Springer LOD Conference Portal. Demo paper

Aliaksandr Birukou[1], Volha Bryl[1], Kai Eckert[2], Andrey Gromyko[3], and Markus Kaindl[4]

[1]Springer Nature, Heidelberg, Germany
[2]Stuttgart Media University, Stuttgart, Germany
[3]Net Wise, Trento, Italy
[4]Springer Nature, Munich, Germany

Abstract. Despite many efforts for making data about scholarly publications available on the Web of Data, information about academic conferences is still contained in (at best) free-text format. Availability of this data in a structured format would enable more efficient decision making for researchers, libraries, publishers, funding and evaluation bodies. This demo paper showcases the Springer Linked Open Data (LOD) conference portal (available at `http://lod.springer.com`). We cover the architecture, vocabularies and features of the portal and present usage scenarios.

Keywords: proceedings, linked open data, LOD, LNCS, conferences.

1 Introduction

Linked Open Data about conferences is often provided not directly by the publishers, journals or conferences, but by third-parties, such as DBLP, COLINDA, WikiCfP or EventSeer. A prominent exception is the Semantic Web Dog Food repository[1], now being relaunched as the Scholarly Data project [1]. It provides very detailed information about the conferences, authors, PC members, and papers, and such data is provided directly by the conference organizers. This data is limited to the Semantic Web community and has certain shortcomings, as shown by Nuzzolese et al. [1]. As of early 2017, the repository contained the data about 48 conferences and 235 workshops.

This context led us to launching the Conference Linked Open Data portal, available at `http://lod.springer.com`, on the 30th of March 2015 [2]. In this demo paper we present the current version of the portal, which has significantly advanced since 2015. As of 27th of July 2017, the portal contains data about 8,965 conferences (which are grouped into 1,646 conference series), and 10,093 conference proceedings, published by Springer since 1973. The portal is the first resource providing structured, trusted (directly by the publisher) and continuously updated data about conferences at such a scale. One can download the data described in this paper in [3].

[1] SWDF: `http://data.semanticweb.org`.

2 Dataset description

In the following, we list the main characteristics of the LOD portal. The data in the portal is freely available under CC0 1.0 Universal license.

Topical coverage and data sources. When launched in 2015 the dataset only contained conferences, conference series and proceedings in computer science. As of late 2016, the coverage was extended to include data about newly published conference proceedings in other disciplines, mainly engineering and mathematics. In early 2017 we have started adding recent books to the portal, so there are 57,043 books (10,093 of which are proceedings) and 1,203,912 book chapters and papers. In total, together with all the properties, the portal contains around 20 million triples. New information about the published conferences is loaded to the portal on a daily basis and monthly dumps are provided.

Interlinking and technologies. The conference series in the portal are linked to the conference series in DBLP[2]. Thus, the portal respects the FAIR principles[3] and provides 5 star Linked Open Data [4]. The **SPARQL endpoint** is powered by Apache Jena Fuseki, the URI dereferencing is implemented by means of Pubby with the DM2E enhancements [5].

Vocabularies. We use custom vocabularies as it makes mapping to our internal formats easier. However, in the future we would like to have mappings to the most common vocabularies, similarly to what Nature did in the past[4].

The ontology files are included in the dumps at `http://lod.springer.com/data/dumps`. The complete description of the entities and their properties is available at `http://lod.springer.com/wiki/bin/view/Linked+Open+Data/Data+Description`.

In the following, we list the main entities of the dataset:

- **conference series**, which has a unique conference series ID, a conference series ID at DBLP, a conference series name, and links conferences from different years. For instance, *International Semantic Web Conference*, `http://lod.springer.com/data/html/confseries/semweb`;
- **conference**, which is characterized by a conference acronym, name, year, and number, the city and country, the start and end dates. The conference is also linked to the conference series and all books, which contain the proceedings of the conferences. For instance, *ISWC 2016*, `http://lod.springer.com/data/html/conference/semweb2016`;
- **book**, which is characterized by usual bibliographic attributes, such as ISBN, DOI, editors; the acronym and name of the book series, the volume number, the title and subtitle. The book also contains information about whether it is indexed in Scopus and Ei Compendex and is linked to the conference and lists all the chapters (papers in the proceedings). See `http://lod.springer.com/data/html/book/978-3-319-46523-4`.

[2] see, for instance, `http://lod.springer.com/data/html/confseries/semweb`

[3] `https://www.force11.org/group/fairgroup/fairprinciples`

[4] `https://www.nature.com/ontologies/datasets/linksets/`

– **chapter**, which has a DOI and a title, authors, the publication date and copyright year, page numbers. See `http://lod.springer.com/data/html/bookchapter/978-3-319-46523-4_1` for an example.

An important feature of the portal is that conference series together with conferences are treated as first-class objects, similar to scientific journals and issues. This addresses the fact that in computer science, conference series are comparable to journals regarding their impact and importance for publishing.

