linked to issues of control and responsibility. One question raised by Campbell, DeBlois & Oblinger (2007) in this regard is who decides what information can be shared with students and staff and how should this be done? Swenson (2014) asks more specifically who has the power to:

- make decisions about the learning analytics model and data
- legitimise some student knowledge or data and not others
- focus on potential intervention strategies and not others
- give voice to certain students and not others, and
- validate some student stories and not others

Pardo and Siemens (2014) ask who owns the data: institutions, students or the companies which might use them to enhance their products? A relevant legal issue here is the intellectual property rights (IPR) in the data relating to a student. As these have been collected and possibly enhanced by the institution it is the owner of the IPR. However, it must ensure that the data are accurate or the student can request that they are deleted. At any time the student can request a copy of their personal data though the IPR remains with the institution which may prevent the student further transferring it. This has implications for the portability of learning analytics as a student moves from one institution to the next, taking their own data with them (Kay, Korn & Oppenheim 2012).

Clow (2012) argues that metrics used for analytics should be open and transparent, and that this removes potential barriers to the effective use of analytics, increasing the social acceptability of the process. He suggests that misapplications are more likely to be noticed by stakeholders if the data is open and therefore able to be corrected. Clow does not advocate openness for all learners' data and metrics but suggests that restrictions should not be added unnecessarily.

9: Using publicly available data

Publicly available data - question	
40. Should learners involved in an open course be required to give consent for data collection and analysis?	Pardo & Siemens 2014

There may be something fundamentally different about learner data when individuals choose to engage in public environments or mass environments such as Massive Open Online Courses (MOOCs). Clearly the responsibilities of institutions are different in relation to data in systems such as Twitter than they are with data held in the institutional "walled gardens" of the student information system and the VLE. Esposito (2012) asserts that privacy seems to be less of a concern to researchers when the data is already publicly available. If the forums of a MOOC are visible to the public as well as enrolled participants there may be a greater acceptance that user activity and comments are subject to scrutiny and analysis. The fact that a MOOC can take place in multiple systems may add to this. However, there is a registration process for MOOCs in Coursera and FutureLearn, for example, and non-participants cannot view forum postings and other data on learners so there would appear to be a greater responsibility on those MOOC providers to steward the data appropriately.

Esposito believes that when learners post a message in a forum there is an assumption that the content will be read or archived. However, users may be less comfortable with their messages being subject to analysis by researchers and it may be better to obtain their informed consent for this in advance. The example is given of learners feeling "violated if they saw their posts de-contextualised and highlighted in a publication". There is the usual problem with quoting people here too in that some prefer their comment to be anonymised while others feel they should be acknowledged as the author.

The ethics of using data from social networking sites are discussed by Rivers & Lewis (2014). They argue that it is generally regarded as appropriate to collect information without consent from physical public spaces where there is a reasonable expectation of observation by strangers. However, they suggest that tweets, while readable by anyone with internet access, are individually attributable which makes them fundamentally different from observations on aggregate populations in a physical space. Twitter users, they propose, can expect a level of "anonymity of the crowd" to help manage their privacy; they give an example of someone who discusses his mental health with his digital community but does not expect his comments to be used subsequently by researchers.

As well as obtaining informed consent from students for use of their data from external sites, Slade & Prinsloo (2013) note two additional concerns to address. Firstly, the institution has no control over the varied data protection policies of those sites. Secondly, it may be impossible to authenticate student identity properly.

10: Accuracy of data

Greller & Draschler (2012) believe that flawed data is the biggest technical challenge for analytics. They point out that users frequently "pollute" databases with erroneous or incomplete data. One example is a teacher who wishes to view their VLE from a student's perspective so sets up a test account which is then included in the analytics for the course. Another problem is "enmeshed identities" where the data does not differentiate between an authenticated individual and a group. Students working together on a device may unwittingly leave enmeshed fingerprints in their data. Meanwhile when data is collected against identifiers such as IP addresses or cookies and attributed to an individual there is a danger that it does not actually relate to that person at all. This could represent a privacy risk when access is provided to data relating to someone else (ICO 2010).

Bischel (2012) in an EDUCAUSE study of research into institutional use of learning analytics, mainly in the US, notes that many participants thought that problems with quality of the data should be tackled before an analytics problem is initiated. However, others felt the data would never be perfect and that this should not stop an institution from beginning to develop its analytics capabilities. Bollier (2010) quotes Jesper Andersen, a computer scientist and statistician, who warns that drawing conclusions from a single data source can be dangerous and that it is better to use data from multiple sources. Meanwhile Swenson (2014) is concerned about inaccurate or incomplete data used for predictive modelling. She suggests that lack of accountability is an issue when predictions are made without the statistical methods being verified and without the accuracy and completeness of the data being checked.

Obtaining the right data for learning analytics can be complex: Bischel (2012) mentions that ownership of data is an issue for many institutions, with it being held in "silos" and individual departments unwilling to make their data available for analytics. They suggest that senior management puts in place policies which oblige the sharing of data for learning analytics, balancing this with requirements for security and privacy.

Greller & Draschler (2012) point out a contradiction in that learning analytics requires publicly available datasets to advance the methods of researchers while the protection of learner data is a high priority of IT departments in institutions. They find it strange that in the commercial sector users are happy to click a "register" button, giving entire ownership of their data to a company, while in educational institutions "everything is protected from virtually everyone". Perhaps the ethical requirement to do something with the data if it could help students should be extended to sharing the data you "own" with others who may be able to use it for the benefit of learners.

Slade & Prinsloo (2013) propose that students should be given the opportunity to prove the predictions about their likely performance wrong or incomplete. Implementing this from a procedural and technical perspective may however be complex.

11: Considering personal circumstances

Personal circumstances - questions	
41. Does [a student profile] bias people's expectation and behaviours?	Campbell et al. 2010
42. Should the institution even create profiles that lead to generalisations about students?	Campbell et al. 2010
43. Are there profile uses that should be prohibited?	Campbell et al. 2010
44. Will the data influence perceptions of the student and the grading of assignments?	Hoel et al. 2014
45. Will the process become overly deterministic?	Reilly 2013

A working party at Charles Sturt University (2014) notes that "learning is a complex social activity and that technical methods do not fully capture the scope and nuanced nature of learning". Reducing the complexity of student behaviour to a number or a traffic light is pointed out by Campbell et al (2007) to result potentially in oversimplified or insensitive conclusions. Any algorithm or method will be reductive in that it attempts to create a manageable set of metrics which do not necessarily reflect reality (Greller & Draschler 2012). No prediction can take into consideration all possible factors such as problems at home or financial difficulties. Did the student fall in love or was a death in the family the reason for academic failure (Contact North 2012)?

Slade & Prinsloo (2013) point out that as much of the data related to learning is held in systems outside the control of the institution (e.g. in cloud-based public services) it is impossible to obtain a holistic picture of student life. Moreover the data itself is temporal and may only afford a view of an individual at a specific place and time, not allowing for the changing and multiple identities of learners as they progress through their studies.

A number cannot represent the personal growth or development of relationships that arise from attending an educational establishment. Johnson (2014) worries that data mining can treat a subject as a collection of attributes rather than an individual. He discusses course recommendation systems where students are encouraged to do what people like them have done before. Arizona State's eAdvising system aims to identify students whose skills do not match up to their ambitions. For Johnson it appears that learners are being thought of as mere collections of skills to be matched to an outcome rather than individuals. He thinks such systems undermine students' autonomy, and condemns Arizona State's processes to compel struggling students to change their major as "coercive", denying students the opportunity to take their own decisions. Meanwhile the "softer" approach of the course recommendation system at Austin Peay he feels encourages students to conform to the values and behaviours that the University considers to be most likely to result in success.

Such interferences may be valid on the basis of preventing wastage of taxpayer's money or guiding students who are not mature or informed enough to take sensible decisions. But every violation of autonomy should, Johnson feels, be justified. A way forward, he suggests, may be to design systems which encourage autonomy and help students to make decisions for themselves without institutional paternalism.

Campbell, DeBlois & Oblinger (2007) even question whether institutions should be creating individual profiles which lead to generalisations about students. Learning analytics solutions are largely technical and do not usually take into account the cultural and behavioural contexts of the learner, the learning, the discipline and the institution. Tertiary education can be an "individualised artisanal craft" where the standardised metrics and interventions of learning analytics may not fit easily (Contact North 2012).

Draschler & Greller (2012) quote one participant in their survey of 156 educational practitioners and researchers into the use of and attitudes to learning analytics:

> *It would be easy for learning analytics to become a numbers game focused on QA, training/instruction and rankings charts, so promoting its creative and adaptive potential for lifelong HE/professional-life learning is going to be key for the sector - unless learning analytics people want to spend all their lives doing statistical analysis?*

Siemens (2012) proposes that human-contributed feedback and corrective options can help to improve personalisation. This is already policy in some UK institutions such as Derby University and Bridgwater College where staff are able to contextualise automated email interventions before they are directed to students (Sclater 2014b).

12: Respecting privacy

Privacy - questions	
46. Will funders such as employers track progress?	Reilly 2013
47. Is informed consent a sine qua non or are there circumstances in which other principles override the need for informed consent?	Slade & Prinsloo 2013
48. Is it unethical for administrators to do whatever possible to help ensure student success, even if it means stretching the meaning of privacy?	Willis & Pistilli 2014
49. Is the improvement of the overall learning environment a valid reason to record the exact location of students within the institution and share it with peers to facilitate collaborative learning?	Pardo & Siemens 2014

Various leading figures from the IT world including Google and Facebook have declared that privacy is no longer a valid concept. However, the public outcry over Edward Snowden's revelations around the surveillance carried out by the US National Security Agency has shown that many people still regard privacy as extremely important. This has had knock-on effects for major IT companies who are concerned that unless the public can be reassured that their privacy is being protected they may stop using their services (Richards & King 2014). King & Richards (2014) argue that it is unrealistic to consider data as wholly in either the public or private domains. They also stress (in Richards & King 2014) that private information can remain confidential even after it has been shared and that privacy is not a simplistic, binary concept. However, information has to be shared in the era of big data in order for it to be useful.

