Toward an Open Learning Analytics Ecosystem

Mohamed Amine Chatti, Arham Muslim, Ulrik Schroeder

Abstract In the last few years, there has been a growing interest in learning analytics (LA) in technology-enhanced learning (TEL). LA approaches share a movement from data to analysis to action to learning. The TEL landscape is changing. Learning is increasingly happening in open and networked learning environments, characterized by increasing complexity and fast-paced change. This should be reflected in the conceptualization and development of innovative LA approaches in order to achieve more effective learning experiences. There is a need to provide understanding into how learners learn in these environments and how learners, educators, institutions, and researchers can best support this process. In this paper, we discuss open learning analytics as an emerging research field that has the potential to deal with the challenges in open and networked environments and present key conceptual and technical ideas toward an open learning analytics ecosystem.

1 Introduction

In recent years, learning analytics has attracted a great deal of attention in technology-enhanced learning (TEL) research as practitioners, institutions, and researchers are increasingly seeing the potential that learning analytics has to shape the future TEL landscape. Learning analytics represent the application of "big data" and analytics in education (Siemens et al., 2011). Generally, learning analytics deals with the

---

Mohamed Amine Chatti
Informatik 9, RWTH Aachen University, e-mail: chatti@informatik.rwth-aachen.de

Arham Muslim
Informatik 9, RWTH Aachen University, e-mail: muslim@cil.rwth-aachen.de

Ulrik Schroeder
Informatik 9, RWTH Aachen University, e-mail: schroeder@cil.rwth-aachen.de

development of methods that harness educational data sets to support the learning process.

In the past few years, the discussion about technologies for learning has moved away from only institutionally managed systems (e.g. LMS) to open and networked learning environments (e.g. PLE, MOOC) (Chatti, 2010). In fact, learning is increasingly distributed across space, time, and media. Consequently, a large volume of data - referred to as big data - about learners and learning is being generated. This data is mainly traces that learners leave as they interact with increasingly complex and fast changing learning environments.

The abundance of educational data and the recent attention on the potentiality of efficient infrastructures for capturing and processing big data have resulted in a growing interest in big learning analytics among researchers and practitioners (Dawson et al., 2014). Big learning analytics refers to leveraging big data analytics methods to generate value in TEL environments (Chatti et al., 2014). Harnessing big data in the TEL domain has enormous potential. Learning analytics stakeholders have access to a massive volume of data from learners activities across various learning environments which, through the use of big data analytics methods, can be used to develop a greater understanding of the learning experiences and processes in the new networked learning environments.

The research field of learning analytics is constantly developing new ways to analyze educational data. However, most of the learning analytics approaches to date are restricted to analytics tasks in a narrow context within specific research projects and centralized learning settings. Little research has been conducted so far to understand how learners learn in todays open and networked learning environments and how learners, educators, institutions, and researchers can best support this process. Operating in these environments requires a shift towards learning analytics on more challenging datasets across a variety of different sites with different standards, owners, and levels of access (Ferguson, 2012; Fournier et al., 2011) by applying mixed-method approaches to address a wide range of participants with diverse interests, needs, and goals. Further, there is a need for a new learning analytics model as an ongoing process across time and environments, where everyone can be producer and consumer of the learning analytics exercise.

A central aspect of this discussion is the concept of open learning analytics. Siemens et al. (2011) provide an initial proposal expressing the importance of an integrated and modularized platform to integrate heterogeneous learning analytics techniques. The concept of open learning analytics represents a significant shift toward a new learning analytic model that takes "openness" into account. This leads to questions about how should "open" be interpreted in relation to learning analytics? What are the challenges in open learning analytics? What are the components of an open learning analytics ecosystem? What are the requirements for an effective open learning analytics platform? What are the technical details (i.e. architecture and modules) of an open learning analytics platform?

In this paper, we address these questions and present the theoretical, conceptual, and technical details toward an open learning analytics ecosystem that aims at supporting learning and teaching in fragmented, diverse, and networked learn-

ing environments. Research on open learning analytics is still in the early stages of development. Our endeavor is to foster a common understanding of key conceptual and technical ideas in this research area that will support communication between researchers and practitioners as they seek to address the various challenges and opportunities in this emerging field towards sustainable practical open learning analytics.

2 Learning Analytics

Different definitions have been provided for the term Learning Analytics (LA). The most commonly cited definition of learning analytics which was adopted by the first international conference on learning analytics and knowledge (LAK11) is "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" (as cited in Siemens & Long, 2011, Learning Analytics section, para. 2). Ferguson (2012) and Clow (2013) compile a list of LA definitions and provide a good overview on the evolution of LA in recent years. Although different in some details, LA definitions share an emphasis on converting educational data into useful actions to foster learning. Furthermore, it is noticeable that these definitions do not limit LA to automatically conducted data analysis. In this paper, we view LA as a TEL research area that focuses on the development of methods for analyzing and detecting patterns within data collected from educational settings, and leverages those methods to support the learning experience.

Learning analytics is not a genuine new research area. It reflects a field at the intersection of numerous academic disciplines (e.g. learning science, pedagogy, psychology, Web science, computer science) (Dawson et al., 2014). It borrows from a variety of related fields (e.g. academic analytics, action analytics, educational data mining, recommender systems, personalized adaptive learning) and synthesizes several existing techniques (e.g. machine learning, data mining, information retrieval, statistics, and visualization) (Chatti et al., 2014; Ferguson, 2012).

Chatti et al. (2012, 2014) provide a systematic overview on LA and its key concepts through a reference model based on four dimensions. The authors further build on this model to identify a series of challenges and develop a number of insights for LA research in the future. As depicted in Figure 1, the four dimensions of the proposed model are:

– **What?** What kind of *data* does the system gather, manage, and use for the analysis? This dimension refers to the data used in the LA task. It also refers to the *environments* and *contexts* in which learning occurs. Educational data comes from formal as well as informal learning channels. It can also come in different formats, distributed across space, time, and media.
– **Who?** Who is targeted by the analysis? The application of LA can be oriented toward different *stakeholders*, including students, teachers, (intelligent) tutors/mentors, educational institutions (administrators and faculty decision mak-
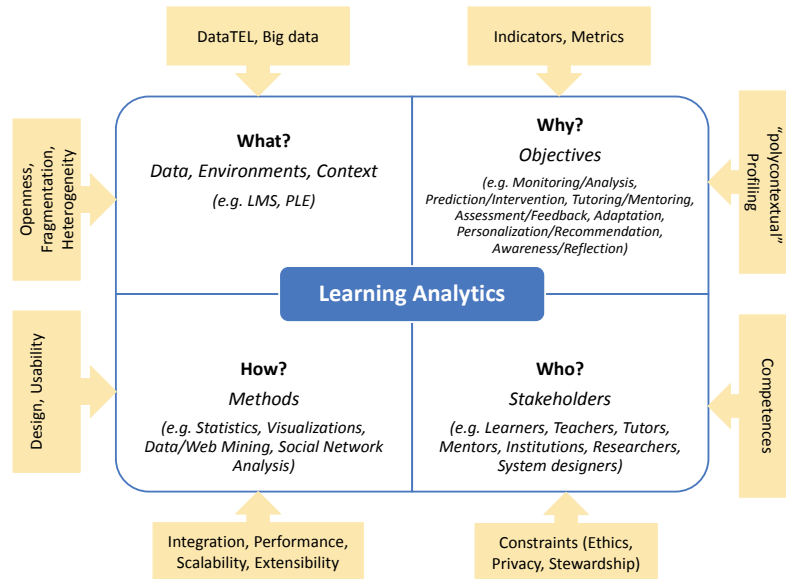
Fig. 1 Learning analytics reference model (Chatti et al., 2014)

ers), researchers, and system designers with different perspectives, goals, and expectations from the LA exercise.

