Action plan for the collection of learning data in the Vital project: legal & ethical aspects

Introduction

VITAL is an Erasmus+ project on Learning Analytics for higher education (Oct. 2015-Sept. 2017) coordinated by the Center Applied Linguistics of Hasselt University. Partners are the University of Central Lancashire, the University of Amsterdam and HT2 (Oxford).
Learning analytics is a new and promising research field which can be defined as "the measurement, collection, analysis and reporting of data about learners in their context, for purposes of understanding and optimizing learning and the environment in which it occurs" (Siemens et. Al).

From a pedagogical point of view, Learning analytics can help to better understand concrete learning behaviour and uses of electronic courseware and online resources, based on which educational processes can be improved and better support can be provided to learners.
Students' learning data in 4 courses of 3 European universities (UHasselt, UCLAN, UvA) will be collected (self-study activities in online learning environments).
For their analysis by the UHasselt research team, the data will be anonymised and collected in a Learning Record Store and statistical and process mining techniques will be used. The project will adopt a bottom-up learning analytics approach by taking the perspective of the learning process, focusing on the courses used and the students' learning trails through the courses.

The goal is to map existing learning patterns and to feed back this new knowledge to the most important educational actors themselves, being the students and their lecturers. The results will be presented to the students and their instructors by means of learning dashboards. These will consist of a selection of graphs and progress overviews allowing students to understand how they learn online but also to compare their profile to user patterns of their peers. Educators will get overviews of how their students are progressing, which students might be at risk of dropping out or of failing for the course and which parts of the courses cause difficulties/require more feedback.

Legal & ethical aspects

Learning analytics presents opportunities for positive engagements and impacts on learning. However, the project consortium is aware of the risk of misunderstandings, misuse of data and adverse impacts on students linked with learning analytics in general.
The project is committed to complying with necessary ethical and privacy requirements linked with the data collection. Therefore, one of the project's outputs is this action plan concerning the legal & ethical aspects of the collection of learning data. This document is based on JISC[1]'s ethical Code of Practice for Learning Analytics (https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics) and other Learning Analytics policies.
At every stage, VITAL foregrounds the importance of ownership, consent, transparency, privacy, validity, access, action, minimizing adverse impact, stewardship of data and security. This action plan describes all of these legal, ethical and logistical principles aiming at complete transparency and clear institutional policies regarding the purposes of learning analytics, the data collected, the processes involved, and how they will be used to enhance the educational experience.
This document will be submitted to each university's privacy committee for their information and to ensure a clear and transparent communication.

[1] UK higher education, further education and skills sectors' not-for-profit organisation for digital services and solutions

The project will also adopt an open and transparent communication to the students, reassuring them that any information we may analyse either has already been provided directly by them or has been collected in the course of their studies, and any conclusions drawn from the analysis will be for their benefit. Indeed, Learner analytics can assist students individually or using aggregated and anonymised data can help other students or improve the educational experience more generally. It is distinct from assessment, and will be used for formative rather than summative purposes. The students will be informed that they and their instructors will be provided with information about their learning. The data collection and interpretation of the results will have no consequences on their continuation of the course or on their exam result.

Ownership and control

Within the VITAL project, each university acts as data controller of the collected data of its students. The overall responsibility for the data collection at UHasselt for example is with the UHasselt project team. They are responsible for the appropriate and effective use of learning analytics. They also decide which data are collected and used for analytics.
Each university project partner is responsible for collecting and anonymising their own collected data before sharing the data with the project partners. Each partner thereby has to conform to their own national legislation and the European legislation.
Students remain in control of their data and will be offered the chance to learn from their personal learning data and from the benchmarks possibly offered to them. Students are given the possibility of accessing the collected data at all time.

Consent

In the context of a policy for learning analytics, informed consent refers to the process whereby the student is made aware of the purposes to which some or all of their data may be used for learning analytics and provides consent. Informed consent applies at the point of the first login in the online learning environment. At that point, students are informed

- that data regarding their online learning activities will be collected,
- which type of data are collected (accesses to and clicks into the virtual learning environment)
- that these data will be analysed in combination with university data already collected by the university from enrollment such as age, study record),
- what the aims of the data collection are,
- that the online learning data will be transferred to a project partner outside the university only in an anonymised form.

Students will be informed that the data will be used for purely statistical analyses where individual students cannot be identified and that the data will be used for improvement of teaching and learning. The dashboards will present individual information only to the student himself and to his instructor. Requests to students to participate in educational research will follow existing university practices.
Students have the option to opt out of the online learning data collection and/or out of the delivery of personalised interventions. However, they will be informed that once they opt out, they will not be able anymore to have their data analysed and to receive interventions/dashboards based on their online learning. Students who are well informed in advance, know that the interventions taking place based on the learning analytics have the aim of supporting their learning and are thus advantageous for them.

Transparency

Hasselt University has a clear plan to communicate with students about the approach to learning analytics by informing them at the point of the first login in the online learning environment as described above:

a) in order that students understand the approach and feel reassured that data is used responsibly
b) where we can share our interpretation of data with students via the learning dashboards. We will do so unless there are good pedagogical reasons to do otherwise

c) to support students in making informed decisions about their learning path via the learning dashboards
d) to reassure that this is about enhancing support services in order to improve qualification completion which will benefit all students of the University.

Privacy

Access to student data and analytics is restricted to those identified by the University as having a legitimate need to view them. Data will be coded and used anonymously. A coding table will be used assigning randomly generated identification numbers to students in order to keep students anonymous. The coding table will be generated and stored only locally by the UHasselt VITAL project developers. Only assigned UHasselt VITAL project collaborators will have access to the table. Re-identification of individuals by aggregating multiple data sources will thereby be made possible only for internal use in the University.
Students will be informed that data will be stored by a project partner outside the University, but that before storage, data are made completely anonymous. Data will furthermore not be outsourced to other parties outside the project consortium.

