PAUL PRINSLOO AND SHARON SLADE

STUDENT DATA PRIVACY AND INSTITUTIONAL ACCOUNTABILITY IN AN AGE OF SURVEILLANCE

INTRODUCTION

Modern higher education is accountable to a range of stakeholders – typically to national governments, sponsors and other authorities and, increasingly, to employers and students (Burke, 2005; Lauen & Gaddis, 2012; Marope, Wells, & Hazelkorn, 2013). This accountability extends to the use of resources, quality of teaching, and also higher education responses to meet socio-economic demands and needs. Higher education has always used a range of historical student data sources, such as school leaving marks, to plan institutional responses to support student learning. In the context of student retention and success, higher education institutions progressively rely on the real-time information trails left on institutional learning management systems and other platforms to support student intervention strategies and determine the subsequent allocation of resources (Long & Siemens, 2011). The harvesting and analysis combined with dynamic data sets from disparate sources offers huge potential to both provide increasingly complete student profiles, and to offer specific and relevant real-time guidance and customized support (Bichsel, 2012; Booth, 2012; Crow, 2012; Diaz & Brown, 2012; Siemens, 2011). As a result, accountability in higher education is likely to extend to more explicit information regarding which data are harvested and how they are analyzed, but also how they are used within institutions to influence decisions which aim to make teaching and learning more effective and appropriate.

The harvesting and use of individuals' data may be compared to an Orwellian Big Brother or a Foucaultian Panopticon where individuals are watched and then profiled, resulting in a Kafkaesque labyrinth where students may not know what data are harvested, for what purposes and by whom (Prinsloo, 2013; Solove, 2004). This may lead to a perception of students as producers of data and passive recipients of services. In such an unequal relationship, those institutions grant themselves unrestrained rights to harvest, analyze and employ student data. Further, based on a belief that bigger data must be, by definition, better data, higher education institutions often assume that the more data they collect, the more reliable and complete are the resulting student profiles.

While the notion of surveillance is an integral element in the discussion of the harvesting and use of student data, the issue of student data privacy can be approached from a number of different discourses such as institutional accountability, and legal and social definitions of privacy. Our intention is not to

M. E. Menon et al. (eds.), Using Data to Improve Higher Education, 197–214.

analyze and compare (inter)national legal frameworks and legislation, nor to consider the implications of different institutional policies relating to the uses of student data.

While many of the discourses on surveillance focus on concerns regarding an individual's right to privacy (Haggerty & Ericson, 2006; Lanier, 2013; Lyon, 2006; Morozov, 2013a; Solove, 2004), and the moral justification of surveillance (Bauman & Lyon, 2013; Haggerty & Ericson, 2006; Marwick, 2014; Mayer-Schönberger & Cukier, 2013), surveillance is much more than just watching and monitoring; it is also a "calculated practice for managing and manipulating human behavior" (Henman, 2004, p. 176). Surveillance should therefore also be understood within the discourses on governmentality. Focusing exclusively on surveillance is "too narrow" and an exploration of "surveillance as governance" provides a more "analytics capacity for understanding the rationale, operation, effects and transformations of surveillance" (Henman, 2004, p. 177).

In this chapter we aim to explore some of the assumptions and approaches to the use of student data in the context of the discourses of surveillance and specifically raise a number of concerns in the broad area of privacy. Having mapped some of the key issues, the chapter goes on to review a selected number of frameworks regarding the use and analysis of personal data. Based on this analysis, we explore a number of elements that could form the basis for a *student-centered* learning analytics.

PROBLEMATIZING THE HARVESTING, ANALYSES AND USE OF DATA

In any system with noisy data and underdeveloped theory, one of the many dangers associated with data analysis and the resulting predictions is to mistake the noise for the signal. Noise has the ability to pollute "journals, blogs, and news accounts with false alarms, undermining good science and setting back our ability to understand how the system really works" (Silver, 2012, p. 162).

In the wider context, the increasing amount of data noise often results in the creation of "an electronic collage … a life captured in records, a digital person composed in the collective computer networks" (Solove, 2004, p. 1). "Shards of data" are used, often out of context and without reference to timeframe, to create digital dossiers, and these may be used by a huge range of interested parties without necessarily taking into consideration the original purpose and context of data harvested. The producers and owners of these data often know very little about how that personal information is used, and lack the power to do much about it. This "elaborate lattice of information networking" (Solove, 2004, p. 3) consists of information flows between the different computer databases of both private sector and public sector organizations.

Many individuals willingly share personal information on an unprecedented scale, contributing to this "elaborate lattice of information networking" (Solove, 2004, p. 3) without knowing how the information will be used and with very little power to affect its use. Such information is unthinkingly shared through social networking, loyalty cards and online purchasing and browsing. This creates both the opportunity for various stakeholders to create individual “digital biographies” (Solove, 2004, p. 44), and a presumption that these digital biographies may be taken as complete, up-to-date and reliable sources of information. Solove (2004, p. 46) warns that “we are more than the bits of data we give off as we go about our lives. Our digital biography is revealing of ourselves but in a rather standardized way.” These personal overviews are not explicitly authorized, and are to some extent reductive, partial and often inaccurate (Solove, 2004).

Disturbingly, not only do commercial and geopolitical entities increasingly harvest and share data sets (Marwick, 2014; Mayer-Schönberger & Cukier, 2013) but consumers and individuals also increasingly voluntary give up certain elements of privacy for commercial or egoistical gain (Datoo, 2014). This results in “big data-as-a-service” becoming a very lucrative enterprise (Datoo, 2014; Marwick, 2014). “Data about your online and offline behavior are combined, analyzed, and sold to marketers, corporations, governments, and even criminals” (Marwick, 2014, para. 1). “A stupendous amount of information about our private lives is being stored, analyzed and acted on in advance of a demonstrated valid use for it” (Lanier, 2013, p. 69).

While this chapter will focus on the role of the institution in harvesting and analyzing student data, it is important to note that present day surveillance has changed from being uni-directional to a “mutual, horizontal practice” (Albrechtslund, 2008, para. 46). The changed nature of surveillance now also includes social and “playful aspects.” In the typology of surveillance developed by Knox (2010), it is clear that surveillance has evolved from being panoptic to synoptic – where everyone is engaged, in one way or another, in watching another (also see Lyon, 2001, 2006, 2007; Varvel, Montague, & Estabrook, 2007).

Boyd and Crawford (2013, p. 2) suggest that

> With the increased automation of data collection and analysis – as well as algorithms that can extract and inform us of massive patterns in human behavior – it is necessary to ask which systems are driving these practices, and which are regulating them.

They moot the following six propositions regarding the use of Big Data:

- Automating research changes the definition of knowledge
- Claims to objectivity and accuracy are misleading
- Bigger data are not always better data
- Not all data are equivalent
- Just because it is accessible, it doesn’t make it ethical
- Limited access to big data creates new digital divides.

Danaher (2014) and Morozov (2013a, 2013b) explore the dangers of being ruled by algorithm and the threat of algocracy. Morozov (2013a) points to the fact that both capitalism and bureaucratic administrations “thrive on information flows” and that legislation, technology and markets are active participants in maintaining the demand for data and sustaining capitalism (para. 8). Morozov (2013a) therefore petitions that there is more at stake than protecting the privacy of individuals. The solution to addressing the concerns regarding big data does not lie in more laws, or tools (ensuring privacy), but in placing the interrogation of big data into the political arena and linking "the future of privacy with the future of democracy in a way that refuses to reduce privacy either to markets or to laws" (para. 46). We therefore need to "politicize the debate about privacy and information sharing," learning to "sabotage the system" by refusing to share information through loyalty cards or participation in the "quantified self" movement and employ "provocative digital services" that reveal who benefits from tracking our digital footprints (Morozov, 2013a, para. 49). Such technologies and services may "help to equalize the balance of power between ordinary humans and epistemologically elite humans" (Danaher, 2014, para. 29).

Despite or amidst the hype around Big Data, Johnson (2013, p. 2) warns that

> the constructed nature of data makes it quite possible for injustices to be embedded in the data itself …. Whether by design or as unintended consequences, the process of constructing data builds social values and patterns of privilege into the data.

Sadowsky (2013, para. 4) therefore advises that "when data is used to allocate resources or anticipate needs, it can perpetuate injustices by over representing privileged groups of people." Solove (2004, p. 48) agrees and states that "databases do not cause the disempowering effects of bureaucracy; they exacerbate them – not merely by magnifying existing power imbalances but by transforming these relationships in profound ways that implicate our freedom." Rosen (2000, in Solove, 2004, p. 48) observes

> Privacy protects us from being misdefined and judged out of context in a world of short attention spans, a world in which information can easily be confused with knowledge. True knowledge of another person is the culmination of a slow process of mutual revelation.

According to Solove (2004, p. 51), privacy also "involves the ability to avoid the powerlessness of having others control information that can affect whether an individual gets a job, becomes licensed to practice in a profession, or obtains a critical loan."

Crawford (2013) warns against the inherent biases in Big Data often ignored by proponents in "data fundamentalism": "Data and data sets are not objective; they are creations of human design. We give numbers their voice, draw inferences from them, and define their meaning through our interpretations" (Crawford, 2013, para. 2). Crawford (2013) continues to warn that we stand the risk of misunderstanding and misallocating resources, because we presume that "big data's numbers …speak for themselves" (para. 5). She closes with the statement "raw data is an oxymoron" (para. 7).

New technologies that "transcend the physical, liberty-enhancing limitations of the old" (Marx, 1998, p. 171) are fast emerging, resulting in permeable boundaries around issues such as privacy, the public good, national security and corporate interests. Current frameworks which guide surveillance in terms of data protection and/or human rights do "not necessarily protect privacy" (Pounder, 2008, p. 1).

Current European legislation applying to surveillance, is spread over “a minimum of three separate pieces of legislation – data protection, human rights and the surveillance legislation,” leaving aggrieved individuals with “three possibly divergent routes of redress” (Pounder, 2008, p. 2).

Student Data: Issues to Consider

There is considerable hype and excitement surrounding the potential of learning analytics in higher education, with mounting claims made of how it will assist institutions in, inter alia, making more informed choices in resource allocation and improving student success (Booth, 2012; Long & Siemens, 2011; May, 2011; Oblinger, 2012; Siemens, 2011; Wagner & Ice, 2012). Learning analytics has the potential to assist higher education institutions to find patterns in random noise and to isolate significant signals amidst the increasing levels of data “noise” (e.g. Silver, 2012). Some authors though have flagged concerns regarding the preparedness of current institutional policies and frameworks to provide an enabling and ethical environment for the institutionalization of learning analytics (Bichsel, 2012; Ferguson, 2012; Prinsloo & Slade, 2013), and raise a number of ethical dilemmas associated with the move toward an unthinking automation of the harvesting, analysis and use of student data (Bollier, 2010; Ess, Buchanan, & Markham, 2012; Slade & Prinsloo, 2013).

Diaz and Brown (2012) state that in the broader genre of learning analytics, learners generate “digital footprints, or digital breadcrumbs” as they study and that these may be “supplemented or augmented by data about the learner, such as previous coursework, demographics, and other data that might exist in the student information system” (p. 2). The combination of these data trails and other datasets then allows analysts to “detect patterns and make predictions” (p. 2). Further, these patterns of sensemaking may be informed by comparing individual learners’ activities to others in a current or previous cohort, and also to their own activity in earlier courses, whether at the same institution or at a different institution. This raises a number of interesting points, namely:

- The implication that there be inter and intra-institutional integrated course platforms and data that allow comparisons within and between different student cohorts.
- An individual learner’s activity in one context is transferable to a different (disciplinary) context.
- Consideration of the different stages of students’ learning and life trajectories and the validity of comparing students at different stages.
- The need to understand student identity as a transient concept which ought not to be fully defined by current or earlier activity or data.

In contemplating the ethical dimensions of learning analytics, we find ourselves in the nexus (or liminal space) between a number of debates and discourses such as surveillance studies, the promise and perils of Big Data, and issues of governmentality and privacy. In this nexus, the various discourses often overlap, and the issues raised in one discourse often constitute a response in another.

Reflecting on the issue of ethics in learning analytics can take, as points of departure, elements from bioethics and patient privacy and the ethics in clinical trials; the debates and voices in ethics and morality; national security or legal perspectives on privacy and the ownership of data (Slade & Prinsloo, 2013). None of these approaches are necessarily mutually exclusive.

A review conducted by Prinsloo and Slade (2013) found that, in the case of two mega distance education institutions, the current policy frameworks regarding student data and privacy did not address crucial ethical issues with regard to the harvesting, storage, use and governance of student data and thus did not create an enabling environment for the institutionalization of learning analytics.

FRAMEWORKS FOR REALIZING THE POTENTIAL OF LEARNING ANALYTICS

In this section we explore a selected number of frameworks in an attempt to create a discursive space around the harvesting, and use of, personal information.

In 1973, a Code of Fair Information Practices was formulated which contained the following five principles (in Solove, 2004, p. 104):

- There may be no personal-date record-keeping systems whose existence is secret.
- There must be a way for an individual to find out what information about him (sic) is in the record and how it is used.
- There must be a way for an individual to prevent information obtained about him for one purpose from being used or made available for other purposes without his consent.
- There must be a way for an individual to correct or amend a record of identifiable information about him.
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take reasonable precautions to prevent misuse of the data.

Based on the 1973 Code of Fair Information Practices, Marx (1998) developed a set of 29 questions (see Table 1) dealing with the approach, context and purposes of information gathering, with an emphasis on “the watchers rather than the watched, on avoiding harm rather than doing good, on the individual more than the group, and on the short rather than the long run” (Marx, 1998, p. 173).

Marx recognized that the questions would not satisfy those who “lust after a Rosetta stone of clear and consistent justifications” but aimed to provide the basis for “an imperfect compass than a detailed map” (Marx, 1998, p. 182). Due to the complex nature and interrelationships of the issues involved in the harvesting and use of data in particular contexts, Marx (1998) states that a detailed map “can lead to the erroneous conclusion that ethical directions can be easily reached or to a statement so far in the stratosphere that only angels can see and apply it” (p. 182) and proposed “simple coordinates and rough estimates” to guide ethical data harvesting and use (p. 174).

Table 1. Questions to help determine the ethics of surveillance (Marx, 1998, p. 174).

Means	1. Harm: Does the technique cause unwarranted physical or psychological harm? 2. Boundary: Does the technique cross a personal boundary without permission (whether involving coercion or deception or a body, relational, or spatial border)? 3. Trust: Does the technique violate assumptions that are made about how personal information will be treated, such as secret recordings? 4. Personal relationships: Is the tactic applied in a personal or impersonal setting? 5. Invalidity: Does the technique produce invalid results?
Data collection context	6. Awareness: Are individuals aware that personal information is being collected, who seeks it, and why? 7. Consent: Do individuals consent to the data collection? 8. Golden rule: Would those responsible for the surveillance (both the decision to apply it and its actual application) agree to be its subjects under the conditions in which they apply it to others? 9. Minimization: Does a principle of minimization apply? 10. Public decision-making: Was the decision to use a tactic arrived at through some public discussion and decision-making process? 11. Human review: Is there human review of machine-generated results? 12. Right of inspection: Are people aware of the findings and how they were created? 13. Right to challenge and express a grievance: Are there procedures for challenging the results, or for entering alternative data or interpretations into the record? 14. Redress and sanctions: If the individual has been treated unfairly and procedures violated, are there appropriate means of redress? Are there means for discovering violations and penalties to encourage responsible surveillant behavior? 15. Adequate data stewardship and protection: Can the security of the data be adequately protected? 16. Equality-inequality regarding availability and application: (a) Is the means widely available or restricted to only the most wealthy, powerful, or technologically sophisticated? (b) Within a setting is the tactic broadly applied to all people or only to those less powerful or unable to resist? (c) If there are means of resisting the provision of personal information, are these means equally available, or restricted to the most privileged? 17. The symbolic meaning of a method: What does the use of a method communicate more generally? 18. The creation of unwanted precedents: Is it likely to create precedents that will lead to its application in undesirable ways? 19. Negative effects on surveillants and third parties: Are there negative effects on those beyond the subject and, if so, can they be adequately mediated?

Uses	20. Beneficiary: Does application of the tactic serve broad community goals, the goals of the object of surveillance, or the personal goals of the data collector? 21. Proportionality: Is there an appropriate balance between the importance of the goal and the cost of the means? 22. Alternative means: Are other, less costly means available? 23. Consequences of inaction: Where the means are very costly, what are the consequences of taking no surveillance action? 24. Protections: Are adequate steps taken to minimize costs and risk? 25. Appropriate vs. inappropriate goals: Are the goals of the data collection legitimate? 26. The goodness of fit between the means and the goal: Is there a clear link between the information collected and the goal sought? 27. Information used for original vs. other unrelated purposes: Is the personal information used for the reasons offered for its collection and for which consent may have been given, and do the data stay with the original collector, or do they migrate elsewhere? 28. Failure to share secondary gains from the information: Are the personal data collected used for profit without permission from, or benefit to, the person who provided it? 29. Unfair disadvantage: Is the information used in such a way as to cause unwarranted harm or disadvantage to its subject?

