JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

A Data Protection Framework for Learning Analytics

Andrew Cormack, Jisc, UK
Andrew.Cormack@jisc.ac.uk

ABSTRACT: Most studies on the use of digital student data adopt an ethical framework derived from human-subject research, based on the informed consent of the experimental subject. However, consent gives universities little guidance on using learning analytics as a routine part of educational provision: which purposes are legitimate and which analyses involve an unacceptable risk of harm. Obtaining consent when students join a course will not give them meaningful control over their personal data three or more years later. Relying on consent may exclude those most likely to benefit from early intervention. This paper proposes a new framework based on the approach used in data protection law. Separating the processes of analysis (pattern-finding) and intervention (pattern-matching) gives students and staff continuing protection from inadvertent harm during data analysis. Students have a fully informed choice whether or not to accept individual interventions. Organizations obtain clear guidance: how to conduct analysis, which analyses should not proceed, and when and how interventions should be offered. The framework provides formal support for practices already being adopted and helps with several open questions in learning analytics, including its application to small groups and alumni, automated processing, and privacy-sensitive data.

Keywords: Learning analytics, privacy, data protection, consent, legitimate interests

1 INTRODUCTION

Learning analytics has been defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long & Siemens, 2011, p. 33). Analyzing learner data could inform a university's management of finances, resources, and enrollment; enhance its course presentation and materials; or enable personalized guidance and intervention for individual students and staff (Leece, 2013). Analytics could improve both the provision of learning to future students and the support and guidance offered to current teachers and students, both at the cohort and at the individual level.

Examples such as Purdue University's Course Signals show how data from past students can help new undergraduates make the transition from school to university learning — "like sitting next to somebody who has been through the class before" (Mathewson, 2015). Tickle (2015) reports that analysis of current interactions can trigger timely support before a student's problem becomes insurmountable. The National Union of Students (2015) identify the "massive power and potential" to tackle challenges facing UK higher education, but concerns are also raised about "students under surveillance" (Warrell, 2015). Analytics can, indeed, use of a very wide range of both historic and current personal data "from formal transactions (such as assignment submissions) to involuntary data exhaust (such as building access, system logins, keystrokes and click streams)" (Kay, Korn, & Oppenheim, 2012, p. 9). As in other applications of Big Data,

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

the technology itself may be ethically neutral but the use of it is not (Wen, 2012). Universities and their students need an answer to Tickle's question: "How comprehensive and intrusive should data collection be?"

To date, learning analytics has largely been conducted within an ethical framework originally developed for medical research. Individuals' informed consent provides the foundation for all analysis and intervention. However, as learning analytics becomes a normal part of educational provision, this paper suggests that a different ethical basis is required. Treating large-scale learning analytics as a form of human-subject research may no longer provide appropriate safeguards. Instead the paper suggests that the broader approach taken by European law to protect personal data offers clearer guidance, helping both organizations and students assess when and how analytics should, and should not, be used. Different measures may be needed at different stages of an activity. In particular separating the processes of analysis and intervention provides clearer guidance and stronger safeguards for both. Examples are given of how this new ethical framework can inform questions commonly raised in learning analytics. This approach should help learning analytics processes to be trusted by both students and organizations, delivering the greatest benefit for current and future students, teachers, educational organizations, and society.

2 THE NEED FOR A LEARNING ANALYTICS FRAMEWORK

Learning analytics shares many characteristics with the well-established processes of educational data mining (Prinsloo & Slade, 2013) and website personalization (Pardo & Siemens, 2014). Like them, it seeks to "understand and optimize learning and the environment in which it occurs" (Pardo & Siemens, 2014, p. 443). However, both the nature and scale of data available and the expectations of students and society are changing to create new opportunities and new risks. This may require a new approach to maintain trust between educational organizations, staff, and students.

Prinsloo and Slade see "the increasing digitisation of education, technological advances, the changing nature and availability of data" (2013, p. 244) as creating important opportunities. Analyzing these data could help researchers understand the factors that contribute to "the effectiveness of learning, student success and retention" (ibid.) and lead to improvements in the future provision of education. The data trails created by students interacting with digital systems also raise the possibility of analysis in near real-time. Detailed information about individual students' learning and engagement might be used to offer "personalized and customized curricula, assessment and support to improve learning and retention" (2013, p. 240), thus benefitting current students as well.

Such uses of personal data may be increasingly accepted, even expected, both by individual students and by society. According to Kay et al., "[u]sers, especially born digital generations, appear increasingly to expect personalized services that are responsive to profile, need and interest and are therefore more likely to be content for their data to be used to those ends" (2012, p. 4), while Pardo and Siemens find

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

"society seems to be evolving toward a situation in which the exchange of personal data is normal" (2014, p. 440).

