Drachsler, Hendrik

Trusted Learning Analytics

Synergie (2018) 6, S. 40-43

Trusted Learning Analytics

HENDRIK DRACHSLER

The use of data to inform decision-making in education and training is not new but the scope and scale of its potential impact for teaching and learning has increased by orders of magnitude over the last years. We are now at a stage where data can be automatically harvested at previously unimagined levels of granularity and variety. Once all the data would be combined in an ideal way, the analysis of these data holds the potential to provide evidence-based insights into learner abilities and patterns of behavior that in turn can provide crucial insights to guide curriculum design, to improve outcomes for all learners, change assessment from mainly summative to more formative assessments.

Data science in education has been coined as "Learning Analytics" (LA), an umbrella term for research questions from overlapping research domains such as psychology, educational science, computer and data science. Despite the great enthusiasm currently surrounding LA, there are substantial questions for research and organizational development that have brought the implementation of LA to a hold, and in some prominent cases have even reversed it due to privacy and ethics concerns (Singer 2014). The new General Data Protection Regulation 2018 (GDPR) adds another layer of the complexity to the application of LA in Europe that forces LA researchers to thoroughly think about their LA applications and especially provide fundamental new rights to the target users. This all raises a climate that demands a different approach to LA in Europe and especially in Germany. This climate is not only stimulated by the GDPR 2018, it is also grounded in serious concerns of the society about data and analytics in general. These concerns result in a lack of trust in LA that needs to be addressed in order to unfold the expected benefits of it. Within this article, we shortly describe what LA is, highlight the concerns that come with the application of LA, and finally draw a new approach towards LA that we call: Trusted Learning Analytics (TLA).

The legal ground for Learning Analytics in Europe

Very recently on May 25th 2018 the EU renewed its data protection legislation. The GDPR (General Data Protection Regulation) is more up to date to meet the challenges of a digital world (rules regarding breach notification, automated decision making and profiling, data portability, etc.). It also promotes the principle of Privacy by design as the main design principle for data driven applications. Among these design principles, the GDPR 2018 grants the data subject a set of new rights like the following not exclusive list shows:

- Right to be informed …
  how the software works and how personal data is processed.
- Right to access …
  forces the data controller to provide a copy of the personal data in an electronic format.
- Right to object …
  to processing of the data subjects' personal data, the data subject can at any time stop processing on illegitimate grounds.
- Right to erasure …
  entitles the data subject to have the data controller erase his / her personal data.

These regulations among others are hard requirements, LA system designers need to take into account in order to be compliant with the law. In our TLA research program we aim to enroll a modern TLA architecture which is not only GDPR compliant but enforces user guided privacy control.

Learning Analytics in a nutshell

LA has been defined by many different authors in the past. One frequent used approach has been published by Greller and Drachsler (2012). They clustered LA within their LA framework into the following six dimensions that we will shortly introduce:

- **Stakeholders: contributors and beneficiaries of learning analytics**
  The stakeholder dimension includes data clients as well as data subjects. Data clients are the beneficiaries of the LA process who are entitled and meant to act upon the outcome (e. g. students & teachers). Conversely, the data subjects are the suppliers of data, normally through their browsing and interaction behavior.

- **Objectives: set goals that learning analytics applications aim to support**
  The main opportunities for LA as a domain are to unveil and contextualise so far hidden information out of the educational data and prepare it for the different stakeholders. Here, we mainly talk about supporting reflection and making predictions and personalisation.

- **Data: educational datasets and the environment in which they occur**
  LA uses datasets from different educational systems. Most of the data produced in institutions is protected from external access or usage. There is, however, an increasing amount of open and linked data sources from governments and organizations like OECD that can be used to further investigate target groups for certain courses or programs (Berg et al. 2016).

