Guest Editorial: Ethics and Privacy in Learning Analytics

Rebecca Ferguson
IET, The Open University, UK
Rebecca.Ferguson@open.ac.uk

Tore Hoel
Oslo and Akershus University College of Applied Sciences, Norway
Tore.Hoel@hioa.no

Maren Scheffel
Open Universiteit, The Netherlands
Maren.Scheffel@ou.nl

Hendrik Drachsler
Open Universiteit, The Netherlands
Hendrik.Drachsler@ou.nl

ABSTRACT: The European Learning Analytics Community Exchange (LACE) project is responsible for an ongoing series of workshops on ethics and privacy in learning analytics (EP4LA), which have been responsible for driving and transforming activity in these areas. Some of this activity has been brought together with other work in the papers that make up this special issue. These papers cover the creation and development of ethical frameworks, as well as tools and approaches that can be used to address issues of ethics and privacy. This editorial suggests that it is worth taking time to consider the often interwoven issues of ethics, data protection, and privacy separately. The challenges mentioned within the special issue are summarized in a table of 22 challenges used to identify the values that underpin work in this area. Nine ethical goals are suggested as the editors’ interpretation of the unstated values that lie behind the challenges raised in this paper.

Keywords: Data protection, ethics, LACE, privacy, values

1 INTRODUCTION

In autumn 2013, we outlined our plans for a Learning Analytics Community Exchange (LACE) that would bring together people involved in learning analytics across Europe and beyond. Among other aims, LACE was to integrate communities, inform future research and policy agendas, and accelerate the identification of viable and effective solutions to real problems. One of the problems we identified was the challenge of developing and applying a clear set of ethical and privacy guidelines for learning analytics (Ferguson, 2012; Greller & Drachsler, 2012). This was not straightforward; the ethical and privacy aspects of learning analytics are varied, and they shift as the use of data reveals information that could not be accessed in the past. In order to identify viable solutions to the problem, we set out to encourage explicit discussion of this topic.
Since then, the LACE project has organized or co-organized a series of six workshops on ethics and privacy in learning analytics (EP4LA1). These have taken place in Utrecht, The Netherlands (October 2014); Rotterdam, The Netherlands (November 2014); Washington, DC, USA (November 2014); Paris, France (February 2015); London, UK (February 2015); and Poughkeepsie, NY, USA (March 2015). The next workshop in the series will take place in April 2016 at the Learning Analytics and Knowledge conference (LAK ’16) in Edinburgh, UK.

These workshops have been responsible for driving and transforming work in this area. As Hoel and Chen note in this issue (2016), in 2014 privacy was only superficially recognized in a few papers at the major annual conference in this area, LAK ’14. However, by the time that papers were submitted for LAK ’15, which included an EP4LA workshop, there were signs of a new approach that not only identified privacy as a concern but also pointed to privacy solutions at different levels.

As early as the first EP4LA workshop in Utrecht in 2014, the LACE project brought together a heterogeneous group of experts including representatives from the Dutch organization SURF and the UK organization Jisc, lawyers from national organizations in the Netherlands, and consultants from big consulting companies. After that, SURF and LACE deeply investigated European laws related to the benefits and limitations of the use of learning analytics in higher education. In 2015 those efforts resulted in a national report on learning analytics and its relation to privacy law in the Netherlands (Engelfriet, Jeunink, & Manderveld, 2015) as well as in the DELICATE checklist that operationalizes key concepts and questions for ethics and privacy in learning analytics (Drachsler & Greller, 2016). The DELICATE checklist helps to investigate the obstacles that could impede the rollout of learning analytics (see figure 1) and the implementation of trusted learning analytics for higher education. The LACE project recently published a review on the state of the art of ethics and privacy as part of its learning analytics review series.1

In the same period, Jisc, working closely with stakeholders on an initiative to promote the effective use of learning analytics, identified tackling issues relating to privacy and ethics as a priority. The Code of

1 http://www.laceproject.eu/ethics-privacy-learning-analytics/
2 http://www.laceproject.eu/learning-analytics-review/
Practice for Learning Analytics that emerged as a result (Sclater & Bailey, 2015) was informed and shaped by the EP4LA workshops (Sclater, 2016). That development process continues in this special issue, which proposes amendments to the Code of Practice when applied in school settings (Rodríguez-Triana, Martínez-Monés, & Villagrá-Sobrino, 2016). These are currently under consideration by the Code’s authors.

