

Linked Data for Science and Education

Editor(s): Name Surname, University, Country

Solicited review(s): Name Surname, University, Country

Open review(s): Name Surname, University, Country

Carsten Keßler^a, Mathieu d’Aquin^b and Stefan Dietze^c

^a *Institute for Geoinformatics, University of Münster, Germany*

E-mail: carsten.kessler@uni-muenster.de

^b *The Open University, UK*

E-mail: m.daquin@open.ac.uk

^c *L3S Research Center, Germany*

E-mail: dietze@l3s.de

Abstract. Sharing of resources and metadata is a central principle in scientific and educational contexts. With the emergence of the Linked Data approach as the most recent evolution of the Semantic Web, scientific and educational practitioners have started to adopt those principles. The communities working on Linked Data for science and education have since developed common schemas used for describing scientific or educational resources, substantial collections of structured data, bibliographic collections, domain-specific vocabularies capturing vast amounts of scientific domain knowledge, as well as baseline technologies used to expose and integrate linked datasets. In this paper, we give an overview of the current landscape related to the use of Linked Data in the academic sector. We look at the common challenges, prominent datasets, tools and applications, and conclude on the major directions for research in this area.

Keywords: Survey, Linked Science, Linked Learning, Education

1. Introduction

Sharing of resources, resource metadata, and data across the Web is a central principle in scientific and educational contexts. Scientific collaboration has long been striving for wider reuse and sharing of knowledge and data. Likewise, the Open Educational Resources community has promoted the widespread exploitation of public and reusable educational resources throughout the last decade. Hence, technologies to enable interoperability of shared resources and data have long been at the centre of scientific and educational information systems. However, due to the lack of a shared technology stack and joint principles, the landscape of developed and utilised standards is very fragmented and covers an increasing variety of heterogeneous technologies, such as repositories with proprietary interfaces and query mechanisms. Moreover, a broad range of largely incompatible metadata schemas

and taxonomies have been developed to describe and expose educational resources, and scientific workflows and data. Due to the prevailing heterogeneity of deployed approaches and technologies, interoperability remains an open challenge.

At the same time, the Linked Data (LD) [10] approach has emerged as the most recent evolution of the Semantic Web [8] and has widely established itself as the de-facto standard for sharing data on the Web. While the LD approach provides a set of well-established principles and (W3C) standards, such as the use of URIs as identifiers, RDF, SPARQL [91], aiming at Web-scale data interoperability, it has produced an ever growing amount of data sets and schemas available on the Web. Given the proven capabilities of LD technologies towards realising Web scale data sharing and reuse, scientific and educational practitioners have started to adopt those principles. Results of such activities cover joint schemas used for

describing scientific or educational resources, vast collections of structured data about, for instance, cultural or historic artifacts, bibliographic collections, domain-specific vocabularies capturing extensive amounts of scientific domain knowledge – where the life sciences are particularly well represented – as well as baseline technologies used to expose and integrate linked datasets.

In this paper, we give an overview of the current landscape related to the use of LD for science and education, looking at the common challenges, prominent datasets, tools and applications, to conclude on the major directions for research in this area.

2. Knowledge and data sharing in Science & Education: challenges

With technologies evolving, communities of practitioners in the science and education sectors have produced a variety of standards and approaches, especially to facilitate information sharing. These standards have however shown limited success in their concrete adoption, and generally seem not to be suitable for the Web-scale distribution of information of educational and scientific relevance. We discuss here some of their shortcomings, identifying the challenges where LD technologies and principles can prove of valuable use.

2.1. Educational data and metadata sharing

Throughout the last decade, research in the field of technology-enhanced learning (TEL) has focused fundamentally on enabling interoperability and reuse of learning resources and data. That has led to a fragmented landscape of competing metadata schemas, i.e., general-purpose ones such as Dublin Core [31] or schemas specific to the educational field, like IEEE Learning Object Metadata (LOM) [46] or ADL SCORM¹, but also interface mechanisms such as OAI-PMH² or SQI³. These technologies are exploited by educational resource repository providers to support interoperability. In addition, social data as well as so-

called attention metadata, capturing the perception of learning resources by learners, has become increasingly useful for tailoring learning experiences to particular user needs and requirements. To this end, although a vast amount of educational content and data is shared on the Web in an open way, the integration process is still costly as different learning resources are isolated from each other and based on different implementation standards [24].

2.2. Scientific data and metadata sharing

The scientific community has been developing new ways of sharing (meta-)data along with the move to digital research environments. However, similar to the field of TEL, the developed approaches were characterized by a segregation of the research process into different aspects. Tools such as Taverna⁴ or MyExperiment⁵ support researchers in managing workflows for their experiments. Electronic lab notebooks help keeping track of progress. Data repositories and data management systems, such as laboratory information management systems, provide means to store and access research data through proprietary APIs. The way they handle and enable interaction with data is often domain-specific due to the varying requirements across disciplines. Finally, once a finding is published, the publication and its metadata is provided through an electronic library catalogue. Even though these tools and systems support and document parts of the research process, there is still a significant lack of integration of different information sources. The documentation of a research project is segregated, placing each aspect into a different silo, and the entry point to that documentation – the publication – is decontextualized [6].

2.3. Challenges

While there already is a large amount of educational and scientific data available on the Web via proprietary and/or competing schemas and interface mechanisms, the main challenge is to (a) start adopting LD principles and vocabularies while (b) leveraging on existing data available on the Web by non-LD compliant means [30]. In the following, we list the major research challenges which are currently approached by adopting LD-principles within science and education.