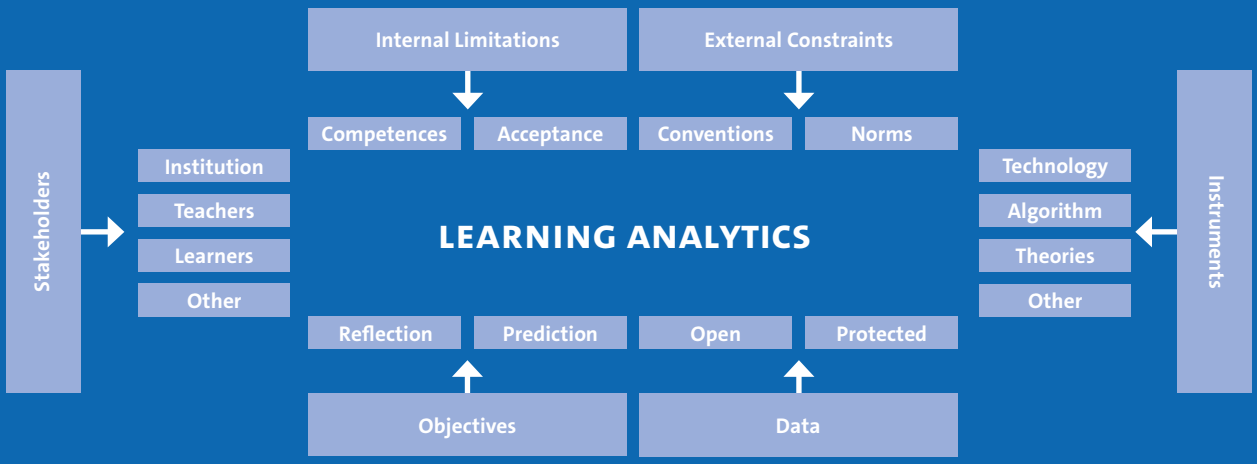

Figure 1: The Learning Analytics Framework by Greller & Drachsler 2012.

- **Instruments: technologies, algorithms, and theories that carry learning analytics**
  Different technologies can be applied in the development of educational services and applications that support the objectives of the different educational stakeholders (Drachsler et al. 2015). LA takes advantage of machine learning, social network analysis, or classical statistical analysis techniques in combination with visualization techniques (Jivet, Scheffel, Specht & Drachsler 2017).

- **External Constraints: restrictions or potential limitations for anticipated benefits**
  The LA community is highly aware of the ethical challenges for LA this resulted in first policies and guidelines regarding privacy, legal protection rights are implemented by universities that have a LA unit. Although guidelines and policies are provided for ethics and privacy, fundamental research questions and innovative technological solutions are needed to find answer to the issues mentioned in these policies.

- **Internal Limitations: user requirements to exploit the benefits**
  In order to make LA an effective tool for education, it is important to recognize that LA does not end with the presentation of algorithmically attained results. Those results need interpretation by the educational stakeholders. Therefore, the exploitation of LA requires some high-level competences, such as interpretative and critical evaluation skills (Jivet, Scheffel, Specht & Drachsler 2017).

- **Putting all six dimensions together**
  All six dimensions are equally important for a sustainable implementation of LA, as we envision the Trusted Learning Analytics (TLA) approach. Within the TLA research program, we believe there are ways to design and provide Privacy by design LA that can be beneficial to all stakeholders and allow the user to stay in control of the data themselves, all within the established trusted relationship between them and the institution.

Fears of learning analytics

That there are major concerns towards the digitized society of the future is evident from a series of very prominent dystopias that are showing the dark side of the digitization. These examples range from very prominent examples like the movie series "Black Mirror", as well as online web services like: "Will a robot take your job?[1]".

A grounded identification of the fears towards LA has been published by Drachsler and Greller in 2016, where they listed the most common fears and the propositions for privacy and ethics towards LA. We will shortly summaries some of the findings from the article in the following subsection that have been framing the need for the TLA research program that we are working on.

- **Uncertainty**
  One of the first fears is the lack of knowledge and a feeling of uncertainty of people and institutions dealing with LA. Many people are not aware of the legal boundaries and ethical limits to what they can do within the sphere of privacy protection. Institutions, on the one hand, have a fiduciary duty and need to demonstrate care for the well-being and positive development of students, leading them to success in their studies. On the other hand, there is widespread fear of negative consequences from the application of LA, such as negative press or loss of reputation and even financial fees as specified in the GDPR 2018.