In scope

Categories of data captured by the university as part of its interaction with students and potentially available as individual or combined data sets for use in learning analytics:

- personal information provided by the student at registration
- the student's study record held by the university
- system-generated data such as the date and frequency of accessing the learning module

Out of scope

In its adoption of a learning analytics approach to provide student support, the university does not intend to use the following types of data. This list is subject to review.

- Data on student complaints.
- Data that identifies individuals created on external sites, e.g. social networking sites not owned by the University, third party sites where there is no permission to employ shared information, etc.
- Data relating to enquirers and informal learners rather than registered students or alumni.
- Sensitive information on religious belief and sexual life will not be used as part of the analytical models. Any combinations of data or derived data that may contravene an individual's right to respect for their private and family life will not be used.

Validity

The techniques used in learning analytics are based on standard statistical methods, but typically involve the development of complex models, the full working of which will only be apparent to those familiar with the data and with the statistical methods employed.
It is therefore vital that the Universities monitor the quality, robustness and validity of their data and analytics processes in order to develop and maintain confidence in learning analytics and ensure it is used to the benefit of students. UHasselt ensures that:

- Inaccuracies in the data are understood and minimised
- The implications of incomplete datasets are understood
- The optimum range of data sources is selected
- Spurious correlations are avoided

All algorithms and metrics used for predictive analytics or interventions will be understood, validated, reviewed and improved by appropriately qualified staff.
Data and analytics may be valid but should also be useful and appropriate; the data from the learning analytics will therefore be seen in their wider context and combined with other data and approaches as appropriate.

Access

Through the project's end products, students will be able to access the learning analytics performed on their own data and on group averages in online meaningful, accessible formats. No individual data or learning analytics will be made available for peers. University instructors will have access to students' group and individual learning data.

Action

Where data indicates that there is potential for action to be taken that might better support students in achieving their study goals, the University will act on this. This needs to be balanced against consideration of effective use of resources. If many more potentially helpful interventions are identified, priorities will be established. The first aim of the learning analytics is to make suggestions for action in the form of dashboards to improve the learning path of the students. However, students are not obliged to take action based on these suggestions, ultimately it is their choice whether to take up any additional resources or support offered as a result of learning analytics.

Minimising adverse impact

The VITAL partners recognize that analytics can never give a complete picture of an individual's learning and may sometimes ignore personal circumstances. They will take steps to ensure that trends, norms, categorization or any labeling of students do not bias staff, student of institutional perceptions and behaviours towards them.
Analytics systems and interventions will be carefully designed and regularly reviewed to ensure that:

- Students maintain appropriate levels of autonomy in decision making relating to their learning, using learning analytics where appropriate to help inform their decisions
- Opportunities for "gaming the system" or any benefit to the student from doing so are minimised

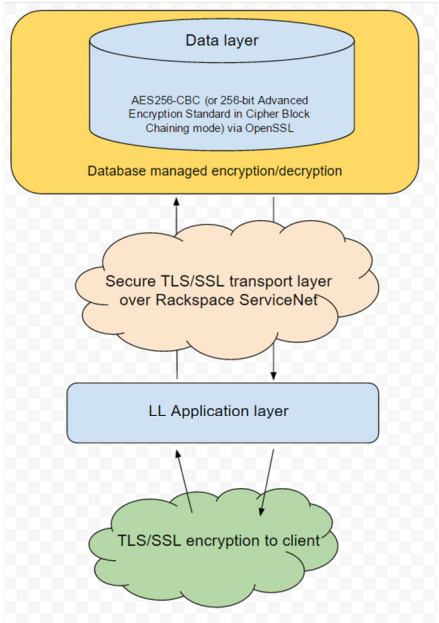
- Knowledge that their activity is being monitored does not lead to non-participation by students or other negative impacts on their academic progress or wellbeing
- Adverse impacts as a result of giving students and staff information about the students' performance or likelihood of success are minimised
- Staff have a working understanding of legal, ethical and unethical practice

Stewardship of data

Data collection for learning analytics will comply with existing institutional data policies, and will in particular be:

- Kept to the minimum necessary to deliver the purposes of the analytics reliably
- Retained only for appropriate and clearly defined periods, more specifically the average HE study term. This will be necessary to make any predictive conclusions from past data. A study term (Bachelor + Master) typically takes 5 years but in order to factor a study delay, a storage term of 8 years will be used.

On request by students any personal data used for or generated by learning analytics will be destroyed or anonymised, with the exception of certain, clearly specified data fields required for educational or statutory purposes.



Security

The data will be secured in transit, as well as at rest, as guaranteed by HT2, ISO9001 & ISO27001 accredited company in the UK, provider of an open source Learning Record Store (HT2.co.uk).

Data in transit to the Learning Record Store and within the HT2 ServiceNet environment is secured using SSL.

Security at rest is using encryption. The encryption mode is the AES256-CBC (or 256-bit Advanced Encryption Standard in Cipher Block Chaining mode) via OpenSSL. AES-256 uses a symmetric key; i.e. the same key to encrypt and decrypt text.

The encryption occurs transparently in the storage layer; i.e. all data files are fully encrypted from a file system perspective, and data only exists in an unencrypted state in memory and during transmission.

Bibliography

Commissie voor de bescherming van de persoonlijke levenssfeer (2011). Privacy: een vademecum voor de onderzoeker, Brussel.
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
Pardo, A. & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45* (3), 438-450.
Sclater, N. (2015). *Code of practice for learning analytics: A literature review of the ethical and legal issues.* Jisc. Retrieved from http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf