The 3rd LAK data competition

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

Maren Scheffel  
Welten Institute, Open University of the Netherlands  
maren.scheffel@ou.nl

Abstract
The LAK Data Challenge 2015 continues the research efforts of the previous data competitions in 2013 and 2014 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

General Terms

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, a higher level of interoperability is being provided. This makes it more convenient to use Linked Data for research purposes and makes the outcomes of the research more comparable.

The LinkedUp project has created a dataset catalogue [1] of educationally relevant (linked) datasets that are freely accessible [4]. The main aim of LinkedUp was 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 2014, the LinkedUp project [2] organized the 2nd LAK Data Challenge based on the LAK13 dataset [3]. A range of interesting applications and analytical research has been contributed to the previous challenges that provided new insights into the development of the emerging research field [1]. The submissions provide evidence for the main differences as well as the relationships between the EDM and the LA research community. As the LinkedUp project has been successfully completed in October 2014, we are glad to announce that the LACE project [4] was interested to host the LAK data challenge in 2015 in cooperation with SoLAR.

---

The challenge submissions should 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:

- What are the main differences between the EDM and LAK community over time?
- What are new and emerging topics?
- Do different visualisations of the data lead to different insights?
- How can the LAK dataset be used to point to related research efforts in other scientific disciplines such as educational science, ethics and legal aspects but also new technologies?
- Which of the articles in the data provide evidence for positive or negative effects of learning analytics research on the educational practices?

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](http://lak.linkededucation.org).

Throughout the past year and following on the LAK Data Challenge 2014, the dataset has been improved and expanded as a joint effort by SoLAR, ITD-CNR5, and the LinkedUp project. Latest publications from the LAK2014 and EDM2014 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 previous editions of the LAK Data Challenge (2013-14)
- Proceedings of the International Conference on Educational Data Mining (2008-14)
- 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.
- Journal of Educational Data Mining (2009-14)
- Journal of Learning Analytics (2014)

3. Workshop Organization

3.1 Workshop Facilitators

The workshop is organized jointly by SoLAR, the LACE project and associated partner CNR-ITD. In addition the special interest groups Linked Education ([http://linkededucation.org](http://linkededucation.org)) and SIG dataTEL ([http://ea-tel.eu/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 foster the cooperation between different Learning Analytics research units and to offer reference datasets for data-driven research.

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

- Linked Learning 2014 – 4th International Workshop on Learning and Education with the Web of Data (LILE2014) at ISWC2014, Riva del Garda, Italy.
- Linked Learning 2013 – 3rd International Workshop on Learning and Education with the Web of Data (LILE2013) at WWW 2013, Rio de Janeiro, Brazil.
- Learning Analytics and Linked Data (LADA12) at 2nd at LAK12, Vancouver, Canada.
- Linked Learning 2012 - 2nd International Workshop on Learning and Education with the Web of Data (LILE2012) at WWW2012, Lyon, France.

3.2 Evaluation of submissions

The submissions will be reviewed by members of the challenge committee to pre-select submissions for presentation. The accepted submissions will be published in online proceedings and presented during an interactive session at the LAK 2015 conference. During the LAK conference and based on the presentations, the challenge winners are identified according to the LinkedUp Evaluation Framework [9]. Finally, the best three submissions will be presented in a panel at the main stage of the conference and receive an award by SoLAR.

References


---

5 [www.itd.cnr.it/](http://www.itd.cnr.it/)