Power-relationship, data / user exploitation

One of the criticisms levelled against analytics and Big Data in general is the asymmetrical power relationship it entails between the data controller and the data subject (Slade & Prinsloo 2015). This can lead to a feeling of being powerless and exploited. In fact, this concern reaches wider than LA and also applies to data and analytics in other disciplines. With the increasing usage of data of the digitised society educational institutes need to develop into "Trusted Knowledge Organisations" that can demonstrate a responsible, transparent, and secure treatment of learning data also in LA systems.

- **Data ownership**
  There has been no clear regulation for data ownership of any party, i. e. neither the student, the university or a third-party provider. This changed since the GDPR 2018 is empowered and

THE DELICATE CHECKLIST

to implement trusted Learning Analytics

D

DETERMINATION—Why you want to apply Learning Analytics?
- What ist the added value (Organisational and data subjects)?
- What are the rights of the data subjects? (e.g., GDPR, 2018)

E

EXPLAIN—Be open about your intentions and objectives
- What data will be collected for which purpose?
- How long will this data be stored?
- Who has access to the data?

L

LIGITIMATE—Why you are allowed to have the data?
- Which data sources you have already (aren't they enough)?
- Why are you allowed to collect additional data?

I

INVOLVE—Involve all stakeholders and the data subjects
- Be open about privacy concerns (of data subjects)
- Provide access to the personal data collected (about the data subjects)
- Training and qualification of staff

C

CONSENT—Make a contract with the data subjects
- Ask for a consent from the data subjects before the data collection
- Define clear and understandable consent questions (Yes / No options)
- Offer the possibility to opt-out of the data collection without consequences

A

ANONYMISE—Make the individual not retrievable
- Anonymise the data as far as possible
- Aggregate data to generate abstract metadata models (Those do not fall under EU Directive 95/46/EC)

T

TECHNICAL—Procedures to guarantee privacy
- Monitor rugularly who has access to the data If the analytics change, update the privacy regulations (new consent needed)
- Make sure the data storage filfills international security standards

E

EXTERNAL—If you work with external providers
- Make sure they also fulfill the national and organisational rules
- Sign a contract that clearly states responsibilities for data security
- Data should only be used for the intended services and no other purposes

Figure 2: Drachsler, H. & Greller, W. (2016). Privacy and Analytics—it's a DELICATE issue. A Checklist to establish trusted Learning Analytics and Knowledge Conference 2016, April 25–29, 2016, Edinburgh, UK.

specifies rights towards data collections of users. Due to the GDPR, the data subject has the right to know about all the information that has been collected about them, as well as get access to them and demand deleting this data as well. This raises requirements for the software development that are not properly addressed so far and are in the core of the TLA program we are working on.

– **Transparency and trust**
It is often said that lack of transparency can cause unease and concern with data subjects. But just providing access to data subjects raw data is not very helpful but at least a first step. The focus of analytics should be put on providing information for human decision making, prediction and self-reflection rather than accountability.

Related Work

In the next section, we will briefly introduce some related projects and available tools that contribute to the vision of TLA in the higher education field.

– **DELICATE checklist**
Another important resource that was derived from the intensive study of the ethical and legal texts, as well as from a thorough literature review is the DELICATE checklist by Drachsler and Greller (2016). The DELICATE checklist aims to provide a practical tool that can be used by LA developers and implementers to quickly check the privacy risks that are associated with the introduction of data processing in an educational institution. It can be a helpful instrument for any educational institution to demystify the ethics and privacy discussions around LA. It can be downloaded from the EU LACE project website. On the basis of the DELICATE checklist the new SHEILA framework to support policy making in LA has recently been announced (Tsai et al. 2018).

– **Trusted Learning Analytics Infrastructure**
The heart of the technology side of the TLA program is the Trusted Learning Analytics Infrastructure that can not only be connected to open source systems such as learning management systems or interactive tools for learning and teaching such as audience response systems. The TLA infrastructure is developed as an open source project. The infrastructure is the first big data system that takes the GDPR into account and provides practical functions to empower the data subject in the LA process. By giving the data subjects more control over their data, we want to increase their agency, critical reflection and engagement of the data subjects with LA.

