Chapter 30: Linked Data for Learning Analytics: Potentials and Challenges

Amal Zouaq[1], Jelena Jovanović[2], Srećko Joksimović[3], Dragan Gašević[2,4]

[1] School of Electrical Engineering and Computer Science, University of Ottawa, Canada
[2] Department of Software Engineering, University of Belgrade, Serbia
[3] Moray House School of Education, University of Edinburgh, United Kingdom
[4] School of Informatics, University of Edinburgh, United Kingdom

DOI: 10.18608/hla17.030

ABSTRACT

Learning analytics (LA) is witnessing an explosion of data generation due to the multiplicity and diversity of learning environments, the emergence of scalable learning models such as massive open online courses (MOOCs), and the integration of social media platforms in the learning process. This diversity poses multiple challenges related to the interoperability of learning platforms, the integration of heterogeneous data from multiple knowledge sources, and the content analysis of learning resources and learning traces. This chapter discusses the use of linked data (LD) as a potential framework for data integration and analysis. It provides a literature review of LD initiatives in LA and educational data mining (EDM) and discusses some of the potentials and challenges related to the exploitation of LD in these fields.

Keywords: Linked data (LD), data integration, content analysis, educational data mining (EDM)

The emergence of massive open online courses (MOOCs) and the open data initiative have led to a change in the way educational opportunities are offered by shifting from a university-centric model to a multi-platform and multi-resource model. In fact, today's learning environments include not only diverse online learning platforms, but also social media applications (e.g., SlideShare, YouTube, Facebook, Twitter, or LinkedIn) where learners connect, communicate, and exchange data and resources. Henceforth, learning is now occurring in various forms and settings, both at the formal (university courses) and informal (social media, MOOC) levels. This has led to a dispersion of learner data across various platforms and tools, and brought a need for efficient means of connecting learner data across various environments for a comprehensive insight into the learning process. One salient example of the need for data exchange across platforms is the connectivist MOOC (cMOOC). In cMOOCs, learning, by definition, does not take place in a single platform, but relies on a range of dedicated online learning applications as well as social media and networking applications for sharing information and resources among learners (Siemens, 2005). These developments led to new requirements and imposed new challenges for both data collection and use.

From the perspective of data collection, the emergence of cloud services and the rapid development of scalable web architectures allow for pulling and mashing data from various online applications. This is supported by the development of large-scale interfaces (APIs) by major Web stakeholders such as Facebook, LinkedIn, or Twitter, and by MOOC providers such as Coursera and Udacity. From the perspective of data use, the plethora of resources and interactions occurring in educational platforms requires analytic capabilities, including the ability to handle different types of data. Various kinds of data are generated, some of which capture learners' interactions in learning and social media platforms (learners' logs/traces), whereas others take the form of unstructured content, ranging from course content and learners' blogs to discussion forum posts. This multitude of kinds and sources of data provides fertile ground for the field of learning analytics and its overall objectives to better understand learners and the learning process, provide timely, informative, and adaptive feedback, and foster lifelong learning (Gaešvić, Dawson, & Siemens, 2015).

Challenges associated with the collection, integration, and use of data originating from heterogeneous sources are often dealt with, in the educational community, by developing a standardized data model that allows for integration and leveraging of heterogeneous data (Dietze et al., 2013). This chapter focuses on linked data (LD) as one potential approach to the development and use of such a data model in both formal and informal online learning settings. In particular, the use of LD principles (Bizer, Heath, & Berners-Lee, 2009) allows for establishing a globally usable network of information across learning environments (d'Aquin, Adamou, & Dietze, 2013), leading to a global educational graph. Similar graphs could be created at the individual level, for each particular learner, connecting all the data and resources associated with their learning activities. The educational potentials and benefits of such graphs have already been examined and discussed. For instance, Heath and Bizer (2011) propose an *educational graph* across UK universities, comprising knowledge extracted from the content of learning resources. Given the development and use of knowledge graphs by an increasing number of major companies such as Google, Microsoft, and Facebook, the potential and possibilities opened up by such graphs for learning should be examined (Zablith, 2015).

This chapter describes the current state of the art of LD usage in education, focusing primarily on existing and potential applications in the learning analytics (LA)/educational data mining (EDM) field. After a brief introduction to LD principles in the next section, the chapter analyzes the potential of LD along two particular dimensions: 1) the data integration dimension and 2) the data analysis and interpretation dimension. Finally, we discuss some potentials and challenges associated with the use of LD in LA/EDM.

