Towards an Open Architecture for Learning Analytics in Open and Distance Education

Tore HOEL[a]* **& Jun XIAO**[b]
[a]*Oslo Metropolitan University, Norway*
[b]*Shanghai Open University, China*
*tore.hoel@oslomet.no

Abstract: A characteristic of open and distance education is that all transactions between the learner and the university leave a digital footprint. This data trail gives open universities a golden opportunity to be in the forefront developing solutions for learning analytics and utilizing of educational big data. This paper explores development of an open architecture for learning analytics, from clarifying the concept of openness in the context of analytics to suggesting specific components a technical architecture for open and distance education. A literature review of the concept of openness in relation to learning analytics identifies a need for conceptual clarification, in particular in the context of open and distance education. The paper applies two conceptual frameworks to support the development of requirements for an open architecture, the European interoperability framework, and the ISO/IEC reference model for learning analytics. Components of an open architecture defined at Shanghai Open University are presented, and a process for validation of this architecture is suggested.

Keywords: Open learning analytics architecture, open learning analytics, open platform, open and distance education

1. Introduction

Open learning and distance education (ODE) give access to education and training provision for all learners with a smart device and willingness to learn, freeing them from constraints of time and place. The online mode of exchange leaves an enormous amount of data that could be put to good use both for the learner and the educational institution. The principles of ODE invite to a development of an open architecture for use of big data. However, the issue of big data is ripe with thorny questions related to privacy, transparency, and ethical and pedagogical use of analytical insights; therefore, there is a need for careful definition of the constraints of an open architecture for support of learning analytics (LA) within ODE.

ODE institutions offer courses, some of them as massive open online courses (MOOCs), supported by a varying degree of open content (e.g., Open Educational Resources (OER) or open textbooks), often hosted in some kind of open knowledge cloud. Globally, ODE has various backing by the governments of the institutions' country of origin; and some institutions, but not all, see themselves as part of an international open and distance learning (ODL) community that share common values and engage in exchange of ideas, human and other resources.

The open in ODE is usually understood as highlighting the open access to education, and the openness related to mode of access (time, place, platform, etc.). When defining openness related to LA architectures there is a need to revisit the ideas that give ODE its strength and examine the needs of national or regional open university systems. To develop open learning analytics architecture (OLAA) beyond a marketing term we need to understand what openness means in terms of big data and analytics. This is the first research question in this short workshop paper. The next questions will ask what openness will mean in terms of technical architectures, and what components of an OLAA we have designed for an open university in China. The paper will also reflect upon the needs for validation of OLAA, based on Chinese and international practice.

2. Related work

In 2011, as one of the first actions of Society for Learning Analytics Research (SoLAR), Siemens et al. (2011) called for development of an open platform to integrate heterogeneous learning analytics techniques. Three "critical beliefs" underpinned the proposal:

1. Openness of process, algorithms, and technologies is important for innovation and meeting the varying contexts of implementation. 2. Modularized integration: core analytic tools (or engines) include: adaptation, learning, interventions, and dashboards. The learning analytics platform is an open architecture, enabling researchers to develop their own tools and methods to be integrated with the platform. 3. Reduction of inevitable fragmentation by providing an integrated, expandable, open technology that researchers and content producers can use in data mining, analytics, and adaptive

content development.   (Siemens et al., 2011, p. 2)

The SoLAR proposal asked for funding of tool development, in particular, a LA engine; an adaptive content engine; an intervention engine for recommendations and automated support; and a dashboard, including reporting and visualization tools (Figure 1).

Figure 1. SoLAR proposal for an integrated learning analytics system (Siemens et al., 2011)

The SoLAR call for open architectures, when the LA research community was still in its infancy, did not get an immediate response. An April 2014 press release from an Open Learning Analytics Summit at the US Marist College (Solar, 2014) demonstrated interest in the idea from key researchers in the field. A few months earlier the Apereo foundations launched its LA initiative (www.apereo.org/communities/learning-ana-lytics-inititive). Later in 2014 the EU supported networking initiative, LACE, organized Open Learning Analytics Network Summit Europe to develop a shared European perspective on the concept of open LA (Cooper, 2014). The main message coming out of these summits was an emphasis on "the need for open source software, open standards, and open APIs to address the interoperability challenge in this field as well as how important tackling the ethical and privacy issues is becoming for a wide deployment of LA" (Chatti et al., 2017, p. 5).