PCAST (2014) outlines some of the main potential threats to privacy from big data. While this comes from a US perspective it is useful in outlining some possible privacy infringements which might arise from learning analytics applications in particular:

» **Invasion of private communications**: an individual's right to private communication may need to be re-established in the digital era

» **Invasion of privacy in a person's home**: one's "virtual home" now includes Internet access and storage of documents in the cloud

» **Public disclosure of inferred private facts**: analytics may infer facts from apparently harmless data sources. Examples are given of inferring sexual preference or Alzheimer's disease from user input, a private fact of which the individual may not even be aware. Presenting such information publicly would be widely regarded as unacceptable

» **Tracking, stalking and violations of locational privacy**: tracking is increasingly trivial using mobile devices and location-based services. Building a track of an individual's movements is a potential invasion of privacy

- **Harm arising from false conclusions about individuals, based on personal profiles from big-data analytics**: while false conclusions are possible about anyone, there are particular concerns here for groups such as racial minorities or the elderly
- **Foreclosure [inhibiting] of individual autonomy or self-determination**: people should be able to choose a course of action which is not necessarily predicted for them
- **Loss of anonymity and private association**: individuals should have the right to remain anonymous; re-identification of them from the data potentially infringes that right. People should also be able to associate with one another privately

"Information rules" might be a more appropriate and less emotive term than "privacy". Meanwhile, as has been discussed earlier, "data protection" is the term generally used in the EU. The crisis in privacy is around our ability to manage the uses of information that are held about us, most of which is in an intermediate state between completely private and completely public (Richards & King 2014).

The majority of respondents in Draschler & Greller's (2012) survey thought that only appropriate staff members should be allowed to view student data on a "need to know" basis, with many emphasising the need to control access in compliance with the law and ethics. Data from multiple sources should be preserved, secured and shared appropriately. Decisions need to be made on who can access the data (Campbell, DeBlois & Oblinger 2007). Should teaching staff be given access to data collected on other courses for which they are not responsible for example (Pardo & Siemens 2014)? If so should the identity of students for whom they are not responsible be masked?

At Oxford Brookes University there is a hierarchy of permissions for analytics data which is passed up through various levels of the administration with individuals unable to be identified except by those directly responsible for teaching or supporting them (Sclater 2014b). At Charles Sturt University the policy is to control access by roles and to set privileges based on the individual's position and how sensitive the data is. In addition audit trails are expected to be kept on who has accessed what data (Charles Sturt University 2014).

Despite the restrictions on access, how learning analytics systems are used in practice can potentially infringe students' privacy. At the University of Michigan it was found that academic advisors were regularly sharing dashboards designed for their own use with individual students. The screens showed data about other learners; a button was hastily added for advisors to hide the data about other students when required (Aguilar, Lonn & Teasley 2014).

As has been discussed earlier, the DPA makes a distinction between *personal data* and *sensitive personal data* which includes information on racial or ethnic origin, political opinions, religious beliefs, trade union membership, physical or mental health, sexual life and criminal records, and must be handled with particular care. It may be that the compilation of student profiles for learning analytics from a variety of data which are in themselves not especially sensitive, could nevertheless result in a finding which falls into this sensitive category (IWGDPT 2014).

One concern which institutions or individuals may have around confidentiality is whether they are required to release information attributable to specific students in response to a freedom of information request. However, the Freedom of Information Act specifies that no personal data should be disclosed in response (Kay, Korn & Oppenheim 2012). Meanwhile, institutions should be aware that the Data Protection Act continues to apply even when personal data is already in the public domain (Ministry of Justice 2011). And while some organisations have attempted to avoid data protection legislation restrictions by sharing metadata instead of personal data, the power of big data algorithms means re-identification is possible which would therefore breach the DPD (and DPA) (Richards & King 2014).

13: Opting out

Opting out - questions	
50. Does an individual have an option to "opt out" of an analytics project?	Campbell et al. 2010
51. Should students be allowed to opt out of having their personal digital footprints harvested and analysed?	Prinsloo 2013
52. Can students opt out of having their information used?	Slade & Galpin 2012
53. What are the consequences of [students opting out]?	Slade & Galpin 2012

As has been noted the DPD implies that students have a right to opt out of data collection:

> *any data subject should … be entitled, on legitimate and compelling grounds relating to his particular situation, to object to the processing of any data relating to himself*

While enabling students to opt out of both data collection and interventions ostensibly complies with the legislation, a number of negative consequences of this have been discussed. For the purposes of general educational research and the "greater good" of learning analytics, having gaps in the dataset is unhelpful. Researchers should seek to minimise harm to individuals, however, there may be situations where they have to balance the risks to participants against the possible benefits for others (ESRC 2012). A key moral problem for institutions of allowing students to opt out is whether by "harming" one learner, perhaps by breaching their privacy, they are potentially benefitting students overall. Pardo and Siemens (2014) provide a hypothetical example of an institution which records the location of its students and shares them with others to encourage collaborative learning. They point out that in medical research analysing patients' records can have benefits for society as a whole and that in learning contexts too absolute confidentiality may not always be for the best of the wider group.

Slade & Prinsloo (2013) think that the benefits for many outweigh the rights of individuals who wish to withhold their data for reporting purposes e.g. to funding bodies. However, students should, they believe, be able to opt out of having their learning personalised, assuming they are made aware of the consequences. Such decisions may be highly contextual though - personalised learning may evolve to such an extent that opting out on an individual basis simply does not make sense in the context of the pedagogy.

The Open University (2014b) does not currently enable students to opt out of data collection but proposes this as a possibility for the future. However, it suggests that "opt out would most likely relate to the delivery of personalised interventions rather than the removal of individual student data items from the complete set of data for analysis". A "frequently asked question" which it is suggested that students might ask is:

> *I understand that the University needs to collect a certain amount of personal data, such as my ethnicity, age and previous educational experience. Can I choose not to have my data included in any analysis that links to learning analytics?*

The answer given is:

> *In order to have a complete dataset, the University will use all student data to analyse patterns of behaviour. The analysis stage works on the dataset as a whole, that is, it does not identify an individual student by name or PI. It is important to maintain a full dataset here as any significant loss in student data may mean that the remaining dataset is not representative of the whole.*

The implication here is that sensitive information of this nature will not be used to support individual students. It is not clear how this will be reconciled with the assertion in another document that ethnic origin and disability are regarded as in scope for learning analytics (The Open University 2014a).

14: Interpretation of data

Interpretation of data - questions	
54. What are the potential ethical consequences of stripping data of personally identifiable information?	AoIR 2012
55. How might removal of selected information from a dataset distort it such that it no longer represents what it was intended to represent?	AoIR 2012
56. How complete and permanent a picture do our data provide about students?	Prinsloo 2013
57. [Are] bigger data sets always better or [do they] provide more complete pictures?	Prinsloo 2013
58. How reliable and robust are the models?	Slade & Galpin 2012

Siemens (2012) points out that the two main data sources for learning analytics, virtual learning environments and student information systems, represent only a fraction of the learning that takes place, and that work is needed to integrate data from other sources such as libraries, mobiles and social media profiles. These will provide "analytics opportunities that far exceed single data points."

Bollier (2010) says that visualisation tools for analytics can make it very easy to find spurious correlations – relationships that do not really exist. However, it can be more difficult to draw objective conclusions from more than one data source as they are all prone to errors and you may simply be magnifying the inaccuracies when you combine multiple datasets. The challenge is to discover what the most relevant data is for you to be collecting in order to make the right decisions. Less may turn out to be more; as opinion pollsters and market researchers know, small samples can be very reliable proxies for large populations. Prinsloo, Slade & Galpin (2012) find that the sheer volume of student data collected in their large institutions (the UK Open University and the University of South Africa) can prove problematic in developing understanding of student and institutional behaviours. Ellis (2013) argues that when carrying out learning analytics using assessment data it is better to base the analysis on pedagogical principles rather than on what data is available.

One of the key stated benefits of big data is that patterns can be discovered and conclusions drawn which were never anticipated. This tends to conflict with traditional scientific method where a theory is postulated and the data is then sought to confirm or contradict it. It also challenges the privacy principle in the Directive that data cannot be used for purposes incompatible with the original purpose. As big data is also about maximisation it conflicts with the principles of relevance and data minimisation, principles which are intended to ensure that only the data required is stored, and that it is deleted when it is no longer of use for its original purpose. The point about big data is that its value is related to possible future uses as well as current purposes. Organisations will not wish to delete data which could be a future source of insights and revenue; it may thus become increasingly difficult for authorities to enforce the requirement to delete personal data (IWGDPT 2014).

A *Business Week* journalist, Stephen Baker, believes that predictions based on big data do not necessarily have to be correct – they just have to be better than the status quo. Revenue streams for companies can be based on imprecise data methods, and finding the truth is less important than what works (Bollier 2010). Is it the same in education or are we under a greater obligation to establish the "objective truth" before taking action? The algorithms and student profiles should be regularly assessed to ensure that analytics result in responsible, fair and ethical decisions (IWGDPT 2014).

