INTERNATIONAL COUNCIL FOR OPEN AND DISTANCE EDUCATION

Global guidelines: Ethics in Learning Analytics

MARCH 2019

Authors: Sharon Slade and Alan Tait

ISBN: 978-82-93172-43-7

Add a little bit of body text

Executive Summary

Learning Analytics systems and practices have developed over 10 years or more in a wide range of educational contexts, collecting and analysing large amounts of data to analyse and predict student behaviour and success. The ethical issues that are inherent in this range of practices have however been slower to be understood. This Report proposes a number of Core Issues that are important on a global basis for the use and development of Learning Analytics in ethics-informed ways . They include

- Transparency
- Data ownership and control
- Accessibility of data
- Validity and reliability of data
- Institutional responsibility and obligation to act
- Communications
- Cultural values
- Inclusion
- Consent
- Student agency and responsibility

The Report does not intend nor is it able to answer all ethical questions about the use of Learning Analytics in all countries, and accepts there are legitimate issues of differentiation deriving from national legislation and broader cultural issues. The Report does however aim to identify a range of Core Issues which it proposes are of global relevance.

Introduction

1.1
The use of clues derived from student behaviour to adapt a teacher's practice, and in particular the organisation of learning and teaching, is nothing new. However the growth over the last 10 years or so in both campus-based and online educational programmes using so-called big data in real time – the practice known as Learning Analytics - offers both new opportunities and challenges. The ethical implications of such practices in terms of ownership of data, privacy, and the use and abuse of data for understanding the individual student and her or his progress have been slower to be understood than the development of sophisticated Learning Analytics systems. This report aims to provide a platform for consideration of the ethical issues of what is a valuable area of development. We aim to do so in ways that take nothing away from the value of Learning Analytics in supporting student success, but also to support recognition of the risks by commission or omission in their improper use.The Report does not claim to answer all the questions that arise from the ethical issues raised here. To some considerable extent these will depend, it is acknowledged, on a wide range of local factors.

Definition and purpose of the use of Learning Analytics

2.1
Within the broader context of education, educational technology, and student support, learning analytics is a relatively recent new discipline. The New Media Consortium's Horizon Report (NMC, 2011) flagged learning analytics as an emerging technology. Perhaps the earliest established definition of learning analytics was established by Siemens (2011) who stated that learning analytics are "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs".

2.2
In a higher education context, the broadly accepted purpose of learning analytics is to improve the chances of student success (Gasevic, Dawson & George Siemens, 2015), as well as to inform pedagogy, allocate resources and inform institutional strategy (Rienties, Boroowa, Cross, Kubiak, Mayles, & Murphy, 2016).