Although there is still much to say and much to investigate on the topic of ethics and privacy, the three main outcomes of the EP4LA workshop series (1. SURF Privacy & Analytics, 2. DELICATE checklist, 3. Jisc Code of Conduct) form the cornerstones of an approach that will enable us to solve some of the most pressing challenges for learning analytics. Any educational organization that considers rolling out learning analytics can take advantage of those three practical tools and adapt them for their needs.

In addition to those tools, the papers in this special issue provide a range of responses and deep thinking about sub-topics related to issues of ethics and privacy in learning analytics. They also identify a series of problems and challenges related to ethics, privacy, data protection, data sharing, and data anonymization. These areas are often bundled together because they deal with related issues. However, it is worth taking the time to consider each area separately, because they have different priorities.

2 ETHICS

Although the word “ethics” is used increasingly frequently within the learning analytics literature, there have been few attempts to define the term and to specify what it means in this context. Ferguson (2012) identifies the need for ethical guidelines as one of four major challenges for the field, and implies it relates to ownership and stewardship of data and to informed consent. Drachsler and Greller (2016) clearly differentiate between ethics and privacy:

> Ethics is the philosophy of morality that involves systematizing, defending, and recommending concepts of right and wrong conduct. In that sense, ethics is rather different from privacy. In fact, privacy is a living concept made out of continuous personal boundary negotiations with the surrounding ethical environment.

Slade and Prinsloo (2013) consider the area in more detail and identify three broad classes of ethical issues: the location and interpretation of data; informed consent, privacy, and the de-identification of data; and the management, classification, and storage of data. They also show that the ethical issues differ depending on perspective (see also Kay, Kom, & Oppenheim, 2012) and that, beyond the practicalities of consent and data storage, ethical choices in this field relate to much broader issues around power, surveillance, and the purpose of education. They note that an institution’s use of learning analytics is going to be based on its understanding of the scope, role and boundaries of learning analytics and a set of moral beliefs founded on the respective regulatory and legal, cultural, geopolitical and socio-economic contexts. Any set of guidelines concerned with the ethical dilemmas and challenges in learning analytics will necessarily also be based on a set of epistemological assumptions. (Slade & Prinsloo, 2013)
Although this is clearly the case, it is rare for authors in the field to acknowledge this or to attempt to set out the assumptions that underpin their ethical decisions and prompt their questions about ethics. Campbell in his doctoral work (2007) and later with colleagues (Willis, et al., 2013) provides a framework for making these assumptions explicit. He and his colleagues reference Samuel M. Thompson’s *The Nature of Philosophy* when they note that ethics is concerned with two practical problems of human life:

- What is worth seeking — that is, what ends or goals of life are good?
- What individuals are responsible for — that is, what duties should they recognize and attempt to fulfill? (Willis, Pistilli, & Campbell, 2013)

In relation to learning analytics, learner success is typically what is judged to be worth seeking — the ethical goal. How we interpret learner success — and who we believe to be responsible for that success — are therefore important. Success could be seen in terms of an individual or a group completing a course, passing a programme of study, meeting personal goals, becoming more employable or gaining the competencies to lead a fulfilling life. It could be seen in terms of an educator gaining qualifications or experience, taking on more students (or perhaps fewer students), or guiding more students successfully towards specific goals. It could also be seen in terms of an institution, a nation, or a group of nations. The “No Child Left Behind” policy in the US, for example, had ethical as well as political implications.

Willis et al. (2013) suggest some philosophical frameworks to help determine what ends or goals are good. They reference Immanuel Kant’s categorical imperative (“Act on the maxim that you wish to have become a universal law”), John Stuart Mill’s principle of utility (“Seek the greatest happiness for the greatest number”), and John Rawls’ veil of ignorance (“Justice emerges when negotiations are without social differentiation”). In practice, the values underpinning ethical practice are rarely set out clearly, justified, or interrogated. The implication is that these values are universal and can go unstated. As, the EP4LA workshops have shown, though, when comparing experiences across Europe and beyond, values are not consistent from country to country, from institution to institution or even from classroom to classroom. It is worth, therefore, considering the values that are implicit within discussion about learning analytics. In the following sections, we examine the learning analytics challenges with the ethical dimensions identified within this special issue (the full list of challenges is set out and numbered in Table 1).

**Challenges 1–6:** For those used to regarding ethics as a restriction on action rather than a call to action, the first challenge — the duty to act (Kay, et al., 2012) — may come as a surprise. If what is worth seeking is learner success, and if it is the responsibility of institutions and educators to help learners achieve success, then it follows that data should be used to support that aim whenever possible. This in turn places a responsibility on learners. Learning analytics algorithms are less likely to work effectively if they are fed incomplete, inaccurate, or out-of-date information, so learners need to ensure that their information is both complete and correct. To motivate this additional work, everyone needs to be convinced that learning analytics promote student success and that they do so at least as effectively as...
other approaches requiring similar resources. This requires that the institution meet its responsibilities to ensure that the analytics are both valid and reliable, that they are corrected when necessary, and that they are presented clearly and in ways that support learning.