An unwieldy raw data dump provides little value to staff supporting students (Dringus 2012) or indeed to the students themselves. Greller & Draschler (2012) refer to conceptual instruments including theories, algorithms and weightings which help to develop information from the raw data. The methods chosen will have a significant influence on the quality of the information and could produce different outcomes with consequences for decision making. A common issue arising from the interpretation of learning analytics is that correlation is often mistaken for causality. Moreover, correlation analysis, while potentially proving accurate for a group, can be incorrect or misleading for an individual (IWGDPT 2014).

15: Stewardship, preservation and deletion of data

Stewardship questions	
59. How is the data preserved, secured and shared?	Campbell et al. 2010
60. What happens to learner data once their relationship with the provider has ended?	Reilly 2013
61. How long is data kept for?	Slade & Galpin 2012

Learners should be able to develop without records of past experiences "becoming permanent blemishes on their development history". Such data should have an agreed lifespan and expiry date, and students should be able to request the deletion of data relating to them according to agreed criteria (Slade & Prinsloo 2013). The Open University's Retention of Student Data and Records Policy states that there is an expectation by students, employers and Government agencies that students' names, modules, qualifications and outcomes will be retained permanently, and also that some data needs to be kept while a student is studying with the institution, which can be for many years (Prinsloo & Slade 2013). It would seem however that most learning analytics data is unlikely to be regarded as appropriate for permanent retention.

The key dilemma for institutions in this area is described by Pardo & Siemens (2014): guaranteeing that personal data will be deleted will build trust among student, however, keeping that data allows institutions to refine their models, track performance over multiple years and cohorts and assist with quality assurance processes. Applying big data techniques to large datasets regarded as worthless could result in valuable insight to the institution (PCAST 2014). Arguably though many of these functions could still be carried out using anonymised data, minimising risks to privacy. Useful guidance from the UK Information Commissioner (ICO 2014) is that:

> *big data analytics is not an excuse for stockpiling data or keeping it longer than you need for your business purposes, just in case it might be useful. Long term uses must be articulated or justifiable, even if all the detail of the future use is not known.*

PCAST (2014) points out that it may not be possible for data controllers to discover all the information they hold about an individual or to be absolutely sure that they have deleted all data relating to that person, and not immediately as might be expected by the data subject. As data is increasingly distributed it is difficult to prove that it has been completely erased. Moreover, as soon as data has been presented to an individual's eyes or ears in an analogue way, it can be "re-digitised". Rogue computer programs may also obtain data and copy it illegally elsewhere. Meanwhile, metadata may be stored separately and thought of as fundamentally different but may be as important in identifying individuals as the data itself. The only realistic position, PCAST suggests from its US-based stance, is to assume that as soon as data is created it is permanent; policy should therefore concentrate on the use of data rather than its collection. The emphasis should be on preventing inappropriate use of the data rather than resorting to anonymisation.

16: Interventions and the “obligation to act”

“Obligation to act”- questions	
62. What is the responsibility of faculty, students and institutions to act on [predictive analytics]?	Campbell et al. 2010
63. With whom does the obligation to act lie? How is the responsibility shared among different groups?	Campbell et al. 2010
64. What responsibility comes with "knowing"?	Prinsloo 2013
65. Once administrators "know" something about a student (via statistical regression), are institutions or individuals compelled to act? What happens if no action occurs?	Willis 2014
66. What is the role of "knowing" a predictive analytic — once something is known, what are the ethical ramifications of action or inaction?	Willis & Pistilli 2014
67. Is it unethical for an institution not to readily offer support when it can identify students who might benefit from various resources?	Willis & Pistilli 2014
68. Once an administration "knows" something about student performance, what ethical obligations follow?	Willis, Campbell & Pistilli 2013
69. Once a college's administration has the tools to "know" with statistical significance those who might be in jeopardy of failing, who is compelled to act on that knowledge?	Willis, Campbell & Pistilli 2013

A key ethical issue mentioned by various authors (e.g. Campbell, DeBlois & Oblinger 2007; Kay, Korn & Oppenheim 2012) is around the “duty of care, “obligation of knowing” or “obligation to act”. What obligations do staff, students and institutions have to take action on the basis of predictive analytics? One of Slade & Prinsloo’s (2013) principles is that universities “cannot afford to not use learning analytics” and that to ignore data which might help achieve an institution’s goals seems “short-sighted in the extreme”. There may also be a risk that a failed student could take legal action against an institution that had information that they were at risk but did not provide additional support (Kay, Korn & Oppenheim 2012). This principle is made concrete in The Open University’s (2014a) policy where it is stated that:

> *Where data indicates that there is potential for action to be taken which might better support students in achieving their study goals or in reaching their potential, the University has a responsibility to act on this. For example, if there is evidence that a student is not engaging with essential learning activities, we should consider making an appropriate intervention.*

Prinsloo, Slade & Galpin (2012) argue that while student data, and the resulting analytics, can help with decision making, institutional decisions, processes and *non-action* by the institution can have a major effect on student choices and actions.

Interventions - questions	
70. How should teachers react to the data? Should the teacher contact the student?	Hoel et al. 2014
71. [What are digital dossiers] used for?	Prinsloo 2013
72. What recourse do institutions have when students provide false or incomplete information which may provide them with additional support at a cost to the institution (and to other students)?	Slade & Prinsloo 2013
73. What happens when something unexpected turns up in the data (either as a single previously unknown data point or as a correlation of aggregate data)? What infrastructure exists to handle it?	Willis 2014
74. How should the faculty member react to the data?	Willis, Campbell & Pistilli 2013
75. Should the faculty member contact the student?	Willis, Campbell & Pistilli 2013
76. What action is appropriate based on the information learned as a result of the analysis?	Willis, Campbell & Pistilli 2013

But whose responsibility is it to take action and how is the action distributed across groups? Are the goals of instructors, managers and senior management aligned? Information derived through learning analytics should not be applied simultaneously for competing purposes (Contact North 2012).

Campbell, DeBlois & Oblinger 2007) wonder if some actions, based on student profiles which may be generalised (and therefore inaccurate) should be prohibited. If so it will be important to establish what types of action could be inappropriate. Willis & Pistilli (2014) suggest being guided by the question of what *should be done* as opposed to what *can be done*. However, even if appropriate interventions can be put in place it may be difficult for staff to find the time to take them and they may not be rewarded adequately for doing so (Contact North 2012).

Swenson raises a concern about whether intervention strategies might favour one group of students over another e.g. campus-based over distance students. The algorithms themselves may actually reinforce discriminatory attitudes and actions, selecting at-risk students on the basis of race or gender (MacCarthy 2014) or wrongly directing them into pathways designed for those with high or low potential (PCAST 2014). This in turn could further stratify society, and also limit students' possibilities of learning through failure and experimentation. However, learning analytics and big data, despite potentially creating discrimination, could also help it to be identified and addressed (Polonetsky & Tene 2014).

Data can very easily be misinterpreted and interventions ill-thought out. At Rio Salado College a correlation was identified between logins on day one of a course and subsequent student success. An assumption was made that if students were encouraged to log in on their first day they would be more likely to succeed. The welcome email that was sent to students as a result turned out to have no impact. An alternative theory of the relationship between the early logins and success is that both are related to the learner's motivation (Johnson 2014).

Student responsibility - questions	
77. What obligation does the student have to seek assistance?	Willis, Campbell & Pistilli 2013
78. What is the obligation for the student to either accept explicit guidance or to seek support which may be in conflict with their own preferences or study goals?	Ferguson 2012
79. At what point are college students to be treated as independent agents who are, as adults, responsible for their own successes and failures?	Willis & Pistilli 2014

How students should respond to the guidance or direction provided by learning analytics systems is another issue. Dashboards and automated interventions should not give students the impression that the academic environment is completely controlled; ultimately it is their choice whether to take up any additional resources or support offered as a result of learning analytics (Willis, Campbell & Pistilli 2013). Slade & Prinsloo (2013) ask what obligation learners have to take recommended actions that are contrary to their own aims or preferred ways of studying. If students are allowed to opt out, the consequences of missing out on additional support should be made clear to them and other stakeholders. Meanwhile, there is arguably an obligation on the institution to prevent students from continuing on a particular pathway when analytics demonstrate that it is neither in their interests nor the institution's to continue. (Slade & Prinsloo 2014).

The Open University's Data Protection Policy states that the institution may use data about the student's ethnic background or disability to identify those requiring additional support. They "consider disclosure of this information as explicit consent to use this information for this purpose". Not using such data alongside educational performance data is arguably immoral in an educational system where there are historic injustices in relation to characteristics such as race and gender (Prinsloo & Slade 2014).

17: Impacts on student behaviour

Impacts on student behaviour - questions	
80. What affect will [unmediated access to internal analysis of learners' performance] have on learning outcomes?	Reilly 2013
81. Could comparable learning analytics visualised for ready consumption become part of job applications?	Reilly 2013
82. Are there some labels which should be prohibited?	Willis & Pistilli 2014
83. Will the data affect student motivation in any quantifiable way?	Willis, Campbell & Pistilli 2013
84. Will the data influence perceptions of the student and the grading of assignments?	Willis, Campbell & Pistilli 2013
85. Who is affected by the analysis or application of big data, and how should they be affected by it?	Willis, Campbell & Pistilli 2013
86. Does the institution provide a calculated probability of academic success or just a classification of success (e.g., above average, average, below average)?	Willis, Campbell & Pistilli 2013

There are dangers in using correlations to make predictions. Interventions can "incentivise behaviour" (Johnson 2014), and if users know their behaviour is being monitored they may alter it to "game" the system – or do so subconsciously (Bollier 2010). Clow (2012) discusses how learners' behaviour is influenced by the assessment, and how in the same way learning analytics systems bring a risk of optimising decisions based on a metric which does not reflect the fundamental outcome desired.