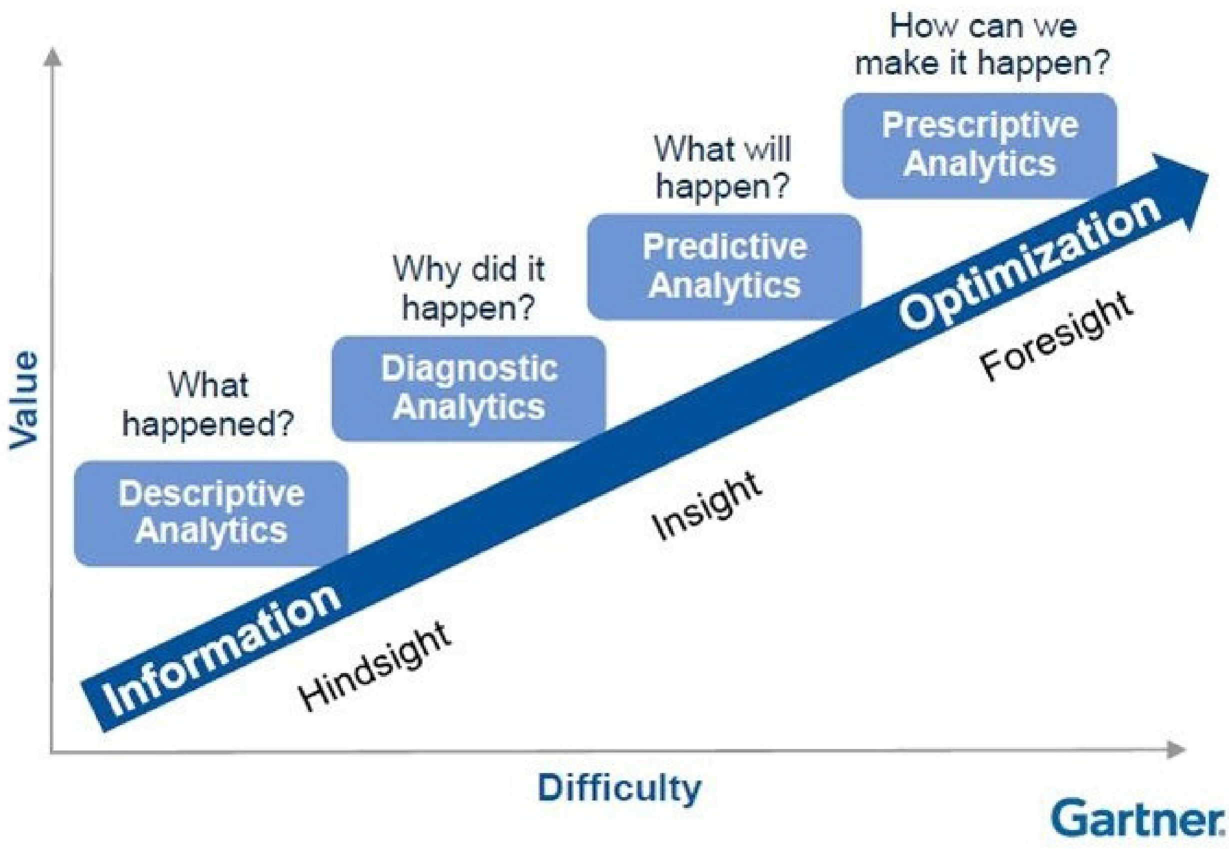
2.3

The NMC Horizon Report: 2015 Higher Education Edition highlights the increase in the measurement of student learning through data-driven practice and assessment. The report suggests that institutions increasingly seek to gather and analyse ‘large amounts of detail about individual student interactions in online learning activities. The goal is to build better pedagogies, empower students to take an active part in their learning, target at-risk student populations, and assess factors affecting completion and student success.’ (p.12). Some might argue that the data typically collected includes not only that recorded as part of online learning activity, but may also include broader data sets, such as that collected as part of the enquiry or registration process, assessment data and data shared by students as part of their daily social and study lives (see Sharples et al,. 2014).

The UK-based Higher Education Academy (HEA, 2015) states that learning analytics offers the potential to provide educators with quantitative intelligence to make informed decisions about student learning.

Data is drawn from a broad range of sources including behavioural data from online learning systems (discussion forums, activity completion, assessments) and functional data extracted from registration systems and progress reports (Sharples et al,. 2014).

As well as basic tracking, a range of interpretative approaches can be applied. These include predictive models (which suggest potential completion rates, for example), social network analyses (which examine possible relationships between networks of individuals and groups), relationship mining (which analyses links between sets of data patterns such as student success rates), and dashboards (data visualisation which provides a means of delivering feedback to educators and learners). The range of analytic practices along the axes of Value and difficulty has been set out helpfully in graphic form (Davis 2013).



2.4 Justification of the need for global guidelines

As the field of learning analytics has developed, consideration of related ethical issues has developed in tandem. It is now broadly accepted that consideration ought to be given to the purpose of, and issues related to, the collection, analysis, and use of student data. Any such consideration should recognise and reflect, where feasible, the potentially conflicting interests of different stakeholder groups, primarily the students themselves and their educational institution.

2.5

Over the past 5 years or so, a number of guidelines, codes of practice and policies have been developed in response to this. Slade and Prinsloo (2013) established one of the earliest frameworks with a focus on ethics in learning analytics. Others have followed, including the Open University's policy on the ethical use of student data for learning analytics in 2014, JISC's code of practice in 2015 and the Learning Analytics Community Exchange (LACE) framework in 2016 (Drachsler & Greller, 2016).

2.6
Although well received, it is fair to say that many of these guidelines have been shaped by their specific geo-political context. Given that several of the underlying issues will relate to legislative policy around, for example, data protection, it is easy to understand why this has been the case. A global code of conduct would allow for a broader set of approaches from which senior practitioners and managers within higher education might take a view.