More recently, Pounder (2008) proposed a number of principles to support the harvesting and use of personal data. The principles range from considering the rationale for the surveillance (principle 1), the process and elements required in getting approval to collect data (principle 2), the need to separate the authority who does the surveillance and the authority who oversees the surveillance (principle 3), adherence to agreed upon principles by those who do the surveillance (principle 4), rules and guidelines to ensure transparent and accountable reporting (principle 5), the need for independent supervision of surveillance activities (principle 6), the protection of the privacy rights of individuals (principle 7), procedures to ensure compensation should surveillance activities cause harm (principle 8), and the principle that ensures that surveillance is ceased if conformity to the previous eight principles is compromised (principle 9).

Table 2. A summary of Pounder's (2008) nine principles.

Principle 1: The justification principle	An assessment must be possible to ensure that "surveillance can be justified in terms of pressing social needs and measurable outcomes" (p. 11-12). Information regarding the surveillance policy (e.g. justification, complaints procedures) "should be made proactively available by the public authority performing the surveillance (e.g. on an appropriate web-site)" (p. 12).
Principle 2: The approval principle	Surveillance must be limited to lawful purposes based on legislation/policy that "has been thoroughly scrutinized" and

	"where appropriate, informed public debate has taken place" (p. 13). The first two principles "are likely to draw out any alternatives to the surveillance, and thereby strengthen the justification for, and the public acceptability of, any surveillance that is eventually authorized" (p. 13).
Principle 3: The separation principle	The authority that authorizes the surveillance cannot be the same authority that sets the procedures for surveillance and monitors the surveillance – "the more invasive the surveillance, the wider the degree of separation" (p. 14).
Principle 4: The adherence principle	Surveillance should be managed in a professional way and audited; staff should be adequately trained and the training assessed; and "any malfeasance in relation to a surveillance activity can be identified and individuals concerned suitably punished" (p. 15). Should individuals raise legitimate concerns, the appointed regulator should "possess sufficient clout to resolve and investigate any problem" (p. 16).
Principle 5: The reporting principle	An appointed regulator shall "determine what records, … are retained and maintained concerning a surveillance activity to ensure transparency and accountability" to appropriate structures (e.g. the public, Parliament) (p. 16).
Principle 6: The independent supervision principle	This principle emphasizes that the surveillance activity is independent of, e.g. the Government, "well-financed, and has effective powers of investigation and can delve into operational matters." Also, "the more invasive the surveillance, the more important it is for the powers of the Regulator to be available" (p. 17).
Principle 7: The privacy principle	This principle protects individuals' right to the privacy of personal data and includes "the right to object to the processing of personal data in appropriate circumstances" (p. 18).
Principle 8: The compensation principle	In the case where individuals suffer damage, distress or detriment caused by surveillance, individuals have the right to compensation.
Principle 9: The unacceptability principle	In the event where the previous eight principles cannot be complied with, the surveillance should cease, or alternative measures should be taken to ensure conformity, or an appropriate regulatory or legislative body should approve non-compliance.

While the above two frameworks do help to highlight a number of relevant issues relating to surveillance, Solove (2004) warns that current structures represent an "architecture of vulnerability, one with large holes, gaps, and weak spots" (p. 119). The harm is not only due to the gaps and holes, but "caused by the architecture itself" (p. 119). The only way out of this impasse is through implementation of the two general aims of the Fair Information Practices, namely participation and responsibility (Solove, 2004). Solove (2004) suggests as a requirement the participation of individuals and groups in the harvesting and use of their own personal information, and secondly, that the "collection and use of personal information is an activity that carries duties and responsibilities" (Solove, 2004, p. 121).

In following Diller (1996), Gilligan (1982) and Held (2005), we propose an ethics architecture or an "ethics of care" as a basis for a moral approach to learning analytics that may stand in stark contradiction to the governmental and "technocratic predictive logic" inherent in much of the current discourses in learning analytics. An ethics of care is a counter-narrative to the dominant neoliberal discourses providing a basis for the hegemony of managerialism and performativity in higher education (Hennessy & McNamara, 2013; Peters, 2013). Many of the discourses on the impact of surveillance practices focus on individuals' right to privacy, often juxtaposed to the moral justification of surveillance by state agencies (Bauman & Lyon, 2013; Marwick, 2014; Mayer-Schönberger & Cukier, 2013). An ethics of care focuses on "the responsibility of people to others in caring for relationships" (Patton, 2000, para. 24). Seeing learning analytics as a *relational* practice means that we should never separate the practice of learning analytics from the consequences of conflating information about a person with the inherent worth and future potential of an individual (Bauman & Lyon, 2013). The technocratic logic in Big Data and learning analytics often results in disembodied information profoundly shaping, often irrevocably, the futures of individuals and groups of people.

An ethics of care will involve individuals and groups in the gathering and use of their personal information; providing individuals and groups with access to the stored information and insight into how it may be used, and would go some way to addressing many concerns. Indeed, it might be suggested that a more responsible approach would be to offer a default option to opt in to the harvesting and use of personal information rather than the more usual default option to opt out. "The architecture should empower people with an easy, quick, and convenient way to challenge inaccuracies about their personal information as well as fraudulent entries …" (Solove, 2004, p. 121).

The second element of an ethics architecture is the principle that the harvesting, storage and use of personal information entails clear duties of stewardships and responsibilities. Ethics architectures for learning analytics should address specific issues in different contexts in response to an institution's "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 socioeconomic contexts" (Slade & Prinsloo, 2013, p. 9). Slade and Prinsloo (2013) therefore developed a set of principles that are, on the one hand, broad enough to allow for context and institution-specific responses, and on the other, offer sufficient clarity on foundational issues regarding ethical issues in learning analytics. They list the following principles:

1. Learning analytics as moral practice: Amidst the increasing technocratic pressures to harvest and report on data, education, per se, should focus on appropriate and desirable outcomes, not only on those interventions that prove effective. Interventions can be effective, but neither appropriate nor desirable (e.g. Biesta, 2007). Learning analytics as moral practice is a counter-narrative to attempts to justify the harvesting and analysis of data, often without consent, on the basis that the end justifies the means.

2. Students as agents: Students are much more than passive recipients of services and/or producers of data. Students can and should be seen as agents making informed decisions regarding, inter alia, the number of courses they take, and decisions to dropout. Stage and Hossler (2000) suggest that students are active agents in the whole process of making choices regarding persisting with or cancelling their studies. Students are therefore not "passive recipients of experiences" (Stage & Hossler, 2000, p. 172). Students' self-efficacy is furthermore not linear and only progressive, but often spiral and cyclical (Prinsloo, 2009; Stage & Hossler 2000; Subotzky & Prinsloo, 2011).

   Learning analytics as moral and student-centered practice should involve students in proactively sharing information as well as co-interpreting and updating outdated data (see also Kruse & Pongsajapan, 2012). Students should therefore be equal partners in learning analytics as a discursive and disclosive space (Stoddart, 2012) and be enabled to see learning analytics as serving their learning and development. Diaz and Brown (2012, p. 7) emphasize that we must realize "that students are not univariate actors. Students engage in various activities related to learning, making it difficult to arrive at definitive conclusions, especially for the middle band of students, as opposed to the high- and low-performing students."
3. Student identity and performance as temporal dynamic constructs: Slade and Prinsloo (2013) point to student identity and their trajectories of becoming as central to learning analytics as moral practice. "Students should be allowed to evolve and adjust and learn from past experiences without those experiences, due to their digital nature, becoming permanent blemishes on their development history" (p. 11). While digital dossiers and biographies (Mayer-Schönberger, 2009; Solove, 2004) do not have expiry dates and often function as "digital tattoos" (Mayer-Schönberger, 2009, p. 14), student records should have "an agreed-upon life span and expiry date, as well as mechanisms for students to request data deletion under agreed-upon criteria" (Slade & Prinsloo, 2013, p. 11).
4. Student success is a complex and multidimensional phenomenon: Student success is often portrayed as a one-sided affair where the full responsibility for success (or failure) rests on students. Student success is the result of mostly non-linear, multidimensional, interdependent interactions at different phases in the nexus between student, institution and broader societal factors (Subotzky & Prinsloo, 2011). One of the dangers in learning analytics as an act of harvesting and analysis of student data is that we can forget that success or failure often results from a mismatch between personal and institutional dispositions and processes and the dynamic interaction with macro-societal factors (Subotzky & Prinsloo, 2011). This is in stark contrast with the proposal made by Willis, Campbell and Pistilli (2013, para. 13) that emphasizes ethical principles that are "actionable for the student."
5. Transparency: A recurring theme of the metaphors discussed in this paper, is the lack of transparency not only in the methods used for the collection of data, but also its use. Slade and Prinsloo (2013, p. 11) state that "higher education

institutions should be transparent regarding the purposes for which data will be used, under which conditions, who will have access to data and the measures through which individuals' identity will be protected."

6. Higher education cannot afford to not use data: The first five principles suggested by Slade and Prinsloo (2013, p. 12) form the basis for the use of student data in learning analytics. The sixth principle "makes it clear that higher education institutions cannot afford to *not* use learning analytics." Learning analytics allows higher education institutions to be more accountable to all stakeholders (including students) and allows institutions to be more transparent with regard to the allocation of resources and pedagogical decisions taken. Willis et al. (2013, para. 1) also highlight the responsibility that comes from "knowing" – "once an administration 'knows' something about student performance, what ethical obligations follow?" (also see Diaz & Brown, 2012).

Based on these principles, Slade and Prinsloo (2013) continue to moot a number of considerations for learning analytics as moral practice, namely:

- Who benefits and under what conditions? This question is the most important question to clarify and forms the basis for considering the ethical implications in learning analytics' regimes. This does not, however, mean that because learning analytics is seen as serving the interests of students, other considerations with regard to consent, privacy, etc. are disqualified. Trust between the institution and students is of extreme importance and Stoddart's (2012) proposal of a discursive-disclosive regime is appropriate.
- Conditions for consent, de-identification, and opting out, including considerations regarding vulnerability and harm. Slade and Prinsloo (2013) ask whether there are any conditions where the notion of informed consent as default may be waived, and what would be the criteria for doing so. This also raises the issue whether the waiving of the (current) default position of "opting out" should not be changed to "opting in." Students may then be consulted with regard to which data are included and excluded by "opting in" and made to understand the implications of "opting out."
- Vulnerability and harm involve the consequences of being profiled as, for example, a student who is "at risk" based on a number of characteristics and harvested data. This also raises the question regarding the validity of the combination of criteria and the epistemologies on which certain algorithms are based (see, for example, Hickman, 2013). Based on the dangers of stereotyping or classifications based on incomplete data, Slade and Prinsloo (2013, p. 14) suggest that "institutions should provide additional opportunities for these students either to prove the initial predictive analyses wrong or incomplete, or to redeem themselves despite any initial institutional doubt regarding their potential." There is also a need to provide mechanisms for redress for students and institutions alike (Slade & Prinsloo, 2013).
- Data collection, analyses, access and storage. Student learning and progress involves much more than the digital breadcrumbs left on an institutional learning management system (LMS), and there is a need to review expectations regarding the predictive value of such data. Should institutions also harvest and

combine data from sources outside of the LMS, students have not only the right to know, but also the right to provide consent and have access to the analyses and profiling in their digital dossiers. Slade and Prinsloo (2013) stress that data harvested in one context do not necessarily transfer to other contexts. Diaz and Brown (2012) warn that "…while the LMS may be the starting point for much early learning analytics work, learning is a complex and highly contextualized activity" (p. 7-8). Online algorithms furthermore also create the possibility for "autopropaganda" resulting in an incestuous cycle of recycling past actions and search histories (Pariser, 2011).