– **Why?** Why does the system analyze the collected data? There are many *objectives* in LA according to the particular point of view of the different stakeholders. Possible objectives of LA include monitoring, analysis, prediction, intervention, tutoring/mentoring, assessment, feedback, adaptation, personalization, recommendation, awareness, and reflection.

– **How?** How does the system perform the analysis of the collected data? LA applies different *methods* to detect interesting patterns hidden in educational data sets. Possible methods include statistics, information visualization (IV), data mining (DM), and social network analysis (SNA).

3 Open Learning Analytics

A particularly rich area for future research is open learning analytics. The concept of open learning analytics was introduced in 2011 by a group of leading thinkers on LA in an initial vision paper published by the Society for Learning Analytics Research (SoLAR) (Siemens et al., 2011). A first summit was then held in Indianapolis, Indiana in March 2014 to promote networking and collaborative research and "to bring together representatives from the learning analytics and open source

software development fields as a means to explore the intersection of learning analytics and open learning, open technologies, and open research" (Alexander et al., 2014). This summit initiated discussion towards the idea of open learning analytics as a conceptual and technical framework around which different stakeholders could network and share best practices. From a technical perspective, the summit focused on open system architectures and how open source communities can provide new open source learning analytics services and products. Building on the first summit, the Learning Analytics Community Exchange (LACE) project organized in December 2014 the Open Learning Analytics Network Summit Europe to develop a shared European perspective on the concept of an open learning analytics framework (Cooper, 2014a). Sclater (2014) provides a good summary of this summit. He notes that the most obvious aspect of open in the context of learning analytics is the reuse of code and predictive models.

So far, from the initial vision paper through the last summit, the development of the concept of open learning analytics was restricted to a discussion 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. The concept of open learning analytics is, however, still not well defined and concrete conceptual and development plans are still lacking. Several important questions remained unanswered. These include:

– How should "open" be interpreted in relation to learning analytics?
– How can open learning analytics be leveraged to foster personalized, networked, and lifelong learning?
– What are the challenges in open learning analytics in addition to interoperability and privacy?
– What are the components of an open learning analytics ecosystem?
– What are concrete user and system scenarios that an open learning analytics platform should support?
– What are the requirements for an effective open learning analytics platform?
– What are the technical details (i.e. architecture and components) of an open learning analytics platform?

In the next sections, we attempt to give answers to these questions. We start by providing a clarification of the term open learning analytics and then present the conceptual and technical details toward an open learning analytics ecosystem.

What is open learning analytics? The term openness has received a great deal of attention from TEL community, due to the growing demand for self-organized, networked, and lifelong learning opportunities. "The two most important aspects of openness have to do with free availability over the Internet and as few restrictions as possible on the use of the resource, whether technical, legal or price barriers" (OECD, 2007, p. 32). According to Wiley (2009), at its core, openness is sharing and education is a relationship of sharing. Open education has been evolving over the past century (McNamara, 2012). From the late nineteenth century and during the twentieth century, open education has been explored in the development of distance

education along with other open learning initiatives, such as the open classroom, open schooling, and the open university (Peters, 2008). Open educational resources (OER) and open courseware (OCW) represent a further important advancement in the open education movement over the past decade (Downes, 2007; McNamara, 2012). With the introduction of the Massive Open Online Course (MOOC) term in 2008, MOOCs have been in the forefront of the open education movement. MOOCs have been considered as an evolution of OER and OCW (Yuan and Powell, 2013).

Driven by the different perspectives on openness as discussed in the literature on open education, OER, OCW, and MOOCs, several suggestions can be made as to how "open" should be interpreted in relation to learning analytics.

- **Open learning** by providing understanding into how learners learn in open and networked learning environments and how learners, educators, institutions, and researchers can best support this process (Chatti et al., 2014).
- **Open practice** that gives effect to a participatory culture of creating, sharing, and cooperation.
- **Open architectures, processes, modules, algorithms, tools, techniques, and methods** that can be used by following the four R's "Reuse, Redistribute, Revise, Remix" (Wiley, 2009; Hilton et al., 2010). Everyone should have the freedom to use, customize, improve, and redistribute the entities above without constraint.
- **Open access** to learning analytics platforms granted to different stakeholders without any entry requirements in order to promote self-management and creativity.
- **Open participation** in the LA process by engaging different stakeholders in the LA exercise. Daniel and Butson (2014) state that in LA, "there is still a divide between those who know how to extract data and what data is available, and those who know what data is required and how it would best be used" (p. 45). Therefore, it is necessary to bring together different stakeholders to work on common LA tasks in order to achieve useful LA results. Further, it is essential to see learners as the central part of the LA practice. This means that learners should be active collaborators, not just mere data subjects (Sclater, 2014) and recipients of interventions and services (Slade and Prinsloo, 2013). Learner and teacher involvement is the key to a wider user acceptance, which is required if LA tools are to serve the intended objective of improving learning and teaching.
- **Open standards** "to reduce market fragmentation and increase the number of viable products" (Cooper, 2014a). Open standards and specifications can help to realize the benefits of better interoperability (Cooper, 2014b).
- **Open Research** and **Open science** (Fry et al., 2009) based on **open datasets** with legal protection rules that describe how and when the dataset can be used (Verbert et al., 2012). Sclater (2014) points out that datasets "from one environment can be connected to that in another one, not only across the different systems in one institution but potentially with other institutions too". Following an open dataset approach, a group of interested researchers started an initiative around "dataTEL". The main objective was to promote exchange and interoperability of educational datasets (Duval, 2011; Verbert et al., 2011). Examples of open

datasets include PSLC datashop as a public data repository that enables sharing of large learning datasets (Koedinger et al., 2010).

– **Open learner modeling** based on user interfaces that enable refection, planning, attention, and forgetting and that can be accessed by learners to control, edit, update, and manage their models (Kay and Kummerfeld, 2011). This is important to build trust and improve transparency of the LA practice.

– **Open assessment** to help lifelong learners gain recognition of their learning. Open assessment is an agile way of assessment where anyone, anytime, anywhere, can participate towards the assessment goal. It is an ongoing process across time, locations, and devices where everyone can be assessor and assesse (Chatti et al., 2014).