Learning analytics, however, by consuming increasing quantities of varied and real-time data, also shares characteristics with Big Data. It creates the possibility, identified by Mayer-Schönberger and Cukier, "to extract new insights" (2013, p. 6) that change organizations and their relationship with individuals. In the past, data have been analyzed to test prior hypotheses that have been assessed against ethical guidelines. Before analysis (or even data collection) begins, all possible outcomes have already been checked for ethical acceptability. But "big data thrives on surprising correlations and produces inferences and predictions that defy human understanding" (Ohm, 2014, p. 100). According to the Article 29 Working Party of European data protection authorities (2014b) the possibility of analysis revealing entirely unexpected information about people "raises important social, legal and ethical questions, among which [are] concerns with regard to … privacy and data protection rights." For example, combining datasets from different sources can result in accidental re-identification of individuals (Ohm, 2010) or an unexpected correlation between two factors might suggest a causal link that does not in fact exist. This may be a particular challenge in education. Contrasting with Big Data collected in the retail sector, for example through loyalty cards, Pardo and Siemens consider that

> Educational institutions pose a new scenario with specific requirements. Students interact very intensively with the university (or its computational platforms) during a concrete time frame, carrying out very specific tasks, and produce highly sensitive data during the process. These special conditions prompt the need for a revision of the privacy principles with respect to analytics and their application in educational settings. (2014, p. 448)

But the relationship between the individual and the organization is also very different. Universities and their students have a long-term relationship and a strong mutual interest in improving learning processes. Given a suitable ethical framework, it should be possible for universities to use learner data safely, in ways that benefit both student and organization far more than a simple economic bargain exchanging shopping habits for discount vouchers.

It is clear that, as Pardo and Siemens suggest, in learning analytics "a delicate balance between control and limits needs to be achieved" (2014, p. 440). However, Prinsloo and Slade worry that "current policy frameworks do not facilitate the provision of an enabling environment for learning analytics to fulfil its promise" (2013, p. 240). Those frameworks have largely been based on practice in human-subject research. Thus Kay et al. find the Nuremberg Code's principle that "research participants must voluntarily consent to research participation" is "highly relevant" to analytics. They regard consent as "fundamental" (2012, p. 20). According to Sclater, "informed consent is recognized as key to the analysis of learner data by many commentators and is a basic principle of scientific research on humans" (2014, p. 16). At the University of South Africa "the Policy on Research Ethics makes informed consent and anonymity non-negotiable" (Prinsloo & Slade, 2013, p. 242).

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

However Kay et al. note that in other fields different ethical approaches are considered "self-evident good practice" (2012, p. 26). For example,

> The practice adopted by leading business to consumer players provides a clear and legally grounded approach that is likely to be readily understood by the public in much of the world. In particular, the development of a sense of mutual gain, recognized and shared by a service organization and its customers, is something to be learned from such [organizations] as Amazon and Nectar. (p. 24)

They conclude,

> The challenge is whether the education community, not least in the emerging field of learning analytics, should revise its ethical position on account of the widespread changes of attitude in the digital realm from which learners and researchers are increasingly drawn. (p. 26)

The following sections of the paper will consider what problems are likely to arise from continuing to rely on "informed consent" as learning analytics techniques transfer from university research to university operations and whether a more fruitful ethical framework can be derived from European data protection law.

3 LIMITATIONS OF "INFORMED CONSENT"

To date learning analytics has largely been a subject for educational research. However, the techniques are increasingly being adopted as part of the routine operation of universities and colleges (for example, Open University [n.d.]). Such processes may affect all current and future students and staff — not just those who participate in research studies — through changes to how education is provided in general and through specific individual interventions. With this significantly increased impact, "informed consent" may no longer provide adequate protection and guidance either for individuals or for organizations.

Both law and ethics require that for consent to be valid it must be both informed and freely given.[1] However, in Big Data approaches "[t]he potential value of the gathered data becomes clear only after they are subjected to analysis by computer algorithms, not beforehand" (Van der Sloot, 2015, p. 46). When using an inductive, data-driven approach rather than a deductive, hypothesis-driven one, it is hard for the organization to predict even what correlations will emerge from the data, let alone what their impact on the individual will be. Richards and King see a significant change:

> Before big data, individuals could roughly gauge the expected uses of their personal data and weigh the benefits and the costs at the time they provided their consent. Even if they guessed wrong, they would have some comfort that the receiving party would not be able to make additional use of their personal data. The growing adoption of big data and its ability to make extensive, often unexpected, secondary uses of personal data changes this calculus. (2014, p. 414)

[1] Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive) [1995] OJ L281/281, Article 2(h).