– **Learning Analytics Indicator Repository**
The Learning Analytics Indicator Repository (LAIR) lists and visualizes the findings of the literature review in an interactive web-application . Visualizations of LA metrics and indicators are valuable for TLA dashboards, where they can help to explain the stakeholders which metrics are used, how are they combined into indicators, and how the analytics results have

been achieved (Biedermann, Schneider & Drachsler 2018). With the LAIR we can directly address multiple aims of the TLA approach: Provide transparency of the algorithm; better inform the data subject and increase their agency in dealing with the outcomes of a TLA system; it address rights given to the data subjects by the GDPR such as the "right to be informed", or the "right to object".

In conclusions or Towards Trusted Learning Analytics

The widespread rollout and adoption of LA in educational institutions in Europe has lately stagnated due to concerns about privacy and ethics with regards to personal data and the new GDPR. In this ongoing discussion, fears and realities are often indistinguishably mixed up, leading to an atmosphere of uncertainty among potential beneficiaries of LA as well as institutional managers who aim to innovate their institution's learning support but now hesitate to implement analytics.

The TLA research program aims to renew the "contract" between learners and their educational providers not only to reach a high level of trust but also to release the full potential of LA with practical tools. The design of the TLA tools is done according to value-sensitive design processes, which allows considering ethical and privacy values on the same level as functional requirements. Thereby, the aforementioned ethical considerations help to develop a TLA system that achieves its aims not only in a technical but also in an ethical, humane and therefore trustful manner. In order to reach this level of trust, it is crucial to "white box" the so far "black box" analytics systems. Within the TLA approach, we aim to be open about the algorithms applied, being transparent about the metrics used, and indicators computed; the main goal of a TLA system is providing feedback rather than being used for automated decision-making. This non-exclusive list describes core functionalities that are not in the center of current Big Data system neither of LA systems until today. They lay for us the foundation for a new manner to treat educational data and provide meaningful services for its stakeholders. The TLA approach is therefore fundamentally different in its design compared to other LA systems that have been developed so far. For sure, the TLA approach will also have its limitations as the result of a complex algorithm like a neural network is very hard to explain and make transparent to its users. It also demands high level competences such as self-regulations skills, information agency, and critical thinking at the data subject side to properly work and act on information provided by a TLA system.

PODCAST

PROF. DR. HENDRIK DRACHSLER
Goethe Universität Frankfurt am Main
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
drachsler@dipf.de

Comment

1 https://uhh.de/blh30 [11.07.18]

Literature

Berg, A., Scheffel, M., Drachsler, H., Ternier, S. & Specht, M. (2016). Dutch Cooking with xAPI Recipes: The Good, the Bad and the Consistent. *Proceedings of the International Conference on Advanced Learning Technologies,* ICALT '16, p. 234–236. Available under: https://uhh.de/oig40 [11.07.2018].

Biedermann, D., Schneider, J. & Drachsler, H. (2018). The Learning Analytics Indicator Repository. In Drachsler, H., Pammer-Schindler, V. & Pérez-Sanagustín, M. (Eds.), *Lifelong technology enhanced learning: Dealing with the complexity of 21st century challenges.* 13th European Conference on Technology-Enhanced Learning (EC-TEL 2018), Leeds, UK.

Drachsler, H. & Greller, W. (2016). Privacy and Analytics—it's a DELICATE issue. A Checklist to establish trusted Learning Analytics. *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge, LAK '16,* p. 89–98, New York, NY, USA. ACM. Available unter: https://uhh.de/ub0c9 [11.07.2018].

Greller, W. & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society, 15(3),* p. 42–57. Available under: https://uhh.de/21lzv [11.07.2018].

Jivet, I., Scheffel, M., Specht, M. & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. *Proceedings of International Conference on Learning Analytics and Knowledge, Sydney, NSW, Australia, March 7–9, 2018 (LAK '18),* 10 pages. Available under: https://uhh.de/8gwjl [11.07.2018].

Singer, N. (2014). InBloom Student Data Repository to Close. *The New York Times,* April 21, 2014. Available under: https://uhh.de/2rgnb [11.07.2018].

Slade, S. & Prinsloo, P. (2015). Student vulnerability, agency and learning analytics: an exploration. *Journal of Learning Analytics,* Special Issue on Ethics and Privacy.

Tsai, Y.-S., Moreno-Marcos, P. M., Tammets, K., Kollom, K. & Gašević, D. (2018). SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18).* ACM, New York, NY, USA, p. 320–329. Available under: https://uhh.de/e9prn [11.07.2018].