The concept of open learning analytics covers all the aspects of "openness" outlined above. It refers to an ongoing analytics process that encompasses diversity at all four dimensions of the reference model introduced in Section 2:

– What? It accommodates the considerable variety in learning data, environments, and contexts. This includes data coming from traditional education settings (e.g. LMS) and from more open-ended and less formal learning settings (e.g. PLE, MOOC, social web).
– Who? It serves different stakeholders with very diverse interests and needs.
– Why? It meets different objectives according to the particular point of view of the different stakeholders.
– How? It leverages a plethora of statistical, visual, and computational tools, methods, and methodologies to manage large datasets and process them into indicators and metrics which can be used to understand and optimize learning and the environments in which it occurs.

4 Open Learning Analytics Platform

The aim of open learning analytics is to improve learning efficiency and effectiveness in lifelong learning environments. In order to understand learning and improve the learning experience and teaching practice in todays networked and increasingly complex learning environments, there is a need to scale LA up which requires a shift from closed LA tools and systems to LA ecosystems and platforms where everyone can contribute and benefit.

An open learning analytics ecosystem encompasses different stakeholders associated through a common interest in LA but with diverse needs and objectives, a wide range of data coming from various learning environments and contexts, as well as multiple infrastructures and methods that enable to draw value from data in order to gain insight into learning processes.

In the following sections, we provide our vision for an open learning analytics platform through a detailed discussion of possible user scenarios, requirements,

technical architecture, and components. Our goal is to form the technical foundation of an ecosystem for open learning analytics.

4.1 User Scenarios

This section presents three possible user scenarios that the open learning analytics platform will support.

4.1.1 Teacher Scenario

Rima is a lecturer at ABC University where she uses the university LMS to administer her courses. She uses personalized dashboard of the open learning analytics platform which gives her an overview of her courses using various indicators to augment and improve her teaching process. On the dashboard she has various predefined indicators such as participation rate of students in lecture, students involvement rate in discussion forum, most viewed/downloaded documents, and the progress of her students in assignments.

Recently, Rima came up with the requirement to see which learning materials are more discussed in discussion forums. She looked in the list of available indicators but did not find any indicator which can fulfill this requirement. She opened the indicator editor which helps her in generating the new indicator and defining the appropriate visualization for this indicator. The newly generated indicator is also added to the list of available indicators for future use by other users.

4.1.2 Student Scenario

Amir is a computer science student at ABC University. He is interested in web technologies. He uses the open learning analytics platform to collect data from his learning activities related to this subject on the university LMS, the edX MOOC platform, Khan Academy, his blog, Facebook, YouTube, Slideshare, and various discussion forums.

What Amir likes most about the open learning analytics platform is that it provides him the possibility to select which learning activities from which application can be collected in his profile. For Amir privacy is one of the big concerns. By default all the logged activity data are only available to him. He has, however, the option to specify which data will be publicly available to whom and for how long.

Amir is interested in monitoring his performance across the different platforms. He uses the indicator editor to generate a new indicator which aggregates marks from the university LMS, the peer-review feedback from the edX MOOC platform, and open badges from Kahn Academy. He specifies to visualize his marks compared to his peers as a line chart, his peer-review feedback in a textual format, and

his badges as a list view. The platform then generates the visualization code that Amir can embed in the assessment module of the university LMS. Further, Amir is interested in getting recommendations related to web technologies in form of lecture slides, videos, online articles, blog posts, and discussion forums. He generates a new indicator which recommends him learning resources from different sources. He then embeds the generated indicator in the learning materials module of the university LMS.

4.1.3 Developer Scenario

Hassan is a researcher at ABC University. He developed a mobile application for collaborative annotation of lecture videos. He is interested in using the open learning analytics platform to analyze the social interactions of the applications users. Based on the data model specification and guidelines provided by the open learning analytics platform, he develops a new collector to collect activity data from his mobile application and send it to the platform. Further, he uses the indicator editor to define a new indicator which should apply the Gephi social network analysis method on the collected data. Unfortunately, this method is not available in the platform yet. Therefore, he uses the platform API to register Gephi as a new analysis method. Hassan goes back to the indicator editor and selects the newly registered analysis method to be applied in his indicator.

4.2 Requirements

Open learning analytics is a highly challenging task. It introduces a set of requirements and implications for LA practitioners, developers, and researchers. In this section, we outline possible requirements which would build the foundation for an open learning analytics platform.

4.2.1 Data Aggregation and Integration

As pointed out in the "what?" dimension of the LA reference model in Section 2, educational data is distributed across space, time, and media. A key requirement here is to aggregate and integrate raw data from multiple, heterogeneous sources, often available in different formats, to create a useful educational data set that reflects the distributed activities of the learner; thus leading to more precise and solid LA results.

4.2.2 Interoperability

The heterogeneity of data must be reduced to increase interoperability. Interoperability addresses the challenge of efficiently and reliably moving data between systems (Cooper, 2014b). A widely used definition of interoperability is the "ability of two or more systems or components to exchange information and to use the information that has been exchanged" (Benson, 2012, p. 21; Cooper, 2013). Interoperability benefits include efficiency and timeliness, independence, adaptability, innovation and market growth, durability of data, aggregation, and sharing (Cooper, 2014b). Interoperability is needed to do comparable analyzes (Daniel and Butson, 2014) and test for broader generalizations, for instance, whether a predictive model is still reliable when used in a different context (Romero and Ventura, 2013).

4.2.3 Specifications and Standards

It is important to adopt widely accepted specifications and standards in order to achieve interoperability of datasets and services. LA has stimulated standardization activities in different consortia, organizations, bodies, and groups, resulting in a number of specifications and standards that could be adopted or adapted for LA (Hoel, 2014). There are numerous existing specifications and standards that contribute elements of interoperability (Cooper, 2014b). Cooper (2014c) provides a technical-level summary of the range of existing work, which may be relevant to LA system developers. The summary lists specifications and standards related to data exchange (e.g. ARFF, CSV, GraphML), models and methods (e.g. PMLL), logging (e.g. Activity Streams, CAM, xAPI), assessment (e.g. IMS QTI, Open Badges), privacy (e.g. UMA).

As stated by Cooper (2014c) and Hoel (2014), there is no organized attempt to undertake pre-standardization work in the open learning analytics domain yet. Currently, there is only preliminary work to raise awareness of existing technical specifications and standards that may be of relevance to implementations of open learning analytics. None of the available specifications are fit for use as they stand. It is expected that the focus of activity in the near future is likely to be sharing experiences in using various candidate specifications and standards, and tentatively moving towards a set of preferred specifications and standards to be used in open learning analytics practices.

4.2.4 Reusability

It is necessary to follow the four Rs "Reuse, Redistribute, Revise, Remix" in the conceptualization and development of open learning analytics architectures. Adopting agreed upon specifications and standards would promote the reuse of data, services, and methods which is of vital practical importance in open learning analytics.

4.2.5 Modularity

An open learning analytics model requires new architectures that make it easy to accommodate new components developed by different collaborators in order to respond to changes over time. A modular and service-oriented approach enables a faster, cheaper, and less disruptive adaptability of the open learning analytics architecture. This is particularly relevant for LA where the methods are not yet mature (Cooper, 2014b).

