Guidelines on the Ethical Use of Student Data: A Draft Narrative Framework

Paul Prinsloo
University of South Africa (Unisa)
prinsp@unisa.ac.za

Siyaphumelela
we succeed

"If you have come to help us, you can go home. If you have come to accompany us, please come. We can talk"[1]

"The question of whose interests are served is central. And of course, there is clear advantage for those who collect and control the data and information over those who provide the data and seek to benefit from that contribution"[2]

"The governance of data – that is, who has the power and authority to make rules and decisions about the design, interpretation, validation, ownership, access to and use of data – has emerged as a site of contestation between indigenous peoples and the colonial settler states within which they reside."[3]

"When institutions use race, ethnicity, age, gender, or socioeconomic status to target students for enrolment or intervention, they can intentionally, or not, reinforce… inequality"[4]

Acknowledgements

In drafting this draft narrative framework I relied and made use of a number of other policies and frameworks[5][6]; my presentation[7] at the Siyaphumelela workshop hosted on 27 October 2016 at an Ethics Symposium as part of the Siyaphumelela Project (Kopanong Hotel & Conference Centre, Johannesburg, South Africa); and the report and draft guidelines that were drafted after the workshop of 27 October.

This draft narrative framework serves as an invitation to institutions and individuals to consider the different aspects and proposed guiding principles as basis for developing context-appropriate ethical approaches to the collection, analysis and use of student data.

I am indebted to Jan W. Lyddon, (Ph.D), (Principal, Organizational Effectiveness Consultants, Columbus, OH) for her guidance, input, generous support, and copy editing.

This framework is subject to and should be read in conjunction with the South African Protection of Personal Information Act (2013)[8]

Paul Prinsloo
Research Professor in Open and Distance Learning
University of South Africa
27 June 2017
prinsp@unisa.ac.za

Contents

Background

The five Siyaphumelela Partner Institutions (Durban University of Technology, Nelson Mandela Metropolitan University, University of the Free State, University of Pretoria and University of Witwatersrand) have committed to collect, use and analyse student data to drive an evidence-based approach to enhancing student success. This draft narrative framework and guidelines attempt to provide a broad basis for consideration in the ethical collection, analysis and use of student data, specifically in a South African higher education context. The ethical collection, analysis and use of data goes *beyond* privacy issues, and involve a balancing act between institutional risks (funding constraints, operational inefficiencies, socioeconomic downturn, etc.), personal risks (bias, stereotyping, disclosure, unintended harm, etc.), care and context.

Siyaphumelela Partner Institutions acknowledge that we cannot ignore the use of data during the colonisation of Africa and in apartheid South Africa where the definition and selection of data, data collection, analysis and data use were utilised as political acts and served declared and hidden assumptions.

There are different approaches to consider the ethical implications in the collection, analysis and use of student data. There are also various lenses on determining the scope and content of ethical considerations such as (1) an utilitarian approach (deciding on an action that "provides the greatest balance of good over evil"); (2) a rights approach (referring to basic, universal rights such as the right to privacy, or not to be injured); (3) a fairness or justice approach; (4) the common-good approach (where the welfare of the individual is linked to the welfare of the community); and (5) the virtue approach (based on the aspiration towards certain shared ideals).[9]

Should we furthermore understand learning analytics as Research (with a capital 'R') or as an emerging form of research (see Principle 5); both positions have implications for how we approach this framework. In the context of Research (with a capital 'R'), the 'procedural' notions of informed consent, confidentiality and preventing unintended harm are well-established but it is unclear how and if these traditional procedural elements of ethics need to be considered in the context of learning analytics.[10] [11]

In light of the recent student protests and demands to decolonise the curriculum in the broader context of transforming South African higher education, we also have to consider the implications of decolonising methodologies and ethics in learning analytics. Responding to this need, recent theoretical work on researching indigenous peoples and vulnerable populations[12] [13] [14]may hold huge potential to (re)consider the ethical implications of learning analytics in the South African context.

The need to consider the right of marginalised and vulnerable populations to control any knowledge produced about them informed by this narrative framework. As such, this framework falls into the broad category of *emancipatory* research with its "reliance on the ethics of human rights and equal power, and acknowledgement of the ways in which 'the academy and academic knowledge in particular are deeply implicated in the operations of power."[15] Emancipatory research aims to prevent quantified, coded, "shallow, monocled gazes" and embraces an ethics of reciprocation that "gives back ownership of knowledge and material benefit to those participating in research."[16] In the context of the "increasing algorithmic processing and data as an emergent regime of power/knowledge", feminist, emancipatory approaches suggest, "we can mobilise data mining in practice, not in pursuit of universally valid truth claims or the discovery of law-like disembodied generalisations, but in an ethical, self-reflexive and situated attempt to achieve multiple partial views on everyday life practices and experiences." [17]

In this narrative framework I therefore propose to see the collection of student data as a *moral* act, warranting an approach that goes *beyond* a rules-based (or deontological) approach to a teleological approach that considers the potential for harm, critically engages with the scope of consent and allocation of processes to facilitate recourses and appeal in cases of unintended harm. Such an approach also considers the potential vulnerabilities of those affected by the intervention or opportunity and the danger of pathogenic vulnerability.[18]

Key considerations in this narrative framework include a recognition that data[19] often reflect skewed allocation of values and resources based on sexual orientation, gender, culture in the context of the lasting and intergenerational impact of colonialism and apartheid. This proposal also requires a consideration of the purpose, processes, tools, staff and governance that accompany the ethical use, collection and analysis of student data.

Introduction

Ethical behaviour, in general, refers to behaving according to standards of human behaviour to which humans in a particular community and context embrace and aspire to realise.

Various factors inform an approach to ethics such as cultural norms and practices (including a range of religious orientations and beliefs) as well as broader legal frameworks on the protection of privacy and human rights. This proposed framework provides a broader basis for reflection than just adherence to legal rules. We accept that ethics may at times and in a particular context require higher standards from us than what the law may require of us.

Traditional ways of defining ethical behaviour often distinguish between two approaches; this framework combines both and includes utilitarian considerations. One traditional approaoch, *deontological*, is 'rule-based' that forms the basis for contracts and Terms and Conditions. Another, a *teleological* approach, embraces ethics as a discursive space where the potential for harm and the scope of consent, and recourses in cases of unintended harm are negotiated and agreed upon. The framework presented here aims to achieve the most good while at the same time to minimise harm, respect for the rights and dignity of data subjects, a commitment to treat all human beings equitably.[20]

Learning analytics is "a structuring device, not neutral, informed by current beliefs about what counts as knowledge and learning, coloured by assumptions about gender/race/sexual orientation/class/capital/literacy and in service of or perpetuating existing or new power relations"[21] As structuring device, learning analytics allow us multiple opportunities to redress the historical and persisting inequalities and injustices. Any consideration of ethics in the collection, analysis and use of student data cannot ignore the legacy of colonialism and apartheid and the continued intergenerational legacy of these systems of oppression and dehumanisation.

In the light of the fact that most of the current literature (both scholarly and popular) on learning analytics "…originates from the Global North" and does not, necessarily, speak to "a number of specificities of the Global South"[22] we need to critically engage with scholarship and models of learning produced in the Global North and formulate context-appropriate responses. [23]

This narrative framework therefore aims to provide a broad basis and principles for the consideration of the ethical implications of the collection, analysis and use of student data *specific in the context of South Africa*.

A Critical Sense of Location

Central to this proposal is the question: "What does a contextualised, South African perspective on the ethical collection, analysis and use of student data entail?" The basis for this question is an acknowledgement of the role played by a particular definition of data as well as the intergenerational legacy of the use of data in the colonial period and later in Apartheid. We would be disingenuous if we did not acknowledge that data collection, analysis and use are, *per se*, political acts, informed by (often contesting) ideologies and serve declared and hidden assumptions about the purpose of higher education in the South African context.

South African higher education shares a number of characteristics and trends with international higher education such as the massification of higher education, increasing funding constraints and competition, persistent concerns about student attrition and failure, as well as the challenges and opportunities offered by technological advances.[24] Intersecting and amplifying these trends are context-specific political, economic, social, technological, environmental and legal factors.

The collection, analysis and use of student data play an increasingly important role in South African higher education's ability to make sense of the impact of these trends and how to respond appropriately, effectively, but also, ethically.

Framing context-appropriate, effective and ethical responses therefore have to consider that student data are not, necessarily, indicators of students' potential, merit or even necessarily engagement but often, the result of the inter-generational impact of the skewed allocation of value and resources based on race, gender, sexual orientation, class, language and culture.

This framework therefore proposes that the collection, analysis and use of student data:

- Acknowledges the lasting, inter-generational effects of colonialism and Apartheid
- Collects, analyses and uses student data with the aim of addressing these effects and tensions between ensuring quality, sustainability and success
- Critically engages with the assumptions surrounding data, identity, proxies, consequences and accountability
- Responds to institutional character, context and vision
- Considers the ethical implications of the purpose, the processes, the tools, the staff involved, the governance and the results of the collection, analysis and use of student data.

Realities

Each of the partner institutions has a particular context, culture, value system, vision and mission. It is important to consider and align the collection, analysis and use of student data to institutional infrastructure, capacities and maturity in educational data mining. *Some* general realities that exist are:

Realities regarding (students) data

- Student demographic and behavioural data collected by different departments and at different times within the university are not integrated

- There is often a duplication of research foci and requests for analysis with the same questions being asked by different departments/units
- There are variances in student data accuracy, integrity and comprehensiveness
- Student data still comprises a variety of forms from analogue to digital with many forms of historical data under threat of being in formats that are increasingly obsolete
- Some institutions have more data than others
- Despite having access to data collected, institutions do not use the data
- With the complexity, velocity, variety and amount of data increasing, institutions may not have the capacity or skill sets/expertise to collect and analyse student data
- Student data are scattered across institutions in a variety of formats, governed by distinctive and often exclusionary regimes
- A lot of student data actually serve as proxies for understanding their behaviour, risk, need for support and potential, e.g. the socioeconomic category of their home addresses
- The amount, content and quality of student data depends on the digital maturity of the institution, the digitisation of learning, pedagogical and assessment strategies

Realities regarding sources of data

- There are silos of data in various units/departments across the institution, and silos of analysis
- Some data are stored on systems by outside vendors
- Different data systems do not “talk to each other”
- With the increase in use and complexity in algorithmic decision-making in institutions, there is increasing concerns regarding the ‘black box” of automated, algocratic systems
- Data gathered through Research (with a capital ‘R’) is not fed back into the system and is lost to future sense-making and analysis

Realities regarding quality of data

- Students often provide inaccurate data, intentionally or unintentionally, thereby compromising data quality
- The quality of data is also affected by the quality of questions

- Quality of data depends on who provides the data, the epistemologies, expertise and identity of the person who samples, captures, and analyses the data

Realities regarding processes of data collection and analysis

- Different sections/departments/stakeholders across institutions collect, analyse and use data in a variety of ways, using a variety of software and tools. There is a need for a standardised and consistent approaches across the institution
- There are too many points of data collection, which results in duplication and negatively impacts return rates
- Different people come up with different conclusions from analysis of same data, and these are too seldom compared and reconciled

Realities regarding tools

- Data collection tools contain items that are often vague and/or culture specific resulting in unreliable/inaccurate data that is inconsistent from subject to subject (e.g. item on members in a household)
- Data collection tools, despite best intentions, are often too long – not minimal
- There are often too many surveys or data collections performed, particularly of students

Realities regarding the use of analyses by staff and students

- There is, in general, too little evidence of the impact of many efforts in which student data were collected and analysed
- Communication of the findings to stakeholder groups in *formats* that would be understandable is poor
- There is a general lack of sharing on data collected and analysed across the institution
- There are varying levels of expertise and infrastructure to support ethical and appropriate data collection and analyses. Many of the stakeholders at the coalface of teaching and learning lack the necessary expertise to engage with the interpretation and analyses of data
- Students are often not informed about the usefulness of providing correct data at various points in their learning journeys. Sharing research findings and the outcomes of initiatives with students could help them to understand the importance of the

data they provide, and better understand how the university functions. This could also improve the quality of the data

Realities regarding data governance (access & storage)

- Institutions, in general, do not have integrated systems, policies and processes to manage data well resulting in unauthorised access and use
- There is not sufficient tracking of who accesses data, changes made to the formats and scope of data, and how the data are used
- Policies on data handling and access are often not clear or not in place
- There is often a lack of oversight and training of staff specifically responsible for the governance of data
- Not enough consideration given to including a clause in all new employment contracts regarding the ethical handling of student or staff data. For existing staff good governance of personal data could be covered by getting them to sign confidentiality agreements

Rationale

Collecting, analysing and using student data have always formed an integral part of South African higher education, whether on the institutional or national level. This has served a variety of purposes such as funding allocations, projections of growth and quality assurance frameworks (e.g., student retention and success rates as indicators of quality).

Student data have become an invaluable resource in light of increasing funding constraints, the impact of the massification of higher education on enrolments and resource allocation, the systematic digitisation of higher education, and persistent concerns about student attrition and failure. -

Increasing volumes, variety and velocity of student data combined with increasing capacity (e.g., technological as well as human resources) enable higher education to collect not only more data but also to collect finer-grained data than ever before. This affects not only the immense potential of using these data, but also significantly deepens our responsibility in line with higher education's fiduciary duty.

The current collection, analysis and use of student data *for research purposes* are subject to stringent institutional ethical review approval processes and procedures ensuring not only ethical conduct but also accountability, transparency and recourse for redress in case of malpractice or adverse effects. Learning analytics as the collection, analysis and use of student data *to inform teaching and learning* currently falls outside the jurisdiction of ethical review policies, processes and procedures.[25]

Problem Statement

Learning analytics as an *emerging,* specific form of research currently falls outside the scope of institutional ethical review policies, processes and procedures. Therefore this necessitates a framework containing broad principles for consideration. Defining learning analytics in the light of where it falls with regard to traditional approaches to research on students opens up spaces to consider the ethical implications in learning analytics. Learning analytics can be seen as (1) the scholarship of teaching and learning; (2) as dynamic, synchronous and asynchronous processes of sense-making; (3) learning analytics as a potential automated process and, lastly, (4) as a participatory process and collaborative sense-making.[26] Each of these possibilities may entail unique but also overlapping ethical considerations.

Different stakeholders in higher education institutions may have disparate, contesting (if not contradictory) views on the factors and variables affecting student retention and success. The scope, institutionalisation and success of learning analytics depends on a shared, critical and informed understanding of student success as the result of a range of intersecting, often interdependent and mutually constitutive variables and generative mechanisms in the nexus between students, the institution, disciplines and schools, and broader societal factors.[27] [28]

Though learning analytics focuses on the collection, analysis and use of *student* data, this data need to be considered in the context of and in relation to other data such as student: staff ratios, staff engagement, institutional operational efficiencies and macro-societal trends and impacts (political, economic, societal, technological, environmental and legal).