Now, six or seven years after the first call for open learning analytics (OLA) we start to see proposals for architectures that could be implemented, e.g., the Jisc Open Learning Analytics Architecture (analytics.jiscinvolve.org/wp/category/architecture; Sclater, Berg, & Webb, 2015), and the Open learning analytics platform (OpenLAP) proposed by Muslim, Chatti, Bashir, Barrios Varela, and Schroeder (2018).

In his book Learning Analytics Explained Sclater (2017) describes the Jisc LA architecture as building upon Apereo vision for an open LA platform (with the elements collection, storage, analysis, action, and communication). The Jisc framework sets out according to Sclater (2017) the key components for carrying out LA, providing the data to stakeholders, obtaining students' consent for data collection and intervention, and managing any resulting interventions with learners. The Jisc architecture harvests data from library systems, learning management systems, student information systems, and from students themselves (self-declared data).

Muslim et al. (2018) describe the Apereo and Jisc solutions as "addressing the goals of a limited set of stakeholders, working with a specific set of data, answering a predetermined set of objectives, and relying on a predefined set of analytics methods" p. 1); and their OpenLAP is targeting to address these limitations. The openness in their solutions lies in the flexible definition and dynamic generation of indicators to meet the needs of different stakeholders with diverse goals (Muslim, Chatti, Mughal, & Schroeder, 2017). The OpenLAP focuses on the modularity and extensibility of the LA platform. "OpenLAP adapts a modular and extensible architecture that allows the easy integration of new analytics modules, analytics methods, and visualization techniques" (Muslim et al., 2018, p. 93).

2.1 What is openness?

In the OpenLAP architecture described above openness was being able to define one's indicator set and have a framework that allows for the dynamic addition of new analytics methods and visualization techniques at runtime. Modularity and extensibility are just two of a number of 'abilities' that could define openness, e.g., interoperability, reusability, flexibility, scalability, usability.

Chatti, Muslim, and Schroeder (2017) discuss in detail the concept of openness related to LA and seem to conclude that the full range of 'openness issues' should be considered: Open learning; open practice; open architectures, processes, modules, algorithms, tools, techniques, and methods that can be used to reuse, redistribute, revise, or remix objects; open access to learning analytics platforms; open participation in the LA process by engaging different stakeholders in the LA exercise; open standards, open research; open science; open learner modeling; and open assessment. Chatti et al. (2017) claim that 'open' should be interpreted in relation to these conceptualizations of openness; however, their discussion leaves much to be done in defining exactly what openness means in the context of LA.

In summary, the conceptualization and development of LA as an open approach are still to be done. There is a need to understand how learning is supported in these environments and how learners, educators, institutions and researchers interact in order to utilize educational big data. In the following sections of this paper we will focus on requirements for open LA in ODL.
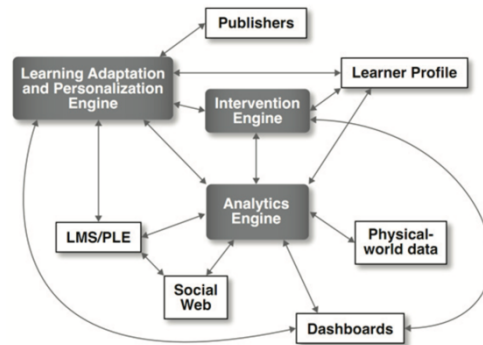
3. Openness in Open and Distance Education

In general, and as interpreted in many countries around the world, ODE has a broad mandate to provide education to a wide group of learners, often without institutional affiliation, and with different kinds of disadvantages, socially, culturally, and economically. According to Commonwealth of Learning (2000) there is no universally accepted definition on ODL; however, most scholars agree that there is an important characteristic of separation of teacher and learner in time or place, or in both time and place. As new technologies and teaching methods emerge, ODL branches into new approaches. LA will contribute to this development. Therefore, when focusing on the role of openness in ODL related to LA we need to use an analytical framework that captures the multifaceted and dynamic characteristics of ODL.