4.2.6 Flexibility and Extensibility

Daniel and Butson (2014) note that the best platforms harnessing the power of big data are flexible. "They also blend the right technologies, tools, and features to turn data compilation into data insight" (p. 41). Thus, an open learning analytics platform should be fully flexible and extensible by enabling a smooth plug in of new modules, methods, and data after the platform has been deployed.

4.2.7 Performance and Scalability

Performance and scalability should be taken into consideration in order to allow for incremental extension of data volume and analytics functionality. This is a technical requirement which can be achieved by leveraging big data solutions which provide powerful platforms, techniques, and tools used for collecting, storing, distributing, managing, and analyzing large datasets with diverse structures such as Apache Hadoop, MapReduce, NoSQL databases, and Tableau Software (Daniel and Butson, 2014).

4.2.8 Usability

For the development of usable and useful LA tools, guidelines and design patterns should be taken into account. Appropriate visualizations could make a significant contribution to understanding the large amounts of educational data. Statistical, filtering, and mining tools should be designed in a way that can help learners, teachers, and institutions to achieve their analytics objectives without the need for having an extensive knowledge of the techniques underlying these tools. In particular, educational data mining tools should be designed for non-specialists in data mining (Romero and Ventura, 2010).

4.2.9 Privacy

It is crucial to build ethics and privacy into the LA solutions right from the very beginning. As Larry Johnson, CEO of the New Media Consortium (NMC) puts it "Everybodys talking about Big Data and Learning Analytics, but if you dont solve privacy first it is going to be killed before it has really started" (as cited in Bomas, 2014).

4.2.10 Transparency

Data and interpretations in LA might be used in other than the intended ways. For instance, learners might fear that personal data will not be used for constructive feedback but for monitoring and grading. This could lead to the unintended effect that learners are not motivated to use LA tools and participate in analytics-based TEL scenarios. Transparency is vital to drive forward the acceptance of LA. It provides an explicit definition of means how to achieve legitimacy in the process of learning analytics. It should be applied across the complete process, without exceptions. This means that at all times, there should be easily accessible and detailed documentation of how is the data collected, who has access to the data, which analytics methods are applied to the data, how long is the data valid and available, the purposes for which the data will be used, under which conditions, and which measures are undertaken to preserve and protect the identity of the learner (Bomas, 2014, Chatti et al., 2014, Pardo and Siemens, 2014; Sclater, 2014; Slade and Prinsloo, 2013). Further, it is important to increase institutional transparency by clearly demonstrating the changes and the added-value that LA can help to achieve (Daniel and Butson, 2014; Dringus, 2012).

4.2.11 Personalization

It is important to follow a personalized and goal-oriented LA model that tailors the LA task to the needs and goals of multiple stakeholders. There is a need to adopt a user-in-the-loop LA approach that engages end users in a continuous inquiry-based LA process, by supporting them in setting goals, posing questions, interacting with the platform, and self-defining the indicators that help them achieve their goals.

4.3 Conceptual Approach

In the following sections, we discuss in detail the building blocks of an open learning analytics platform, as depicted in Figure 2.
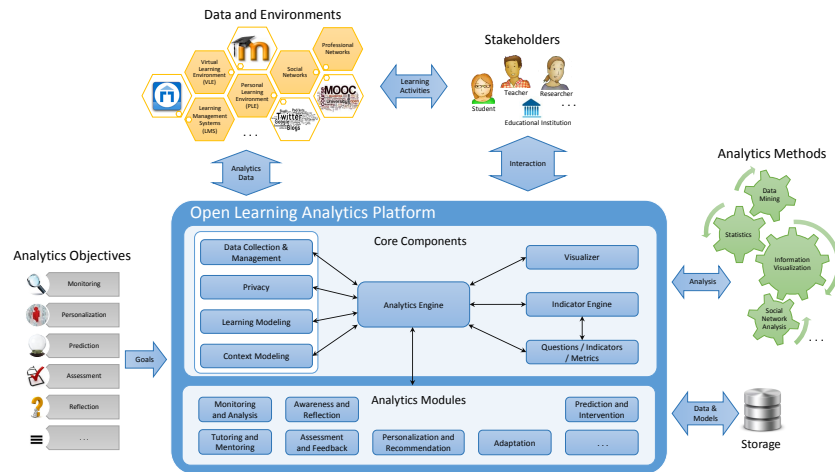
Fig. 2 Open learning analytics platform abstract architecture

4.3.1 Data Collection and Management

LA is focused on how to exploit "big data" to improve education (Siemens and Baker, 2012). The possibilities of big data continue to evolve rapidly, driven by innovation in the underlying technologies, platforms, and analytic capabilities. The McKinsey research report defines big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al, 2011). Gartner analyst Doug Laney uses the 3Vs model for describing big data; i.e. increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources) (Laney, 2001). Gartner defines big data as "high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" (Laney, 2012). Generally, the literature presents a number of fundamental characteristics associated with the notion of big data including in addition to volume, velocity, variety veracity (biases, noise and abnormality in data generated from various sources and questions of trust and uncertainty associated with the collection, processing and utilization of data), verification (data corroboration and security), and value (ability of data in generating useful insights and benefits) (Daniel and Butson, 2014).

Following these characteristics, data from learning processes can be characterized as big data:

– Volume - A single online learning platform can generate thousands of transactions per student.
– Velocity - The data that is collected should be processed and analyzed in real time to e.g. provide accurate and timely feedback.

– Variety - The data that needs to be analyzed comes from a variety of sources, such as LMS log files, assessment scores, social web etc.
– Veracity and Verification - Quality of data, privacy, and security issues need to be resolved in order to build trust and achieve legitimacy in the LA process.
– Value - The main aim of LA is to harness the educational data to provide insight into the learning processes.

LA is a data-driven approach. The first step in any LA effort is to collect data from various learning environments. Gathering and integrating this raw data are nontrivial tasks and require adequate data collection and management tasks (Romero and Ventura, 2013). These tasks are critical to the successful discovery of useful patterns from the data. The collected data is heterogeneous, with different formats (e.g. structured, semi-structured, unstructured documents, videos, images, HTML pages, relational databases, object repositories) and granularity levels, and may involve many irrelevant attributes, which call for data pre-processing (also referred to as data preparation) (Liu, 2006). Data pre-processing mainly allows converting the data into an appropriate format that can be used as input for a particular LA method. Several data pre-processing tasks, borrowed from the data mining field, can be used in this step. These include data cleaning, data integration, data transformation, data reduction, data modeling, user and session identification, and path completion (Han and Kamber, 2006, Liu, 2006; Romero & Ventura, 2007). After the data collection and pre-processing steps, it is necessary to carry out data integration at the appropriate level to create a complete dataset that reflects the distributed activities of the learner.