3 Portal architecture and main functionalities

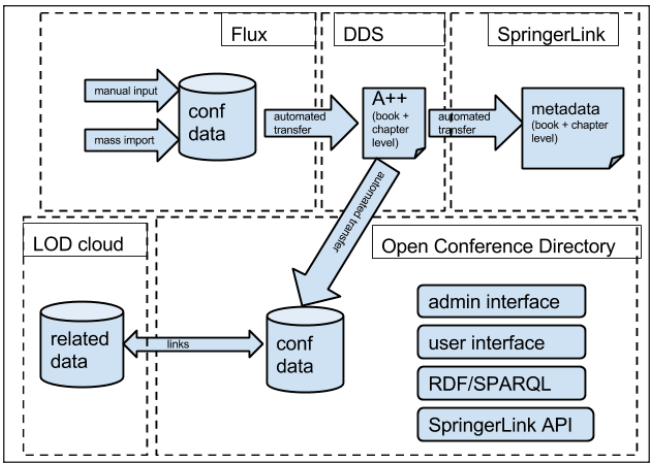

Fig. 1: Conference metadata workflow

Conference names, dates, locations etc. have previously only been available as the unstructured subtitles of our proceedings volumes. During the development of the LOD conference portal this information has been transformed to a well-defined and structured RDF format and integrated into the Springer internal workflow system. Figure 1 shows how the metadata about conferences flows through Springer systems: 1) conference-related information is entered during production; 2) it is stored in the product data; 3) it is included in the chapter and conference series pages on SpringerLink [5] as HTML-embedded microdata and 4) imported to the LOD portal as RDF.

Google Scholar uses the data about the conference where a specific paper is published for improving Google Scholar Metrics[6], a rating of publication venues in different disciplines.

[5] see `https://link.springer.com/conference/semweb`, for an example of the ISWC conference series page

[6] `https://scholar.google.de/intl/en/scholar/metrics.html`

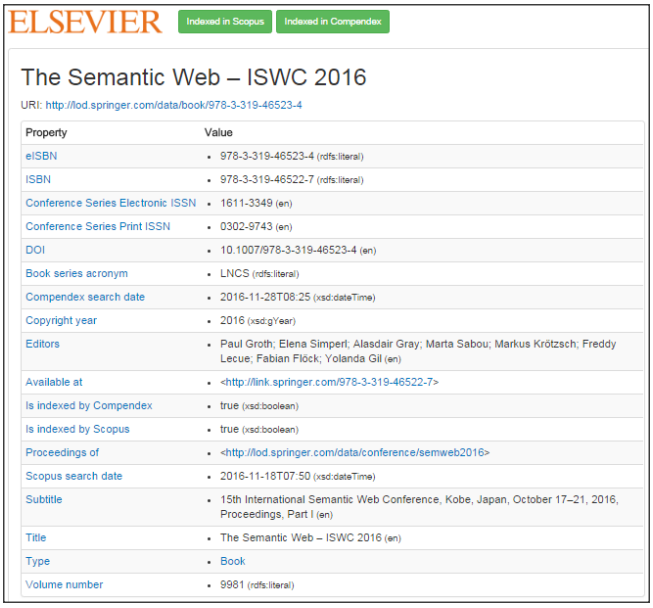

Fig. 2: Sample data: Indexing in Scopus and EI Compendex

Users can check if their conference papers published in proceedings are indexed in **Scopus** or **Ei Compendex**. Conference organizers or authors can search ISBN, DOI, conference acronym or the volume number of the book series and see if the proceedings are indexed or not (see Figure 2). Such services are enabled by the Scopus and Ei Compendex APIs.

Semantic web researchers can also access the portal via the SPARQL endpoint at `http://lod.springer.com/sparql-form/index.html`.

References

1. Nuzzolese, A.G., Gentile, A.L., Presutti, V., Gangemi, A.: Conference linked data: The scholarlydata project. In: The Semantic Web – ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II, Cham, Springer International Publishing (2016) 150–158
2. Bryl, V., Birukou, A., Eckert, K., Kessler, M.: What's in the proceedings? Combining publisher's and researcher's perspectives. In: Proceedings of the 4th Workshop on Semantic Publishing co-located with the 11th Extended Semantic Web Conference (ESWC 2014), Anissaras, Greece, May 25th, 2014. (2014)
3. Birukou, A., Bryl, V., Eckert, K., Gromyko, A., Kaindl, M.: Dump of lod.springer.com - 27 july 2017. doi 10.6084/m9.figshare.5368747 (2017)
4. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web. Morgan & Claypool Publishers (2011)
5. Baierer, K., Dröge, E., Eckert, K., Goldfarb, D., Iwanowa, J., Morbidoni, C., Ritze, D.: DM2E: A linked data source of digitised manuscripts for the digital humanities. Semantic Web **8**(5) (2017) 733–745