Background for the establishment of the Working Group from ICDE

3.1
In the light of the rapid development of Learning Analytics on a global basis, ICDE took the initiative to produce a set of guidelines for ethically-informed practice that would be valuable to all regions of the world. The aim of the Guidelines is to identify which principles relating to ethics are core to all, and where there is legitimate differentiation due to separate legal or more broadly cultural environments. A Working Group was established with members from a range of countries with the following tasks:

- To suggest global guidelines for ethics in Learning Analytics based on a brief review of existingregional/national/institutional guidelines
- To establish a broad set of principles and recommended best practice
- To inspire and support facilitation of the use of LA broadly in education worldwide within a framework for sound ethics

3.2
The primary intended audiences for these Guidelines are educational policy makers and regulators in government departments at national and regional levels, International Governmental Organisations, and practitioners as well as middle and senior management in higher education institutions.

3.3
Membership of the Working Group was:

- Zibuyile Aftaz, University of South Africa, South Africa
- Tian Belawati, Universitas Terbuka, Indonesia
- Torunn Gjelsvik, ICDE, Norway (until August 2018)
- Serine El Salhat, Hamdan Bin Mohammed Smart University, United Arab Emirates,
- Sharon Slade, Open University, UK
- Christoph Stueckelberger, Globethics.net Foundation, Switzerland
- Alan Tait, (Chair), ICDE, and Open University UK

3.4 The Working group met through 2018, with the final report written by Sharon Slade and Alan Tait

Core Issues

4.1

The Report proposes a number of Core Issues, that is to say, issues that we suggest need consideration in all world regions where Learning Analytics are in use or in development.

Data ownership and control

4.2

The issue of data ownership will be impacted to some extent by relevant national and international legislation. For example, the European Union implemented the General Data Protection Regulation in 2018 which aims to harmonise privacy laws across Europe. It sets out principles establishing ways in which data may be accessed, stored and used, and defines as special categories personal and sensitive data.

4.3

Institutions should be aware of and make transparent issues around third party sharing (which might typically include sharing student data which service providers for marketing purposes, for example). In such case, the third party would typically be bound by the data protection rules which apply to the institution, for example, that data may not be passed on nor used for unintended additional purposes.

4.**4**

In a learning analytics context, the presumption is often that data collected is owned by the institution, whereas students themselves might be expected to take another view. The emphasis on 'data as property' overestimates individuals as autonomous and rational agents (e.g., Lazaro & Le Métayer 2015). Others propose an ontological understanding of data where, for example, "student data is not something separate from students' identities, their histories, their beings. ... data is an integral, albeit informational part of students' being. Data is therefore not something a student owns but rather is. Students do not own their data but are constituted by their data" (Prinsloo 2017). And although there is often clear existing legal protection relating to personal data, the lack of clarity around who owns the data muddies principles of meaningful consent.

4.5

One way to view the issue then, might be to suggest that the institution does not own the student data that it holds, but has temporary stewardship. Subject to legislation and policy, the institution may store datasets under certain conditions and for specified periods, but within a higher education context the issue of ownership could remain open. In this situation, the institution would be able to collect, analyse and apply from the learning analytics outputs, are theirs to exploit for unrelated gain.

4.6

For personal and sensitive data, it might be argued that students should have some input to determine which data can be collected, how that data can be used, who is able to access it, and for what purposes (Prinsloo, 2017).

4.7

In addition, it might be argued that institutions should grant students the ability to correct and/or add context to their raw data, and to review and make a case for choices which appear to be limited as a result of a learning analytics application.

Transparency

5.1

Institutional transparency might best begin by making clear to students and to other stakeholders the purpose of learning analytics. Although this may appear trite, it is fair to say that uses of learning analytics are often made for the benefit of the institution (maximising completion rates across the whole cohort) rather than the individual student (providing the 'best' outcome, whether that relates to improving course scores or allowing free study choices).