LINKED DATA IN EDUCATION

Linked data has the potential to become a de facto standard for sharing resources on the Web (Kessler, d'Aquin, & Dietze, 2013). It uses URIs to uniquely identify entities, and the RDF data model[1] to describe entities and connect them via links with explicitly defined semantics. In particular, LD relies on four principles:

1. Use URIs as names for things; for instance, historical novel "Paris" is uniquely identified by its ISBN (a kind of URI): 0385535309
2. Provide the ability to look up names through HTTP URIs; while an ISBN does uniquely identify a book, it cannot be used to provide direct access to it on the Web, so HTTP URIs should be used instead; the book from our example could be looked up via the following HTTP URI: <http://www.worldcat.org/oclc/827951628>
3. Upon URI look up, return useful information using the standards RDF and SPARQL[2]; for instance, we can state, in a machine-processable manner, that the resource identified by the <http://www.worldcat.org/oclc/827951628> URI is of the *type* book and belongs to the genre of historical fiction: *<http://www.worldcat.org/oclc/827951628> rdf:type schema:Book* ; *schema:genre "Historical fiction"*.
4. Include links to other entities uniquely identified by their URIs; for instance, we can connect the book from our example with its author: *<http://www.worldcat.org/oclc/827951628> schema:author <http://viaf.org/viaf/34666>* where the latter URI uniquely identifies the writer Edward Rutherfurd.

Thanks to the simplicity of these principles, LD represents an elegant framework for modelling and querying data at a global scale. It is usable in various applications and domains, and can constitute a response to the interoperability and data management challenges that have long faced the educational community (Dietze et al., 2013).

Billions of data items have been published on the web as linked data, forming a global open data space – the linked open data cloud (LOD)[3] – that includes open data from various domains such as government data, scientific knowledge, and data about a variety of online communities. Huge cross-domain knowledge bases have also emerged on the LOD such as DBpedia[4], Yago[5], and Wikidata[6]. As such, LD has the potential to enable a global shift in how data is accessed and utilized, offering access to data from various sources, through various kinds of data access points, including Web services and Web APIs, and allowing for seamless creation of dynamic data mashups (Bizer et al., 2009). In fact, one salient feature of LD is that it establishes semantic-rich connections between items from different data sources, and thus opens up data silos (e.g., traditional databases) for more seamless data integration and reuse.

Despite all these potential benefits, the LD formalism and technologies have had a slow adoption in the area of technology-enhanced learning; initiatives that employ LD technologies have only emerged recently (Dietze et al., 2013). We can identify several application scenarios

[1] Resource Description Framework, http://www.w3.org/RDF/

[2] https://www.w3.org/TR/sparql11-query/

[3] http://lod-cloud.net/

[4] http://lod-cloud.net/

[5] http://bit.ly/yago-naga

[6] http://bit.ly/wikidata-main

in the LA/EDM field that would benefit from the LOD, including 1) resource discovery (e.g., faceted search) and content enrichment (e.g., augmenting content with data from LOD datasets) (Maturana, Alvarado, López-Sola, Ibáñez, & Elósegui, 2013); 2) content analysis based on semantic annotation (Joksimovi et al., 2015); 3) resource and service integration (Dietze et al., 2012); 4) personalization (Dietze, Drachsler, & Giordano, 2014); and 5) interpretation of EDM results (d'Aquin et al., 2013).

DATA INTEGRATION USING LINKED DATA

One of the most salient benefits of LD lies in its data integration potential. This is particularly relevant for the LA/EDM field since it requires the collection and management of learner and content data from a variety of sources (applications and services) used in informal and life-long learning (Santos et al., 2015). In particular, to build a comprehensive learner model, one needs to integrate learner data recorded in different learning platforms/tools the learner has interacted with (Desmarais & Baker, 2012). Therefore, the challenges associated with handling multiple data formats and the overall lack of data interoperability, are becoming a key issue (Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Duval, 2011). More generally, the ease of data transfer, pre-processing, use, combination and analysis without loss of meaning across learning platforms are becoming important factors for the efficiency of LA/EDM (Cooper, 2013).