**Table 1: Learning Analytics Challenges with Ethical Dimensions**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Use data to benefit learners</td>
</tr>
<tr>
<td>2</td>
<td>Provide accurate and timely data</td>
</tr>
<tr>
<td>3</td>
<td>Ensure accuracy and validity of analyzed results</td>
</tr>
<tr>
<td>4</td>
<td>Offer opportunities to correct data and analysis</td>
</tr>
<tr>
<td>5</td>
<td>Ensure results are comprehensible to end users</td>
</tr>
<tr>
<td>6</td>
<td>Present data/results in a way that supports learning</td>
</tr>
<tr>
<td>7</td>
<td>Gain informed consent</td>
</tr>
<tr>
<td>8</td>
<td>Safeguard individuals’ interests and rights</td>
</tr>
<tr>
<td>9</td>
<td>Provide additional safeguards for vulnerable individuals</td>
</tr>
<tr>
<td>10</td>
<td>Publicize mechanisms for complaint and correction of errors</td>
</tr>
<tr>
<td>11</td>
<td>Share insights and findings across digital divides</td>
</tr>
<tr>
<td>12</td>
<td>Comply with the law</td>
</tr>
<tr>
<td>13</td>
<td>Ensure that data collection, usage, and involvement of third parties are transparent</td>
</tr>
<tr>
<td>14</td>
<td>Integrate data from different sources with care</td>
</tr>
<tr>
<td>15</td>
<td>Manage and care for data responsibly</td>
</tr>
<tr>
<td>16</td>
<td>Consider how, and to whom, data will be accessible</td>
</tr>
<tr>
<td>17</td>
<td>Ensure data are held securely</td>
</tr>
<tr>
<td>18</td>
<td>Limit time for which data are held before destruction and for which consent is valid</td>
</tr>
<tr>
<td>19</td>
<td>Clarify ownership of data</td>
</tr>
<tr>
<td>20</td>
<td>Anonymize and de-identify individuals</td>
</tr>
<tr>
<td>21</td>
<td>Provide additional safeguards for sensitive data</td>
</tr>
</tbody>
</table>

**Challenge 7:** Informed consent is a practice that originated relatively recently in the medical sciences, partly because the medical model became less paternalistic and more patient centred, and partly because doctors became increasingly aware that they could be open to charges of assault and battery (Murray, 1990). As the medical guidelines set out in the 1947 Nuremberg Code were taken up in other areas, it became more common for social scientists to seek informed consent or, at least, to consult with ethical advisers about whether consent would be needed. The ethical goal that underpins informed consent in these cases is, broadly speaking, to “do no harm.”

In the case of learning analytics, though, the ethical basis of informed consent is different because there is no physical danger to learners and no deception involved. Slade and Prinsloo (2013) consider that learning analytics should involve students as collaborators, thus learners should give informed consent and collaborate in providing access to data so that analytics can support their learning. In that case, as with Challenges 1–6, the ethical goal would be learner success.

Another perspective, implied by learner comments (Slade & Prinsloo, 2014) is that informed consent forms part of a process of establishing a trusting relationship between learners and their educational
institutions, so the ethical goal might be that educational institutions can be trusted by learners. Or, if data are considered property or assets, then informed consent has to do with the assignation of property rights, and the ethical goal is a society in which we respect private and group ownership of assets. These different understandings of why we seek to gain informed consent all have implications for how we go about that process.

Challenges 8–10. These three challenges (safeguard individuals’ interests and rights, provide additional safeguards for vulnerable individuals, and publicize mechanisms for complaint and correction of errors) all relate to the ethical goal that we should safeguard those in our care. Educational institutions should protect the rights and interests of their staff and students; educators should safeguard their students. How this safeguarding is carried out depends on context, but it should include some mechanism for complaint and the correction of errors (Rodríguez-Triana, et al., 2016).

Challenge 11 relates to an ethical goal of a society in which everyone has equal access to education, while Challenge 12 is associated with a goal of a just society in which laws are fair and equally applied, and citizens abide by those laws.