Student-centered learning analytics (as suggested by Kruse & Pongsajapan, 2012; Slade & Prinsloo, 2013) enshrines access to the processes involved in the harvesting of data, the epistemologies, assumptions and evidence on which algorithms are based, as well as the scope and content of students' digital dossiers. Students have the right to know how data are stored, who has access to their digital dossiers (and under which conditions) and be informed regarding the process of redress in case of a breach of trust and violation.

- Governance and resource allocation. The adequate allocation of resources and providing an enabling and protective policy environment is an essential requirement for learning analytics as moral practice.

TOWARDS A STUDENT CENTERED LEARNING ANALYTICS CONTRACT

The idea of a social contract is not new – although applied to the context of learning analytics, the application of the notion of a social contract between students and the institution is novel. Since its origins in the classical Greek and Roman periods, throughout the work of Hobbes, Locke and Rosseau, the social contract between the individual and the state entails, in broad strokes, mutual agreement that the individual agrees to forego or cede some of his or her rights for a range of other protections and services.

A number of authors (e.g. Bauman, 1998, 2011, 2012; Bauman & Lyon, 2013; Henman, 2004; Mayer-Schönberger & Cukier, 2013) discuss how the nature of the social contract between state and corporate capital and individuals has irrevocably changed with ubiquitous surveillance as one of the defining characteristics of the 21st century (Lyon, Haggerty, & Ball, 2012). In this new social contract, the collection, and analysis of information and the resulting segmentation of populations and profiling of individuals have become naturalized (Henman, 2004; Marwick, 2014; Webster 2004).

Against this backdrop, we petition for a new social contract based on an ethics of care. The following elements are, in our opinion, the minimum requirements for a fair and student-centered approach to learning analytics.

a. The harvesting and analysis of aggregated, non-personalized data is essential for the rendering of effective and appropriate teaching and learning opportunities and pathways, and is therefore a justified and integral part of the mandate of a higher education institution (e.g. see, Pounder, 2008, first principle; Prinsloo & Slade, 2013; Slade & Prinsloo, 2013). Even though higher education institutions have access to data of individual students, students should be given an opportunity to make informed decisions regarding whether to opt in to specialized, customized services and support (see, Marx, 1998, questions 6-19). It is crucial that the institution and student understand the role of students as agents, and not merely as the producers of data and passive recipients of services (Diaz & Brown, 2012; Slade & Prinsloo, 2013; Stoddart, 2012).

b. Students are informed regarding the scope of data harvested, the purpose of the harvesting, the conceptual models informing the algorithms used to analyze the data, the individuals and departments that have access to the data, the life span of their digital records (Slade & Prinsloo, 2013) or "digital tattoos" (Mayer-Schönberger, 2009), and procedures to have access to their digital dossiers (Kruse & Pongsajapan, 2012; Marx, 1998, question 12; Pounder, 2008, principle 6).

c. Students accept the responsibility to inform the institution of any change in life circumstances that affects not only the relevance/accuracy of students' digital dossiers, but also the effectiveness and appropriateness of the services which the institution renders (Diaz & Brown, 2012). Slade and Prinsloo (2013) emphasize that student identities are temporal and dynamic constructs necessitating procedures to ensure the relevance and accuracy of student data.

d. The institution should ensure that the surveillance of activities and the harvesting of data will not harm, in any way, students' progress in their studies (see, Marx, 1998, questions 1-5). Students do not have the option to opt in to the harvesting and analyses of aggregated student data, but can opt out of having their personalized data shared with some stakeholders. In choosing to opt in or out, students understand and accept the implications of their choices. The university commits itself to adequate data stewardship and data security.

e. The university commits to having human review of machine-generated results and should there be any possibility of different interpretation of the data, students themselves will have the opportunity to confirm the analysis (Boyd & Crawford, 2013). There are furthermore established procedures for challenging the result of an analysis. In the event of unfair treatment or violation of trust or procedures, students have access to redress (see, Marx, 1998, questions 11-14).

f. Higher education institutions accept the reality that the available data, algorithms employed, and analyses provide context and time-specific, provisional and incomplete pictures of students (Crawford, 2013; Johnson, 2013). Institutions will, therefore, take reasonable steps to ensure that the algorithms used to analyze data are frequently reviewed and validated (Boyd & Crawford, 2013; Slade & Prinsloo, 2013).

CONCLUSIONS

Exploring learning analytics against the backdrop of increased accountability in an age of surveillance opens a necessary discursive space offering a much-needed critical lens on the issue of student data privacy in higher education.

Current frameworks and architectures exploring the complexities and ethical dilemmas in the harvesting and analyses of data are mostly incomplete with a number of "large holes, gaps, and weak spots" (Solove, 2004, p. 119). Acknowledging the gaps and vulnerabilities in these architectures can be addressed through participation and responsibility (Solove, 2004). Should higher education realize the immense potential of learning analytics (Booth, 2012; Long & Siemens, 2011; Oblinger, 2012; Siemens, 2011), the effort must be driven through as a result of the active engagement of students – not as producers of data but as full participants – with rights as well as duties (Kruse & Pongsajapan, 2012).