The collection, analysis and use of student data in learning analytics may also have different foci, such as informing curriculum development, adjusting pedagogical strategies, providing feedback to faculty and students, allowing for the personalisation of student support (whether pastoral, technical or cognitive) and/or predicting student success.[29]

There is a range of institutional stakeholders involved in the collection, analysis and use of student data support to increase student retention and success. Most institutions do not have a composite view of exactly what data the institution has access to, but also who accesses the data for what purposes.

Higher education institutions also have access to an increasingly wide range of student data, scattered across the institution in disparate databases, in different formats, and possibly governed by different policies, processes and procedures. Though there is talk about a “data revolution”, there is an awareness of the impact of insufficient data, the impact of limited technical capacity, different norms and standards, an insufficient coordination on institutional, national and international levels.[30]

Much, if not all of current learning analytics reports rely almost exclusively on *quantitative* analysis involving a range of statistical methodologies and the use of a range of increasingly sophisticated software. While there is a general assumption of objectivity and neutrality in the use of statistical models and software, we cannot ignore the ethical implications of, for example, assuming that small differences are meaningful, equating statistical significance with real-world significance, mistaking correlation for causation, and the allure of thinking big(ger)/more data are better data.[31] Finding ways to include *qualitative* research data in learning analytics may result in a more holistic, if not also deeper understanding of students’ learning journeys.[32] [33]

We can also not ignore the fact that due to the (increasing) complexities of data collection and analysis, that institutional capacity and individual researcher competencies will increasingly play an important role in the scope of ethical considerations in learning analytics. Equally, we should also not underestimate the fact that many of the end-users may lack (and in all probability will lack) the statistical and/or data literacy to critically evaluate the analysis of data and/or misinterpret the analysis with possible unintended detrimental effects on students.

In light of increased sharing of analysis with students with the help of student-facing dashboards, we have to be critically aware of students' information literacy and consider the ethical implications of our nudges.

Definitions

Analysis: in the context of learning analytics refers to descriptive analytics (what happened?); diagnostic analytics (why did it happen?); predictive analytics (what will happen?); and prescriptive analytics (how can we make it happen?)[34]

Biometrics: A technique of personal identification that is based on physical, physiological or behavioural characterisation including blood typing, fingerprinting, DNA analysis, retinal scanning and voice recognition[35]

Cohort: In the context of learning analytics, this term encompasses any group of students who registered for the same module in a particular registration period; or registered for the same qualification in a particular registration period. Because learning analytics primarily occurs in the context of a specific module, the use of the term 'cohort' differs from the use of the word 'cohort' in the context of reporting on student graduation rates.

Data: Data in the context of learning analytics typically include three basic broad categories, namely data '*volunteered*' by students, data collected as part of *automated* collection processes, as well as data collected originating from a *human* need for more information, and collected, analysed and used in different in different combinations of human and algorithmic decision-making. These data can include, among other things, demographic data and behavioural data. Data protection and research policies normally govern access to data.

Data subject: Means the person to whom the personal information relates.

Electronic information: Means any text, voice, sound or image message sent over an electronic communications network which is stored in the network or in the recipient's terminal equipment until it is collected by the recipient.[36]

Fiduciary duty: A duty that arises from a legal and/or ethical relationship of trust with another entity, person or group of persons.

Learning analytics: The measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs[37] Learning analytics differs from, but possibly overlaps with *academic analytics*, the latter dealing with student aggregated data reporting on cohort analysis and retention on institutional and program level.

Obligation: Arises from a contractual or legal basis or moral grounds

Personal information: Refers to information relating to an identifiable, living, natural person, and where it is applicable, an identifiable, existing juristic person, including, but not limited to—

> *(a)* information relating to the race, gender, sex, pregnancy, marital status, national, ethnic or social origin, colour, sexual orientation, age, physical or mental health, well-being, disability, religion, conscience, belief, culture, language and birth of the person;
>
> *(b)* information relating to the education or the medical, financial, criminal or employment history of the person;
>
> *(c)* any identifying number, symbol, e-mail address, physical address, telephone number, location information, online identifier or other particular assignment

to the person;

(d) the biometric information of the person;

(e) the personal opinions, views or preferences of the person;

(f) correspondence sent by the person that is implicitly or explicitly of a private or confidential nature or further correspondence that would reveal the contents of the original correspondence;

(g) the views or opinions of another individual about the person; and

(h) the name of the person if it appears with other personal information relating to the person if the disclosure of the name itself would reveal information about the person.[38]

Public record: Refers to a record that is accessible in the public domain and which is in the possession of or under the control of a public body, whether or not it was created by that public body[39]

Record: Any recorded information—

(a) regardless of form or medium, including any of the following:

(i) Writing on any material;

(ii) information produced, recorded or stored by means of any tape-recorder, computer equipment, whether hardware or software or both, or other device, and any material subsequently derived from information so produced, recorded or stored;

(iii) label, marking or other writing that identifies or describes anything of which it forms part, or to which it is attached by any means;

(iv) book, map, plan, graph or drawing;

(v) photograph, film, negative, tape or other device in which one or more visual images are embodied so as to be capable, with or without the aid of some other equipment, of being reproduced;

(b) in the possession or under the control of a responsible party;

(c) whether or not it was created by a responsible party; and

(d) regardless of when it came into existence[40]

Resources: The purpose of learning analytics to inform departmental, faculty, student support and student agency to improve learning in the context of a module, resources encompasses infrastructure, financial, systems and processes, and capacity.

Special personal information: Refers in general parlance to sensitive information and is defined in terms of Section 26 of the Protection of Personal Information Act (2013)[41]. It includes the collection, analysis and use of personal information regard the religious or philosophical beliefs, race or ethnic origin, trade union membership, political persuasion, health or sex life or biometric information of the data subject; or criminal behaviour of a data subject to the extent that such information relates to (1) the alleged commission by a data subject of any offence; or (2) any proceedings in respect of any offence allegedly committed by a data subject or the disposal of such proceedings.

Student: Refers to an individual registered to study a module or qualification. This does not include enquirers or informal learners. In the context of learning analytics, the term refers to *currently* registered students.

Scope

General

Though this narrative framework cannot prescribe to South African higher education institutions how to institutionalise an ethical and context-appropriate learning analytics praxis, we propose the following broad scope statement for this ethical framework:

> The ethical and context-appropriate collection, analysis and use of student data in learning analytics include
>
> (1) acknowledging our assumptions about data;
>
> (2) being clear about the reasons for the collection of data and what the benefits are to whom;
>
> (3) ensuring the appropriate methods and tools used in the collection and analysis of data;

(4) developing the expertise and competencies as well as institutional capacity to analyse the data;

(5) considering how the analyses and findings are disseminated and used;

(6) confirming institutional responsiveness to act on the findings; and

(7) providing oversight and accountability for the whole process.

The scope includes using the aggregated findings on a *macro* level – by departments, schools and the institution as a whole to inform policy development, revision and implementation, curriculum development, operational strategies on departmental, school and institutional levels to provide effective and appropriate, general and discipline-specific student support.

On a *micro* level the information should be used by students, faculty (tenured and/or contract teaching staff), course administrators and advisors, and student support staff to offer increasingly personalised forms of support and feedback (whether human or algorithm), and peer support

In scope

Categories of data that fall in scope include three basic broad categories, namely data *volunteered* by students, data collected as part of *automated* collection processes, as well as data collected originating from a *human* need for more information, and collected, analysed and used in different combinations of human and algorithmic decision-making.[42]

Sources of data include

- Records (audio, email) of *pre-registration* personal inquiries
- Application-to-register data
- Data provided or required as part of the registration process
- Provided records (digitised) of students' *prior* learning (informal or forma) whether at other institutions or the same institution
- Sensitive information that the institution has obtained the right to use
- Interactive content generated by enquirers or students; for example: completing diagnostic tests, student responses to marketing surveys and research

- System-generated data such as the date and frequency of accessing the institutional learning management system (LMS), library records or requests, etc.
- *Anonymised* data from discussion forums or where explicit consent is received from students to use these data to personalise support
- Third party data held by the institution or data generated internally in combination with data provided by third parties where there is agreement to do so from the third party and students concerned
- *Anonymised* data from external sites, e.g., social networking sites not owned by the University on *cohort* level. In the context of the collection, analysis and use of such *non-anonymised* data in the context of identifying students at risk or in need of personalised support *explicit permission must be obtained from students*
- Any other information explicitly provided by students for the purposes of rendering more effective and personalised support

Communication of the above to students is non-negotiable.

Though current definitions of learning analytics deal explicitly with student data it is crucial to understand student data as part of a bigger picture that includes factors and data such as course design, student: instructor ratios, lecturer participation and responsiveness data, and data relating to institutional (in)efficiencies. Focusing *exclusively* on student data as basis for descriptive, diagnostic, predictive and prescriptive analytics result in partial understanding, bias and increases the potential for skewed analysis or findings that do not consider the immediate and broader contexts.

Out of scope

The following pointers refer to data and uses of data that may fall outside the scope of this Framework depending on the institutional context, data collection and analysis capacities, infrastructure, software, skill sets and ethical and oversight processes:

- Student complaints
- Data collected and/or shared by the individuals concerned or shared with or without the knowledge of the individuals concerned that were located on sites external to the institution such as social networking sites

- Data from third-party sites where there is no permission to employ the shared information
- Historical data of non-registered students and alumni
- The notion of 'sensitive data' fall under the provision of the Protection of Personal Information Act (2013)[43] as 'special information' (Section 26) and include data such as religious or philosophical beliefs, race or ethnic origin, trade union membership, political persuasion, health or sex life or biometric information of data subjects; or historical criminal records. Data classified as 'special information' may be used under provisions of Section 26 which include, for example, research or where the individual concerned has made the data public. Should such data be required for the purpose of *learning analytics*, consent will have to be obtained by a suitable means, such as through explicit Ethical Review.
- Any combinations of data or derived data that may contravene an individual's right to respect for their private and family life should be considered in terms of Section 27.[44]
- Institutions may have a number of policies and guiding frameworks that relate to the collection, analysis and use of student data for purposes other than learning analytics as defined in this narrative framework.

Assumptions

A critical reflection on underlying assumptions about student data, sources, and quality of data, processes, tools, people and governance (access and storage) is fundamentally at the heart of an ethical approach to student data. Some assumptions include:

a. Students trust us implicitly with their data
b. All staff have access to the student data
c. All staff can get data; and all staff understand the student data
d. Informed consent is understood
e. Data are aligned to the mission and vision of the university
f. Data are accurate and correct
g. Data are interpreted the same

h. Students do not mind if their data are shared or accessed
i. The interpretation of the Protection of Personal Information Act (POPI) will vary if even known by both staff and students
j. The standards required in Protection of Personal Information Act (POPI) are implemented
k. Processes are consultative and/or at the very least participatory
l. We are gathering data for student success
m. The tools we use work
n. The tools are relevant
o. The tools and/or systems are aligned with the aims of the data being collected
p. The tools and/or systems talk to each other
q. Tools are biased as they are designed as a certain ideology, therefore subjective

Draft Policy Statement

The draft policy statement provides a high-level view of broad principles on how institutions involved in the Siyaphumelela project will collect, analyse and use student data. The principles may have implications for policy development and implementation in each of the institutions. It falls outside of this narrative framework to presume that all the institutions will implement the principles in the same way. Despite the differences in implementation, each of these principles should significantly shape the collection, analysis and use of student data. These principles are not regulatory in nature but are intended to inform and guide the ethical collection, analysis and use of student data.

Overview of the principles

The following set of principles are intentionally phrased in very broad terms as to encompass the different nuances and specificities in the different types of data (e.g., structured and unstructured, qualitative and quantitative), databases, data infrastructures and holdings, assemblages; the different forms of learning analytics (e.g., descriptive, diagnostic, predictive and prescriptive); and the vast array of possible uses and applications of learning analytics. We also have to consider the specific ethical implications in the reality that an increasing number of institutions make use of Artificial Intelligence, machine learning and various combinations of human-algorithmic decision-making processes.

The principles are also not isolated statements and often overlap with other principles, though each of them offers a distinct perspective and principle.

The proposed seven principles are as follows:

Principle 1: The moral relational duty of learning analytics

Principle 2: Defining student success in the nexus of student, institution and macro-societal agencies and context

Principle 3: Understanding data as framed and framing

Principle 4: Student data sovereignty

Principle 5: Accountability

Principle 6: Transparency

Principle 7: Co-responsibility

Principle 1: The moral relational duty of learning analytics

In the South African context in the face of the intergenerational legacy of colonialism and Apartheid, it is unthinkable to focus on learning analytics other than as a moral practice. Not withholding the need for more effective teaching and learning practices, learning analytics should not only focus on what is effective, but also aim to provide relevant pointers to decide what is *appropriate* and *morally* necessary.

Learning analytics as moral relational practice underlines the whole project, from how we define and see the purpose of data, determining what data to collect and what data are but proxies for complex and intergenerational layers of injustice and inequality, the identity, skills sets and capacities of those who collect, analyse and use student data, and ensuring accountability, transparency and oversight. As moral project, learning analytics is *per se* relational and emancipatory [45] [46]

Principle 2: Defining student success in the nexus of student, institution and macro-societal agencies and context

Student success flows from a range of variables, many of which are non-linear, interdependent and mutually constitutive, in the nexus between student, institutional and macro-societal contexts. Research in education has to consider the implications of the fact that education is an open and recursive system, unlike research in medicine with its evidence-based approach based on research in controlled environments. This has implications for claims of correlation, causality and evaluating the effectiveness of interventions.

Student success in light of institutional reporting cycles is determined by established criteria of what constitutes student success in the eyes of accrediting, quality assurance and ranking regimes and organisations. Learning analytics , in international and national institutional contexts often define, use, and classify student success according to these prescribed and sanctioned epistemologies. In these epistemologies, students are classified according to their risks of failing, using predominantly a deficit model of what students lack and need in order to succeed. This narrative of student success is permeated with descriptions of students as lurkers, not engaged, dropouts, and stop-outs, and of not fitting in.

Despite our best intentions, there is a danger that our commitment to operational efficiencies, measurements and quality assurance regimes, shapes our vocabulary to frame students' agency, capital and potential in terms of *our* epistemologies and ways of seeing performance.