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

An educational organization may, and probably should, tell students and staff when it collects their information that such secondary uses will be limited to improving educational provision. However the law requires that for consent to be valid individuals should have "an appreciation and understanding of the facts and implications of an action … includ[ing] also an awareness of the consequences of not consenting to the processing in question" (Article 29 Working Party, 2011, p. 19). A simple statement of purpose is unlikely to provide this. But attempting to make consent valid by being more specific before collection or analysis has taken place involves second-guessing the results. This could stop both the organization and its students benefiting from unexpected findings. Consent obtained at the time of joining the organization will only cover uses and processing that were "reasonable expectations" at that time (Article 29 Working Party, 2011, p. 17). Since expectations of how data will be used are changing very rapidly, organizations might even have to use different processes for individuals who gave consent at significantly different times.

Furthermore, the law increasingly presumes that consent is not freely given in situations where the party requesting consent has significant power over the individual granting it. The European Commission's draft General Data Protection Regulation states that "[c]onsent should not provide a valid legal ground for the processing of personal data, where there is a clear imbalance between the data subject and the controller" (2012, Recital 39). Although the Nuremberg Code requires that research subjects "be allowed to discontinue their participation at any time," Kay et al. observe that in education it may be hard to opt out; once analytics has become part of university operations the only way to prevent your data being processed may, in fact, be to leave the university (2012, pp. 20, 27). It is hard to reconcile this with the Article 29 Working Party's test that "[c]onsent can only be valid if the data subject is able to exercise a real choice, and there is no risk of deception, intimidation, coercion or significant negative consequences if he/she does not consent" (2011, p. 12).

At a practical level, the use of consent may well bias the results of learning analytics, potentially excluding those who have most to gain from the process. Kay et al. note that whether the choice is presented as opt-in or opt-out will affect the outcome because "the user's inertia" (2012, p. 18) will mean that the majority simply accept the default. To this is added the risk of self-selection: that different groups will opt in or out in different proportions. This could skew the results either in favour of or against those groups (Clarke & Cooke, 1983). For example, analytics often aim to help those disengaged from the educational process but will get little information about this group from data collected on an opt-in basis.

Consent requires individual students to "take responsibility for technologies and business practices that they do not themselves create but find themselves increasingly dependent upon" (Richards & King, 2014, p. 430). The only responsibility the law assigns to organizations is to ensure that individuals have the information required to make their consent valid. From an ethical perspective, this is unsatisfactory: organizations should also behave responsibly in their adoption and use of the new practices they ask students to agree to. However using consent as a basis provides little guidance on what responsible

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

behaviour might be. The law does require that processing of personal data be "fair"[2] but that is, again, largely a matter of the information and controls provided to individuals (Information Commissioner, n.d.). A consent approach runs the risk of discussions focusing on whether new analytics processes and results are within "the scope of the data subject's consent" (Article 29 Working Party, 2014a, p. 13). An ethical framework should instead be considering whether inferences may perhaps be "inaccurate, discriminatory or otherwise illegitimate" (Article 29 Working Party, 2013, p. 45) and should not be used at all.

These doubts about the use of consent are not limited to the education context. In discussing the development of Big Data over the next five years, the European Data Protection Supervisor also sees the focus moving to the behaviour of organizations:

> Big data that deals with large volumes of personal information implies greater accountability towards the individuals whose data are being processed. People want to understand how algorithms can create correlations and assumptions about them, and how their combined personal information can turn into intrusive predictions about their behaviour. (2015, p. 10)

This is a particular concern in relationships where voluntarily provided information may be supplemented by that "observed and inferred without the individual's knowledge" (ibid.). Here there is a tendency for "opaque privacy policies, which encourage people to tick a box and sign away their rights" (2015, p. 11). Rather than relying on prior consent, organizations should be providing clear information on how and why information can be used and providing individuals with continuing opportunities to detect and challenge "mistakes in the assumptions and biases [data analytics] can make about individuals" (ibid.).

Such an approach seems more suitable for education where, according to Prinsloo and Slade,

> It is accepted that there are certain types of information and analyses (e.g., cohort analyses) that fall within the legitimate scope of business of higher education. There is though an urgent need to approach personal data differently when it is used to categorise learners as at-risk, in need of special support or on different learning trajectories. (2013, p. 244)

Kay et al. call for a framework that recognizes the differences between these two uses of personal data and provides appropriate protections for both. This would "enable Further and Higher Education institutions to progress their use of analytics whilst managing risk," providing "benefit to the individual, the institution and the wider mission of education" (2012, p. 8). The next sections propose a new source for such a framework and identify some of the questions it answers and poses about the operational use of learning analytics.

[2] Data Protection Directive, Art. 6(1)(a).

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

4 A DATA PROTECTION FRAMEWORK

If human-subject research no longer provides a suitable ethical model for the operational use of learning analytics, then a new reference model is needed. This paper proposes that learning analytics practice could instead be guided by European law's approach to protecting personal data.