Several domains have been successful in exploiting LD for data integration issues such as the biomedical domain (Belleau, Nolin, Tourigny, Rigault, & Morissette, 2008), pharmacology (Groth et al., 2014), and environmental sciences (Lausch, Schmidt, & Tischendorf, 2015). All of this suggests that LD technologies could provide the solid data integration layer that LA/EDM necessitates.

Previous Initiatives in Data Integration in the Educational Community

The technology enhanced learning research community has long recognized the importance of data integration, which eventually resulted in multiple standardization efforts. Cooper (2013) provides a valuable overview of various standards related to learning. Mainly, these standards relate to the representation of data about learners and their activities, as well as learning content and services.

At the learner level, standards focus on facts about individuals and their history, their connections and interactions with other persons, and interactions with resources offered by learning environments (person and learning activities dimensions). Various specifications exist to model learners (e.g., FOAF[7]), and learner activities and interactions (e.g., Contextualized Attention Metadata [Schmitz, Wolpers, Kirschenmann, & Niemann, 2011], Activity Streams[8], or ADL xAPI[9]).

At the content level, previous initiatives such as IEEE Learning Object Metadata (LOM)[10] and ADL SCORM[11] attempted to create vocabularies and standards that would unify the description of online educational resources or the specification of computer-based assessment (e.g., IMS QTI[12]). Other efforts targeted the mapping between various data models, such as the work of Niemann, Wolpers, Stoitsis, Chinis, and Manouselis (2013) who aimed at aggregating sets of social and interaction data. Finally, several interfaces were proposed to provide guidelines for the implementation of services compliant with these standards (Dietze et al., 2013).

Based on different viewpoints, these efforts led to multiple competing projects and thus created sub-communities with various technologies, languages, and models, and very little interoperability among them. The LD philosophy provides a solution to these interoperability issues by allowing a multiplicity of models on the Web, bridging these models using Web-accessible semantic links. Thus semantically similar models that are differently represented can still be aligned using typed links that establish meaningful connections between concepts originating from different models; for instance, equality connections (*owl:sameAs*), or hierarchical connections (*rdfs:subclassOf or skos:broader*).

Current Data Integration Initiatives Using Linked Data

Integration based on LD requires the availability of Web-accessible LD vocabularies that describe the types of entities in specific subject domains, entities' attributes, and the kinds of connections among the entities. It also depends on the availability of services that allow for exploiting multiple datasets for a given task, as well as services that expose data as LD. This section introduces some of the available vocabularies in the educational domain, and efforts aimed at exposing educational data as LD. A more comprehensive overview of education-related vocabularies can be found in Dietze et al. (2014). The section also gives examples of services exploiting the integration of multiple LD datasets.

An increasing number of educational institutions have been exposing their data following LD principles, such as the Open University in the UK or the University

[7] http://www.foaf-project.org/
[8] http://activitystrea.ms/
[9] http://www.adlnet.gov/tla/experience-api
[10] http://ieeeltsc.org/wg12LOM/
[11] http://www.adlnet.org/
[12] http://www.imsglobal.org/question/

of Münster[13] in Germany. One prominent effort in exposing educational data as LD was the LinkedUp project[14], which resulted in a catalog of datasets related to education and encouraged the development of competitions such as the LAK Data Challenge[15], whose aim was to expose LA/EDM publications as LD and promote their analysis by researchers. While these initiatives represent a step in the adoption of LD by the educational community, their impact remains limited. For example, the data representation and use in MOOC platforms – one of the most striking developments in today's technology-enhanced learning – has not been based on LD principles or technologies to date. Still, few recent initiatives (Kagemann & Bansal, 2015; Piedra, Chicaiza, López, & Tovar, 2014) showed some interest in describing and comparing MOOCs using an LD approach. For example, MOOCLink (Kagemann & Bansal, 2015) aggregates open courseware as LD and exploits these data to retrieve courses around particular subjects and compare details of the courses' syllabi. Recently, there has also been an initiative that relies on schema.org[16] to create a vocabulary for course description[17] with the purpose of facilitating the discovery of any type of educational course. Schema.org is a structured data markup schema (or vocabulary) supported by major Web search engines. This schema is then used to annotate Web pages and facilitate the discovery of relevant information. Given its adoption by major players on the Web, this is a welcome initiative that might have some long-term impact in the educational community. Similarly, some authors worked on providing an RDF representation (binding) of educational standards. For example, an RDF binding of the Contextualised Attention Metadata (CAM) (Muñoz-Merino et al., 2010) and an RDF binding of the Atom Activity Streams[18] were developed. This enabled data integration and interoperability both at syntax and semantic levels.