A student-centered learning analytics contract based on an ethics of care acknowledges that higher education institutions cannot afford not to harvest, integrate and analyze disparate data sets in order to plan more effective and appropriate learner support. Realizing the potential of learning analytics can be enhanced when students know the scope and purpose of data harvesting as well as have access to their digital dossiers. Where students' personalized data are used and shared, students furthermore have a right and responsibility to make informed decisions to consciously opt in or out of personalized and customized support.

Student identities are dynamic, temporal, context and time-specific constructs. A student-centered learning analytics contract therefore requires students to accept a responsibility to contribute relevant and correct information and institutions of higher learning to use this information with care.

While Big Data clearly offer considerable opportunities for tailored and directed support, the provision of a student-centered learning analytics approach should also provide the necessary checks and balances to protect both students and higher education institutions from data fundamentalism, the dominance of technocratic predictive logic, and from confusing noise as signal.

REFERENCES

Albrechtslund, A. (2008). Online social networking as participatory surveillance. *First Monday, 13*(3). Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2142/1949

Bauman, Z. (1998). *Globalization: The human consequences*. Cambridge, U.K.: Polity Press.

Bauman, Z. (2011). *Collateral damage: Social inequalities in a global age*. Cambridge, U.K.: Polity Press.

Bauman, Z. (2012). *On education: Conversations with Riccardo Mazzeo*. Cambridge, U.K.: Polity Press.

Bauman, Z., & Lyon, D. (2013). *Liquid surveillance*. Cambridge, U.K.: Polity Press.

Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress and recommendations*. Louisville, Colorado: EDUCAUSE Center for Applied Research.

Biesta, G. (2007). Why ''what works'' won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory, 57*(1), 1-22. Retrieved from http://www.vbsinternational.eu/files/media/research_article/Evidencebased_education_Biesta1.pdf

Bollier, D. (2010). *The promise and peril of big data*. Washington, DC: The Aspen Institute. Retrieved from http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

Booth, M. (2012, July 18). *Learning analytics: The new black*. Retrieved from http://www.educause.edu/ero/article/learning-analytics-new-black

Boyd, D., & Crawford, K. (2013). *Six provocations for big data*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431

Burke, J. S. (2005). *Achieving accountability in higher education: Balancing public, academic and market demands.* San Francisco, CA: John Wiley & Sons.

Code of Fair Information Practices. (1973). Retrieved from http://simson.net/ref/2004/csg357/handouts/01_fips.pdf

Crawford, K. (2013, April 1). The hidden biases in big data. *Harvard Business Review*. Retrieved from http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html

Crow, M. M. (2012, July 18). No more excuses: Michael M. Crow on Analytics. Retrieved from https://net.educause.edu/ir/library/pdf/ERM1241P.pdf

Danaher, J. (2014, January 7). *Rule by algorithm? Big data and the threat of algocracy*. Retrieved from http://ieet.org/index.php/IEET/more/danaher20140107

Datoo, S. (2014, January 14). *Big data: 4 predictions for 2014*. Retrieved from http://www.theguardian.com/technology/datablog/2014/jan/14/big-data-4-predictions-for-2014

Diaz, V., & Brown, M. (2012). *Learning analytics: A report on the ELI focus session*. Retrieved from http://net.educause.edu/ir/library/PDF/ELI3027.pdf

Diller, A. (1996). Ethics of care and education: A new paradigm, its critics, and its educational significance. In A. Diller, B. Houston, K. P. Morgan, & M. Ayim (Eds.) *The gender question in education: Theory, pedagogy and politics*. Boulder, Colorado: Westview Press.

Ess, C., Buchanan, E., & Markham, A. (2012). *Ethical decision-making and internet research: 2012 recommendations from the AOIR ethics working committee*. Unpublished draft.

Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges* (Technical Report KMI-12-0). Milton Keynes, U.K.: Knowledge Media Institute, The Open University. Retrieved from http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Haggerty, K. D., & Ericson, R. V. (Eds.). (2006). *The new politics of surveillance and visibility*. Toronto, Canada: University of Toronto Press.

Held, V. (2005). *The ethics of care: Personal, political, and global*. New York: Oxford University Press.

Henman, P. (2004). Targeted! Population segmentation, electronic surveillance and governing the unemployed in Australia. *International Sociology, 19*, 173-191.

Hennessy, J., & McNamara, P.M. (2013). At the altar of educational efficiency: Performativity and the role of the teacher. *English Teaching: Practice and Critique, 12*(1), 6-22.

Hickman, L. (2013, July 1). How algorithms rule the world. *The Guardian*. Retrieved from http://www.guardian.co.uk/science/2013/jul/01/how-algorithms-rule-world-nsa

Johnson, J. A. (2013, April). *From pen data to information justice*. Paper presented at Midwest Political Science Association Annual Conference, Chicago, Illinois, USA. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2241092

Knox, D. (2010, October). *Spies in the house of learning: A typology of surveillance in online learning environments*. Paper presented at Edge2010, Memorial University of Newfoundland, St. Johns, Newfoundland, Canada. Retrieved from http://www.mun.ca/edge2010/wp-content/uploads/Knox-Dan-Spies-In-the-House.pdf

Kruse, A., & Pongsajapan, R. (2012). *Student-centered learning analytics*. Retrieved from https://cndls.georgetown.edu/m/documents/thoughtpaper-krusepongsajapan.pdf

Lanier, J. (2013). How should we think about privacy? Making sense of one of the thorniest issues of the digital age. *Scientific American, 309*(5) 64-71.