There is, however, another side to the definition of student success. How do we talk about student success when most of our indicators may hold them responsible for intergenerational injustices and inequalities? How do we frame student success not in terms of what how *they* don't fit into our definitions, epistemologies and performance criteria, but in terms of how we may have mistaken how exclusionary and possibly our epistemologies and criteria for engagement, agency and 'fitting in' are? What happens if we replace terms like 'dropouts' with 'forced removal's, and how they are pushed into 'exile'?[47]

How and where do we meet students to explore and map how our understandings of their learning journeys, aspirations, challenges and potential resulted in a voice-over where they lost control and the ability to self-define them and not having us occupying their potential? How can we collaboratively craft a shared understanding of what engagement mean for them, for us, and then determine what criteria is most useful and descriptive of where they are coming from, where they are and where they want to go? How do we stop demonising their words, their experiences, their ways of knowing and understanding the world?

Principle 3: Understanding data as framed and framing

Data are not neutral and should not be accepted at face value, or treated as neutral, objective, and pre-analytic in nature. Data are framed "technically, economically, ethically, temporally, spatially and philosophically. Data do not exist independently of the ideas, instruments, practices, contexts and knowledges used to generate, process and analyse them."[48] Data are framed and framing.[49]

We should also accept that our data never provide the complete or full picture and that the data we do have of our students are glimpses and often proxies of complex, layered lives.

Principle 4: Student data sovereignty

Student data are not something separate from students' identities, their histories, their *beings*. This framework accepts that data are an integral, albeit informational part of students' being. Data are therefore not something students own but rather are. Students do not own their data but are constituted by their data.[50]

This principle therefore enshrines the right of students to determine how to share their data, the purposes for which their data will and can be used and the conditions of use and storage. This principle also implies that students are the primary owners of bestowing meaning to whatever data institutions may have and collect from them.[51] Though this principle acknowledges the right and duty of institutions to collect, analyse and use student data, this principle establishes the need to establish what information students will need in order to make more informed decisions regarding their learning journeys as basis for all collection, analysis and use of student data. When we gather data, we do not "enter data deserts, but existing systems" of sense making and ways of seeing agency, knowledge, learning and the world.[52]

Central to data sovereignty is not only access and control of data, but also defining and agreeing on the use of data. While it is accepted that the data are also of use to the institution, this principle clearly establishes the ownership and final control of data. The collection, analysis and use of student data therefore needs to primarily reflect the interests, values and priorities of students.[53] As such students have the final say over who has access to their data, for what purposes, and assurances that their data will be stored and protected against unauthorised access and/or use.

We have to acknowledge that quantification of learning functions as a technology of power.[54] As such, students have the right to assert their sovereignty over what indicators they need to inform their choices regarding the learning process. Asserting student sovereignty over the choice of indicators and categories result in a destabilising of universal truths such as degrees of 'economic engagement', the notion of 'household', 'dependents', and 'active economic participation'. In indigenous contexts the notion of 'dependency' is, for example, "a more complex phenomenon than in the Global North – it is not a one-way relationship – and chronological age is not necessarily a good indicator of dependence."[55]

The formulation of indicators is based on an assertion of power to "produce knowledge and to define or shape the way the world is understood" – it is never neutral and/or objective and depend on "culturally specific categorisations that determine what it is 'significant' to measure."[56] It is crucial to realise that our indicators not only shape how students' learning is understood, but it also contain "embedded value judgements."[57] As such, our categorisations make certain things visible and make others invisible.

In multicultural contexts student data sovereignty means considering indigenous definitions of engagement, of respect, of voicing opinions – often in terms where objective statistics do not adequately reflect the nuances. It is possible that due to the fact that we do not know how to measure indigenous definitions of authority, respect, engagement and participation, that we dismiss these expressions as irrelevant.[58]

Student data sovereignty therefore means finding ways to acknowledge “culture-smart data” – “information that can be produced locally, captures local social units, conditions, priorities and concerns and is culturally informed and meaningful”[59]

Foundational to student data sovereignty are ‘ownership’, ‘control’ and ‘access.’[60] Of specific interest to ethical considerations in learning analytics is considering the notion of ownership. it is important to distinguish it from stewardship. “Ownership is distinct from stewardship. The stewardship or custodianship of data or information by an institution that is accountable to the group is a mechanism through which ownership may be maintained”.[61] *This implies that institutions of higher learning have stewardship of student data but never ownership.* “While ‘ownership’ identifies the relationship between a people and their data, possession reflects the state of stewardship of data” (emphasis added).[62]

Control of data means that students have a right to control not only what *personalised*[63] data is collected from them, but also the scope of the collection, how it will be used, who will have access to this data, for what purposes, and under what conditions, to the ultimate destruction of the data. Student data sovereignty also implies that students will have *access* to information and data about themselves. Access implies that students should be able to access all learning analytics performed on their data in meaningful, accessible formats, and to obtain copies of these data in a portable digital format.[64]

Student control of their data also implies that students will have access to how institutions categorise (label) and describe them and the rationale and criteria for these descriptions. Learning analytics produces “regimes of truth”[65] and categories of ‘others’, individuals who are often pathologised and problematised.[66] In the event where students would contest the label, rationale and criteria for the categorisation, such a disagreement provides an exceptionally valuable opportunity to engage with not only students’ narratives about their learning journeys and agency, but also point to deficiencies in our categorisations or the need for more nuance.

Student data sovereignty implies that institutions provide students with the choice *to opt out* of the collection, analysis and use of their demographic and/or behavioral data and information for offering them personalised support, curricula and assessment. We should not think in binary terms of either opting in or out, but opting out can be structurally nuanced with students opting out/in depending on the purpose of the collection, analysis and use of their data, the disciplinary module or context, the variety of possible data that can be collected, analysed and used, and understanding the risks of opting in/out.

Finally, students should have access to supported and transparent recourse when (1) they allege harm as a result of the collection, analysis and use of their data; (2) they did not have an opportunity to provide context or more information on the data collected and used for the alleged infringement; (3) their choices are limited without a clear explanation on the rationale and appropriateness of the limitation as well as a how the limitation will affect their learning journey; and (4) when they have not been informed of the collection, analysis and use of their data outside of the original consent provided and original purpose for the collection of their data.

Principle 5: Accountability

An etymology of the word 'accountability' points not only to the need to be *answerable* and *responsible*, but also to being *response-able* and the having the *obligation to act*. Accountability also entails an obligation to adhere to national legislation such as the protection of human rights and the Protection of Personal Information Act (2013). Higher education institutions *account for* their actions, their policies, the effectiveness and quality of the learning they offer to a range of stakeholders such as quality assurance and accrediting bodies, funding regimes.

In the context of learning analytics, higher education institutions are also accountable to *students*, not only for the quality and for the content of the curricula and validation of learning, but also for the appropriate and effective provision of learning. Learning analytics allows institutions to grow in their understanding of the complexities of learning, but also insights how to optimise "learning and the environments in which it occurs" (see the definition of learning analytics). Accountability therefore means accepting the legal and ethical duty of being response-able and to allocate sufficient resources and capacity to act in appropriate and consultative ways to the collection, analysis and use of student data.[67]

Finally, accountability in learning analytics also implies providing oversight. Oversight in learning analytics is unresolved. If learning analytics is classified as Research (with a capital 'R') then there are clear processes and policies to ensure ethical research conduct and the dissemination of findings. Evidence[68] suggests that learning analytics is not currently defined as Research and therefore falls outside the scope of oversight by Ethical Review Boards and research policies. If learning analytics does not 'qualify' or is treated as Research, it would seem as if there are four distinct possibilities namely

- Learning analytics as the scholarship of teaching and learning;
- Learning analytics as dynamic, synchronous, and asynchronous sense-making;
- Learning analytics as an automated process;
- Learning analytics as a participatory process and collaborative sense making.[69]

In none of these four forms is there, at present, any clarity on the oversight of learning analytics. The following pointers therefore need consideration:

If we consider learning analytics as the scholarship of teaching and learning, it would make sense that *individual faculty, departmental administration and module support teams* take responsibility for reporting on the collection, analysis and use of student data. Of specific concern is the level of technical or statistical expertise that many of these individuals and/or departments may lack in engaging in learning analytics in appropriate and ethically responsible ways. Of specific concern is that in the context of the increasing pressure on academics to publish research findings that individuals may present findings at academic conferences and submit for publication without having ever received ethical clearance. Most institutions will also refuse to grant retrospective ethical clearance for such research. Complicating matters is the fact that at the time of the collection, analysis and use of student data, faculty, administrative and/or support staff may not foresee the possibilities of presenting at conferences or of publishing the findings.

This alludes to the need to consider who will provide oversight to learning analytics as scholarship of teaching and learning, as dynamic, synchronous, and asynchronous sense-making and/or as a participatory process and collaborative sense-making.

Of specific interest in the ethical dimensions and need for oversight in learning analytics as automated process.[70] [71] Considering as a basis the four dimensions of sensing, processing, acting and learning, there are a number of possibilities that arise when we consider the use of algorithmic decision-making in conjunction with human responsibility and oversight (see Figure 1).

Figure 1: Human-algocratic decision-making grid[72]

	(1) Humans perform the task	**(2) Task is shared with algorithm**	**(3) Algorithms perform task with human oversight/ supervision**	**(4) Algorithms perform the task independently/ autonomously – no human input**
Sensing	Yes or no?	Yes or no?	Yes or no?	Yes or no?
Processing	Yes or no?	Yes or no?	Yes or no?	Yes or no?
Acting	Yes or no?	Yes or no?	Yes or no?	Yes or no?
Learning	Yes or no?	Yes or no?	Yes or no?	Yes or no?

With regard to oversight, as stated above, there is no clarity regarding the collection, analysis and use of student data when it falls, as it currently does, outside the scope of 'Research'. Therefore, institutions will have to establish guidelines and processes governing human action in the four processes of sensing, processing, acting and learning (the first vertical column in Figure 1). On the other extreme end is the need to consider the ethical implications, processes and oversight when algorithms sense, process, act and learn without any human oversight (the extreme right-hand column in Figure 1). When we consider the ethical considerations when humans and algorithmic decision-making systems share the process of sensing, processing, acting and learning, then the ethical considerations get more complex. For example, what happens when we use an algorithmic agent/code to sense (e.g., which students have not logged onto the institutional learning management system in a particular week), process this data with student demographic and other behavioural data and then send the report to a human to act and learn (the third vertical column in Figure 1)?

Figure 1 illustrates 256 logically possible procedures[73] with each procedure having possibly unique ethical dimensions or sharing ethical dimensions with other procedures.

In 2014, the White House[74] issued a position paper on the opportunities but also the challenges in Big Data and warned of "the potential of encoding discrimination in automated decisions." In 2016 a new position paper titled "Big Data: A report on Algorithmic Systems, Opportunity, and Civil Rights"[75] specifically addresses the challenges in using algorithmic systems in, inter alia, education. "The opportunities to use big data in higher education can either produce or prevent discrimination—the same technology that can help identify and serve students who are more likely to be in need of extra help can also be used to deny admissions or other opportunities based on the very same characteristics"[76]

> As learning itself is a process of trial and error, it is particularly important to use data in a manner that allows the benefits of those innovations, but still allows a safe space for students to explore, make mistakes, and learn without concern that there will be long-term consequences for errors that are part of the learning process.[77] [78]

Principle 6: Transparency

Institutional transparency regarding the criteria and assumptions about enrolment, allocation of resources to those students identified as being at-risk, and the success of interventions aimed to increase student success and retention is often intimately linked to institutional reputation, aspirations and business models. As such, the criteria and assumptions may not be transparent to ensure competitive advantage. There is also the reality that "Institutions cannot be transparent about what may not even be transparent to them"[79]

Despite and amid the above context, transparency, as a principle in the ethical collection, analysis and use of student data underpins the other principles in this framework aiming to institutionalise ethical data collection, analysis and use. For example, transparency flows from learning analytics as moral practice (Principle 1) and the commitment to understanding student success as flowing from a range of intersecting variables in the complex nexus of student, institution and macro-societal factors (Principle 2). A commitment to transparency also involves making our assumptions about data known (Principle 3), acknowledge how the data we have provides us with but partial glimpses of student learning and their life-worlds and that our data are incomplete and tentative. "Students should not be wholly defined by their visible data or our interpretation of that data."[80]

Transparency also underlies the acceptance of students to have and claim sovereignty of their data and to bestow us not ownership, but temporary possession and stewardship (Principle 4). A commitment to transparency (Principle 6) provides the basis and assurances for the willingness to be held accountable (Principle 5).

- A commitment to transparency in the collection, analysis and use of student data is underpinned by the following:
    - Acknowledging student sovereignty of their data as basis for consultation with students and opening spaces for crafting consensus regarding the collection, analysis and use of students' data
    - Negotiating meanings, access and security issues
    - Considering how feedback on learning analytics will be given and by whom
    - Ensuring open, effective communication and feedback through inclusive engagement
    - Committing to fiduciary compliance
    - Considering the implications of third party vendors who refuse to release details of their algorithms and metrics
    - Communicating the options and consequences of opting out
    - Committing to transparency about compliance, breaches and consequences

- Sharing rules and arrangements regarding how the governance of and access to student data
- Committing to increase staff competencies and understanding of the complexities and ethical implications in the collection, analysis and use of student data
- Keeping records of all requests for and analysis of data and ensuring auditing of the processes, analysis and use of student data
- Students should have access to a mirror of what academics see

Principle 7: Co-responsibility

The concluding and capstone principle in this framework emphasises the interdependency between institutions and students in facilitating effective and appropriate learning experiences.

Higher education institutions need to collect, analyse and use student data in order to account to various stakeholders for their ability to provide effective, quality and appropriate learning experiences. In light of the asymmetrical power-relationship between institutions and students, the fiduciary duty and responsibility of institutions to ensure the ethical and appropriate collection, analysis and use of student data is non-disputable.

On the other hand, in order to allow institutions to make ethical and informed decisions regarding optimising resources to ensure effective and appropriate learning experiences, students have co-responsibility not only for their learning, but to ensure that the data they provide to institutions are correct and updated.

Learning analytics as student-centred practice does not only involve that the aspirations and well-being of students are of foremost concern, but they are invited to participate in learning analytics as collaborative, democratic sense making.

Towards implementation

There is general agreement that policy and legal frameworks often lag behind technological developments and as such always seem to become obsolete very soon after their publication. Critical policy studies also point to the fact that there is a "difference between policy rhetoric and practiced reality" and that policies often enforce and perpetuate dominant understandings of phenomena that are in line with the comfort zones of institutional hierarchies of decision-making.[81] Policies and their implementation are then used to distribute power, resources and knowledge and create categories of 'winners' and 'losers' and result in a social stratification, whether in institutions, or as a result of the implementation of the policy to those who are on the receiving end of the implementation.[82]

It is therefore crucial to ask how this framework will reinforce and/or reproduce social injustices and inequality or whether this policy may contribute to the breaking of intergenerational cycles of inequality and injustice.

This draft narrative framework was presented at a workshop on 27 June 2017 as part of the third Siyaphumelela Conference that took place from 27-29 June at the Wanderers Club, Johannesburg. The workshop allowed delegates from individuals and institutions involved and/or interested in the Siyaphumelela project to provide feedback on this narrative framework.

Based on the feedback and the scope of consensus on the principles contained in this narrative framework, a final framework was presented at the conference on 28 June 2017.

(In)conclusion

We acknowledge that students and the institution carry responsibility to ensure the success of students; we cannot discount the impact of the asymmetrical power-relationship between students and the institution. Institutions have access to increasing amounts;

variety and velocity of student data, and as such, have a unique and immense responsibility to ensure the ethical and appropriate collection, analysis and use of student data.

It will also be disingenuous to fail to acknowledge the immense pressures on the higher education sector in South Africa to respond to increasing demands for responsiveness to student demands, institutional transformation and increasing funding, cost and capacity constraints. Higher education also has the unique responsibility to contribute to efforts to address the persisting intergenerational effects of colonialism and Apartheid. Our definitions of student data, the words and categories we use to define students, our understanding of the data we collect, our critical interrogation of our own assumptions and epistemologies regarding student success will irrevocably shape the potential of learning analytics to address generations of inequality and injustice.

Finally, the future sustainability of higher education institutions is intimately linked to collecting, analysing and using student data in appropriate and ethical ways.

References

[1] Glesne, C. (2016). Research as solidarity. In T. Kukutai and J. Taylor. (Eds), *Indigenous data sovereignty. Toward an agenda* (pp. 169-178). Canberra, Australia: Australian National University Press. Retrieved from https://press.anu.edu.au/publications/series/centre-aboriginal-economic-policy-research-caepr/indigenous-data-sovereignty
[2] First Nations Information Governance Centre (FNIGC). (2016). Pathways to First Nations' data and information sovereignty. In T. Kukutai and J. Taylor. (Eds), *Indigenous data sovereignty. Toward an agenda* (pp. 139-156). Canberra, Australia: Australian National University Press. Retrieved from https://press.anu.edu.au/publications/series/centre-aboriginal-economic-policy-research-caepr/indigenous-data-sovereignty
[3] Smith, D.E. (2016). Governing data and data for governance: the everyday practice of Indigenous sovereignty. In T. Kukutai and J. Taylor. (Eds), *Indigenous data sovereignty. Toward an agenda* (pp. 117-135). Canberra, Australia: Australian National University Press. Retrieved from https://press.anu.edu.au/publications/series/centre-aboriginal-economic-policy-research-caepr/indigenous-data-sovereignty
[4] Ekowo. M., & Palmer, I. (2016). The promise and peril of predictive analytics in higher education. New America. Retrieved from https://www.newamerica.org/education-policy/policy-papers/promise-and-peril-predictive-analytics-higher-education/
[5] Open University. (2014). Policy on ethical use of student data for learning analytics. Retrieved from http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf
[6] Sclater, N. (2014). Code of practice for learning analytics: A literature review of the ethical and legal issues. Jisc, November 5. Retrieved from http://sclater.com/papers/JR0022_Learning_Analytics_A-_Literature_Review_v1.3.pdf
[7] Prinsloo, P. (2016). Mapping the ethical implications of using student data – A South African contextualised view. Presentation on 27 October 2016 at an Ethics Symposium as part of the Siyaphumelela Project, Kopanong Hotel & Conference Centre, Johannesburg, South Africa. Retrieved from https://www.slideshare.net/prinsp/mapping-the-ethical-implications-of-using-student-data-a-south-african-contextualised-view
[8] Government of South Africa. (2013).Protection of personal information Act. Government Gazette, 26 November 2013. Retrieved from http://www.justice.gov.za/legislation/acts/2013-004.pdf
[9] Velasquez, M., Andre, C., Shanks, T.S.J., & Meyer, M.J. (2015, August 1). Thinking ethically. Retrieved from https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/thinking-ethically/
[10] Prinsloo, P., & Slade, S. (2014). Educational triage in open distance learning: Walking a moral tightrope*. The International Review of Research in Open and Distributed Learning*, 15(4), 306-331.
[11] Prinsloo, P., & Slade, S. (2017). Big Data, higher education and learning analytics: Beyond justice, towards an ethics of care. In Ben Kei Daniel, (Ed.), *Big Data and Learning Analytics in Higher Education* (pp. 109-124). Springer International Publishing.
[12] Swartz, S. (2011). 'Going deep' and 'giving back': strategies for exceeding ethical expectations when researching amongst vulnerable youth. *Qualitative Research,* 11(1), 47-68.
[13] Sunseri, L. (2007). Indigenous voice matters: Claiming our space through decolonising research. *Junctures: The Journal for Thematic Dialogue*, (9), 93-106.
[14] South African San Institute (2017). San code of research ethics. Retrieved from http://trust-project.eu/wp-content/uploads/2017/03/San-Code-of-RESEARCH-Ethics-Booklet-final.pdf
[15] Swartz, 2011, p. 48
[16] Ibid, p. 49
[17] Leurs, K. (2017). feminist data studies: using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review*, 115(1), 130-154.
[18] Cole, A. (2016). All of us are vulnerable, but some are more vulnerable than others: the political ambiguity of vulnerability studies, an ambivalent critique. *Critical Horizons,* 17(2), 260-277.
[19] Throughout this document I refer to data as phenomenon as singular except where quoting another source where data are referred to as plural.
[20] Velasquez, M., Andre, C., Shanks, T.S.J., & Meyer, M.J. (2015, August 1). Thinking ethically. Retrieved from https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/thinking-ethically/

[21] Prinsloo, P., & Slade, S, (2017, March 17). Building the learning analytics curriculum: Should we teach (a code of) ethics? Retrieved from https://www.slideshare.net/prinsp/building-the-learning-analytics-curriculum-should-we-teach-a-code-of-ethics

[22] Gašević', D. (2017). Include us all! Directions for the adoption of learning analytics in the Global South. Foundation for Information Technology Education and Development. In press.

[23] Prinsloo, P. (2017). Include us all! Directions for the adoption of learning analytics in the Global South: An African perspective. Foundation for Information Technology Education and Development. In press.

[2424] Altbach, P.G., Reisberg, L., & Rumbley, L., E. (2009). Trends in global higher education: Tracking an academic revolution. Retrieved from http://unesdoc.unesco.org/images/0018/001831/183168e.pdf

[25] Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. *Educational Technology Research and Development*, *64*(5), 881-901.

[26] Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. *Educational Technology Research and Development*, *64*(5), 881-901.

[27] Prinsloo, P. (2009). Modelling throughput at Unisa: the key to successful implementation of ODL. University of South Africa. Retrieved from http://uir.unisa.ac.za/handle/10500/6035

[28] Subotzky, G., & 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.

[29] Gašević', D. (2017). Include us all! Directions for the adoption of learning analytics in the Global South. Foundation for Information Technology Education and Development. In press.

[30] UNESCO. 2017. The data revolution in education. Information Paper n. 39, March. Retrieved from http://uis.unesco.org/sites/default/files/documents/the-data-revolution-in-education-2017-en.pdf

[31] Prinsloo, P., Archer, L., Barnes, G., Chetty, Y., & Van Zyl, D. (2015). Big(ger) data as better data in open distance learning: Some provocations and theses. *International Review of Research in Open and Distributed Learning* (IRRODL), 16(1), 284-306.