Optimism is expressed by Ellis (2013) about the positive impact on students when presenting them with information about where they are placed relative to their cohort (or previous groups of students), "motivating them to improve and aspire to higher levels of achievement". She feels that the attitudes and behaviours of successful students can be used to guide and encourage the lesser-achieving ones. While some learners who are labelled "at risk" may be motivated to do better, particularly if they are equipped with the skills and life circumstances to be able to do so, increased awareness may have adverse consequences for less fortunate students (Swenson 2014). Willis & Pistilli (2014) also mention the possibility of predictions given to students becoming a self-fulfilling prophecy which cause them to give up on a module or course they are predicted to fail. As the algorithms and metrics become more fine-tuned and trusted will this effect intensify?

Ellis (2013) suggests making the learning pathways and strategies that are most likely to lead to success more explicit; perhaps this would reduce the potential negative motivational impact of direct comparisons with peers.

User behaviour may be influenced by the knowledge that data is being gathered or metrics produced as a result of it but perhaps more seriously learners and teachers may be put off using the systems completely if they feel they are being monitored. The interests of some stakeholders may even be contradictory: managers want to analyse the effectiveness of teaching but teachers see this an intrusion on their privacy or autonomy (Chatti et al. 2012). The involvement of students and different types of staff in developing policies around the use of learning analytics is therefore suggested as essential. Greller & Draschler (2012) also suggest that judgements about a learner based on a limited set of parameters could limit their potential. They think prejudices regarding race, class or gender may be reconfirmed with the data, resulting in restrictions imposed on certain groups of learners.

Not only might behaviour be influenced by learning analytics but also status. Swenson (2014) asks whether predictive categories may unintentionally reinforce social power differentials and learners' status vis-à-vis each other. She also wonders to what extent a student being described as "at risk" corresponds with them actually being at risk or feeling at risk. A further danger exists that analytics will infantilise students or spoon feed them with automated suggestions by making the learning process less demanding (Ellis 2013).

Crawford & Schultz (2013) note that organisations rarely inform individuals about potentially harmful predictive systems until they are implemented, and are generally unwilling to share the rationale for any predictions that are made. They argue that institutions are under no legal obligation to record audit trails around predictions. However, this may be indefensible in the European context. Meanwhile, such audit trails would potentially increase accuracy and enable students to question how their data is being used.

IWGDPT (2014) expresses another concern: where increasing numbers of decisions are based on algorithms we will be judged more on what are expected to be our likely future actions than on the basis of our actual actions. The group also mentions the concepts of "echo chambers" or "filter bubbles" where intelligent software exposes us to content which confirms our own attitudes or beliefs. There is a potential danger that learning analytics could channel students in convenient pathways which do not sufficiently challenge us. The exchange of different opinions and viewpoints among people of different backgrounds is arguably a vital aspect of higher education to preserve.

18: Targeting resources appropriately

Triage – questions	
87. Who receives priority if resources are limited?	Campbell et al. 2010
88. Will access to support services be limited to those with the greatest need, or will anyone who has interest be able to receive help?	Campbell et al. 2010
89. What amount of resources should the institution invest in students who are unlikely to succeed in a course?	Hoel et al. 2014
90. How do we make moral decisions when resources are (increasingly) limited?	Prinsloo & Slade 2014
91. Do we have a responsibility to ensure equitable treatment of students based on what we know? (or despite what we know)	Slade & Galpin 2012
92. What amount of resources should the institution invest in students who are unlikely to succeed in a course?	Willis, Campbell & Pistilli 2013

With limited resources institutions may wish to target activity at those students who are likely to benefit the most from interventions. Campbell, DeBlois & Oblinger (2007) question whether this is fair and ask whether those of lower priority should still be able to receive assistance. Slade & Prinsloo (2014) explore the question "how do we make moral decisions when resources are (increasingly) limited?" by applying the concept of "triage" to education. Triage in medicine is where decisions are taken on how to prioritise the treatment of patients based on the seriousness of their condition and the available resources. The aim is usually to save the greatest number of lives. As with medical triage however, in education it is impossible always to target resources accurately: students cannot simply be categorised as "not needing help", "may pass with additional support" and "destined to fail whatever additional support is provided". Transparency is key to justifying the provision of or exclusion to additional services for individual students (Slade & Prinsloo 2013).

Another issue is one of opportunity cost: the benefits for students of learning analytics may be minimal, or of lower impact than spending on other activities (Slade & Prinsloo 2013). A question for many institutions will be whether they should be investing in learning analytics at all when there may be more pressing needs for funding. The Open University (2014a) makes clear that its responsibility for action based on learning analytics needs to be balanced against available resource. It suggests that many more helpful interventions will be identified than can be funded, and that priority groups or areas of the curriculum will need to be identified for initial targeting.

Ellis (2013) notes that the majority of learning analytics work is directed towards learners who are struggling with course materials or at risk of drop out, with the literature also pointing to potential benefits for excellent students who need further challenges. She says that learning analytics is in danger of ignoring the needs of all those students who fall in the category between failing/struggling and excelling. Ellis suggests that there is not nearly enough detail stored about student aptitudes and behaviours; two students receiving the same grade for example may have demonstrated very different strengths and weaknesses in achieving it. The continued use of paper-based assessments makes it extremely difficult to analyse assessment data at deep enough levels of granularity such as measuring student achievement against learning outcomes.

19: Anonymisation

Anonymised data is not subject to the DPD or the DPA assuming it does not enable the identification of a living individual (Kay, Korn & Oppenheim 2012). The data and its uses should be assessed and documented appropriately, perhaps through the formal procedure of a Privacy Impact Assessment.

Anonymisation or "de-identification" of data can be achieved through a number of techniques. The robustness of the technique should be assessed on the basis of whether:

- an individual can still be singled out
- records relating to an individual can still be linked, or
- information relating to an individual can be inferred (IWGDPT 2014)

Anonymisation can be difficult however because the size of the datasets can make "re-identification" easier i.e. linking information to an individual (Bollier 2010). As the dataset gets bigger and is derived from a greater number of sources it becomes easier to re-identify individuals (PCAST 2014). It may also be possible for institutions to identify characteristics about us and even define who we are before we have decided that for ourselves (Richards & King 2014). Conversely very small datasets can enable individuals to be singled out. One example given by Kay, Korn & Oppenheim (2012) is that of a student on a small Masters course who borrows a large print version of a book from the library.

At the UK's Open University the Retention of Student Data and Records Policy states that some data will be anonymised and retained for use in management, development and research. It also states that the University may share personal data with third parties and require them to follow the University's policies relating to data retention (Prinsloo & Slade 2013). This is necessary to comply with the legislation.

If the data controller (or anyone else) can somehow identify who the data relates to then it can still be regarded as personal data (Ministry of Justice 2011). Pseudonymised data, where the student's name or identifier is altered consistently across their records is not equivalent to anonymised data and is subject to the same protections as personal data. If an institution makes such data available to other organisations it should ensure that contracts stop them from trying to re-identify individuals. Aggregated data however, where there is a minimal risk of linking it to other data sets, is less of a risk (IWGDPT 2014).

The Information Commissioner (ICO 2014) pragmatically suggests that:

> *The issue is not about eliminating the risk of re-identification altogether, but whether it can be mitigated so it is no longer significant. Organisations should focus on mitigating the risks to the point where the chance of re-identification is extremely remote. Organisations using anonymised data need to be able to demonstrate that they have carried out [a] robust assessment of the risk of re-identification, and have adopted solutions proportionate to the risk. This may involve a range and combination of technical measures, such as data masking, pseudonymisation, aggregation and banding, as well as legal and organisational safeguards.*

Finally, AoIR (2012) raises another issue of potential relevance to learning analytics which institutions may wish to consider: when stripping personally identifiable data from a dataset this might distort the dataset so that it no longer represents what it was intended to represent.

20: Taking data outside the institution

Outsourcing - question	
93. If we outsource the collection (and analysis) of student digital data to companies, do students need to give consent?	Prinsloo 2013

There are various potential scenarios for transferring personal data for learning analytics outside the student's institution which have particular ethical and legal implications:

- Institutions sharing data with each other to refine learning analytics algorithms and metrics
- Institutions using a third-party data hosting or analytics service
- Students taking their data with them when transferring to other institutions
- Employers wishing to view detailed records of job applicants' educational participation
- Requests for data from external agencies e.g. educational authorities or security agencies

Berg (2014) asks what happens when an external agency asks to be provided with historic student activity data. What is the responsibility of the institution if individual users can be identified? The Directive makes provisions for exceptions in the case of situations such as crime prevention or threats to national security.

Meanwhile, we may discover advantages to students in being able to merge datasets with those from other institutions. Should we be obliged ethically to collaborate in this way or are the increased risks to privacy the greater concern? What should students be told about the possibility of their data being transferred outside the institution (Berg 2014)? Pardo & Siemens (2014) suggest that if guarantees are given to students that their data will not be transferred outside the institution they may develop a higher level of trust in the process.

Around half of the participants in Draschler & Greller's (2012) study (of staff) thought anonymisation technologies would be effective in reducing the abuse of data. However 24% did not. Greller & Draschler (2012) list a set of challenges to be overcome in order for data to be shared:

- The lack of a common dataset
- The need for version control and a common reference system to distinguish and point to different datasets
- Methods to anonymise and pre-process data according to privacy and legal protection rights
- A standardised documentation of datasets so that others can make proper use of them
- Data policies (licences) that regulate how users can use and share certain datasets. For instance, the Creative Commons licensing rights could be considered as a standard way to grant permissions to datasets

These issues are compounded by the growing number of proprietary learning analytics products which do not meet the needs of researchers such as openness, accessibility and customisable tools and algorithms (Siemens 2014). The algorithms and metrics may be the key intellectual property in a learning analytics system, of most interest to researchers but also of most value to the vendors who may therefore be unlikely to welcome opening them up to scrutiny.