5.2

In practical terms, transparency relates primarily to how student data is collected, analysed and used to shape students' potential learning journeys. Such transparency also includes making clear what data is collected (and what is not) and any assumptions made about that data (where it may be incomplete or acting as a proxy for another measure, for instance). Making students and stakeholders more aware of the uses of data may pose additional challenges, but brings with it an opportunity to engage with stakeholders to gain greater insight and involvement. Allowing students access to information regarding their study choices, and how and why these may be impacted by analytics, also facilitates both an opportunity to amend or correct the dataset (or the interpretations gained from it) and to add to institutional understanding of relevant factors which impact on student success. Having said that, it is not always possible to be completely transparent – models built around regression approaches for example can be difficult to understand and interrogate – it is not always clear why one student may be identified as being potentially more vulnerable than another. Nor is it always in the best interests of students to communicate a predicted poor outcome (although this might be balanced against the moral duty of the institution to act in a student's best interests and perhaps advising alternative study paths).

Accessibility of data

6.1

Accessibility of data can relate to both the determination of who has access to raw and analysed data, and to the ability of students to access and correct their own data (as above). Further, it might also be interpreted as making clear which data might typically be included within a learning analytics application, and which might always be assumed to be out of scope. Some of the latter data issues may be established by local data protection legislation, but the institution may also want to give consideration to whether any data categories are irrelevant, or sufficiently sensitive to warrant exclusion.

6.2

In the case of who has access, various categories of staff will have sight of some categories of raw student data (e.g., demographic data, academic history, income, etc.) as a normal aspect of the staff role, depending on their permissions. Where data categories are not required as part of that role, however, data would typically not be made available, and this is likely to be determined by institutional and/or national policy. Typically, within a learning analytics context, we might expect that data is accessed on a 'need-to-know' basis to facilitate the provision of academic and other support services.

Validity and reliability of data

7.1

In order to ensure that outputs from learning analytics applications are both valid and reliable, the institution needs to ensure that data collected and analysed is both accurate and representative of the issue being measured. Datasets should be kept current as far as is possible, with opportunities for students and other stakeholders to refresh and replace existing data.

7.2

Proxy measures should be used with caution. For example, study engagement is commonly measured by tracking online logins. A starting point is to first consider what the institution is trying to measure and to investigate how it might best be represented (rather than looking at available data first and figuring how it might be applied).

7.3

If analytics is predictive and involves statistical calculation, it is also important to ensure that data sets are complete and sufficient to enable robust calculations to be made. Further, the models used to analyse, interpret and communicate learning analytics to stakeholders (support staff, advisers, faculties, students) should be sound, free from algorithmic bias; transparent where possible and clearly understood by the end users. This includes the need to ensure/improve staff competencies and understanding of the complexities and ethical implications in the collection, analysis and use of student data.

Institutional responsibility and obligation to act

8.1

A key principle is to consider whether access to knowing and understanding more about how our students learn brings with it a moral obligation to act. For example, having observed students not submitting summative assignments or having calculated the probabilities of module completion, is the institution obliged to act on what it has identified? Often resources are constrained, and in distance learning institutions in particular where one-to-one conversations are less easy, it is not easy to reach all students who may have been identified as likely to benefit from a support intervention of some type. In these cases, the institution might consider its policy for identifying where support resource is focused, for example, on the group identified as most potentially vulnerable; toward students on high population core modules; at students with particular characteristics (for example, those with known disabilities, etc.). The decision-making process should be transparent and clearly understood by all stakeholders.

Communications

9.1

Care should be taken when communicating directly with students on the basis of their analytics. Contact based on basic tracking of students is potentially less contentious e.g. 'we notice that you have not yet registered for your next module/exam'), but that triggered by predictive analytics needs a great deal more consideration. Predictive analytics are often based around black box regression models and provide predictions often based on 'students like you' and on their previous outcomes. It is key to remember that predictions are only that, a probability generated by a computer. Students are more than the sum of their visible data and whilst predictive analytics are useful in proactively alerting tutors or support staff to issues before they may arise, it is also important to seek additional context. In such cases, communications with students are perhaps most effective if couched in general support terms (e.g. 'we're just checking in with you to see how your studied are going') rather than in probabilistic terms (e.g. 'we think that you have a 10% probability of completing your current module').

9.2

Regular communications to staff should help ensure that they understand the approach: the underlying values linked to the institution's mission and strategy; the anticipated benefits for students; the limitations of data and its interpretation; and guidelines for ethical practice.