We would suggest using a simplified version of the European Interoperability Framework developed by the European Commission (2017) looking at interoperability at three levels: 1) Policy / Legal, 2) Cultural / Organizational, and 3) Semantic / Technical. We will use this framework to drive development of requirements for OLAA.

In Table 1 we have listed some requirements for an open LA architecture grouped by interoperability level. On the two first non-technical levels of interoperability requirements for OLAA share many of the same characteristics as for other educational technologies. For example, ODL institutions will benefit from national policies promoting exchange of technology, methods, training, curriculum development, course design, content, etc between universities and regions. A common approach to the challenges of use and sharing of educational data will also serve ODE. On the cultural and organizational level institutions will benefit from openness in processes and knowledge exchange, in particular in new and emergent fields of practice like LA. However, it is on the semantic and technical level the interoperability challenges of LA are breaking new ground where there is little previous knowledge to build on. The rest of this paper will focus on these challenges and suggest some solutions and ways to test them.

Table 1
OLAA requirements for different levels of interoperability

Interoperability level	OLAA requirements
Policy – Legal	National policies related to open and distance learning to reach out to educational needs National policies regarding open data, open research, open access, etc. Privacy and data protection laws and regulations
Cultural - Organizational	Open processes & knowledge exchange (implementation, quality, assessment, etc.) Open pedagogies (adaptation, interventions, etc.) Open Educational Resources - access and promotion
Semantic – Technical	Open data; Open algorithms; Open learner models; Open consent service; Open Sources code; Open modules; Open inference / analytics engines; Open adaptation and personalization engine; Open Intervention Engine; Open dashboards Use of open standards (e.g., for exchange of activity data, predictive models)

4. Building blocks for an open LA architecture

In the following we will use the ISO/IEC 20748:2016 learning analytics process model (Figure 3) as a template for discussing requirements for OLAA. The first three sub processes of a LA cycle are about selecting and handling data. There need to be a good match between questions and data, otherwise LA becomes irrelevant or in the worst case harmful. In many big data scenarios both questions and types of data are looked upon as business secrets. Even if universities are businesses and there is a strong competition between them there is no need to keep the metrics for LA in education closed. There is a need for continuously discussion on metrics profiles in ODL; these profiles should be openly shared.

Learners' data are imbued with privacy issues, and therefore all proposals to share these data would meet objections. However, there are alternatives to live and openly sharing of activity data that may serve some of the needs for open data. Synthetic data, also known as simulated data, may be useful to limit the risks for illegitimate re-identification of
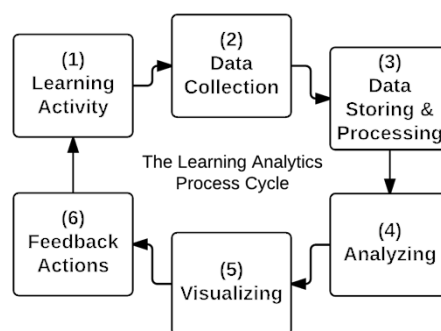

Figure 3. LA process model (ISO/IEC 20748:2016)

personally identifiable information in learning activity data sets. This kind of data may as an example be created by using real example data to train a predictive model, but then generating replacement data from the tuned model.

> Synthetic data enables the rapid prototyping of services before the "real" big data has been amassed or made available to an application. Its availability supports proof of concept, security testing, practicing, and training around data governance processes, boundary testing, user testing of visualizations, and interoperability testing of different architectural components, as well as many other applications. (Berg, Mol, Kismihók, & Sclater, 2016, p. 108).

Trusted data storage is complex as it raises a host of non-technical issues of organizational, legal and even political character. De-identification of data for use in different settings implies combining different techniques, e.g., pseudonymization, aggregation, data reduction, data suppression, data masking etc., weighing risks of re-identification against the risk of depleting the value of the data. This process requires skills and organization. For some communities it may be a good idea to establish trusted services that handle the data storing and give educational stakeholders the opportunity to ask questions to specific data sets and get the answers back without having direct access to the raw data.