Reilly (2013) wonders about the ethical issues of employers potentially tracking the progress of students while they are studying and also speculates that learning analytics could become part of job applications, allowing employers to compare different applicants. In the US the use of student information by third party vendors is permitted without express consent of the student when a university uses educational records for predictive tests or enhancing learning. However, the data must not be released to outsiders and should be destroyed afterwards. The arrangement must be subject to a written agreement between the institution and the vendor. Meanwhile, “Statewide Longitudinal Data Systems” store data about individuals’ education from early childhood until they join the workforce; these are of increasing concern to privacy activists in the US. (Sun 2014).

ICO (2014) recommends if using a third party internet-based computing company that the following should be considered:

- Can it confirm in writing that it will only process data in accordance with your instructions and will maintain an appropriate level of security?
- Can it guarantee the reliability and training of its staff, wherever they are based? Do they have any form of professional accreditation?
- What capacity does it have for recovering from a serious technological or procedural failure?
- What are its arrangements and record regarding complaints and redress – does it offer compensation for the loss or corruption of data entrusted to it?
- If it is an established company, how good is its security track record?
- What assurances can it give that data protection standards will be maintained, even if the data is stored in a country with weak, or no, data protection law, or where governmental data interception powers are strong and lacking safeguards?
- Can it send you copies of your information regularly, in an agreed format and structure so that you hold useable copies of vital information at all times?

If the company cannot provide convincing responses to these questions then an alternative should be sought.

Willis & Pistilli (2014) wonder to what extent education should learn from the analytics taking place in the commercial sector and ask if algorithms developed to increase customers should be used for struggling students. One difference in the approaches they point out is likely to be portability. Business intelligence will be held onto closely by a commercial organisation but there are clear advantages for educational institutions in sharing anonymised data with each other.

Subsequent to the inBloom debacle, California has enacted a law which restricts how schoolchildren’s data can be used by educational technology companies. They are now unable to use students’ text messages, photos, locations or other data relating to them for selling, disclosing or marketing purposes. This updates a key federal law, the Family Educational Rights and Privacy Act, which is now forty years old. In a pledge developed by the Future of Privacy Forum, a Washington-based think tank, fourteen industry players including Microsoft have committed themselves not to use students’ data to target them with advertisements or to compile profiles on individuals - unless authorised by their parents. However, while they will be subject to the new law in California, neither Google nor Apple are currently participating in the Future of Privacy Forum (Singer 2014).

21: Staff awareness and training

Many codes of practice recommend that training is given to staff in ethical and legal issues. At Facebook, subsequent to the Mood experiment debacle training has been introduced into the six week "bootcamp" for new engineers as well as for those carrying out research (Schroepfer 2014). The RESPECT Project (2004) mentions the importance for researchers to critically question assumptions and ensure that outcomes are not predetermined. It is also thought to be important to demonstrate awareness of the limitations of any research and how the values and methods of the researchers may influence the outcomes. This would seem to be directly applicable to the algorithms and processes of learning analytics.

The Open University (2014a) is planning regular communications with staff about learning analytics, ensuring that they understand how it is being carried out, how it aligns to values of the institution, and what the benefits and limitations are. It intends to develop staff skills in the technologies, interpretation and understanding of ethical issues. At Loughborough University, personal tutors are already trained in what they should and should not record about students, and what they should do with sensitive personal information (Sclater 2014b).

22: Principles for a code of practice

Slade & Prinsloo (2013) discuss the potential development of a set of guidelines for learning analytics. They argue that any such document will be based on a limited set of epistemological assumptions and that it would therefore be difficult to produce guidelines that would be applicable in every context. They propose instead some general principles suggesting that institutions could use these to develop their own guidelines. Ethical principles are best applied contextually rather than using a one-size-fits-all approach (AoIR 2012). Kay, Korn & Oppenheim (2012) believe that there are dangers in assuming that ethics are universally applicable across domains and that education is ethically more sensitive than other sectors. Thus they argue principles developed for research, consumer services or social networks may not be directly applicable to learning analytics. Meanwhile, the American Psychological Association APA (2002) draws a distinction between principles and ethical standards. Only the latter, it believes, can be regarded as obligations. Due to rapid technological change, it says, it may be best for a code of practice to concentrate on decision-making processes and questions which can be applied in new contexts.

However, there is no doubt that much valuable thinking has been carried out in other fields which is transferable, and a summary of relevant codes of practice follows.

Chessell (2014) describes an "ethical awareness framework" developed by the UK and Ireland Technical Consultancy Group at IBM requiring organisations to consider:

IBM's *Ethical Awareness Framework*

1. **Context** – for what purpose was the data originally intended and what is it being used for now? How different are the uses?
2. **Consent and choice** – what choices are affected parties given, do they know they are making a choice, do they understand what they are agreeing to, do they have an opportunity to decline and what alternatives are offered?
3. **Reasonable** – is the depth and breadth of the data reasonable for the application it is used for?
4. **Substantiated** – are the data sources appropriate, authoritative, complete and timely?
5. **Owned** – who owns the insight and what is their obligation to act?
6. **Fair** – how equitable are the results of the application to all parties?
7. **Considered** – what are the consequences of the data collection and analyses?
8. **Access** – what access does the data subject have to their data?
9. **Accountable** – how are mistakes and unintended consequences detected and repaired, and can data subjects check the results?

Rayport (2011) proposes four principles:

Rayport's *Code of Ethical Principles for Big Data*

1. **Clarity on practices** – let users know about the data that is being collected about them in real time
2. **Simplicity of settings** – allow users to work out for themselves what level of privacy they want
3. **Privacy by design** – organisations should incorporate privacy protections in everything that they do
4. **Exchange of value** – show users what they get in exchange for sharing their personal information

"Privacy by design" (Cavoukian 2011) encourages organisations to build privacy protection for individuals in all aspects of their operations. She believes that complying with regulatory frameworks is insufficient to protect privacy. There are seven "foundational principles":

Privacy by design *Foundational Principles*

1. **Proactive not reactive; preventative not remedial** – prevent privacy breaches before they happen
2. **Privacy as the default setting** – no action is required by an individual to protect their privacy – it is built in by default
3. **Privacy embedded into design** – of architecture and business systems – not an afterthought – it becomes an essential component of the core functionality of the system
4. **Full functionality** – positive sum not zero-sum – avoid false dichotomies such as security vs privacy – it is possible to have both
5. **End-to-end security** – full lifecycle protection – all data are securely retained and destroyed at the end of the process
6. **Visibility and transparency** – keep it open – all component parts and operations remain visible and transparent to users and providers alike
7. **Respect for user privacy** – keep it user-centric – offer strong privacy defaults, appropriate notice and empowering user-friendly options

Van Rijmenam (2013) suggests four guidelines, three of them almost identical to those of Rayport:

Van Rijmenam's *Guidelines for Big Data Ethics*

1. **Radical transparency** – tell users in real-time what data is being collected about them and how you intend to use it; allow them to understand the data that has been collected and how to delete it
2. **Simplicity by design** – allow users to adjust privacy settings and decide what they want to share
3. **Preparation and data security are key** – define what information you need and what you can do without; develop a crisis strategy in case any data is stolen
4. **Make privacy part of the DNA** – embracing transparency, security and simplicity means that users will embrace your organisation

Richards and King (2014) describe four "normative values" for big data ethics:

Richards and King's *Normative Values for Big Data Ethics*

1. **Privacy** – more than just keeping information secret, this is about defining rules for information flows
2. **Confidentiality** – a kind of privacy based on trust and promises between individuals and other parties
3. **Transparency** – this is about building trust by holding others accountable and is about rebalancing the power being the institutions which hold huge amounts of data on individuals and the secrecy with which they cloak their operations
4. **Identity** – this relates to the fundamental right we have to define who we are, and consequently not to be defined solely by an algorithm or the data held about us

Schwartz (2011) has developed some analytics principles for industry with an intention to "maximize good results and minimize bad ones for individuals whose information is processed." He sidesteps the ethical issue of whether to carry out analytics if they have a negative impact on the individual despite proving beneficial to the majority. However, his principles have direct relevance for individual institutions. He looks at privacy issues in relation to what he considers the four stages of analytics: collection, integration and analysis, decision making, and review and revision.

Schwartz's *Overarching Ethical Standards for Analytics*

1. Comply with legal requirements
2. Assess, beyond legal requirements, whether the process reflects cultural and social norms about acceptable activities
3. Assess the impact on stakeholders' trust in the organisation
4. Use accountable measures and acknowledge that there could be negative as well as positive impacts on individuals; develop policies for information governance and management; designate individuals to oversee data processing operations and decision making
5. Protect the security of information used for analytics
6. Assess whether the analytics involve sensitive areas; if so add safeguards appropriate to the risk

At the collection stage Schwartz suggests that organisations should avoid gathering certain information and consider legal, cultural and social factors, as well as risks both to themselves and to the individuals. In integration and analysis, he advocates considering the quality of the data and anonymising personal information if appropriate. For Decision making he proposes that the greater impact a decision has on an individual, the more accurate the data should be. Finally, organisation should engage in ongoing review and revision of their analytics processes, ensuring that personal information is relevant and being responsive to unforeseen consequences.