Cultural values

10.1

In multicultural contexts, understanding and interpreting data are necessarily more complex. A measure of participation or engagement may differ in different contexts. It is clear also that measures established as being correlated with successful or unsuccessful outcomes are likely to differ in different geographies and cultures. Whilst not all institutions will have the capacity or resource to develop in-house analytics tools developed on knowledge of its own students, care should be taken if purchasing analytics packages from developers to ensure that the approach is fit for purpose and can be adapted if appropriate with local data and with local constraints in mind.

Inclusion

11.1

Where Learning Analytics are conceived predominantly or even solely in the institution's interest, there is a danger that certain categories of students will be identified negatively as particularly at risk. While there are situations where discussion about a student's progress or lack of progress may lead to the outcome of leaving the institution being in the student's best interest, there are ethical issues relating to inclusion and exclusion if such interactions relate predominantly to the institution's desire to protect its success rates. With greater pressure on fee levels and government interest in quality of outcomes in many countries, there is a risk that Learning Analytics can be used in in ways that legitimise exclusion. Learning Analytics should be primarily used to support students, in student centred centred ways.

Consent

12.1

Many existing approaches to consent are flawed. Consent to collect student data for uses beyond those required for institutional reporting and basic student support is sought at the point of registration. At this point, many students will be largely unaware of learning analytics and how it may be used to support them. Consent at this point is certainly most convenient for the institution, but arguable less meaningful for the student. If consent is to be sought at this stage, it should be coupled with transparency (of purpose, of data collected, etc.) and potentially with a later option to withdraw consent.

12.2

In considering the EU's GDPR, Sclater (2017) suggests the following three-tiered response: 1. that consent is not required for the use of non-sensitive data for analytics (on the basis that this may be considered as of legitimate interest); 2. That consent is required for use of sensitive data (which, under the GDPR, will be labelled 'special category data'); 3. That consent would be required to take interventions directly with students on the basis of the analytics. This implies that if the data in question are not considered 'sensitive', and do not form the basis for any intervention, consent is not required.

12.3

There are opposing views to this position. Prinsloo and Slade (2018) suggest that data is not a neutral construct, but is shaped by the ideas and contexts used to generate it. As a result, they question that consent is not required for non-sensitive data. They also point out that what may constitute non-sensitive data in one context or at a particular time, may be considered sensitive in another context and/or time.

12..4

An alternative approach might be to differentiate between initial consent for the collection of data and specific consent when data are used to intervene in the choices students have or/and in adapting their learning experience or access to resources is preferred. Although there are practical difficulties in doing so, an expectation that users should consent to uses of personal data unknown at the point of registration seems to be an unreasonable and unethical one.

12.5

Other approaches might consider more nuanced versions of consent, for example, individuals might be granted additional opportunities to provide or withdraw consent where an intervention might significantly alter their experience. Ideally, consent should not be considered in simple binary terms, but presented to students as a menu of options which depend 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 an understanding of the risks of opting in/out.

12.6

National legislation will influence positions taken here, but generally this principle should be built around a minimum of informed consent (that is, transparency before registration).

Student agency and responsibility

13.1

Where feasible, it is recommended that institutions seek to engage students in applications of learning analytics. Although it is clear that there is an asymmetrical power-relationship between institutions and students, proactive engagement at least seeks to treat students as equal participants in the uses of their data. In this way, students can be more actively involved in helping the institution to design and shape interventions that will support them. By engaging students with the development and implementation of learning analytics, we are better able to

- ensure that students understand their responsibility for keeping personal information up to date (and can give informed/meaningful consent)
- achieve a more accurate interpretation of data relating to student behaviours
- improve understanding of what forms of intervention and support are most appropriate

- understand how to tailor a student’s learning journey to meet their needs, potentially as a personalised learning path
- Produce outcomes that students will find useful and be able to respond positively to, which might include a decision to continue or discontinue with their studies.

Conclusion

14.1

The Core Issues proposed in this Report represent 9 groups of issues that we propose must be discussed wherever around the world Learning Analytics are in development or already in use. The outcomes of such discussion will properly be adapted to local and regional contexts, and may lead to development of national guidelines where none exist.