Lauen, D. L., & Gaddis, S. M. (2012). *Accountability pressure, academic standards, and educational triage*. Society for Research on Educational Effectiveness. Retrieved from http://files.eric.ed.gov/fulltext/ED530122.pdf

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, September/October, 31-40.

Lyon, D. (2001). Facing the future: Seeking ethics for everyday surveillance. *Ethics and Information Technology, 3*, 171-181.

Lyon, D. (Ed.). (2006). *Theorizing surveillance: The panopticon and beyond*. Devon, U.K.: Willan Publishing.

Lyon, D. (2007). *Surveillance studies: An overview.* Cambridge, U.K.: Polity Press.

Lyon, D., Haggerty, K. D., & Ball, K. (2012). *Routledge handbook of surveillance studies*. New York, NY: Routledge.

Marope, P. T. M., Wells, P. J., & Hazelkorn, E. (Eds.). (2013). *Rankings and accountability in higher education: Uses and misuses*. Paris, France: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0022/002207/220789e.pdf

Marwick, A. E. (2014). How your data are being deeply mined. *The New York Review of Books*. Retrieved from http://www.nybooks.com/articles/archives/2014/jan/09/how-your-data-are-being-deeply-mined/?pagination=false

Marx, G. T. (1998). Ethics for the new surveillance. *The Information Society: An International Journal, 14*(3), 171-185. DOI: org/10.1080/019722498128809.

May, H. (2011). *Is all the fuss about learning analytics just hype?* Retrieved from http://www.loomlearning.com/2011/analytics-schmanalytics

Mayer-Schönberger, V. (2009). *The virtue of forgetting in the digital age*. Princeton, NJ: Princeton University Press.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt Publishing Company.

Morozov, E. (2013a). The real privacy problem. *MIT Technology Review*. Retrieved from http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/

Morozov, E. (2013b). *To save everything, click here: Technology, solutionism, and the urge to fix problems that don't exist*. London, U.K.: Penguin Books.

Oblinger, D. G. (2012). Let's talk analytics. *EDUCAUSE Review*, July/August, 10-13.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London: Viking.

Patton, J. W. (2000). Protecting privacy in public? Surveillance technologies and the value of public places. *Ethics and Information Technology, 2*, 181-187.

Pounder, C. N. M. (2008). Nine principles for assessing whether privacy is protected in a surveillance society. *Identity in the Information Society, 1*, 1-22. DOI 10.1007/s12394-008-0002-2.

Peters, M. A. (2013). Managerialism and the neoliberal university: Prospects for new forms of 'open management' in higher education. *Contemporary Readings in Law and Social Justice, 5*(1), 11-26.

Prinsloo, P. (2009). *Modeling throughput at Unisa: The key to the successful implementation of ODL*. Retrieved from http://umkn-dsp01.unisa.ac.za/handle/10500/6035

Prinsloo, P. (2013). *Ethics and learning analytics as a Faustian pact: Between Orwell, Huxley, Kafka and the deep blue sea.* Presented at the LASI13 conference, Pretoria, South Africa. Retrieved from http://www.up.ac.za/telematic/sahela2013/day2-9h30-sahela2013-paul_UNISA_ethics_and_learning_analytics_as_a_fuastian_pact.pdf

Prinsloo, P., & Slade, S. (2013, April). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 240-244). New York: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2460344

Rosen, J. (2000). *The unwanted gaze: The destruction of privacy in America*. New York, NY: Random House.

Sadowsky, J. (2013). *The injustices of open data*. Retrieved from http://www.slate.com/blogs/future_tense/2013/06/28/open_data_can_promote_social_injustice.html

Siemens, G. (2011). *Learning analytics: Envisioning a research discipline and a domain of practice*. Paper presented at 2nd International Conference on Learning Analytics and Knowledge (LAK12), Vancouver, Canada. Retrieved from http://learninganalytics.net/LAK_12_keynote_Siemens.pdf

Silver, N. (2012). *The signal and the noise: The art and science of prediction*. London, U.K.: Allen Lane.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57(*10), 1509-1528. DOI: 10.1177/0002764213479366.

Solove, D. J. (2004) *The digital person: Technology and privacy in the information age*. New York, NY: New York University Press.

Stage, F. K., & Hossler, D. (2000). Where is the student? Linking student behaviors, college choice, and college persistence. In J. M. Braxton (Ed.) *Reworking the student departure puzzle*. Nashville, TN: Vanderbilt University Press.

Stoddart, E. (2012). A surveillance of care: Evaluating surveillance ethically. In K. Ball, K. D. Haggerty, & D. Lyon (Eds.) *Routledge handbook of surveillance studies.* London, U.K.: Routledge.

Subotzky, S., & Prinsloo, P. (2011). Turning the tide: A socio-critical model and framework for improving student success in open distance learning at the University of South Africa. *Distance Education, 32*(2), 177-193.

Varvel, V. E., Montague, R. A, & Estabrook, L. S. (2007). Policy and e-learning. In R. Andrews & C. Haythornthwaite (Eds.) *The Sage handbook of e-learning research.* London: Sage.

Wagner, E., & Ice, P. (2012). *Data changes everything: Delivering on the promise of learning analytics in higher education.* Retrieved from http://www.educause.edu/ero/article/data-changes-everything-delivering-promise-learning-analytics-higher-education

Webster, A. (2004) State of the art, risk, science and policy – Researching the social management of uncertainty. *Policy Studies, 25*(1), 5-18. DOI: 10.1080/0144287042000208206.

Willis, J. E. III, Campbell, J. P., & Pistilli, M. D. (2013). Ethics, big data, and analytics: A model for application. Retrieved from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application

Paul Prinsloo
The Institute for Open and Distance Learning
University of South Africa

Sharon Slade
Faculty of Business and Law
Open University (United Kingdom)