PCAST (2014) presents the Consumer Privacy Bill of Rights, issued by the US Administration in February 2012, comprising a number of obligations on providers:

US Administration's *Consumer Privacy Bill of Rights*

1. **Respect for Context** - Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data
2. **Focused Collection** - Consumers have a right to reasonable limits on the personal data that companies collect and retain
3. **Security** - Consumers have a right to secure and responsible handling of personal data
4. **Accountability** - Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights

There are also "consumer empowerments":

5. **Individual Control** - Consumers have a right to exercise control over what personal data companies collect from them and how they use it
6. **Transparency** - Consumers have a right to easily understandable and accessible information about privacy and security practices
7. **Access and Accuracy** - Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data are inaccurate

The UK Statistics Authority (2009) has developed a Code of Practice for Official Statistics which build on the Civil Service "core values":

UK Statistics Authority *Code of Practice for Official Statistics*

1. **Integrity** – putting the public [student in the context of education] interest above organisational, political or personal interests
2. **Honesty** – being truthful and open about the statistics and their interpretation
3. **Objectivity** – using scientific methods to collect statistics and basing statistical advice on rigorous analysis of the evidence
4. **Impartiality** – acting solely according to the merits of the statistical evidence, serving equally well all aspects of the public [student] interest

“Social justice principles” for managing student learning engagement (MSLE) are the theme of a report sponsored by the Australian Government by Nelson and Creagh (2013). While these principles do not relate so strongly to learning analytics as some of the others they reflect similar values of fairness and openness.

Nelson & Creagh’s *Social Justice Principles for Managing Student Learning Engagement*

Self-determination - Students participate in programme design, enactment and evaluation, and make informed decisions about their individual participation in the programme.

1. **Rights** - MSLE initiatives should ensure that all students are treated with dignity and respect and have their individual cultural, social and knowledge systems recognised and valued
2. **Access** - Programmes are designed to serve as active and impartial conduits to the resources of the institution (for example, curriculum, learning, academic, social, cultural, support, financial and other resources)
3. **Equity** - Programmes are designed to demystify and decode dominant university cultures, processes, expectations and language for differently prepared cohorts
4. **Participation** - MSLE programs lead to socially inclusive practices and students experience a sense of belonging and connectedness

The Economic and Social Research Council (ESRC - 2012) presents six principles of ethical research which it expects those it funds to adhere to “whenever applicable”.

ESRC *Principles of Ethical Research*

1. Research should be designed, reviewed and undertaken to ensure integrity, quality and transparency
2. Research staff and participants must normally be informed fully about the purpose, methods and intended possible uses of the research, what their participation in the research entails and what risks, if any, are involved
3. The confidentiality of information supplied by research participants and the anonymity of respondents must be respected
4. Research participants must take part voluntarily, free from any coercion
5. Harm to research participants and researchers must be avoided in all instances
6. The independence of research must be clear, and any conflicts of interest or partiality must be explicit

The Asilomar Convention for Learning Research in Higher Education (2014) has six principles:

Asilomar Convention for Learning Research in Higher Education Principles

1. **Respect for the rights and dignity of learners** - Data collection, retention, use, and sharing practices must be made transparent to learners, and findings made publicly available, with essential protections for the privacy of individuals. Respect for the rights and dignity of learners requires responsible governance by institutional repositories and users of learner data to ensure security, integrity, and accountability. Researchers and institutions should be especially vigilant with regard to the collection and use of identifiable learner data, including considerations of the appropriate form and degree of consent

2. **Beneficence** - Individuals and organizations conducting learning research have an obligation to maximize possible benefits while minimizing possible harms. In every research endeavour, investigators must consider potential unintended consequences of their inquiry and misuse of research findings. Additionally, the results of research should be made publicly available in the interest of building general knowledge

3. **Justice** - Research practices and policies should enable the use of learning data in the service of providing benefit for all learners. More specifically, research practices and policies should enable the use of learning data in the service of reducing inequalities in learning opportunity and educational attainment

4. **Openness** - Learning and scientific inquiry are public goods essential for well-functioning democracies. Learning and scientific inquiry are sustained through transparent, participatory processes for the scrutiny of claims. Whenever possible, individuals and organizations conducting learning research have an obligation to provide access to data, analytic techniques, and research results in the service of learning improvement and scientific progress

5. **The humanity of learning** - Insight, judgment, and discretion are essential to learning. Digital technologies can enhance, do not replace, and should never be allowed to erode the relationships that make learning a humane enterprise

6. **Continuous consideration** - In a rapidly evolving field there can be no last word on ethical practice. Ethically responsible learner research requires ongoing and broadly inclusive discussion of best practices and comparable standards among researchers, learners, and educational institutions

Kay, Korn and Oppenheim (2012) suggest four principles to be taken into account for educational institutions carrying out analytics:

CETIS *Analytics Principles*

1. **Clarity**, open definition of purpose, scope and boundaries, even if that is broad and in some respects open-ended
2. **Comfort and care**, consideration for both the interests and the feelings of the data subject and vigilance regarding exceptional cases
3. **Choice and consent**, informed individual opportunity to opt-out or opt-in
4. **Consequence and complaint**, recognition that there may be unforeseen consequences and therefore providing mechanisms for redress

Dringus (2012) presents her argument around five "must statements" for learning analytics, which could be regarded as principles:

Dringus' *Must Statements for Learning Analytics*

Effective learning analytics in online courses:

1. MUST develop from the stance of getting the right data and getting the data right. What is meaningful data? If the data trail produces no meaningful evidence of learning or non-learning, or has no impact on changing instructional design or practice then it is ineffective
2. MUST have transparency. What do we see? The data must allow the visualisation of learners' status (e.g. Success, Failure or At-Risk)
3. MUST yield from good algorithms. What are we looking for? Learning analytics can be harmful if naively done by rule e.g. presenting number of postings in a forum is not necessarily good evidence of student participation in learning
4. MUST lead to responsible assessment and effective use of the data trail. What do we do with the data? Again analytics can be harmful if naively carried out by rule e.g. data showing absence is interpreted only as the student being "inactive" or a "no-show". There could be more complex reasons
5. MUST inform process and practice. How do we improve the online experience with good algorithms, the promotion of self-reflection and the development of communities of practice

Pardo and Siemens (2014) provide some principles for learning analytics research (not necessarily practice):

Pardo and Siemens' *Principles for Learning Analytics Research*

1. **Transparency** – all stakeholders should have access to details of what data is being collected, how it is collected, stored and processed, and how the analytics are being carried out
2. **Student control over data** – students need to know what data is collected, when, how and how it has been manipulated; they should be able to correct inaccurate data
3. **Right of access** – a detailed access policy should be developed, specifying the type of operations permitted and the access rights for each type of user
4. **Accountability and assessment** – every aspect of learning analytics should have a person or unit designated as responsible for its proper functioning; the institution should continually evaluate and refine areas such as data collection, security and transparency

The Open University (2014a) has published a Policy on Ethical Use of Student Data for Learning Analytics with eight principles, some of which build on the research of Sharon Slade and Paul Prinsloo.

The Open University's *Policy on Ethical Use of Student Data for Learning Analytics*

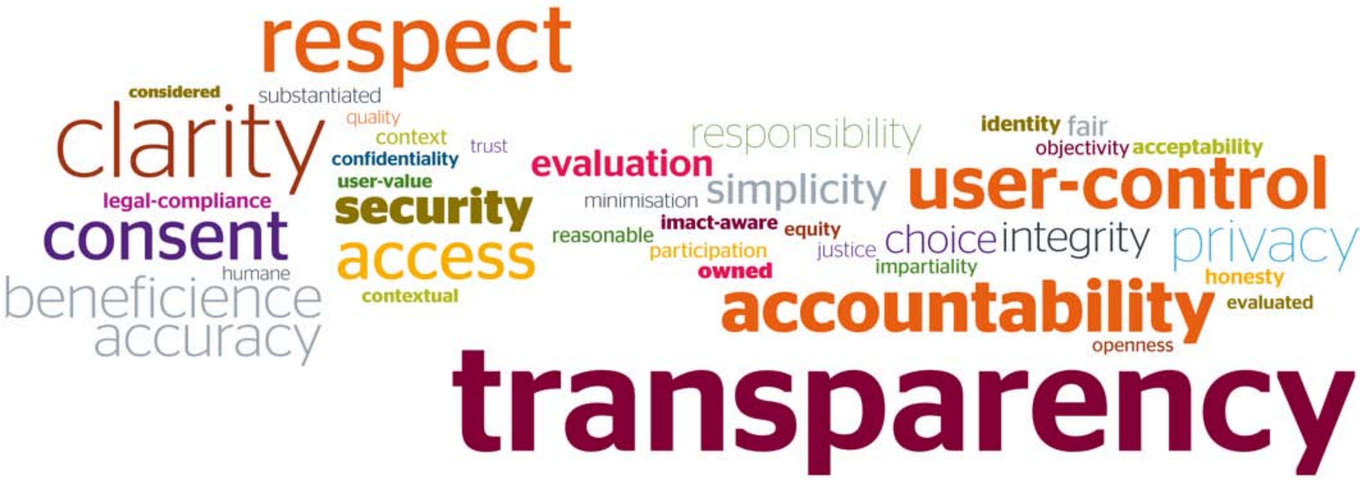
1. Learning analytics is an ethical practice that should align with core organisational principles, such as open entry to undergraduate level study
2. The OU has a responsibility to all stakeholders to use and extract meaning from student data for the benefit of students where feasible
3. Students should not be wholly defined by their visible data or our interpretation of that data
4. The purpose and the boundaries regarding the use of learning analytics should be well defined and visible
5. The University is transparent regarding data collection, and will provide students with the opportunity to update their own data and consent agreements at regular intervals
6. Students should be engaged as active agents in the implementation of learning analytics (e.g. informed consent, personalised learning paths, interventions)
7. Modelling and interventions based on analysis of data should be sound and free from bias
8. Adoption of learning analytics within the OU requires broad acceptance of the values and benefits (organisational culture) and the development of appropriate skills across the organisation