Analyzing data, the fourth sub process of the LA cycle, is the core process of LA that makes use of a number of technologies, some of them well suited to be shared as open models or open technologies. Learner models plays an important role as baseline for analysis. These models are based on data from particular student populations, courses, national and regional settings. Sharing of learner models will require some level of common understanding of their nature. However, collaboration between institutions like the ones involved in ODE will have much to gain by exchanging learner models.

Developing prediction engines involves a lot of training and fine tuning that is highly dependent upon context and example data. To make this process more transparent one could share example data as synthetic data, and one could make the predictive engines open source so that other interested parties can test and build on the good work of others. In addition, the learning algorithms could be provided as an XML-based predictive model using an open standard like Predictive Model Markup Language (PMML) (Guazzelli, Zeller, Lin, & Williams, 2009).

How visualizations are used to report analytical insights for different groups of stakeholders is not a straight forward process. Gasevic, Dawson, and Siemens (2015) have argued that the use of dashboards can lead to the implementation of weak and perhaps detrimental instructional practices. Therefore, there is a need to experiment with many types of visualizations and to evaluate their effect in different contexts. An open library of dashboards would allow for the design effective learning analytics visualizations that consider both instructional, learning and sense-making benefits for learning. Cobos, Gil, Lareo, and Vargas (2016). have suggested open dashboards deployed as native modules to the usual MOOC platforms as one solution to this challenge.

For the last sub process, Feedback Actions (Figure 3), there are a range of actions that will support interactions between the student and the institution, teachers and their intermediators (e.g., learning resources, learning environment, learning platform etc.). These actions are highly dependent on interoperability on both the Policy - Legal level and the Cultural - Organizational level, as well as on a Semantic - Technical level. Openness in terms of open policies will certainly help to develop successful feedback on insights from analytics. However, there is a need for further research to see if there is basis for design of open feedback action models.

Table 2 summarizes the potential open models related to the different sub processes of a LA cycle. The rest of this paper will be focused on suggestions for an open technical architecture for ODE, based on design work at Shanghai Open University (SOU), one of China's leading ODE institutions.

5. Technical architecture for Open LA

In this section we report on an effort of SOU to design an open architecture to support their course portfolio, which includes degree programs and vocational training, as well as 'leisure education' reaching out to a broad group of learners in all age groups. Based on review of existing blueprints for open architectures (Apereo in Sclater, 2017; Chatti et al., 2017; Jisc in Sclater et al., 2015) SOU

Table 2

Potential open models for LA related to sub processes in the LA process cycle

LA sub processes	Open models
Learning Activity	Open Questions - Open Metrics Profiles
Data Collection & Data Storing & Processing	Synthetic Open Data Trusted Service for Open Questions
Analysis	Learner models Open data for training prediction engines Sharing of Prediction engines
Visualization	Open library of visualizations
Feedback Actions	Open policies on LA

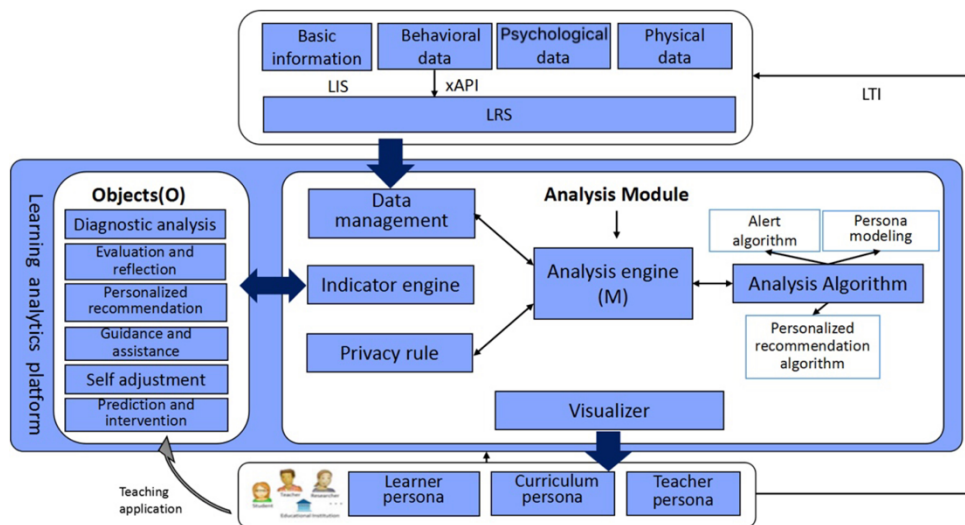has developed an initial model (Figure 4).

