The Learning Analytics & Knowledge (LAK) Data Challenge 2014

Hendrik Drachsler
Welten Institute, Open University of the Netherlands
hendrik.drachsler@ou.nl

Stefan Dietze
L3S Research Center
Leibniz University, Hannover, Germany
dietze@l3s.de

Eelco Herder
L3S, Research Center
Leibniz University, Hannover, Germany
herder@l3s.de

Mathieu d'Aquin
Knowledge Media Institute, The Open University, UK
m.daquin@open.ac.uk

Davide Taibi
Institute for Educational Technologies, National Research Council of Italy
davide.taibi@itd.cnr.it

Abstract
The LAK Data Challenge 2014 continues the research efforts of the second edition by stimulating research on the evolving fields Learning Analytics (LA) and Educational Data Mining (EDM). Building on a series of activities of the LinkedUp project, the challenge aims to generate new insights and analysis on the LA & EDM disciplines and is supported through the LAK Dataset - a unique corpus of LA & EDM literature, exposed in structured and machine-readable formats.

Categories and Subject Descriptors
E.1 [**Data Structures**] *Distributed data structures*; E.2 [**Data Storage Representations**] *Linked representations;* J.1 [**Administrative Data Processing**] *Education*; H.1.1 [**Information Systems**] *Models and principles, Systems and Information Theory*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms
Algorithms, Measurement, Design, Standardization, Experimentation, Human Factors, Theory.

Keywords
Learning analytics, data mining, linked data, visualization.

1. Introduction
A variety of datasets is used in the Learning Analytics field for research on teaching and learning. The available datasets can be roughly distinguished between (a) tracking data that comes from different learning environments [1] and (b) Linked data from the web [2].

Tracking data from different learning environments involves interactions of learners with different tools and resources. The main driver for analyzing these data is the vision of increased awareness of the learning progress, self-regulation support, and personalized learning that offers potential to create more effective learning experiences through new possibilities for the prediction and reflection of individual learning processes. Most of the time the tracking data underlies legal and privacy restrictions that make it difficult to share the data or make it accessible to third parties.

Next to the large amount of tracking data, there is an increasing amount of Linked Data on the Web that covers educational data published by institutions about their courses and learning resources. The Linked Data approach enables the enrichment of learning content and the learning experience by making use of various connected data sources. Through reusing schemas and vocabularies as well as by relying on persistent URIs for data referencing, it provides a higher level of interoperability. This makes it more convenient to use Linked Data for research purposes and make the outcomes of the research better comparable.

Recently, the LinkedUp project has created a dataset catalog[1] of educationally relevant (Linked) datasets that is freely accessible [4]. The main aim of LinkedUp is to identify and promote innovative success stories that exploit Linked data in educational scenarios. Under this objective, it contributed a Linked dataset for the Learning Analytics (LA) and Educational Data Mining (EDM) communities to facilitate research, analysis, and smart explorative applications to gain new insights into the research papers published in this domain [5].

In 2013, the LinkedUp project[2] organized the first LAK Data Challenge based on the LAK13 Dataset [3]. A range of interesting applications and analytical research has been contributed to the first challenge that provided new insights into the development of the emerging research field[3]. The first place of the challenge went to 'Linked Data based applications for Learning Analytics Research' [6] that supports the exploration of the Learning Analytics and Educational Data Mining literature in relation to other online resources. The second place was awarded to Cite4Me, a semantic retrieval and analysis tool for scientific publications [7] and the third place went to an Ontology Learning

LAK'14, March 24–28, 2014, Indianapolis, Indiana, USA.
ACM 978-1-4503-2664-3/14/03…$15.00.
http://dx.doi.org/10.1145/2567574.2567630

[1] http://data.linkededucation.org/ and http://data.linkededucation.org/linkedup/catalog/

[2] http://www.linkedup-project.eu

[3] http://linkedu.eu/event/lak2013-linkeddata-tutorial/?page_id=58

tool that enables the analyze of research trends in the Learning Analytics publications [8]. This submission provided evidences for the main differences as well as the relationships between the EDM and the LAK research community.