Finally, with the current shift towards RESTful (representational state transfer) services on the cloud, education-related services based on LD have started to emerge. At a conceptual level, we can identify two main types of services based on LD currently being investigated in research: 1) services for course interlinking within a single institution and across institutions, and 2) services for integrating learners' log data based on a common model.

For example, Dietze et al. (2012) proposed an LD-based framework to integrate existing educational repositories at the service and data levels. Zablith (2015) suggested the use of LD as a conceptual layer around higher education programs to interlink courses in a granular and reusable manner. Another work links ESCO[19]-based skills to MOOC course descriptions to create enriched CVs (Zotou, Papantoniou, Kremer, Peristeras, & Tambouris, 2014). Interestingly, the authors are able to identify similar skills taught in the Coursera and Udacity MOOC platforms, thus providing implicit links between courses of two different MOOC platforms. One can envisage exciting opportunities for life-long learning based on a cross-platform MOOC course recommendation service.

Another indicator of the growing importance of LD in the realm of education in general, and LA/EDM in particular, is the adoption of LD concepts and technologies into xAPI specifications[20]. With xAPI, developers can create a learning experience tracking service through a predefined interface and a set of storage and retrieval rules. De Nies, Salliau, Verborgh, Mannens, and Van de Walle (2015) propose to expose data models created using the xAPI specification as LD. This proposal provides an interoperable model of learning traces data, and allows for seamless exposing of learners' traces as semantically interoperable LD. Similarly, Softic et al. (2014) report on the use of Semantic Web technologies (RDF, SPARQL) to model learner logs in personal learning environments.

Based on the scalability of the Web as the base infrastructure, and using the interoperability of the W3C standards RDF and SPARQL, we believe that similar initiatives can further contribute to the development of decentralized and adaptable learning services.

DATA ANALYSIS AND INTERPRETATION USING LINKED DATA

Given the rapid growth of unstructured textual content on various online social media and communication channels, as well as the ever-increasing amount of dedicated learning content deployed on MOOCs, there is a need to automate the discovery of items relevant to distance education, such as topics, trends, and opinions, to name a few. In fact, analytics required for the discovery and/or recommendation of relevant items can be improved if the regular input data (e.g., learners' logs) is enriched with background information from LOD datasets (e.g., data about topics associated with the course) (d'Aquin & Jay, 2013). The use of LOD cross-domain knowledge bases such as DBpedia and Yago, alone or in combination with traditional content analysis techniques (e.g., social network analysis, text mining, latent semantic indexing), represent a promising avenue for advancing content analysis

[13] http://lodum.de/
[14] http://linkedup-project.eu/
[15] http://lak.linkededucation.org/
[16] https://schema.org/
[17] https://www.w3.org/community/schema-course-extend/
[18] http://xmlns.notu.be/aair/
[19] European Commission, "ESCO: European Skills, Competencies, Qualifications and Occupations," https://ec.europa.eu/esco
[20] https://github.com/adlnet/xAPI-Spec

and information retrieval in educational settings, as outlined in the following sections.

Content Analysis Using Semantic Annotation

One important development in the LD field has been the rapid expansion and adoption of *semantic annotators* (Jovanović et al., 2014) - services that take unstructured text as input and annotate/tag it with LOD concepts (i.e., entities defined in LOD knowledge bases such as DBpedia, Wikidata,[21] and Yago). The latter are general, cross-domain knowledge bases storing Wikipedia-like knowledge in well-structured formats with explicitly defined semantics. Several of these LD annotators offer interfaces (APIs) that target the extraction of various types of concepts, such as named entities (e.g., people and places), domain concepts (e.g., protein, gene), and themes or keywords, though the diversity of possible annotations is continuously expanding. Examples of these annotators, both from academia and industry, include DBpedia Spotlight,[22] AlchemyAPI,[23] and TagMe.[24]