While research gives consent a unique position in authorizing the use of data about human subjects, under data protection law, consent is one of six justifications for processing personal data. These are set out in Article 7 of the Data Protection Directive and Schedule 2 of the UK Data Protection Act 1998 and, at least under UK and EU law (Article 29 Working Party, 2011), have equal status. Each justification has its own particular mechanisms to protect the interests of data subjects. The UK Information Commissioner suggests that for Big Data analytics the most relevant justifications are "consent, whether processing is necessary for the performance of a contract, and the legitimate interests of the data controller or other parties" (2014, para. 55).[3]

Furthermore the law recognizes (Article 29 Working Party, 2011) that a single transaction may involve processing under several different justifications. This matches Kay et al.'s division of learning analytics into two categories: "[d]ata used for institutional purposes" and "[d]ata used for personal purposes" (2012, p. 10). The Open University (n.d.) give examples: redesigning modules "to take account of topics seen to cause particular issues of understanding" versus "identify[ing] points on the study path where individuals or groups may need additional support." In discussing Big Data, the Article 29 Working Party makes the same distinction between processing to "detect trends and correlations in the information" versus situations where "an organisation specifically wants to analyse or predict the personal preferences, behaviour and attitudes of individual customers, which will subsequently inform 'measures or decisions' that are taken with regard to those customers" (2013, p. 46). Different safeguards are seen as appropriate: for trend analysis, "functional separation is likely to play a key role" in protecting individuals (ibid.) whereas for individual interventions opt-in consent will normally be required.

Data protection therefore suggests that an ethical framework should treat learning analytics as two separate stages, using different justifications and their associated ways of protecting individuals:

- the discovery of significant patterns ("analysis") treated as a legitimate interest of the organization, which must include safeguards for individuals' interests and rights; and
- the application of those patterns to meet the needs of particular individuals ("intervention"), which requires their informed consent or, perhaps in future, a contractual agreement.

The following sections examine the implications of this separation, and how existing guidance on the legitimate interests and consent justifications can further inform the ethical framework and practice of learning analytics.

[3] Respectively, Articles 7(a), 7(b), and 7(f) of the Directive and clauses 1, 2, and 6 of the Act's Schedule 2.

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

4.1 Analysis

Like most other processing of personal data the "legitimate interests" justification requires that the individual be informed what data will be processed, by whom and for what purpose(s) (Data Protection Directive, Art. 10; Data Protection Act 1998, Sch. 1 Part II 2(3)). The purpose of the processing must be a legitimate interest of the organization and the processing must be necessary to achieve that interest: in other words no "less invasive means are available to serve the same end"; unlike the "consent" justification, however, the individual's agreement does not need to be obtained (Article 29 Working Party, 2014a). Instead, individuals are protected by the requirement that the organization's interest must not be overridden by the individual's interests or fundamental rights (Data Protection Directive, Art. 7(f)).[4] An individual with "compelling legitimate grounds relating to his particular situation" may object to processing if a different balance of interests applies in his case (Data Protection Directive, Art. 14(a)).

The Article 29 Working Party describe the legitimate interests justification as involving "an additional balancing test, which requires the legitimate interests of the controller … to be weighed against the interests or fundamental rights of the data subjects" (2014a, p. 48). The Working Party recognizes that a "broad range" of interests may be legitimate, including "the benefit that the controller derives — or that society might derive — from the processing," so long as the claimed interest "corresponds with current activities or benefits that are expected in the very near future" (2014a, p. 24). However, as the Information Commissioner confirms, organizations "need to be able to articulate at the outset why they need to collect and process particular datasets. They need to be clear about what they expect to learn or be able to do by processing that data" (2014, para 73). Since the Working Party quotes as legitimate a business's "interest in getting to know their customers' preferences so as to enable them to … offer products and services that better meet the needs and desires of the customers" (2014a, p. 26), there seems little doubt that the interests of a university or college in improving its educational provision would qualify as legitimate. That these interests are shared by society and current and future students is recognized as "adding weight" to the interest (2014a, p. 35). For example, a university might wish to make one module a pre-requisite for another if analytics indicated that this combination produced better learning outcomes. However legitimate interests do not give *carte blanche* for any kind of analysis: the Working Party warns against "creat[ing] … complex profiles of the customers' personalities and preferences without their knowledge" (2014a, p. 26) since this interference with individual rights could not be justified.

Having identified and articulated a legitimate interest of the organization, the balancing test requires that this be weighed against possible harm to "all relevant interests" of individuals (Article 29 Working Party, 2014a, p. 29). Relevant interests include, but are not limited to, their fundamental rights. The Working Party's guidance on reducing the risk of harm is particularly helpful for how the analysis stage of learning analytics should be conducted.