It is also important to consider the implications for staff and students of the development and rollout of new learning analytics systems. BCS (2002) has some principles which may be helpful:

British Computer Society's *Code of Good Practice*

1. Be aware of the impact of new or changed business solutions on people's working [or study] lives and deal sensitively with them
2. Avoid solutions that impose unacceptable levels of risk on their physical or mental well-being
3. When analysing current practices, show respect for people at all levels in the organisation and assure them that their views will be taken into account
4. Demonstrate an understanding of the business issues; be persuasive and explain to users and management, in language they understand, the benefits of the changes being introduced, as well as identifying any drawbacks and trade-offs
5. Document the results of your analysis in a style that can be understood by the users and the developers
6. Explain your analysis methods to the users and encourage them to understand the results and verify their correctness

The principles from these varied fields have much in common, reflect fundamental human values as well as suggesting pragmatic solutions, and are almost all potentially relevant to a code of practice for learning analytics. Key concepts which appear frequently in the above ethical codes and guidelines as can be seen from the word cloud below are: transparency, clarity, respect for users and user control. Consent, accountability and access also feature prominently. The challenge now is to build on the work in learning analytics and other domains discussed in this review to create a comprehensive code of practice, providing practical guidance to institutions which wish to understand and tackle the complex ethical and legal issues involved in implementing learning analytics.



References

AERA (American Educational Research Association). (2011, February). **Code of Ethics**. Retrieved from: aera.net/AboutAERA/AERARulesPolicies/CodeofEthics/tabid/10200/Default.aspx

Aguilar, S., Lonn, S., & Teasley, S. D. (2014). **Perceptions and Use of an Early Warning System During a Higher Education Transition Program**. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 113–117). New York, NY, USA: ACM.

American Psychological Association. (2002, August 21). **Ethical Principles of Psychologists and Code of Conduct.** Retrieved from: apa.org/ethics/code/index.aspx

AoIR (Association of Internet Researchers). (2012, August**). Ethical Decision-Making and Internet Research**. Retrieved from: http://aoir.org/reports/ethics2.pdf

ASA (American Sociological Association). (1999). **Code of Ethics**. Retrieved from: asanet.org/about/ethics.cfm

BCS (The British Computer Society). (2004, September 1). **Code of Good Practice**. Retrieved from: http://bcs.org/upload/pdf/cop.pdf

BERA (British Educational Research Association). (2011). **Ethical Guidelines for Educational Research**. Retrieved from: bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2011

Berg, A. (2013). **Towards a uniform code of ethics and practices for Learning Analytics**. Retrieved from: http://ict-innovatie.uva.nl/2013/09/13/towards-a-uniform-code-of-ethics-and-practices-for-learning-analytics/

Bertolucci, J. (2014, September 17). **Data Scientists Want Big Data Ethics Standards**. Retrieved October 7, 2014, from: informationweek.com/big-data/big-data-analytics/data-scientists-want-big-data-ethics-standards/d/d-id/1315798

Bichsel, J. (2012, August). **Analytics in Higher Education: Benefits, Barriers, Progress and Recommendations**. Educause Center for Applied Research. Retrieved from: educause.edu/ecar

Bollier, D. (2010). **The Promise and Peril of Big Data**. Retrieved October 7, 2014, from: aspeninstitute.org/publications/promise-peril-big-data

Burton, G. (2014, August 1). **The ethical dilemmas of big data**. Retrieved October 7, 2014, from: computing.co.uk/ctg/feature/2358287/the-ethical-dilemmas-of-big-data

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). **Academic Analytics: A New Tool for a New Era**. EDUCAUSE Review, 42(4), 40–57. Retrieved November 11, 2014, from: educause.edu/ero/article/academic-analytics-new-tool-new-era

Cavoukian, A. (2011, January**). Privacy by Design: The 7 Foundational Principles**. OIPCO (Office of the Information and Privacy Commissioner/Ontario). Retrieved from: privacybydesign.ca/index.php/about-pbd/translations

Charles Sturt University. (2014). **CSU Learning Analytics Strategy v1.3**. Retrieved from: csu.edu.au/__data/assets/pdf_file/0010/543934/201305-CSULearningAnalyticsStrategyv1.3.pdf

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). **A Reference Model for Learning Analytics**. *Int. J. Technol. Enhanc. Learn.*, 4(5/6), 318–331. doi:10.1504/IJTEL.2012.051815

Chessell, M. (2014). **Ethics for big data and analytics**. IBM. Retrieved from: ibmbigdatahub.com/sites/default/files/whitepapers_reports_file/TCG%20Study%20Report%20-%20Ethics%20for%20BD%26A.pdf

Clow, D. (2012). **The Learning Analytics Cycle: Closing the Loop Effectively**. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (pp. 134–138). New York, NY, USA: ACM. doi:10.1145/2330601.2330636

Contact North. (2012). **The learning analytics challenge: Culture of data or culture of evidence?** Contact North. Retrieved from: http://contactnorth.ca/trends-directions/learning-analytics

Crawford, K., & Schultz, J. (2013). **Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms** (SSRN Scholarly Paper No. ID 2325784). Rochester, NY: Social Science Research Network. Retrieved from: http://papers.ssrn.com/abstract=2325784

Drachsler, H., & Greller, W. (2012). **The Pulse of Learning Analytics: Understandings and Expectations from the Stakeholders.** In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 120–129). New York, NY, USA: ACM. doi:10.1145/2330601.2330634

Dringus, L. P. (2012). **Learning Analytics Considered Harmful**. *Journal of Asynchronous Learning Networks*, *16*(3), 87–100.

Duquenoy, P., Dando, N., & Harris, I. (2010, January). **Ethics in the Provision and Use of IT for Business. Institute of Business Ethics**. Retrieved from: ibe.org.uk/userassets/publicationdownloads/ibe_occasional_paper_1_ethics_in_the_provision_and_use_of_it_for_business.pdf

EFPA (European Federation of Psychologists' Associations. (2005, July). **Meta-Code of Ethics**. Retrieved from http://ethics.efpa.eu/meta-code/

Ellis, C. (2013). **Broadening the scope and increasing the usefulness of learning analytics**: The case for assessment analytics. *British Journal of Educational Technology*, *44*(4), 662–664. doi:10.1111/bjet.12028

ESOMAR. (2011). **Guideline for Online Research**. Retrieved from: esomar.org/knowledge-and-standards/codes-and-guidelines/guideline-for-online-research.php

Esposito, A. (2012). **Research ethics in emerging forms of online learning: issues arising from a hypothetical study on a MOOC**. *The Electronic Journal of E-Learning*, *10*(3), 286–296.

ESRC (Economic and Social Research Council). (2012, September). **Framework for Research Ethics**. Retrieved from: esrc.ac.uk/about-esrc/information/research-ethics.aspx

Ess, C., & AoIR (Association of Internet Researchers). (2002, November 27). **Ethical Decison-Making and Internet Research**. Retrieved from: http://aoir.org/reports/ethics.pdf

European Commission. (1995). **Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data**. *Official Journal of the EC*, *23*(6). Retrieved from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1995L0046:20031120:EN:PDF

European Commission. (2012, January 25). **Commission proposes a comprehensive reform of the data protection rules**. Retrieved from: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm

European Commission. (2014a, March 12). **Progress on EU data protection reform now irreversible following European Parliament vote**. Retrieved from: http://europa.eu/rapid/press-release_MEMO-14-186_en.htm

European Commission. (2014b, April 9). **Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC**. Retrieved from: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

Ferguson, R. (2012). **Learning analytics: drivers, developments and challenges**. *International Journal of Technology Enhanced Learning*, 4(5/6), 304. Retrieved from: http://oro.open.ac.uk/36374/

Google. (2012, April 25). **Code of Conduct**. Retrieved from: http://investor.google.com/corporate/code-of-conduct.html

Greller, W., & Drachsler, H. (2012). **Translating Learning into Numbers: A Generic Framework for Learning Analytics**. *Educational Technology & Society*, 42–57. Retrieved from: researchgate.net/publication/234057371_Translating_Learning_into_Numbers_A_Generic_Framework_for_Learning_Analytics/links/0912f50ead2877d5b6000000

Harris, I., Jennings, R., Pullinger, D., Rogerson, S., & Duquenoy, P. (2008). **Helping ICT professionals to assess ethical issues in new and emerging technologies**. Presented at the MINAmI workshop on ambient intelligence and ethics, University of Pavia, Mantua, Italy: British Computer Society. Retrieved from: http://bcs.org/upload/pdf/assessing-ethical-issues.pdf

Hoel, T., Pirkkalainen, H., Clements, K., Richter, T., Kretschmer, T., & Stracke, C. M. (2014). **Learning Analytics – Ethical questions and dilemmas**. Retrieved from: slideshare.net/toreh/learning-analytics-ethical-questions-and-dilemmas

ICO (Information Commissioner's Office). (2010, July). **Personal information online code of practice**. Retrieved from: http://ico.org.uk/~/media/documents/library/Data_Protection/Detailed_specialist_guides/personal_information_online_cop.pdf

ICO (Information Commissioner's Office). (2014). **Big Data and Data Protection**. Retrieved from: http://ico.org.uk/news/latest_news/2014/~/media/documents/library/Data_Protection/Practical_application/big-data-and-data-protection.pdf

IWGDPT (International Working Group on Data Protection in Telecommunications). (2014, May 5). **Working Paper on Big Data and Privacy: Privacy principles under pressure in the age of Big Data analytics.** Retrieved from: datenschutz-berlin.de/attachments/1052/WP_Big_Data_final_clean_675.48.12.pdf