14.2

A range of policies and reports from different parts of the world are listed below in the Appendix. The Working Group would be pleased to receive further references to national legislation and other significant reports, which can be added to the Appendix. These can be sent in the first instance to Professor Alan Tait at alan.tait@open.ac.uk

References

Davis M (2013) Top Ten Moments from Gartners Supply Chain Executive Conference, Gartner Blog Network, https://blogs.gartner.com/matthew-davis/top-10-moments-from-gartners-supply-chain-executive-conference/

Drachsler, H., & Greller, W. (2016, April). Privacy and analytics: It’s a DELICATE issue — a checklist for trusted learning analytics. Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK ′16), 25–29 April 2016, Edinburgh, UK (pp. 89–98). New York: ACM.

Gašević, D., Dawson, S. and Siemens. G. (2015) Let’s not forget: Learning analytics are about learning. TechTrends 59, 1 (Jan. 2015), 64-71.

Jisc (2015) Code of practice for learning analytics https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

NMC (New Media Consortium). (2011). The NMC Horizon Report. http://www.educause.edu/Resources/2011HorizonReport/223122

NMC Horizon Report: 2015. Higher Education Edition. Retrieved from http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf

Open University. (2014). Policy on ethical use of student data for learning analytics. 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

Prinsloo, P. and Slade, S. (2018). Student consent in learning analytics: the devil in the details? In J. Lester, C. Klein, H. Rangwala, and A. Johri (Eds), Learning analytics in higher education: Current innovations, future potential, and practical applications. New York and Abingdon, Oxon: Routledge, pp. 118–139.

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K. and Murphy, S. (2016) Analytics4Action Evaluation Framework: A Review of Evidence-Based Learning Analytics Interventions at the Open University UK. Journal of Interactive Media in Education 2016, 1.

Sharples, M., Adams, A., Ferguson, R., Gaved, M., McAndrew, P., Rienties, B., Weller, M. & Whitelock, D. (2014). Innovating Pedagogy 2014: Open University Innovation Report 3. Retrieved from: http://www.openuniversity.edu/sites/www.openuniversity.edu/files/The_Open_University_Innovating_Pedagogy_2014_0.pdf

Siemens, G., (2011) Learning and Academic Analytics, weblog post 5 August 2011, http://www.learninganalytics.net/?p=131

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(1), 1509–1528.

Sclater, N. (2017). Consent and the GDPR: what approaches are universities taking? Effective Learning Analytics. (2017).

Appendices

Legal Compliance

Legal frameworks such as

1. the EU General Data Protection Regulation (GDPR),
2. OECD's Fair Information Practice Principles
3. The South African Protection of Personal Information Act (POPI Act).
4. From Indonesia (Tian): Ministry of Communication and Informatics Decree No. 20/ 2016 about Electronic Personal Data Protection). This Ministerial Decree is to be elevated to become a Law, which is being drafted right now and targeted to be launched this year. If needed, I can have it translated later on.
5. Also from Indonesia (Tian): ASEAN TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY MINISTERS MEETING (TELMIN) FRAMEWORK ON PERSONAL DATA PROTECTION
6. This article from a global union organisation, published through OECD, might be relevant as a part of the background. Ref https://www.oecd-forum.org/users/75928-dr-christina-j-colclough/posts/32785-not-just-a-facebook-problem-ethical-data-collection-must-be-employed-at-work

7 From United Arab Emirates (Serine): a law which is very similar to the Electronic Personal Data Protection, called Law No. (26) of 2015 “Regulating Data Dissemination and Exchange in the Emirate of Dubai” , which you can find it through this link: http://ogp.dubai.gov.ae/documants/pdf/ltiwndiymjezmdk.pdf

- What are defined as “sensitive data” in various regions/cultures

Staff skills set in using learning analytics tools

- The need for support of staff (institutional examples)'

Resource contraints

- Affordability/costs
- Internal capacity for maintenance and support
- Availability of relevant data
- Complex infrastructure, use of third party system providers
- Lack of policies to address issues of privacy and ethics
- Shortage of leadership and strategic approach to the use of LA

Impact

- The report aims to be point of departure and adaptable to institutional, national and/or regional guidelines for ethics in LA
- The potential of LA for predictive analytics and pedagogical and strategic interventions at macro and micro levels
- Future prospects for the use of LA in the context of new technologies