[4] The UK Data Protection Act transposes this as "prejudice to the rights and freedoms or legitimate interests of the data subject" [Sch. 2 6(1)].

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

First, there must be a clear "functional separation" between the analysis and intervention processes: "data used for statistical purposes or other research purposes should not be available to 'support measures or decisions' that are taken with regard to the individual data subjects concerned (unless specifically authorized by the individuals concerned)" (Article 29 Working Party, 2014a, p. 30). This is likely to involve organizational measures — such as confidentiality agreements and guidance to staff with access to raw data — as well as technical ones to protect users' privacy during analysis.

The quantity of data collected, and access to that data, must be limited to what is required to achieve the stated purpose. The Working Party note that this "is particularly relevant … to ensure that processing of data based on legitimate interests will not lead to an unduly broad interpretation of the necessity to process data" (2014a, p. 29). Technical measures should be put in place to reduce the risk to individual users. Anonymized or statistical data should normally be used for analysis. Where full anonymization is not possible, directly identifying information such as names or student numbers should be replaced with key-codes with the index linking keys to individuals stored separately and, preferably, under separate control (Article 29 Working Party, 2013). The Working Party note that

> The use of less risky forms of personal data processing (e.g., personal data that is encrypted while in storage or transit, or personal data that are less directly and less readily identifiable) should generally mean that the likelihood of data subjects' interests or fundamental rights and freedoms being interfered with is reduced. (2014a, p. 42)

Safeguards that "unquestionably and significantly reduce the impacts on data subjects" may "chang[e] the balance of rights and interests to the extent that the data controller's legitimate interests will not be overridden" (Article 29 Working Party, 2014a, p. 31), thus allowing processing to proceed.

For example, successful combinations of modules cannot be identified using fully anonymized data, as this analysis requires each student's performance to be linked across modules. However, the sequences of modules and results can be constructed and analyzed using key-coded data, or even one-way hashed identifiers, so that there is no index that would allow an identifier to be directly linked to a student. Combined with organizational controls to prevent re-identification, this investigation should easily satisfy the balancing test. If the organization wishes to intervene with those individuals who have chosen a combination likely to be unsuccessful, this should be a separate process from the analysis, conducted after obtaining consent as described in the next section.

Another factor that may tip the balance is whether processing matches users' reasonable expectations of how their data will be used: "the more clearly acknowledged and expected it is in the community and by data subjects that the controller can take action and process data in pursuit of these interests, the more heavily this legitimate interest weighs in the balance" (Article 29 Working Party, 2014a, p. 35). That a university or college would use analytics to improve the educational services it provides — benefitting

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

both the students whose data are analysed and their successors — should not be unexpected. Indeed, as noted by Kay et al. (2012), current students may positively expect that the services they receive will adapt to the individual's activity, so long as their interests and rights are protected. In a Jisc workshop to identify appropriate uses of analytics, students proposed providing personalized suggestions to those who were having difficulty engaging with course material (Sclater, 2015).

An ethical framework may take account of both positive and negative impacts on individuals when assessing the balance of rights and interests. However, data protection law identifies universities' close relationship with their students and employees as creating particular risks of harm from inappropriate processing (Article 29 Working Party, 2014a). Rather than run these risks, if an action may directly affect individuals — even in ways that are expected to be beneficial — it is usually better to seek the consent of those individuals.

4.2 Intervention

The analysis stage can provide a great deal of information to help universities enhance their educational provision. Improvements to university facilities and processes, recruitment, courses, and teaching materials are most likely to be based on aggregated or anonymized data. The discussion above suggests that an ethical framework concentrating on continuing safeguards and the balance of interests will provide both better protection for individuals and more complete data to inform such changes.

However, learning analytics can also be used to support and guide individual students and staff. For example, students might be offered personalized reading lists based on how their progress compares with others' (Sclater, 2015). Here the university's aim is to have a positive effect on the individual so an ethical framework based on minimizing individual impact is no longer appropriate. Instead, intervention will normally be based on the individual's consent. The possibility of intervention being a requirement of a contract or a legal duty is discussed further below. Again, discussion of consent in European law provides helpful ethical guidance for when and how this should be done. In particular, consent must be "freely given, specific and informed" (Article 29 Working Party, 2011, p. 5).

Consent will only be freely given, according to the Article 29 Working Party, if there is "no risk of … coercion or significant negative consequences if [the individual] does not consent" (2011, p. 12). Since a university or college has the ultimate power to grant or withhold its students' qualifications, there is a risk that students will feel compelled to accept. Interventions should therefore be offered as a choice between receiving personalized support or standard educational provision. A virtual learning environment might invite users to enable "people like you…" suggestions; students might be offered classes or seminars appropriate to their learning style. Such choices should avoid "significant negative consequences" and thereby contribute to valid consent.