Johnson, J. A. (2014). **The Ethics of Big Data in Higher Education**. *International Review of Information Ethics*. Retrieved from i-r-i-e.net/inhalt/021/IRIE-Britz-Zimmer.pdf

Kay, D., Korn, N., & Oppenheim, C. (2012, November). **Legal, Risk and Ethical Aspects of Analytics in Higher Education**. CETIS, University of Bolton. Retrieved from: http://publications.cetis.ac.uk/2012/500

King, J. H., & Richards, N. M. (2014, March 28). **What's Up With Big Data Ethics?** *Forbes*. Retrieved from: forbes.com/sites/oreillymedia/2014/03/28/whats-up-with-big-data-ethics/

K.N.C. (2014, April 30). **Withered inBloom**. *The Economist*. Retrieved from: economist.com/node/21601484/print

MacCarthy, M. (2014). **Student Privacy: Harm and Context**. *International Review of Information Ethics*, *21*, 11–24. Retrieved from: i-r-i-e.net/inhalt/021/IRIE-021-MacCarthy.pdf

McDonald, A., & Cranor, L. F. (2008). **The Cost of Reading Privacy Policies**. *I/S: A Journal of Law and Policy for the Information Society*, *Privacy Year in Review issue*. Retrieved from: aleecia.com/authors-drafts/readingPolicyCost-AV.pdf

Mason, R. O., Mason, F. M., & Culnan, M. J. (1995). **Ethics of Information Management (1st ed.).** Thousand Oaks, CA, USA: Sage Publications, Inc.

Ministry of Justice. (2011, April 14). **Public Sector Data Sharing: Guidance on the Law**. Retrieved from: justice.gov.uk/downloads/information-access-rights/data-sharing/annex-h-data-sharing.pdf

Nelson, K., & Creagh, T. (2013). **A Good Practice Guide: Safeguarding Student Learning Engagement**. Queensland University of Technology. Retrieved from: https://safeguardingstudentlearning.net/?page_id=62

OPCC (Office of the Privacy Commissioner of Canada). (2012, August). **The Age of Predictive Analytics: From Patterns to Predictions**. Retrieved from: priv.gc.ca/information/research-recherche/2012/pa_201208_e.asp

Pardo, A., & Siemens, G. (2014). **Ethical and privacy principles for learning analytics**. *British Journal of Educational Technology*, *45*, 438–450.

PCAST (President's Council of Advisors on Science and Technology). (2014, May). **Big Data and Privacy: A Technological Perspective**. Retrieved from: whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

Polonetsky, J., & Tene, O. (2014). **The Ethics of Student Privacy: Building Trust for Ed Tech**. *International Review of Information Ethics*, *21*. Retrieved from: i-r-i-e.net/inhalt/021/IRIE-021-Polonetsky-Tene.pdf

Prinsloo, P. (2013). **Ethics and Learning Analytics as a Faustian Pact: Between Orwell, Huxley, Kafka and the Deep Blue Sea**. Retrieved from: slideshare.net/prinsp/lasi13-za-5-july2013-final-1-paul-prinsloo

Prinsloo, P., & Slade, S. (2013). **An evaluation of policy frameworks for addressing ethical considerations in learning analytics** (pp. 240–244). Presented at the Third Conference on Learning Analytics and Knowledge (LAK 2013), Leuven, Belgium: ACM Press. Retrieved from: http://oro.open.ac.uk/36934/

Prinsloo, P., & Slade, S. (2014). **Educational triage in open distance learning: Walking a moral tightrope**. *International Review of Open and Distance Learning*, *15*(4). Retrieved from: http://oro.open.ac.uk/40903/

Prinsloo, P., Slade, S., & Galpin, F. (2012). **Learning Analytics: Challenges, Paradoxes and Opportunities for Mega Open Distance Learning Institutions**. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 130–133). New York, NY, USA: ACM. doi:10.1145/2330601.2330635

Rayport, J. F. (2011). **What Big Data Needs: A Code of Ethical Practices**. *MIT Technology Review*. Retrieved from: technologyreview.com/news/424104/what-big-data-needs-a-code-of-ethical-practices/

Reilly, M. (2013, August). **Further Education Learning Technology: A horizon scan for the UK Government Foresight Horizon Scanning Centre**. Ariel Research Services. Retrieved from: arielresearchservices.com/wp-content/uploads/2014/03/Further-Education-and-Learning-Technology-Final-Draft.pdf

Research at Facebook | **Facebook Newsroom**. (2014). Retrieved from: http://newsroom.fb.com/news/2014/10/research-at-facebook/

RESPECT Project. (2004). **Code of Practice for Socio-Economic Research. Institute for Employment Studies**. Retrieved from respectproject.org/code/respect_code.pdf

Richards, J. H., & King, N. M. (2014). **Big Data Ethics**. *Wake Forest Law Review*. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174

Rivers, C. M., & Lewis, B. L. (2014). **Ethical research standards in a world of big data**. *F1000Research*. doi:10.12688/f1000research.3-38.v2 Retrieved from: http://f1000research.com/articles/3-38/v2

Rubinstein, I. S. (2013). **Big Data: The End of Privacy or a New Beginning?** *International Data Privacy Law*, ips036. doi:10.1093/idpl/ips036 Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2157659

Schwartz, P. M. (2010). **Data Protection Law and the Ethical Use of Analytics**. The Center for Information Policy Leadership, Hunton and Williams LLP. Retrieved from: privacyassociation.org/media/pdf/knowledge_center/Ethical_Underpinnings_of_Analytics.pdf

Schwartz, P. M. (2011). **Privacy, Ethics, and Analytics**. *IEEE Security & Privacy*, *9*(3), 66–69. Retrieved from: paulschwartz.net/pdf/pschwartz_privacy-eth-analytics%20IEEE%20P-%20Sec%20(2011).pdf

Sclater, N. (2014a, September 18). **Code of practice "essential" for learning analytics**. Retrieved from: http://analytics.jiscinvolve.org/wp/2014/09/18/code-of-practice-essential-for-learning-analytics/

Sclater, N. (2014b, November). **Learning analytics: the current state of play in UK higher and further education**. Jisc.

Shroepfer, M. (2014, October 2). **Research at Facebook | Facebook Newsroom** [Facebook Newsroom]. Retrieved from http://newsroom.fb.com/news/2014/10/research-at-facebook/

Siemens, G. (2012). **Learning Analytics: Envisioning a Research Discipline and a Domain of Practice**. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (pp. 4–8). New York, NY, USA: ACM. doi:10.1145/2330601.2330605

Singer, N. (2014, October 7). **Microsoft and Other Firms Pledge to Protect Student Data**. *The New York Times*. Retrieved from: nytimes.com/2014/10/07/business/microsoft-and-other-firms-pledge-to-protect-student-data.html

Slade, S., & Galpin, F. (2012). **Ethical issues in learning analytics**. Retrieved from: slideshare.net/SharonSlade/ethical-issues-in-learning-analytics

Slade, S., & Prinsloo, P. (2013**). Learning Analytics: Ethical Issues and Dilemmas**. *American Behavioral Scientist*, 0002764213479366. doi:10.1177/0002764213479366 Retrieved from: http://oro.open.ac.uk/36594/

Sun, J. C. (2014). **Legal issues associated with big data in higher education: ethical considerations and cautionary tales**. In *Building a Smarter University: Big Data, Innovation, and Analytics*. SUNY Press.

Swenson, J. (2014). **Establishing an Ethical Literacy for Learning Analytics**. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 246–250). New York, NY, USA: ACM. doi:10.1145/2567574.2567613

The Asilomar Convention for Learning Research in Higher Education. (2014, June). Retrieved from: http://asilomar-highered.info/index.html

The Open University. (2014a, September). **Policy on Ethical use of Student Data for Learning Analytics**. Retrieved from: open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf

The Open University. (2014b, October**). Ethical use of Student Data for Learning Analytics Policy FAQs**. Retrieved from: open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-student-data-faq.pdf

Traxler, J., & Bridges, N. (2005**). Mobile learning - the ethical and legal challenges**. In *Mobile learning anytime anywhere*. Learning and Skills Development Agency. Retrieved from: http://stu.westga.edu/~bthibau1/MEDT%208484-%20Baylen/mLearn04_papers.pdf#page=212

UK Government. (1998). **Data Protection Act**. Retrieved from: legislation.gov.uk/ukpga/1998/29/contents

UK Statistics Authority. (2009, January). **Code of Practice for Official Statistics**. Retrieved from: statisticsauthority.gov.uk/assessment/code-of-practice/

Van Rijmenam, M. (2013, March 11). **Big Data Ethics: 4 Guidelines to Follow by Organisations**. Retrieved from: bigdata-startups.com/big-data-ethics-4-principles-follow-organisations/

Willis, J., III. (2014). **Learning Analytics and Ethics: A Framework beyond Utilitarianism**. *EDUCAUSE Review Online*. Retrieved from: educause.edu/ero/article/learning-analytics-and-ethics-framework-beyond-utilitarianism

Willis, J. E., III, Campbell, J. P., & Pistilli, M. D. (2013**). Ethics, Big Data and Analytics: A Model for Application**. *EDUCAUSE Review Online*. Retrieved from: educause.edu/ero/article/ethics-big-data-and-analytics-model-application

Willis, J. E., III, & Pistilli, M. D. (2014**). Ethical Discourse: Guiding the Future of Learning Analytics**. *EDUCAUSE Review Online*. Retrieved from: educause.edu/ero/article/ethical-discourse-guiding-future-learning-analytics