Figure 4. SOU Open LA Architecture

Using the ISO framework model (Figure 3) as a heuristic we will describe the model as follows:

Process 1 Learning Activity (What are we looking for?): The objectives (in the model "Objects (O)") are described as high-level objectives that can be deducted from the educational activities that are underpinning Diagnostic analysis, Evaluation and reflection, Personalized recommendation, Guidance and assistance, Self-adjustment, and Prediction and intervention. In the Analysis Module there is also indicated different types of algorithms resulting in Alerts, Persona modelling, and Personalized recommendation.

Process 2: Data Collection (What data?): The model identifies four type of data, basic information originating from Learning Information Systems, behavioral data originating from interactions with learning resources, fellow learners, teachers, tools, etc in a learning environment, psychological data, and physical data. Since the model does not describe any relations between the objectives and the data needed, the model is silent about details of the sources of the data.

Process 3: Data Storing & Processing (How to store and share data?): The model is silent about how data are store and processed, only using the more generic concept of a Learning Record Store as the container for data.

Process 4: Analyzing (How are we looking at the data?): In the model, analysis is done by the Analysis engine interacting with the Analysis algorithm. The SOU model is silent about how the algorithms are developed and managed. As described above, in an open architecture it should be possible to test the algorithms by applying own test data or openly shared synthetic data.

The persona modelling is a pivotal part of the analysis process of the model, on which teaching actions are dependent.

Process 5: Visualizing (How will you communicate your findings?): The visualizer component of the model feeds information to different usage scenarios for learners, curriculum designers and teachers.

Process 6: Feedback Actions (What actions will you take?): Students as receivers of feedback actions are only present indirectly through Teaching application in the model. Therefore, the model is silent about how the personalized recommendation, reflection, and self-adjustment would happen.

Based on this architecture SOU has implemented a homework warning system as part of the learning platform of the university. This module is the first of a series of LA applications to be implemented. It sends homework warnings via web, text messages, and chat application.

The SOU Open LA Architecture will go through several design cycles as it will be tested against continued solicitation of user requirements, international experts' evaluation and testing of pilot applications. A first evaluation by a focus group of Chinese and international experts (Xiao, Hoel, Yuan, Sclater, 2017) suggest some parts that needed improvement. In particular, SOU has been doing extensive research in the area of 'learner persona models', a concept that is included in the architecture described in Figure 4. It needs to be explored how this is related to international research on open learner models as described for example in Bull and Kay (2010); and Pavlik Jr, Brawner, Olney, and Mitrovic (2013).

6. Conclusions and further work

This workshop paper marks the start of developing an open architecture for LA for ODE – a somewhat bold claim that at least is defendable if we limit it to ODE in a Chinese context. An open architecture is not just a blueprint for system architecture that is published for open academic discussion. Openness has more distinct meanings related to how he components in the system are interoperable and interchangeable in a way that the community of interest can take advantage of in a transparent and accessible way. In this paper we have seen that the understanding of how openness plays out in the context of learning analytics is still colored by the discourse in related fields, i.e., open research, open data, open educational resources, etc. In future research we will strive to come up with a more coherent and targeted definition of the characteristics of an open LA architecture.

This paper has also demonstrated that the use of the European Interoperability Framework gives conceptual support for design of components of an open LA architecture. Along the same lines, this paper has proved that the use of ISO/IEC's LA process model as analytical tool helps the construction of the technical parts of OLAA. Applied on the first blueprint of an architecture for Shanghai Open University we have been able to pinpoint areas that need further development, e.g., the learner modelling component.

Developing the OLAA further, we plan to make a systematic solicitation of ODE stakeholder requirements in order to define the characteristics of the open modules to be included in the architecture. We will also define a process for testing the design proposal so that we will be able to go through efficient design cycles before implementing the design in running code. We will also do an analysis of data sources available to an ODE institution so that we can deploy a maximum range of analytics services answering to the requirements of the different stakeholders.