The Working Party explains that for consent to be informed "all the necessary information must be given

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

at the moment the consent is requested, and that this should address the substantive aspects of the processing that the consent is intended to legitimise" (2011, p. 9). To be specific, consent "should refer clearly and precisely to the scope and the consequences of the data processing" (Article 29 Working Party, 2011, p. 17). These requirements are much easier to meet after analysis has identified a pattern that may suggest a particular beneficial intervention. Students can now be given detailed and precise information on what the intervention is intended to achieve, and on the implications of either granting or withholding consent for it. The Working Party recognize that consent requested "'downstream', when the purpose of the processing changes" allows information to be provided that "focus[es] on what is needed in the specific context" (2011, p. 19). Thus, rather than obtaining general consent for "appropriate support" when a course begins, waiting till analysis of a student's performance has identified their most effective learning style lets them give fully informed consent to a specific kind of help.

An individual's consent must be signified by some "indication" (Article 29 Working Party, 2011, p. 5). Since the Working Party considers that "[t]he notion of 'indication' is wide, but it seems to imply a need for action" (2011, p. 12), universities should obtain consent through some positive action by the student rather than presuming it from the student's silence or inaction. Students should normally be invited to opt-in to an intervention. If circumstances require an opt-out approach then it might be argued that consent was obtained earlier — for example when the student signed up to the course or module — and that an opportunity to withdraw that consent was being offered at the time of the intervention. This, however, would require the course description to state clearly that personalized interventions would be made. Interventions on this basis could only use personal data that were either stated or obviously relevant at the time of signing up: the information provided to the student must be such that joining "lead[s] to an unmistakable conclusion that consent is given" (Article 29 Working Party, 2011, p. 23). A course might state, for example, that skills would first be assessed and appropriate topics then chosen to address those areas needing improvement.

The Working Party places no specific limit on how long consent may be presumed to last. A single consent could therefore cover a series of interventions (2011, p. 17) so long as the student knows this, the information they were given about the scope and consequences of processing remains accurate, and they do not indicate that they have changed their mind. The Working Party does suggest that organizations should periodically remind each individual "of their current choice and offer […] the possibility to either confirm or withdraw"; how often such reminders are provided will depend on the "context and circumstances" (2011, p. 20) but the rapid pace of developments in learning analytics suggests that they should be relatively frequent. Reminders are particularly important where consent has effect over a long period; for example, where it was obtained at the start of a course. Where a student has actively opted out (as opposed to not opting in) regular reminders may be the best way to inform them of the options available.

If an organization plans to use learning analytics to intervene with individual members of staff, rather than students, then the law warns that consent may not be an appropriate basis. If an employee "might fear

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

that he could be treated differently if he does not consent to the data processing" (Article 29 Working Party, 2011, p. 13), then consent will not be freely given so is not valid. Interventions with staff should therefore be limited to those necessary for the purpose of the (employment) contract or those where there are sufficient safeguards that, despite the strong presumption to the contrary, genuinely free consent can in fact be obtained from the employee (Article 29 Working Party, 2011).

Finally, should a legal duty require a university to intervene with an individual, both ethics and law indicate that any use of personal data must be limited to that strictly necessary to fulfil that duty.

5 APPLYING THE FRAMEWORK

The ethical framework proposed here — particularly the balancing test required by the legitimate interests justification — appears to match the feelings of participants in learning analytics studies. For example, Pardo and Siemens found that "the concerns of users about privacy vary significantly depending on what is being observed, the context and the perceived value when granting access to personal information" (2014, p. 440). Furthermore the balancing test protects all the rights and interests of individuals, not just privacy: the Article 29 Working Party note, for example, that "[t]he chilling effect on protected behaviour, such as freedom of research or free speech, that may result from continuous monitoring/tracking, must also be given due consideration" (2014a, p. 37). This section considers how the framework might inform some specific questions raised by learning analytics practitioners: historic data, small groups, fully automated processing, and sensitive data.

Pardo and Siemens identify issues with historic data: "How long will the data be kept? Will the data be used after students graduate?" (2014, p. 446). They consider long-term information "will be helpful for the university to refine its analytics models, track the development of student performance over multiple years and cohorts or simply for internal or external quality assurance processes" but retaining too much, or for too long, may well damage trust. Applying the balancing test of the framework, the likelihood of direct personal benefit to a student decreases once they have completed their module or course, long-term retention increases the risk of information becoming out-of-date or suffering a security breach. The university must demonstrate that continued processing generates sufficient benefit to balance the decreased benefit and increased risk to the individual's rights and interests. In particular, according to the Information Commissioner, "[i]f organisations wish to retain data for long periods for reasons of big data analytics they should be able to articulate and foresee the potential uses and benefits to some extent, even if the specifics are unclear" (2014, para. 75). They may also need to implement additional safeguards; for example, anonymization rather than pseudonymization (Information Commissioner, 2012). Thus, processing of historic data should still be possible given a sufficiently clear and strong justification but, because the direct benefit to the individual is less, the range of acceptable purposes is likely to be narrower than for current students. Where universities wish to continue to collect data from former students — for example for the Higher Education Statistics Agency's (2015) Destination of Leavers from Higher Education survey — this relies in any case on voluntary participation, so free, informed consent to continued processing of relevant historic data could be obtained at the same time.