[32] Knight, D. B., Brozina, C., & Novoselich, B. (2016). An Investigation of First-Year Engineering Student and Instructor Perspectives of Learning Analytics Approaches. *Journal of Learning Analytics,* 3(3), 215-238.

[33] Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In: *LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, ACM International Conference Proceeding Series, ACM, New York, USA, pp. 56–65. Retrieved from http://oro.open.ac.uk/48992/

[34] Boyer, A., & Bonnin, G. (2016). Higher education and the revolution of learning analytics. ICDE. Retrieved from https://icde.memberclicks.net/assets/RESOURCES/anne_la_report%20cc%20licence.pdf

[35] Government of South Africa. (2013).Protection of personal information Act. Government Gazette, 26 November 2013. Retrieved from http://www.justice.gov.za/legislation/acts/2013-004.pdf

[36] Ibid

[37] Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE* (September/October). Retrieved from http://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

[38] Government of South Africa. (2013).Protection of personal information Act. Government Gazette, 26 November 2013. Retrieved from http://www.justice.gov.za/legislation/acts/2013-004.pdf

[39] Ibid

[40] Ibid

[41] Ibid

[42] Danaher, J. (2014, January 6). Rule by algorithm? Big Data and the threat of algocracy. [Web log post]. Retrieved from http://philosophicaldisquisitions.blogspot.com/2014/01/rule-by-algorithm-big-data-and-threat.html

Kitchen, R. (2013). Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography*, 3, 262-267. SOI: 10.1177/2043820613513388

Knox, D. (2010). Spies in the house of learning: a typology of surveillance in online learning environments. Paper presented at *Edge*, Memorial University of Newfoundland, Canada, 12-15 October.

[43] Government of South Africa. (2013).Protection of personal information Act. Government Gazette, 26 November 2013. Retrieved from http://www.justice.gov.za/legislation/acts/2013-004.pdf

[44] Ibid

[45] South African San Institute (2017). San code of research ethics. Retrieved from http://trust-project.eu/wp-content/uploads/2017/03/San-Code-of-RESEARCH-Ethics-Booklet-final.pdf

[46] Swartz, S. (2011). 'Going deep' and 'giving back': strategies for exceeding ethical expectations when researching amongst vulnerable youth. *Qualitative Research,* 11(1), 47-68.
[47] Tuck, E., & Fine, M. (2007). Inner angles. A range of ethical responses to/with indigenous and decolonising theories, in Denzin, N.K., & Giarnina, M.D., (Eds), *Ethical futures in qualitative research. Decolonising the politics of knowledge,* (pp. 146-168). Walnut Creek, CA: Left Coast Press Inc.
[48] Kitchen, R. (2014). *The data revolution. Big data, open data, data infrastructures and their consequences.* London, UK: Sage, page 2.
[49] Gitelman, L., & Jackson, V. (2013). *'Raw data' is an oxymoron.* Cambridge: MIT Press.
[50] Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology,* 7(4), 185-200.
[51] Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous Data Sovereignty: Toward an agenda* (Vol. 38). ANU Press.
[52] Ibid
[53] Ibid
[54] Morphy, F. (2016). Indigenising demographic categories: a prolegomenon to indigenous data sovereignty. In T. Kukutai and J. Taylor. (Eds). *Indigenous data sovereignty. Toward an agenda* (pp. 99-115). Canberra, Australia: Australian National University Press. Retrieved from https://press.anu.edu.au/publications/series/centre-aboriginal-economic-policy-research-caepr/indigenous-data-sovereignty
[55] Ibid, p. 104
[56] Ibid, p. 104
[57] Ibid, p. 108
[58] Smith, D.E. (2016). Governing data and data for governance: the everyday practice of Indigenous sovereignty. In T. Kukutai and J. Taylor. (Eds). *Indigenous data sovereignty. Toward an agenda* (pp. 117-135). Canberra, Australia: Australian National University Press. Retrieved from https://press.anu.edu.au/publications/series/centre-aboriginal-economic-policy-research-caepr/indigenous-data-sovereignty
[59] Ibid
[60] First Nations Information Governance Centre (FNIGC). (2016). Pathways to First Nations' data and information sovereignty. In T. Kukutai and J. Taylor. (Eds), *Indigenous data sovereignty. Toward an agenda* (pp. 139-156). Canberra, Australia: Australian National University Press. Retrieved from https://press.anu.edu.au/publications/series/centre-aboriginal-economic-policy-research-caepr/indigenous-data-sovereignty
[61] Ibid., p. 149
[62] Ibid., p. 150
[63] Institutions may use student data in *aggregated* formats as part of their contractual, fiduciary duty.
[64] Sclater, N. (2014). Code of practice for learning analytics: A literature review of the ethical and legal issues. Jisc, November 5. Retrieved from http://sclater.com/papers/JR0022_Learning_Analytics_A-_Literature_Review_v1.3.pdf
[65] Foucault, M. (1980). Power/Knowledge: Selected interviews and other writings. New York, NY: Pantheon Books.
[66] Sunseri, L. (2007). Indigenous voice matters: Claiming our space through decolonising research. *Junctures: The Journal for Thematic Dialogue*, (9), 93-106.
[67] Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: the obligation to act. LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference pp. 46-55. Retrieved from http://dl.acm.org/citation.cfm?id=3027406
[68] Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. *Educational Technology Research and Development*, *64*(5), 881-901.
[69] Ibid, p. 893
[70] Prinsloo, P. (2015, December 14). Algorithmic decision-making in higher education: There be dragons there… [Web log post]. Retrieved from https://opendistanceteachingandlearning.wordpress.com/2015/12/14/algorithmic-decision-making-in-higher-education-there-be-dragons-there/
[71] Prinsloo, P. (2017). Fleeing from Frankenstein's monster and meeting Kafka on the way: Algorithmic decision-making in higher education. Special issue of *E-Learning and Digital Media* - Learning in the age of algorithmic cultures. In press.

[72] Danaher, J. (2015 June 15). How might algorithms rule our lives? Mapping the logical space of algocracy. [Web log post]. Retrieved from http://philosophicaldisquisitions.blogspot.co.za/2015/06/how-might-algorithms-rule-our-lives.html
[73] Ibid
[74] White House. (2014). Big data: seizing opportunities, preserving values. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf
[75] White House. (2014). Big Data: A report on algorithmic systems, opportunity, and civil rights. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf
[76] Ibid., p. 16
[77] Ibid., p. 19
[78] Also see Ekowo, M., & Palmer, I. (2017). *Predictive analytics in higher education. Five guiding practices for ethical use*. New America. Retrieved from https://www.newamerica.org/education-policy/policy-papers/predictive-analytics-higher-education/
[79] Ekowo. M., & Palmer, I. (2016). The promise and peril of predictive analytics in higher education. New America. Retrieved from https://www.newamerica.org/education-policy/policy-papers/promise-and-peril-predictive-analytics-higher-education/
[80] Open University. (2014). Policy on ethical use of student data for learning analytics. Retrieved from http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf
[81] Diem, S., Young, M. D., Welton, A. D., Mansfield, K. C., & Lee, P. L. (2014). The intellectual landscape of critical policy analysis. *International Journal of Qualitative Studies in Education*, 27(9), 1068-1090.
[82] Ibid.