Given the plethora of unstructured texts from formal courses, MOOCs, and social media, the capacity of such annotators to produce explicit semantic representations of text makes them valuable for various analytic services. However, very few research works have yet leveraged the power of semantic annotation for learning analytics. Recent research by Joksimović et al. (2015) uses a mixed-method approach for discourse analytics in a cMOOC based on LD and social network analysis (SNA). The aim of the study was to explore the main topics emerging from learners' posts within various social media (i.e., Facebook, Twitter, and blogs) and to analyze how those topics evolve throughout the course (Joksimović et al., 2015). Instead of relying on some of the commonly used topic modelling algorithms (e.g., latent Dirichlet allocation [LDA]), the researchers utilized tools for automated concept extraction (i.e., semantic annotators) along with SNA to identify emerging topics (groups of concepts). Specifically, for each week of the course, concepts were extracted from the posts generated in each of the media analyzed. Further, the authors created graphs based on the co-occurrence of concepts within a single post. Finally, the authors applied modularity algorithm for community detection (Newman, 2006) in order to identify the most prominent groups of concepts (i.e., latent topics). The main advantage of such an approach, over "traditional" topic-modelling algorithms, is possibility to extract compound words (e.g., "complex adaptive systems") that are further linked to knowledge bases (e.g., DBpedia), allowing for easier interpretation of the extracted topics.

Analysis of Scientific Publications in the LA/EDM Field

Another application domain powered by LD and related to the educational context is semantic publishing (e.g., releasing library catalogues as LD) and meta-analysis of scientific publications. In fact, one of the main successes of LD technologies has been their early adoption by various content publishers such as BNF[25] and scientific-based publishing initiatives such as DBLP.[26] This has led to a plethora of LOD vocabularies and datasets related to scientific publications. These datasets provide grounds for various scientometric computations that identify trending topics, influencing researchers, and describe the research community at large (Mirriahi, Gašević, Dawson, & Long, 2014; Ochoa, Suthers, Verbert, & Duval, 2014). They also directly help professionals (researchers, students, librarians, course producers) from the educational sector to locate relevant information.

In the LA/EDM domain, the Learning Analytics and Knowledge (LAK) Dataset (Taibi & Dietze, 2013) represents a corpus of publications from the LA/EDM communities. The LAK Dataset contains both the publications' content and metadata (e.g., keywords, authors, conference). It represents a data integration effort as it relies on various established LOD vocabularies and constitutes a successful application of LD technologies. The analysis of the LAK Dataset has been encouraged since 2013 through the annual LAK Data Challenge, whose goal was to foster research and analytics on the LA/EDM publications. This dataset has been further exploited for the development of data analytics and content analysis applications. One particularly valuable application is the identification of topics and relations between topics in the dataset, per year, per community (LA versus EDM), per publication, and overall. For example, the work of Zouaq, Joksimović, and Gašević (2013) employed ontology learning techniques on the LAK Dataset to identify salient topics and relationships between them. Other techniques applied for discovering topics include latent Dirichlet allocation (LDA; Sharkey & Ansari, 2014) and clustering (Scheffel, Niemann, Leon Rojas, Drachsler, & Specht, 2014). While these approaches offered a text-based content analysis, other works went further in their data integration efforts by relying on the LOD knowledge bases (e.g., DBpedia) and semantic annotators to identify topics of interest. For example, Milikić, Krcadinac, Jovanović, Brankov, and Keca (2013) and Nunes, Fetahu, and Casanova (2013) relied on TagMe and DBpedia Spotlight services, respectively,

[21] https://www.wikidata.org/

[22] https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki

[23] http://www.alchemyapi.com/

[24] https://tagme.d4science.org/tagme/

[25] http://www.bnf.fr/en/tools/a.welcome_to_the_bnf.html

[26] http://datahub.io/dataset/l3s-dblp

to identify topics and named entities in publications. The benefit of LD in this case was highlighted by 1) the ability to enrich the dataset with LOD concepts, keywords, and themes, and 2) the ability to develop advanced services such as potential collaborator detection (Hu et al., 2014), dataset recommendations, or more general semantic searches (Nunes et al., 2013).

Interpretation of Data Mining Results

Several research works have provided insights, patterns, and predictive models by analyzing learners' interaction and discussion data (e.g., identifying the link between learners' discourse and position and their academic performance (Dowell et al., 2015) or course registration data (d'Aquin & Jay, 2013). However, most of these analyses remain limited to a closed or silo dataset, and are often hard to interpret on large datasets.