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

Kay et al. (2012) note that patterns derived from small numbers of individuals represent an increased risk of accidental or deliberate re-identification. They suggest a minimum group size of twenty, though the law, in some circumstances, has considered that averages across groups as small as five individuals can be disclosed safely (Information Commissioner, 2012). The legitimate interests balancing test can take account of the various factors affecting the risk of harmful re-identification and ensure that all groups retain a level of protection appropriate to the risk and benefit of the processing. In most cases, there will be a threshold group size below which the risk outweighs the benefits. This should indicate that such fine-grained processing should not continue.

One paradox highlighted by the balancing test is that, where there is a clear benefit to the individual and a negligible risk, automated processing may be less privacy-invasive than revealing personal data to a human mediator. Re-ordering the choice of topics or modules based on a student's performance in previous courses might be an example of this kind of intervention. Under data protection law, individuals are entitled to know of any fully automated processing that may affect them and the logic used (Data Protection Directive, Art. 12(a)); they also have the right to object to such processing (Data Protection Directive, Art. 15(1)). The law only requires these "profiling" safeguards where there is no human mediation (European Commission, 2012, Art. 20). However, by requiring prior notification and seeking consent at the time of any intervention, the framework extends them to cover both automated and mediated processes.

Pardo and Siemens consider the risks of using privacy-sensitive data in learning analytics:

> For example, in a hypothetical scenario, 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? (2014, p. 439)

The law recognizes various types of data that represent an increased risk to the individual. An ethical framework should reflect these and apply appropriate safeguards. For example, location data can be processed based on legitimate interests, but the e-Privacy Directive warns of significant privacy risks, so clear benefits and strong safeguards must satisfy the framework's balancing test.[5] Anonymized location data might be used, for example, to improve the design of learning spaces or make more efficient use of rooms, provided there was strong protection against re-identification of individuals. However, Pardo and Siemens' (2014) collaborative learning example directly affects individuals so the framework, like Article 9(1) of the e-Privacy Directive, would require the free, informed consent of the individuals involved. If information is categorized by the Data Protection Directive (Art. 8(1)) as Sensitive Personal Data — for example, race or religion (but not location) — then legitimate interests cannot be used even if there is no effect on the individual (Data Protection Act 1998, Sch. 3(4)). A university that wished to use these categories in analysis should obtain the informed consent of each individual before data were collected. Such information could then only be processed for purposes and in ways envisaged at that time. In

[5] Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector [2002] OJ L201/37 Article 9.

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

addition, consent for processing Sensitive Personal Data must be explicit and cannot be inferred from some other action (Data Protection Act, Sch. 3(1)).

6 CONCLUSION

As learning analytics moves from a research context to become a key tool in university operations, the use of prior consent as the sole ethical basis appears inappropriate, as it offers neither the information nor the guidance that students and those conducting analyses need. The Article 29 Working Party warns that

> The use of consent "in the right context" is crucial. If it is used in circumstances where it is not appropriate, because the elements that constitute valid consent are unlikely to be present, this would lead to great vulnerability and … would weaken the position of data subjects in practice. (2011, p. 10)

Instead, the paper proposes a new two-stage ethical framework, based on the approach taken by data protection law. Analysis of learner data is considered a legitimate interest of a university that must be conducted under appropriate safeguards. The university's interests must be continually tested against the interests and rights of individuals; interference with those interests and rights must be minimized; analysis must cease if they cannot be adequately protected. If analysis suggests an intervention that may affect individual students or staff, the consent of those individuals should be sought. Since they can now be provided with full information about the nature and consequences of the intervention, their choice is much more likely to be ethically and legally sound.

The approach should help both universities and students to benefit from developments in learning analytics. It avoids artificially limiting the provision and development of education to what was known or foreseen at the time when consent was originally obtained. This is particularly important in a fast-developing field. It reduces the risk of self-selection bias affecting the data used for planning. It also offers strong protection for individual students and staff by providing both clear guidance on the conduct of current and new analyses and detailed, relevant information when individual interventions are offered. All processing will include safeguards appropriate to the level of risks to privacy and other interests and rights. Students and staff will be offered specific, meaningful, informed choices at the time when these may affect them.