To deal with the interoperability and integration issues, the open learning analytics platform should adopt a standardized data model. Candidate data models for open learning analytics are discussed in Section 4.3.4. Moreover, the platform should provide an API that can be used by different collectors. A collector can be a component in a learning environment which gathers data and push it to the platform in the right format. It can also be an adapter as an intermediate component that enables to get data from a learning environment, map the data from the source format into the format expected by the API, and transform it into the data model used in the open learning analytics platform. In the data collection and management step, privacy issues have to be taken into consideration.

4.3.2 Privacy

Privacy is a big challenge in LA. This challenge is further amplified in open learning analytics practices where learner data is collected from various sources. Therefore, it is crucial to investigate mechanisms that can help develop LA solutions where ethical and privacy issues are considered. Interesting research is being done in order to understand and tackle the ethical and privacy issues that arise with practical use of LA in the student context. Pardo and Siemens (2014), for instance, provide four practical principles that researchers should consider about privacy when working on an LA tool. These practical principles are (1) transparency, (2) student control

over the data, (3) security, and (4) accountability and assessment. Slade and Prinsloo (2013) propose an ethical framework with six guiding principles for privacy-aware LA implementations. These include (1) learning analytics as moral practice, (2) students as agents, (3) student identity and performance are temporal dynamic constructs, (4) student success is a complex and multidimensional phenomenon, (5) transparency, and (6) higher education cannot afford to not use data. Some of the guiding principles are overlapping, or the same as the principles suggested by Pardo and Siemens (2014). Privacy by Design is another framework developed by Ann Cavoukian in the 90s to ensure privacy and gain personal control over ones information based on seven foundational principles, namely (1) proactive not reactive; preventative not remedial, (2) privacy as the default setting, (3) privacy embedded into design, (4) full functionality positive-sum, not zero-sum, (5) end-to-end security full lifecycle protection, (6) visibility and transparency keep it open, and (7) respect for user privacy keep it user-centric (Cavoukian, 2009). It is crucial to embrace all these principles while modeling a learner and her context, as discussed in the next two sections.

4.3.3 Learner Modeling

Learner modeling is the cornerstone of personalized learning. The capacity to build a detailed picture of the learner across a broader learning context beyond the classroom would provide a more personalized learning experience. The challenge is to create a thorough learner model that can be used to trigger effective personalization, adaptation, intervention, feedback, or recommendation actions. This is a highly challenging task since learner activities are often distributed over networked learning environments (Chatti et al., 2014).

A big challenge to tackle here is lifelong learner modeling. Kay and Kummerfeld (2011) define a lifelong learner model as a store for the collection of learning data about an individual learner. The authors note that to be useful, a lifelong learner model should be able to hold many forms of learning data from diverse sources and to make that information available in a suitable form to support learning. Lifelong learner modeling is the continuous collection of personal data related to a learner. It is an ongoing process of creating and modifying a model of a learner, who tends to acquire new or modify his existing knowledge, skills, or preferences continuously over a longer time span. The lifelong learning modeling process may evolve by different means e.g. by education, experience, training, or personal development. The authors further identify different roles for lifelong learner modeling. These roles bring several technical challenges and present a theoretical backbone of a general lifelong learner modeling framework. Driven by these roles, main tasks of the learning modeling module in the open learning analytics platform include:

– Collecting and aggregating learner data from different sources.
– Integrating and managing different parts of a learner model taking into consideration the semantic information.

– Providing interfaces for open learner modeling. Learners should be the ones who own the data they generate. They should have right to control, access, amend, and delete their data. This is important to build trust and confidence in the LA system.
– Sharing the learner model across applications and domains. Thereby, the learner must be able to control which parts of the model can be shared. This helps in making the LA practice more transparent.
– Promoting the reuse of the learner model by different applications by using standard data formats.

In order to achieve these tasks, several issues have to be taken into account, including questions about integration, interoperability, reusability, extensibility, and privacy. Integration and interoperability can be supported by specifications and standards. Reusability and extensibility can be achieved through open APIs that can be used by different external applications. We should always keep in hindsight the ethical and privacy challenges in the learning modeling task. This can be achieved by following the privacy principles as discussed in Section 4.3.2. Moreover, there is a need to implement mechanisms to guarantee that no unauthorized access to a learners data model is possible and that the learner has full control over the data. Technically, this can be achieved by following specifications such as the open User Managed Access (UMA) profile of OAuth 2.0 (Hardjono, 2015). Furthermore, we can have a user interface in the open learner modeling module that enables learners to see what kind of data is being used for which purpose. Furthermore, we need to define access scopes at different granular levels to let the learner decide which data should be taken into account, which applications can collect which data, as well as which data will be publicly available to whom and for how long.

4.3.4 Context Modeling

The six most popular and useful features in (lifelong) learner modeling include the learners knowledge, interests, goals, background, individual traits, and context (Brusilovsky & Millan, 2007). Context is a central topic of research in the area of learner modeling. It is important to leverage the context attribute in the learner model in order to give learners the support they need when, how, and where they need it. Harnessing context in a learning experience has a wide range of benefits including personalization, adaptation, intelligent feedback, and recommendation. A big challenge to tackle here is context capturing and modeling. A context model should reflect a complete picture of the learners context information. The aim is that activity data gathered from different learning channels would be fed into a personal context model, which would build the base for context-aware LA solutions.

A key question here is how to model the relevant data (Duval, 2011). Different specifications for context modeling have been introduced in the LA literature. Lukarov et al. (2014) and Thüs et al. (in review) provide a systematic analysis of what is currently available in this area. They compare and contrast different candidate data models for LA, namely Contextualized Attention Metadata (CAM), Learn-

ing Registry Paradata, NSDL Paradata, IMS Caliper, Activity Streams, and the Experience API (xAPI), based on eight factors which define the general quality of a data model. These factors include correctness, completeness, integrity, simplicity, flexibility, integration, understandability, and implementability (Moody, 2003). The authors note that the studied data models are not user-centered, which is required to support personalized learning experiences. Moreover, they do not preserve the semantic meaning of the stored events (e.g. the verb-ambiguity problem in xAPI), which could lead to misinterpretations and inaccurate LA results. The authors point out that the ideal data model should find a balance between completeness, flexibility, and simplicity and introduce the Learning Context Data Model (LCDM) specification as a modular, simple and easy to understand data model that holds additional semantic information about the context in which an event has been generated. LCDM can be extended by e.g. interests of a learner, thus providing the base for a lifelong learner modeling specification. LCDM further provides a RESTful API that enables the extensibility and reusability of context models. The API encapsulates the complexity of sending context data to be sent to the server in the right format. Currently, there are libraries for the languages Java, PHP, Objective-C and JavaScript. Most important, LCDM provides mechanisms to deal with the privacy issue through OAuth authorization and data access scopes defining what happens with the data and who may have access to it.

4.3.5 Analytics Modules

Each of the analytics modules corresponds to an analytics goal such as monitoring, personalization, prediction, assessment, reflection. They represent components which can easily be added and removed from the open learning analytics platform by the analytics engine. Each analytics module is responsible for managing a list of analytics methods associated with it. Moreover, each module manages a list of user-defined indicators which are generated by the indicator generator in the form of a triad containing a reference to the indicator specification in the questions / indicators / metrics component, the associated analytics method, and the visualization technique to be used for that indicator.

4.3.6 Questions / Indicators / Metrics

The questions / indicators / metrics component is responsible for the management of questions indicators defined by different stakeholders in the open learning analytics platform. Each question is associated with a set of indicators. For each indicator the component stores related queries which are generated in the indicator generation phase. These queries will be used by the analytics engine to fetch the data to be analyzed.

4.3.7 Indicator Engine

The indicator engine is a central component in the open learning analytics platform which enables personalized and goal-oriented LA. The various objectives in LA (e.g. monitoring, analysis, prediction, intervention, tutoring, mentoring, assessment, feedback, adaptation, personalization, recommendation, awareness, reflection) need a tailored set of indicators and metrics to serve different stakeholders with very diverse questions and goals. Current implementations of LA rely on a predefined set of indicators and metrics. This is, however, not helpful in the case of open learning analytics where the set of required indicators is unpredictable. This raises questions about how to achieve personalized and goal-oriented LA in an efficient and effective way. Ideally, LA tools should support an interactive, exploratory, and real-time user experience that enables a flexible data exploration and visualization manipulation based on individual goals of users. The challenge is thus to define the right Goal / Question / Indicator (GQI) triple before starting the LA exercise. Following an inquiry-based LA approach by giving users the opportunity to interact with the platform, define their goal, pose questions, explore the data, and specify the indicator/metric to be applied is a crucial step for effective and personalized LA results. This would also make the LA process more transparent, enabling users to see what kind of data is being used and for which purpose.

The Indicator engine is responsible for the management of the GQI definition process. It can be subdivided into the following four main subcomponents.

Question / Indicator Editor: This component provides a user-friendly interactive interface to set the LA goal, formulate questions, and define indicators associated with those questions. The process starts with a user setting a goal (e.g. monitoring and analysis, awareness and reflection, personalization and recommendation) and formulating the questions which she is interested in. A question can be "How active are my students?". While user is formulating the question, the editor will communicate with the question analyzer component to provide useful suggestions for related questions. The next step is to associate the question with a set of indicators. In our example, possible indicators can be "number of posts in discussion forums", "update rate of wiki pages", "frequency of annotations on lecture videos". Existing indicators can be reused and new indicators can be defined with the help of the indicator generator component. To define a new indicator, the question / indicator editor can be used to specify indicator data objects, choose the analytics method to process the indicator, and select the appropriate visualization technique to render the indicator data.

Question Analyzer: The task of the question analyzer component is to analyze the question as the user is entering it and provide useful suggestions for similar questions. Thereby, information retrieval, term extraction, and Natural Language Processing (NLP) algorithms can be used to infer the list of closely related questions from the questions / indicators / metrics component.

Indicator Generator: This component is responsible for the generation of new indicators. To define a new indicator, the indicator generator communicates with the rule engine component to obtain the list of possible indicator rules and to the ana-

lytics engine to get possible data objects from the storage based on the data model schema used in the open learning analytics platform. Taking the example of the indicator "number of posts in discussion forums", the user first selects the indicator rule "number of X in Y" then assigns the data object "discussion forum" to Y and the data object "post" to X from the list of possible data objects. The indicator generator further communicates with the rule engine to generate the query related to the indicator based on the selected rule and data objects. In our example, in SQL terms, the query "SELECT COUNT(post) FROM table_discussionforum;" will be associated with the indicator "number of posts in discussion forums". After defining the indicator, and based on the LA goal set by the user in the question / indicator editor, the indicator generator communicates with the respective analytics module via the analytics engine to get the list of possible analytics methods. The user can then select the analytics method to be applied on the indicator data. The indicator generator communicates with the visualizer via the analytics engine to get the list of possible visualization techniques that can be applied. After the selection of an appropriate visualization technique by the user, the indicator is processed by the analytics engine. The user can then approve the indicator which is then registered as new indicator in the questions / indicators / metrics component along with the associated query. Moreover, a triad containing the reference to this indicator, the associated analytics method, and the selected visualization technique will be stored in the respective module via the analytics engine. The indicator generator further generates the indicator data request code which can be copied and embedded in the client application (e.g. dashboard, HTML page, widget) to get the indicator visualization code to be rendered on the client.

Rule Engine: This component is responsible for managing indicator rules and their associated queries. Different rule engines can be used to support this task such as Drools, Mandarax, JRuleEngine, InRule.

4.3.8 Analytics Engine

The analytics engine is the backbone of the open learning analytics platform which acts as a mediator between different components in the platform. The major task of the analytics engine is to perform analysis. It is responsible for executing the indicator queries to fetch raw data from the database. The data is transformed into the format acceptable by the selected analytics method before it is analyzed. The results of the analysis are then transformed into a standard format and sent to the visualizer.

The analytics engine is also responsible for handling the indicator requests from external systems. For each request, the analytics engine gets the triad from the respective module. It then gets the related query from the questions / indicators / metrics component, execute the indicator query, apply the specified analytics method, and send the indicator data to the visualizer.

Since the analytics engine is responsible for performing all kind of analysis tasks, its design should support multithreading and multicore processing architecture so that it does not become the bottleneck for the entire platform.

Moreover, the analytics engine is responsible for providing the extensibility of the platform by supporting easy mechanisms to manage, add, and remove analytics modules from the platform as well as managing the repository of analytics methods which can grow as new methods are implemented.

4.3.9 Visualizer

A key step in LA is closing the loop by feeding back the analytics results to learners (Clow, 2012). This requires appropriate representations of the results. Statistics in form of reports and tables of data are not always easy to interpret to the end users. Visualization techniques are very useful for showing results in a way that is easier to interpret (Romero and Ventura, 2013). Mazza (2009) stresses that thanks to our visual perception ability, a visual representation is often more effective than plain text or data. Different information visualization techniques (e.g. charts, scatterplot, 3D representations, maps) can be used to represent the information in a clear and understandable format (Romero and Ventura, 2007). The difficult part here is in defining the representation that effectively achieves the analytics objective (Mazza, 2009).

Recognizing the power of visual representations, traditional reports based on tables of data are increasingly being replaced with dashboards that graphically show different performance indicators. Dashboards "typically capture and visualize traces of learning activities, in order to promote awareness, reflection and sense-making, and to enable learners to define goals and track progress towards these goals" (Verbert et al., 2014, p. 1499). Dashboards represent a helpful medium for visual analytics widely used in the LA literature. They are, however, often not linked to the learning context and they provide more information than needed. LA is most effective when it is an integral part of the learning environment. Hence, integration of LA into the learning practice of the different stakeholders is important. Moreover, effective LA tools are those, which minimize the time frame between analysis and action, by delivering meaningful information in context and without delay, so that stakeholders have the opportunity to act on newly gained information in time. Thus, it is beneficial to view learning and analytics as intertwined processes and follow an embedded LA approach by developing visual analytics tools that (a) are smoothly integrated into the standard toolsets of learners and teachers and (b) foster prompt action in context by giving useful feedback at the right place and time.