In general, pattern discovery in LA/EDM requires a model and a human analyst for the meaningful interpretation of results according to several dimensions (e.g., topics, student characteristics, learning environments, etc.) (d'Aquin & Jay, 2013). The work by d'Aquin & Jay (2013) provides new insights into the usefulness of LD for enriching and contextualizing patterns discovered during the data-mining process. In particular, they propose annotating the discovered patterns with LD URIs so that these patterns can be further enriched with existing datasets to facilitate interpretation. The authors illustrate the idea by a case study of student enrollment in course modules across time. They extract frequent course sequences and enrich them by associating them, via course URIs, with course descriptions, i.e., a set of properties describing the course. The (chain of) properties provide(s) analytical dimensions that are exploited in a lattice-based classification (e.g., the common subjects of frequent course sequences) and as a navigational structure. As illustrated in this case study, LD can help discover new analytical dimensions by linking the discovered patterns to external knowledge bases and exploiting LOD semantic links to infer new knowledge. This is especially relevant in multidisciplinary research where various factors can contribute to a pattern or phenomenon. Given the complexity of learning behaviours, one can imagine the utility of having this support in the interpretation of LA/EDM results.

DISCUSSION AND OUTLOOK

The overall analytical approach to learning experience requires state-of-the-art data management techniques for the collection, management, querying, combination, and enrichment of learning data. The concept and technologies of LD – the latter based on W3C standards (RDF, SPARQL) – have the potential to contribute to all these aspects of data management. First, one of the primary objectives behind LD technologies is to make the data easily processable and reusable, for a variety of purposes, while preserving and leveraging the semantics of the data. Second, LD allows for a decentralized approach to data management by enabling the seamless combination and querying of various datasets. Third, large-scale knowledge bases available as linked open data on the Web provide grounds for a variety of services relevant for the analytic process; e.g., semantic annotators for content analysis and enrichment. Fourth, data exposed as LD on the Web can provide on-demand (just-in-time) data/knowledge input required in different phases of the analytic process, as this knowledge cannot be always fully anticipated in advance. Potential benefits also include representing the resulting analytics in a semantic-rich format so that the results could be exchanged among applications and communicated to interested parties (educators, students) in different manners, depending on needs and preferences (e.g., different visual or narrative forms). Moreover, through its inference capabilities over multiple data sources, originating in semantic-rich representation of data items and their mutual relationships, LD-based methods could be a relevant addition to the existing analytical methods for discovering themes and topics in textual content. More generally, while statistical and machine-learning methods are widespread in the LA/EDM community, other kinds of data analysis methods and techniques – those based on explicitly defined semantics of the data – and open knowledge resources (especially open, Web-based knowledge) can make the traditional analytical approaches even more powerful. Some of the potential enrichments provided by LD include semantic vector-based models (e.g., bags of concepts instead of bags of words), semantic-rich social network analysis with explicitly defined semantics for edges and nodes, or recommendations based on semantic similarity measures.

Finally, LD technologies can be useful in dealing with the heterogeneity of learning environments and social media platforms. In particular, one can query and assemble various datasets that do not share a common schema. This aspect in itself represents a more flexible and practical approach than previous approaches that required compliance to a common model/schema.

However, there are also several challenges related to the use of LD in terms of the following:

1. **Quality:** The quality of the LOD datasets is a concern (Kontokostas et al., 2014), and linking learning resources and traces to external datasets and knowledge bases might introduce noisy data. Although there are some initiatives for data

cleaning, this issue is far from being resolved.

2. **Alignment:** Besides the use of common Web URIs among schemas, there is often a need to semantically align vocabularies and models, which is a challenging task. Current alignment approaches are often based on syntactic matching, which does not deal well with ambiguities. One way to mitigate the alignment issue is to be aware and re-use major LD vocabularies[27] whenever possible (e.g., *foaf:name* is a property depicting the name of a person in the FOAF specification and could be used instead of creating a new property);
3. **Privacy:** Data within MOOCs and learning platforms is often siloed for privacy reasons. Merging information between learning and social platforms would require, for example, that learners grant access to their data and provide log-in information for the different services they use for learning.

Despite the challenges indicated above - and given the use of LOD datasets and knowledge bases in some major initiatives such as Google knowledge graph or Facebook graph search and their increasing adoption in educational institutions - LD is a promising technological backbone for today's learning platforms. It also provides a useful formalism for facilitating the overall learning analytic process, from raw data collection and storage, to data exploitation and enrichment, to interpretation of the analytics results.

REFERENCES

Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. J*ournal of Biomedical Informatics*, 41(5), 706–716.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. Preprint retrieved from http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. International *Journal of Technology Enhanced Learning*, 4(5–6), 318–331.