Above all, as learning analytics moves from the research laboratory into the daily business of education, the framework draws attention to ethically and practically important questions: What topics and methods are appropriate uses of learner data? How can these can be protected? How can interventions best be designed to benefit students? Learning analytics practices that address these issues are much more likely to build and maintain the trust of those who rely on them.

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

REFERENCES

Article 29 Working Party (2011). *Opinion 15/2011 on the definition of consent* (01197/11/EN WP187). Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf

Article 29 Working Party. (2013). *Opinion 03/2013 on purpose limitation* (00569/13/EN WP203). Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

Article 29 Working Party. (2014a). *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC* (844/14/EN WP 217). Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

Article 29 Working Party. (2014b). *Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU* (14/EN WP221). Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf

Clarke, G., & Cooke, D. (1983). *A basic course on statistics*. London: Edward Arnold.

European Commission (2012). *Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)* COM(2012)11 Final.

European Data Protection Supervisor (2015). *Leading by example: The EDPS strategy 2015–2019*. Retrieved from https://secure.edps.europa.eu/EDPSWEB/edps/site/mySite/Strategy2015

Higher Education Statistics Agency (2015). *Destinations of leavers from higher education in the United Kingdom for the academic year 2013/2014*. Retrieved from https://www.hesa.ac.uk/sfr217

Information Commissioner (n.d.). *Processing personal data fairly and lawfully (Principle 1)*. Retrieved from https://ico.org.uk/for-organisations/guide-to-data-protection/principle-1-fair-and-lawful/ - fair-processing

Information Commissioner (2012). *Anonymisation: Managing data protection risk code of practice*. Retrieved from https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdfInformation Commissioner (2014). *Big data and data protection*. Retrieved from https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdfKay, D., Korn, N., & Oppenheim, C. (2012, November). Legal, risk and ethical aspects of analytics in higher education*, JISC CETIS Analytics Series, 1*(6). Retrieved from http://publications.cetis.ac.uk/2012/500

Leece, R. (2013, July). *Analytics: An exploration of the nomenclature in the student experience*. *Proceedings of the 16th International First Year in Higher Education Conference* (FYHE13), 7–10 July 2013, Te Papa Tongarewa, Wellington, New Zealand. Retrieved from http://fyhe.com.au/past_papers/papers13/14E.pdf

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, *46*(5), 31–40.

Mathewson, T. G. (2015, August 21). Analytics programs show "remarkable" results — and it's only the beginning. *Education Dive*. Retrieved from http://www.educationdive.com/news/analytics-

(2016). A data protection framework for learning analytics. *Journal of Learning Analytics*, *3*(1), 91–106.
http://dx.doi.org/10.18608/jla.2016.31.6

programs-show-remarkable-results-and-its-only-the-beginning/404266/

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. London: John Murray.

National Union of Students (2015). *Learning analytics: A guide for students' unions*. Retrieved from http://www.nusconnect.org.uk/resources/learning-analytics-a-guide-for-students-unions

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

Ohm, P. (2014). Changing the rules: General principles for data use and analysis. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, big data and the public good* (pp. 96–111). New York: Cambridge University Press.

Open University (n.d.). *Using information to support student learning*. Retrieved 9 July 2015 from http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/using-information-to-support-student-learning.pdf

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*, 438–450. http://dx.doi.org/10.1111/bjet.12152

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*, 240–244. http://dx.doi.org/10.1145/2460296.2460344

Richards, N., & King, J. (2014). Big data ethics. *Wake Forest Law Review*, *49*, 393–432.

Sclater, N. (2014). *Code of practice for learning analytics: A literature review of the ethical and legal issues*. Jisc. Retrieved from http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

Sclater, N. (2015). *What do students want from a learning analytics app?* [Web log post, April 29]. Retrieved from http://analytics.jiscinvolve.org/wp/2015/04/29/what-do-students-want-from-a-learning-analytics-app/

Tickle, L. (2015, June 30). How universities are using data to stop students dropping out. *The Guardian*. Retrieved from http://www.theguardian.com/guardian-professional/2015/jun/30/how-universities-are-using-data-to-stop-students-dropping-out

Van der Sloot, B. (2015). Privacy as personality right: Why the ECtHR's focus on ulterior interests might prove indispensable in the age of "big data." *Utrecht Journal of International and European Law*, *31*(80), 25–50. http://dx.doi.org/10.5334/ujiel.cp

Warrell, H. (2015, July 24). Students under surveillance. *Financial Times*. Retrieved from http://www.ft.com/cms/s/2/634624c6-312b-11e5-91ac-a5e17d9b4cff.html

Wen, H. (2012). Big ethics for big data. *Radar: Insight Analysis, and Research about Emerging Technologies* [Web log post, June 11]. Retrieved from http://radar.oreilly.com/2012/06/ethics-big-data-business-decisions.html