The visualizer component in the open learning analytics platform is responsible for providing easy mechanisms to manage, add, and remove visualization techniques such as Google Charts, D3/D4, jpGraph, Dygraphs, jqPlot along with the type of visualization (e.g. bar chart, pie chart, line chart, etc.) supported by each technique. An adapter is required for each visualization technique to transform the
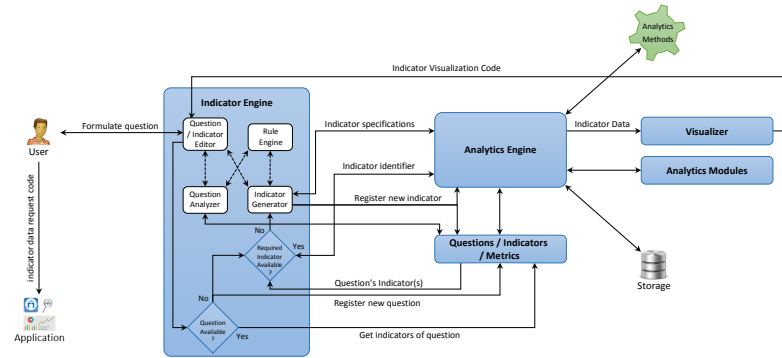
Fig. 3 New indicator generation flow diagram

data format used in the analytics engine to the indicator visualization code to be rendered on the client application (e.g. dashboard, HTML page, widget).

4.4 System Scenarios

In this section, we outline two possible system scenarios to show how the different components of the open learning analytics platform interact with each other.

4.4.1 New Indicator Generation

The new indicator generation process is depicted in Figure 3. The user starts the process by selecting her goal and entering her question using the question / indicator editor. The question analyzer communicates with the question / indicators / metrics component to suggest closely related questions. The user can either select one of the suggested questions or continue to enter a new question. If the user selects one of the suggested questions, the question / indicator editor presents her with all the indicators associated with that question. If the user enters a new question, all available indicators are presented to her from which she can select which indicators to associate with the new question or generate a new indicator. If the user selects one of the available indicators, the analytics engine suggests existing instances of that indicator (i.e. related triads in the respective analytics module). The user can then select one of the instances or associate the indicator with a different analytics method and/or visualization technique using the indicator generator. The user is presented with a different interface in the question / indicator editor where she can define a new indicator using the indicator generator, as discussed in Section 4.3.7. The analytics engine processes the indicator (see Section 4.3.8) and sends the indicator data to the visualizer which generates the indicator visualization code to be
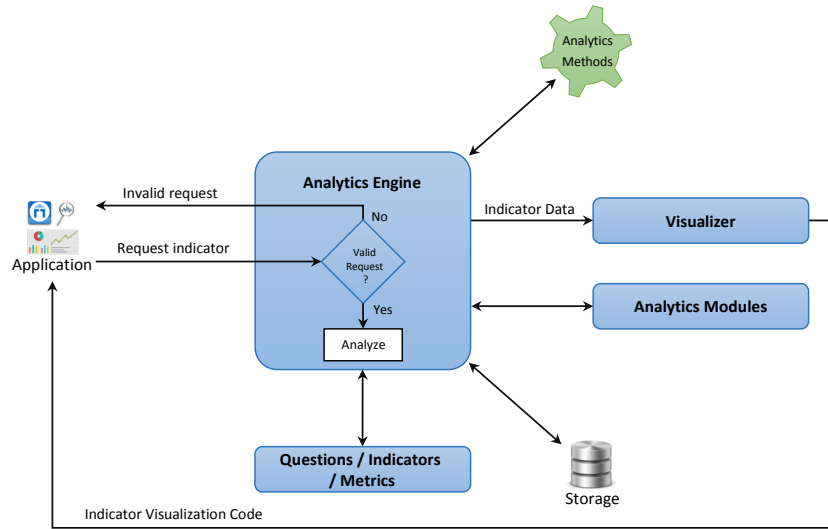
Fig. 4 Indicator data request flow diagram

rendered on the question / indicator editor (see Section 4.3.9). If the user is satisfied with the new indicator, she can copy the indicator data request code generated by the indicator generator and embed it in any client application.

4.4.2 Indicator Data Request

The indicator data request flow is shown in Figure 4. To visualize the indicator on e.g. a dashboard, an indicator data request containing the module identifier, triad identifier (see Section 4.3.5), and additional parameters (e.g. filters) is sent to the open learning analytics platform. The analytics engine intercepts the request and performs the following steps:

1. Check whether the request is valid or not.
2. Communicate with the respective analytics module to get the indicator reference, the associated analytics method, and the visualization technique to be used for that indicator.
3. Communicate with the questions / indicators / metrics component to get the query related to the requested indicator.
4. Execute the query and get the data.
5. Analyze the data using the associated analytics method.
6. Transform the method output data to the data format used in the analytics engine.
7. Send the indicator data to the visualizer.

The visualizer transforms the indicator data to the visualization code to be rendered on the client application (e.g. dashboard, HTML page, widget).

5 Conclusion

In the last few years, there has been an increasing interest in the automatic analysis of educational data to enhance the learning experience, a research area referred to as Learning Analytics (LA). Significant research has been conducted in LA. However, most of the LA approaches to date are focusing on centralized learning settings. Driven by the demands of the new networked and increasingly complex learning environments, there is a need to scale LA up which requires a shift from closed LA systems to open LA ecosystems. In this paper, we discussed open learning analytics as an emerging research field that has the potential to improve learning efficiency and effectiveness in open and networked learning environments. We further presented a vision for an open learning analytics ecosystem through a detailed discussion of user scenarios, requirements, technical architecture, and components of an open learning analytics platform. This paper makes a significant contribution to LA research because it provides concrete conceptual and technical ideas toward an open learning analytics ecosystem, which have been lacking until now.