Cooper, A. R. (2013). Learning analytics interoperability: A survey of current literature and candidate standards. http://blogs.cetis.ac.uk/adam/2013/05/03/learning-analytics-interoperability

d'Aquin, M., & Jay, N. (2013). Interpreting data mining results with linked data for learning analytics: Motivation, case study and directions. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 155–164). New York: ACM.

d'Aquin, M., Adamou, A., & Dietze, S. (2013, May). Assessing the educational linked data landscape. *Proceedings of the 5th Annual ACM Web Science Conference* (WebSci '13), 2–4 May 2013, Paris, France (pp. 43–46). New York: ACM.

De Nies, T., Salliau, F., Verborgh, R., Mannens, E., & Van de Walle, R. (2015, May). TinCan2PROV: Exposing interoperable provenance of learning processes through experience API logs. *Proceedings of the 24th International Conference on World Wide Web* (WWW '15), 18–22 May 2015, Florence, Italy (pp. 689–694). New York: ACM.

Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1–2), 9–38.

Dietze, S., Drachsler, H., & Giordano, D. (2014). A survey on linked data and the social web as facilitators for TEL recommender systems. In N. Manouselis, K. Verbert, H. Drachsler, & O. C. Santos (Eds.), *Recommender Systems for Technology Enhanced Learning* (pp. 47–75). New York: Springer.

Dietze, S., Sanchez-Alonso, S., Ebner, H., Qing Yu, H., Giordano, D., Marenzi, I., & Nunes, B. P. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches. *Program*, 47(1), 60–91.

Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., & Taibi, D. (2012). Linked education: Interlinking educational resources and the web of data. *Proceedings of the 27th Annual ACM Symposium on Applied Computing* (SAC 2012), 26–30 March 2012, Riva (Trento), Italy (pp. 366–371). New York: ACM.

[27] http://lov.okfn.org/dataset/lov/

Dowell, N. M., Skrypnyk, O., Joksimović, S., Graesser, A. C., Dawson, S., Gasevic, D., Hennis, T. A., de Vries, P., & Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 250–257). International Educational Data Mining Society. http://files.eric.ed.gov/fulltext/ED560532.pdf

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 9–17). New York: ACM.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71.

Groth, P., Loizou, A., Gray, A. J., Goble, C., Harland, L., & Pettifer, S. (2014). API-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 29, 12–18.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: Theory and technology*, 1(1), 1–136. Morgan & Claypool.

Hu, Y., McKenzie, G., Yang, J. A., Gao, S., Abdalla, A., & Janowicz, K. (2014). A linked-data-driven web portal for learning analytics: Data enrichment, interactive visualization, and knowledge discovery. In LAK Workshops. http://geog.ucsb.edu/~jano/LAK2014.pdf

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March, Poughkeepsie, NY, USA (pp. 156–165). New York: ACM.

Jovanović, J., Bagheri, E., Cuzzola, J., Gašević, D., Jeremic, Z., & Bashash, R. (2014). Automated semantic annotation of textual content. *IEEE IT Professional*, *16*(6), 38–46.

Kagemann, S., & Bansal, S. (2015). MOOCLink: Building and utilizing linked data from massive open online courses. *Proceedings of the 9th IEEE International Conference on Semantic Computing* (IEEE ICSC 2015), 7–9 February 2015, Anaheim, California, USA (pp. 373–380). IEEE.

Kessler, C., d'Aquin, M., & Dietze, S. (2013). Linked data for science and education. *Journal of Semantic Web*, 4(1), 1–2.

Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R., & Zaveri, A. (2014). Test-driven evaluation of linked data quality. *Proceedings of the 23rd International Conference on World Wide Web* (WWW '14), 7–11 April 2014, Seoul, Republic of Korea (pp. 747–758). New York: ACM.

Lausch, A., Schmidt, A., & Tischendorf, L. (2015). Data mining and linked open data: New perspectives for data analysis in environmental research. *Ecological Modelling*, 295, 5–17.