In the LAK Data Challenge 2014 we aim to stimulate new or more mature approaches to research questions like: *How can we make sense of the emerging field's, historical roots, current state, and future trends, based on how its members report and debate their research?* The challenge submissions exploit the extended LAK 2014 dataset for meaningful purposes. This includes submissions which cover one or more of the following, non-exclusive list of topics:

- Analysis of the emerging LAK & EDM communities in terms of topics, people, citations or connections with other fields
- Innovative applications to explore, navigate and visualize the dataset (and/or its correlation with other datasets)
- Usage of the dataset as part of recommender systems
- Tools to improve the accuracy of the references in the LAK datasets
- Trend analysis within the LAK2014 dataset

2. The LAK Dataset

The LAK Dataset has been extracted to create a structured corpus including full text, references, and metadata including authors, affiliations, titles, keywords and abstracts. The schema used to describe the papers in the dataset is based on two established schemas: the Semantic Web Conference Ontology (already used to describe metadata about publications from the Semantic Web conferences and related events) and the Linked Education schema. The data is accessible under http://lak.linkededucation.org.

Throughout the past year and following on the LAK Data Challenge 2013, the dataset has been improved and expanded as a joint effort by SoLAR, ITD-CNR[4], and the LinkedUp project. Latest publications from the LAK2013 and EDM2013 conferences were added, enriching content and keywords with references to DBpedia and including the actual references of each publication. The current version of the LAK datasets consists of:

- Proceedings of the ACM International Conference on Learning Analytics and Knowledge (LAK) (2011-13)
- The open access journal Educational Technology & Society recently published a 2012 special issue on "Learning and Knowledge Analytics": *Educational Technology & Society* - Special Issue on Learning & Knowledge Analytics.
- Proceedings of the International Conference on Educational Data Mining (2008-13)
- Journal of Educational Data Mining (2008-13)

3. Workshop Organization

3.1 Workshop Facilitators

The workshop is organized jointly by SoLAR, the LinkedUp project and LinkedUp associated partner CNR-ITD. In addition the special interest groups Linked Education (http://linkededucation.org) and SIG dataTEL (http://ea-tel.eu/sig-datatel/) of EATEL will support the competition. All partners aim at advancing data-driven research in Education. The main goals are to promote the re-use of Linked and Open Web data, to foster the cooperation between different Learning Analytics research units and to offer reference datasets for data-driven research.

[4] www.itd.cnr.it/

The partners can look back on an annual workshop series at different conferences, including:

- Linked Learning 2013 – 3rd International Workshop on Learning and Education with the Web of Data (LILE2013) at WWW 2013, Rio de Janeiro, Brazil.
- 1st International Workshop on Learning Analytics and Linked Data at 2nd International Conference on Learning Analytics and Knowledge
- Linked Learning 2012 - 2nd International Workshop on Learning and Education with the Web of Data (LILE2012) at WWW2012, Lyon, France.
- Linked Learning 2011 - 1st International Workshop on Learning and Education with the Web of Data (Linked Learning workshopLILE2011) at the 8th Extended Semantic Web Conference (http://purl.org/linkedlearning).
- dataTEL11 at the 3rd Alpine Rendezvous conference in La Clusaz, France (http://bit.ly/datatel_arv11).

3.2 Evaluation of submissions

The submissions have been reviewed by members of the challenge committee to pre-select submissions for presentation. During the LAK conference and based on the presentations, the challenge winners have been identified according to the LinkedUp Evaluation Framework. The accepted submissions will be published in online proceedings and presented during an interactive session at the LAK 2014 conference in Indianapolis, USA. Finally, the best three submissions will be presented in a panel at the main stage of the conference.

References

[1] 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.

[2] T. Heath and C. Bizer (2011). Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool.

[3] Taibi, D. and Dietze, S. (2013). Fostering Analytics on Learning Analytics Research: the LAK Dataset. CEUR WS Proceedings Vol. 974, Proceedings of the LAK Data Challenge.

[4] d'Aquin, M., Adamou, A., Dietze, S., Assessing the Educational Linked Data Landscape, ACM Web Science 2013 (WebSci2013), Paris, France, May 2013.

[5] d'Aquin, M., Dietze, S., Drachsler, H., Taibi, D. (2014). Using linked data in learning analytics. eLearning Papers, Nr. 36, ISSN: 1887-1542.

[6] Nunes, B. P., Fetahu, B., Dietze, S., Casanova, M. A., Cite4Me: A Semantic Search and Retrieval Web Application for Scientific Publications, 12th International Semantic Web Conference (ISWC2013), Sydney, Australia, (2013).

[7] Maturana, R.A., Alvarado, M. E. López-Sola, S., Ibáñez, M.J., & Ruiz Elósegui, L., (2013). Linked Data based applications for Learning Analytics Research. CEUR WS Proceedings of the LAK Data Challenge, Vol. 974.

[8] Zouaq, Amal, Srećko Joksimović, and Dragan Gašević. (2013). Ontology Learning to Analyze Research Trends in Learning Analytics Publications. CEUR WS Proceedings of the LAK Data Challenge Vol. 974, accessible under http://lak.linkededucation.org.