References

1. Alexander S. et al. (2014) OLA Press Release. Retrieved from, http://solaresearch.org/initiatives/ola/
2. Benson, T. (2012). Why interoperability is hard. In Principles of Health Interoperability HL7 and SNOMED, Health Information Technology Standards, (pp. 21-32). Springer London.
3. Bomas, E. (2014) How to give students control of their data. Retrieved from http://www.laceproject.eu/blog/give-students-control-data/
4. Brusilovsky, P. and Millan, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educationsl Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), The Adaptive Web, LNCS 4321, pp. 3-53. Berlin Heidelberg: Springer.
5. Cavoukian, A. (2009) Privacy by Design The 7 Foundational Principles. Retrieved from https://www.privacybydesign.ca/content/uploads/2009/08/7foundationalprinciples.pdf
6. Chatti, M. A. (2010) The LaaN Theory. In: Personalization in Technology Enhanced Learning: A Social Software Perspective. Aachen, Germany: Shaker Verlag, pp. 19-42.
7. Chatti, M.A., Dyckhoff, A.L., Thüs, H. & Schroeder, U. (2012): A reference model for learning analytics. International Journal of Technology Enhanced Learning (IJTEL). 4(5/6). 318-331.
8. Chatti, M. A., Lukarov, V., Thüs, H., Muslim, A., Yousef, A. M. F., Wahid, U., Greven, C., Chakrabarti, A., Schroeder, U. (2014). Learning Analytics: Challenges and Future Research Directions. eleed, Iss. 10. (urn:nbn:de:0009-5-40350)
9. Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 134-138). ACM.
10. Clow, D. (2013). An overview of learning analytics. Teaching in Higher Education, 18(6), pp. 683-695.
11. Cooper, A. (2013). Learning Analytics Interoperability - a survey of current literature and candidate standards. Retrieved from http://blogs.cetis.ac.uk/adam/wp-content/uploads/sites/23/2013/05/learning-analytics-interoperability-v1p1.pdf
12. Cooper, A. (2014a) Open Learning Analytics Network - Summit Europe 2014. Retrieved from http://www.laceproject.eu/open-learning-analytics-network-summit-europe-2014/

13. Cooper, A. (2014b) Learning Analytics and Interoperability - The Big Picture in Brief, Learning Analytics Review, March 2014, ISSN: 2057-7494.
14. Cooper, A. (2014c) Standards and Specifications - Quick Reference Guide. Retrieved from http://www.laceproject.eu/dpc/standards-specifications-quick-reference-guide/
15. Daniel, B., & Butson, R. (2014). Foundations of Big Data and Analytics in Higher Education. In International Conference on Analytics Driven Solutions: ICAS2014 (pp. 39-47). Academic Conferences Limited.
16. Dawson, S., Gašević, D., Siemens, G. & Joksimovic, S (2014). Current state and future trends: a citation network analysis of the learning analytics field. In: Proceedings of the Fourth International Conference on Learning Analytics & Knowledge (pp. 231-240). NY, USA: ACM New York.
17. Downes, S. (2007) Models for Sustainable Open Educational Resources. Interdisciplinary Journal of Knowledge and Learning Objects, 3, pp. 29-44
18. Dringus, L. P (2012): Learning Analytics Considered Harmful. Journal of Asynchronous Learning Networks, 16(3) pp. 87-100.
19. Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 9-17). ACM.
20. Ferguson, R. (2012) Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6) pp. 304-317.
21. Fournier, H., Kop, R. & Sitlia, H. (2011). The value of learning analytics to networked learning on a personal learning environment. Proceedings of the LAK '11 Conference on Learning Analytics and Knowledge, pp. 104-109.
22. Fry, J., Schroeder, R., & den Besten, M. (2009). Open science in e-science: contingency or policy?. Journal of Documentation, 65(1), 6-32.
23. Han, J. & Kamber, M. (2006). Data Mining: Concepts and Techniques. San Francisco, CA: Elsevier.
24. Hardjono, T. (Ed) (2015) User Managed Access (UMA) Profile of OAuth 2.0. Retrieved from http://docs.kantarainitiative.org/uma/draft-uma-core.html
25. Hilton, J., Wiley, D., Stein, J., & Johnson, A. (2010): The Four R?s of Openness and ALMS Analysis: Frameworks for Open Educational Resources. Open Learning: The Journal of Open and Distance Learning, 25(1), pp. 37-44.
26. Hoel, T. (2014) Standards and Learning Analytics - current activities 2014. Retrieved from http://www.laceproject.eu/blog/standards-learning-analytics-current-activity-2014/
27. Kay, J. & Kummerfeld, B. (2011) Lifelong learner modeling. In Durlach, P.J., Lesgold, A.M., eds.: Adaptive Technologies for Training and Education. Cambridge University Press. 140-164.
28. Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J. (2010) A Data Repository for the EDM community: The PSLC DataShop. In Romero, C.,Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press, pp.43-56.
29. Laney, D. (2012). The Importance of 'Big Data': A Definition. Gartner. Retrieved 6 March 2015 from http://www.gartner.com/resId=2057415
30. Laney, D. (2001): 3D Data Management: Controlling Data Volume, Velocity, and Variety, Application Delivery Strategies published by META Group Inc., 2001, Retrieved 6 March 2015 from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
31. Liu, B. (2006). Web Data Mining. Berlin Heidelberg: Springer.
32. Lukarov, V., Chatti, M. A., Thüs, H., Salehian Kia, F., Muslim, A., Greven, C., and Schroeder, U., 2014. Data Models in Learning Analytics. Proceedings of DeLFI Workshops 2014, pp. 88-95.
33. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A.H., 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.

34. McNamara, T. (2012): Open Education: Emergence and Identity. Retrieved February 5, 2015, from http://oh-institute.org/external_resources/pub/McNamara-OpenEd_Emergence_Identity-CC-by.pdf

35. Moody, D. L. (2003). Measuring the Quality of Data Models: An Empirical Evaluation of the Use of Quality Metrics in Practice. Proceedings of 11th European Conference on Information Systems.

36. OECD (2007): Giving Knowledge for Free: The Emergence of Open Educational Resources. Report. ISBN-978-92-64-03174-6. Retrieved February 5, 2015, from http://www.oecd.org/edu/ceri/38654317.pdf

37. Pardo, A.; Siemens, G.: Ethical and Privacy Principles for Learning Analytics. In: British Journal of Educational Technology, 45(3), 2014, pp. 438-450.

38. Peters, M. A. (2008). The History and Emergent Paradigm of Open Education. Open Education and Education for Openness (pp. 3-16). Rotterdam: Sense Publishers.

39. Romero, C., Ventura, S. (2007). Educational Data Mining: a Survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135-146.

40. Romero, C., Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12-27.

41. Sclater, N. (2014) Examining open learning analytics - report from the Lace Project meeting in Amsterdam. Retrieved from http://www.laceproject.eu/blog/examining-open-learning-analytics-report-lace-project-meeting-amsterdam/

42. Siemens, G., & d Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254). ACM.

43. Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., Duval, E., Verbert, K. & Baker, R. S. J. D. (2011). Open Learning Analytics: an integrated & modularized platform. Open University Press.

44. Slade, S.; Prinsloo, P.: Learning analytics: ethical issues and dilemmas. In: American Behavioral Scientist, 57(10), 2013, pp. 1509-1528.

45. H. Thüs, M. A. Chatti, U. Schroeder (in review) Context Capturing and Modeling in Open Learning Environments. International Journal of Artificial Intelligence in Education Society (IJAIED), IOS Press.

46. Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011, February). Dataset-driven research for improving recommender systems for learning. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 44-53). ACM.

47. Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-Driven Research to Support Learning and Knowledge Analytics. Educational Technology & Society, 15 (3), 133-148.

48. Wiley, D. (2009). Introduction to Open Education. iTunesU. Lecture conducted from BYU, Provo.

49. Yuan, L., Powell, S. (2013): MOOCs and Open Education: Implications for Higher Education. A white paper. Retrieved February 5, 2015 from http://publications.cetis.ac.uk/2013/667