Maturana, R. A., Alvarado, M. E., López-Sola, S., Ibáñez, M. J., & Elósegui, L. R. (2013). Linked data based applications for learning analytics research: Faceted searches, enriched contexts, graph browsing and dynamic graphic visualisation of data. LAK Data Challenge. http://ceur-ws.org/Vol-974/lakdatachallenge2013_03.pdf

Milikić, N., Krcadinac, U., Jovanović, J., Brankov, B., & Keca, S. (2013). Paperista: Visual exploration of semantically annotated research papers. LAK Data Challenge. http://ceur-ws.org/Vol-974/lakdatachallenge2013_04.pdf

Mirriahi, N., Gašević, D., Dawson, S., & Long, P. D. (2014). Scientometrics as an important tool for the growth of the field of learning analytics. *Journal of Learning Analytics*, 1(2), 1–4.

Muñoz-Merino, P. J., Pardo, A., Kloos, C. D., Muñoz-Organero, M., Wolpers, M., Katja, K., & Friedrich, M. (2010). CAM in the semantic web world. In A. Paschke, N. Henze, & T. Pellegrini (Eds.), *Proceedings of the 6th International Conference on Semantic Systems* (I-Semantics '10), 1–3 September 2010, Graz, Austria. New York: ACM. doi:10.1145/1839707.1839737

Newman, M. E. (2006). Modularity and community structure in networks, *Proceedings of the National Academy of Sciences*, 103(23), 8577–8582.

Niemann, K., Wolpers, M., Stoitsis, G., Chinis, G., & Manouselis, N. (2013). Aggregating social and usage datasets for learning analytics: Data-oriented challenges. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 245–249). New York: ACM.

Nunes, B. P., Fetahu, B., & Casanova, M. A. (2013). Cite4Me: Semantic retrieval and analysis of scientific publications. LAK Data Challenge, 974. http://ceur-ws.org/Vol-974/lakdatachallenge2013_06.pdf

Ochoa, X., Suthers, D., Verbert, K., & Duval, E. (2014). Analysis and reflections on the third learning analytics and knowledge conference (LAK 2013). *Journal of Learning Analytics*, 1(2), 5–22.

Piedra, N., Chicaiza, J. A., López, J., & Tovar, E. (2014). An architecture based on linked data technologies for the integration and reuse of OER in MOOCs context. *Open Praxis*, 6(2), 171–187.

Santos, J. L., Verbert, K., Klerkx, J., Duval, E., Charleer, S., & Ternier, S. (2015). Tracking data in open learning environments. *Journal of Universal Computer Science*, 21(7), 976–996.

Scheffel, M., Niemann, K., Leon Rojas, S., Drachsler, H., & Specht, M. (2014). Spiral me to the core: Getting a visual grasp on text corpora through clusters and keywords. LAK Data Challenge. http://ceur-ws.org/Vol-1137/lakdatachallenge2014_submission_3.pdf

Schmitz, H. C., Wolpers, M., Kirschenmann, U., & Niemann, K. (2011). Contextualized attention metadata. In C. Roda (Ed.), *Human attention in digital environments* (pp. 186–209). New York: Cambridge University Press.

Sharkey, M., & Ansari, M. (2014). Deconstruct and reconstruct: Using topic modeling on an analytics corpus. LAK Data Challenge. http://ceur-ws.org/Vol-1137/lakdatachallenge2014_submission_1.pdf

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1), 3–10. http://itdl.org/Journal/Jan_05/article01.htm

Softic, S., De Vocht, L., Taraghi, B., Ebner, M., Mannens, E., & De Walle, R. V. (2014). Leveraging learning analytics in a personal learning environment using linked data. *Bulletin of the IEEE Technical Committee on Learning Technology*, 16(4), 10–13.

Taibi, D., & Dietze, S. (2013), Fostering analytics on learning analytics research: The LAK dataset. LAK Data Challenge, 974. http://ceur-ws.org/Vol-974/lakdatachallenge2013_preface.pdf

Zablith, F. (2015). Interconnecting and enriching higher education programs using linked data. *Proceedings of the 24th International Conference on World Wide Web* (WWW '15), 18–22 May 2015, Florence, Italy (pp. 711–716). New York: ACM.

Zotou, M., Papantoniou, A., Kremer, K., Peristeras, V., & Tambouris, E. (2014). Implementing "rethinking education": Matching skills profiles with open courses through linked open data technologies. *Bulletin of the IEEE Technical Committee on Learning Technology*, 16(4), 18–21.

Zouaq, A., Joksimović, S., & Gašević, D. (2013). Ontology learning to analyze research trends in learning analytics publications. LAK Data Challenge. http://ceur-ws.org/Vol-974/lakdatachallenge2013_08.pdf