3 DATA PROTECTION

Challenges 13–19 relate to data protection, an area often presented as separate from but allied with ethics. Broadly speaking, the ethical issues covered by Challenges 1–12 require a moral response from individuals and institutions. Those covered by Challenges 13–19 require a legal response, which will depend on the relevant area of jurisdiction and regional attitudes towards data security.

From an ethical perspective, the issues here vary, but overlap with those identified above. If the only motivator is compliance with relevant legislation, then these challenges are all aspects of Challenge 12: comply with the law. More broadly, they can be associated with an understanding of data as personal property, or as jointly owned property, in which case the ethical goal might be to respect the property rights of others, or to take good care of the property with which we are entrusted. This could be a safeguarding issue, with institutions working to protect the rights and interests of their staff and students. Or it could be another part of establishing a trusting relationship, with the ethical goal that educational institutions can be trusted by learners. Once again, if we know why we are protecting data, then we will have a better idea of how to go about doing this.

4 PRIVACY

The final set of challenges, Challenges 20–21, are associated with another area usually spoken of as allied with, but in some way separate from both ethics and data protection. As with ethics, there have been few attempts to define privacy in the learning analytics literature. Pardo and Siemens (2014) regard it as “the regulation of how personal digital information is being observed by the self or distributed to other observers” and as something that is handed over in a transaction. This does not
make it clear in what way these data differ from other data and why they should be accorded special treatment. Hoel and Chen (2016) note the need to unpack privacy as a socio-cultural concept and observe that the boundaries around personal and private data are social agreements that depend on who the owner is and in what social settings the data are created and shared.

Privacy can be understood as a freedom from unauthorized intrusion: the ability of an individual or a group to seclude themselves or the information about them, and thus to express themselves selectively. Technological advances have made it almost impossible for individuals or groups to enjoy the same levels of privacy that were routine only decades ago when privacy was not an ethical issue but a default. Learning analytics make it possible to combine data sets to generate insights that would never have been possible in the past, often making use of data that the learner was not aware were being collected or analyzed. Mobile data make it possible to collect details about the learner’s environment — ambient light, temperature, and air pressure — and set them alongside personal data such as blood pressure, heart beat, and perspiration (Hoel & Chen, 2016).

As a result, it would be possible for learning analytics to keep learners under perpetual surveillance and to reveal things about them to others that they are not aware of themselves. A recent LACE report on the future of learning analytics (Griffiths, Brasher, Clow, Ferguson, & Yuan, 2016) revealed deep unease about the monitoring that could be associated with learning analytics. One expert commented that it suggested “a Big Brother scenario, with deep intrusion into the privacy and integrity of students.”

Considered from the perspective of data protection, data are treated as property. From the perspective of privacy, data are much more personal, almost a part of the self and certainly very bound up with the sense of self. If we reveal these data, we reveal ourselves. If we care for and protect these data, we are to some extent caring for and protecting ourselves, particularly if we do so in response to potential threats. These potential threats appear more real and immediate in some parts of the world. Anecdotally, the EP4LA workshops suggest that those European countries under Nazi control during World War II are more acutely aware of the potential for the misuse of personal data and are therefore much more protective of that data.

One ethical goal of privacy may therefore be stated as personal safety or, more broadly, as a society in which individuals and groups can live free from threat. Another goal is the integrity of the self — a society in which we have control over how we are seen and in which we are able to maintain a separation between our private selves and our public personas.

5 ETHICAL GOALS

Considered together, the ethical challenges of learning analytics suggest nine ethical goals:

1. Student success
2. Trustworthy educational institutions
3. Respect for private and group assets
4. Respect for property rights
5. Educators and educational institutions that safeguard those in their care
6. Equal access to education
7. Laws that are fair, equally applied, and observed
8. Freedom from threat
9. Integrity of self

These goals are our interpretation of the unstated values that lie behind the challenges raised in this paper. This may be seen as an optimistic or a biased interpretation. In many cases, the motivator behind apparently ethical behaviour is not a greater good but is instead a fear of the law, a need to fit in, a desire for profit, or a love of personal success. We view these goals from the affluent West, where ethical decisions are not necessarily directly related to a God-given moral code, and where property rights are intertwined with society’s political and economic structure.

This list is clearly not a universal set of human values so obvious that they can go unstated. Each of the goals is open to interpretation and its importance will be understood differently depending on context. The importance of the list is that it offers members of the learning analytics community an opportunity to reflect on the values and principles that underpin our ethical behaviour. If we reach an understanding of what is important about ethics, data protection, and privacy then we can decide how best to work towards them.