References

Berg, A. M., Mol, S. T., Kismihók, G., & Sclater, N. (2016). The role of a reference synthetic data generator within the field of learning analytics. *Journal of Learning Analytics*, 107–128. http://doi.org/10.18608/jla.2016.31.7

Bull S., Kay J. (2010) Open Learner Models. In: Nkambou R., Bourdeau J., Mizoguchi R. (eds) Advances in Intelligent Tutoring Systems. *Studies in Computational Intelligence*, vol 308. Springer, Berlin, Heidelberg

Chatti M.A., Muslim A., Schroeder U. (2017) Toward an Open Learning Analytics Ecosystem. In: Kei Daniel B. (eds) *Big Data and Learning Analytics in Higher Education*. Springer, Cham DOI https://doi.org/10.1007/978-3-319-06520-5_12

Cobos, R., Gil, S., Lareo, A., & Vargas, F. A. (2016). Open-DLAs (pp. 265–268). Presented at the *Third (2016) ACM Conference, New York*, New York, USA: ACM Press. http://doi.org/10.1145/2876034.2893430

Commonwealth of Learning (2000). An introduction to open distance learning. Retrieved from: http://www.col.org/SiteCollectionDocuments/ ODLIntro.pdf

Cooper, A. (2014). Open Learning Analytics Network - *Summit Europe 2014*. http://www.laceproject.eu/open-learning-analytics-network-summit-europe-2014/

European Commission. (2017). European Interoperability Framework - Implementation Strategy. Annex 2. COM(2017) 134 final.

Gasevic, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. http://doi.org/10.1007/s11528-014-0822-x

Guazzelli, A., Zeller, M., Lin, W.-C., & Williams, G. (2009). PMML: An open standard for sharing models. *The R Journal*, 1, 60–65. https://journal.r-project.org/archive/2009/RJ-2009-010/index.html

ISO/IEC 20748-1:2016. Information technology — learning, education, and training — Learning Analytics Interoperability — Part 1: Reference model.

Muslim, A., Chatti, M. A., Bashir, M. B., Barrios Varela, O. E., & Schroeder, U. (2018). A Modular and Extensible Framework for Open Learning Analytics. *Journal of Learning Analytics*, 5(1), 92–100. http://doi.org/10.18608/jla.2018.51.7

Muslim, A., Chatti, M. A., Mughal, M., & Schroeder, U. (2017). The Goal - Question - Indicator Approach for Personalized Learning Analytics (pp. 371–378). Presented at the *9th International Conference on Computer Supported Education*, SCITEPRESS - Science and Technology Publications. http://doi.org/10.5220/0006319803710378

Pavlik Jr, P. I., Brawner, K., Olney, A., & Mitrovic, A. (2013). A Review of Student Models Used in Intelligent Tutoring Systems. In *Design Recommendations for Intelligent Tutoring Systems* (Volume I) https://www.researchgate.net/publication/258883086_A_Review_of_Learner_Models_Used_in_Intelligent_Tutoring_Systems

Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, R., Shum, S. B., Ferguson, R., et al. (2011). Open Learning Analytics: an integrated & modularized platform. *Society for Learning Analytics Research*. http://solaresearch.org/wp- content/uploads/2011/12/OpenLearningAnalytics.pdf

Sclater, N. (2017). Learning analytics explained. Routledge. ISBN : 1-315-67956-6

Sclater, N., Berg, A., & Webb, M. (2015). Developing an open architecture for learning analytics. Presented at the *EUNIS conference*, Dundee, Scotland UK, - June.

Solar (2014). Leaders in Learning Analytics and Open Source Software. Hold Open Learning Analytics Summit. Press release: solaresearch.org/wp-content/uploads/2014/04/OLAPressReleaseFINALv2.pdf

Xiao, J., Hoel, T., Yuan, L, Sclater, N. (2017). The Shanghai Open University Open Learning Analytics Architecture. 1st Report on the Progress of developing of SOU Open LA Architecture – November 2017