6 SPECIAL ISSUE PAPERS

The papers in this special issue set out ways in which the community is already working on these areas. In the first paper, Sclater describes the co-design process involved in Developing a Code of Practice for Learning Analytics. The process included five stages: 1) an extensive literature review; 2) formation of an advisory group; 3) development and validation of a taxonomy of issues; 4) drafting and validation of the code; and 5) population of an accompanying website with guidance and case studies. Development work included both consultation and workshop discussion, and the resulting Code of Practice produced by Jisc is intended as a checklist of issues that can be considered and customized by different institutions. It can therefore be used to spark further consideration of the ethics and values associated with learning analytics.

In paper two, Rodríguez-Triana, Martínez-Monés, and Villagrá-Sobrino show this process of development and consideration in action. In Learning Analytics in Small-Scale Teacher-Led Innovations: Ethical and Data Privacy Issues, they note that current ethical frameworks are institutionally oriented and focus on higher education. In a school setting, teachers are likely to be the main actors in the application of learning analytics processes. Because of their work in different contexts, the authors make recommendations for amending the Jisc Code of Practice in terms of consent, transparency, access, responsibility, privacy, validity, stewardship, and avoiding negative impact.

In the third paper, Steiner, Kickmeier-Rust, and Albert deal with a different context — a European
project covering several countries and diverse settings. In LEA in Private: A Privacy and Data Protection Framework for a Learning Analytics Toolbox, they describe how they moved beyond philosophical ideals, applying these as ethical principles that fed into the design and development of the project’s technologies. This involved drawing up a set of requirements that represented an accountable approach to the ethical and data protection issues associated with the project.

In the fourth paper, Cormack focuses specifically on A Data Protection Framework for Learning Analytics. Some aspects of data protection are necessarily contextual, as legislation varies from country to country. This paper is written from a European perspective. More broadly, Cormack reflects on the assumptions that underlie our treatment of data protection, particularly regarding informed consent. He suggests a distinction between the analysis that reveals significant patterns in the data, which requires safeguards for individuals, and the application of those patterns, which requires informed consent or perhaps even a contractual agreement.

While frameworks and codes of practice are important to the field, we also require tools and approaches that will enable us to put them into practice. In the fifth paper, Berg, Mol, Kismihók, and Sclater tackle The Role of a Reference Synthetic Data Generator within the Field of Learning Analytics. Synthetic data are simulated data that can be used when developing analytics in order to avoid accidental disclosure of reconstructed information. They have uses in such areas as security testing, interoperability testing, benchmarking, and staff training. Synthetic data can be used to protect privacy and to support data protection. More broadly, they help to establish the reliability and validity of the analytics used to support student success.

In the sixth paper, Khalil and Ebner consider the problem of De-Identification in Learning Analytics. One way of protecting privacy is to anonymize data. However, if learning analytics data are entirely detached from the person who generated them, they lose their value since the results cannot be returned to the individual. In addition, educational institutions need to know who has enrolled in a course, who has passed an exam, et cetera — and they may be required to report certain data to the government. De-identification helps to protect personal, or private, data while still making them accessible. This paper proposes a conceptual approach that combines anonymization strategies with learning analytics techniques.

In the seventh paper, Hoel and Chen consider Privacy-Driven Design of Learning Analytics Applications: Exploring the Design Space of Solutions for Data Sharing and Interoperability. Their analysis of over 200 questions generated by the LACE project found that discussion on data sharing and big data for education is still at an early stage. Conceptual issues dominate and there is still a long way to go before we are able to solve issues of technical development and implementation. They propose a Learning Analytics Design Space model to support the required solicitation and design for new learning analytics solutions and to deal with issues related to privacy, control of data, and trust.
In the final paper of this special issue, Prinsloo and Slade move away from frameworks and technical approaches to explore Student Vulnerability, Agency, and Learning Analytics. They suggest that it is impossible for individuals to comprehend, at the time of opting in or out, the scope of data that might be collected, analyzed, re-identified, and aggregated in future — or whether this matters. In the face of this vulnerability, they argue that the overriding responsibility is to restore autonomy to the individual, providing the information needed to make informed and supported choices.

7 CONCLUSION

Work in the area of ethics, data protection, and privacy is proceeding quickly, and we already have frameworks in place that can guide future work. Developing ethical practice is bound up with defining our vision for the future and our aims for what learning analytics could achieve. It is therefore a matter for the entire community, and we hope that this special issue, like the EP4LA workshops, stimulate discussion and action in this area.

8 ACKNOWLEDGEMENTS

The European Commission Seventh Framework Programme funds the Learning Analytics Exchange (LACE) project (www.laceproject.eu): grant number 619424.

REFERENCES