¹Advanced Distributed Learning (ADL) SCORM: <http://www.adlnet.org>

²Open Archives Protocol for Metadata Harvesting <http://www.openarchives.org/OAI/openarchivesprotocol.html>

³Simple Query Interface: <http://www.cen-ltso.net/main.aspx?put=859>

⁴<http://www.taverna.org.uk>

⁵<http://www.myexperiment.org>

Data interoperability—infrastructural: Vast amounts of educational and scientific data have already been available on the Web, however, due to heterogeneity of existing storage and interface approaches, interoperability has been limited [94]. LD offers a technology stack composed of RDF as representation standard and SPARQL as query mechanism and standardised infrastructural HTTP endpoint which facilitates exposing, integrating and sharing of Web data at the infrastructural level. In addition, a number of tools, for instance, for storage and integration of data have been provided, often with domain-specific extensions (see Section 4).

Data interoperability—semantic and syntactic: In addition to infrastructural boundaries, heterogeneity with respect to data representation also hinders wide interoperability of data. This includes, for instance, the use of heterogeneous schemas, vocabularies and representation languages. LD addresses these issues through a range of principles: RDF(S) provides a joint data representation language, the consistent use of dereferencable URIs allows wide reuse of any data or schema item and inherent features of RDF and OWL, such as `owl:sameAs` statements, facilitate the interlinking of heterogeneous datasets. These features are particularly well-suited to support scientific and educational processes, fundamentally based on citing and incrementally extending—i.e., linking—existing work, that has led to the availability of a wide range of schemas and datasets in the area of education and science. In particular the notion of Linked *Open* Data (LOD) emphasizes the public and open nature of data, analogous to the public character of scientific knowledge, where raw data is made available along with publications to ensure reproducibility (see Section 3.2). On the educational side, the open data movement found an equivalent with the Open Educational Resources approach, but also fostered transparency in areas such as enrollment statistics or energy consumption in university facilities.

Formalising domain knowledge. One particular challenge within both, scientific as well as educational contexts, is the lack of formal and well-aligned domain models, which would allow a shared understanding of a discipline, such as the life sciences, and the description of domain-specific data and knowledge resources with consistent vocabularies. Here, the LD movement was particularly successful in generating a wealth of domain-specific vocabularies, where some disciplines, such as biomedicine, are particularly well reflected, while others are still under-represented (see Sections 3.1 and 3.2). Moreover, standards such as

RDFS and SKOS facilitate the combination and alignment of different models making it no longer necessary to subscribe to just *one* domain model.

The wide range of available tools (see Section 4) and applications (see Section 5) which exploit aforementioned capabilities of LD demonstrate its applicability to scientific and educational processes.

3. Datasets

This section gives an overview of existing datasets for education (Section 3.1) and science (Section 3.2). Section 3.3 lists vocabularies specifically developed for science and education, and provides an analysis of the vocabularies actually in use. The range of datasets and vocabularies is steadily growing and has already reached a size that does not permit complete listings in this paper. This section therefore contains a subjective selection that is intended to cover the broad range of thematic areas.

3.1. Educational datasets

Open Educational Resources (OER) are educational material freely available online. The wide availability of educational resources is a common objective for universities, libraries, archives and other knowledge-intensive institutions raising a number of issues, particularly with respect to Web-scale metadata interoperability or legal as well as licensing aspects. Several competing standards and educational metadata schemas have been proposed over time, including IEEE LTSC LOM (Learning Object Metadata), one of the widest adopted, IMS⁶, ISO/IEC MLR - ISO 19788 Metadata for Learning Resources (MLR)⁷ and Dublin Core. The adoption of a sole metadata schema is usually not sufficient to efficiently characterize learning resources. As a solution to this problem, a number of taxonomies, vocabularies, policies, and guidelines (called application profiles) are defined [32]. Some popular examples are: UK LOM Core⁸, DC-Ed⁹ and ADL SCORM.

⁶<http://www.imsglobal.org/metadata/>

⁷http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50772

⁸<http://zope.cetis.ac.uk/profiles/uklomcore/>

⁹<http://www.dublincore.org/documents/education-namespace/>

Due to the diversity of exploited standards, existing OER repositories offer very heterogeneous datasets, differing with respect to schema, exploited vocabularies, and interface mechanisms. For example, OpenLearn¹⁰ is the UK Open University's contribution to the OER movement and it is a member of the MIT Open Courseware (OCW) Consortium¹¹. Video material from OpenLearn, distributed through iTunesU has reached more than 40 million downloads in less than 4 years. One of the largest and diverse collections of OER can be found in the GLOBE (Global Learning Objects Brokered Exchange)¹² where jointly, nearly 1.2 million learning objects are shared.

Regarding the presence of educational information in the LD landscape, two types of LDsets need to be considered: (1) datasets directly related to educational material and institutions, including information from open educational repositories and data produced by universities; (2) datasets that can be used in teaching and learning scenarios, while not being directly published for this purpose. This second category includes for example datasets in the cultural heritage domain, such as the ones made available by the Europeana project,¹³ as well as by individual museums and libraries (such as the British Museum, which has made their collection available as LD,¹⁴ representing more than 100 million triples, or the Bibliothèque Nationale de France,¹⁵ which made available information about 30,000 books and 10,000 authors in RDF, representing around 2 million triples). It also includes information related to research in particular domains, and the related publications (see PubMed¹⁶ which covers more than 21 million citations, in 800 million triples), as well as general purpose information for example from Wikipedia (via DBPedia.org).

Regarding category (1), initiatives have emerged recently using LD to expose, give access to and exploit public information for education. The Open University in the UK was the first education organization to create an LD platform to expose information from across its departments, and that would usually sit in many different systems, behind many different interfaces (see <http://data.open.ac.uk>

which includes around 5 million triples about 3,000 audio-video resources, 700 courses, 300 qualifications, 100 Buildings, 13,000 people). Many other institutions have since then announced similar platforms, including in the UK the University of Southampton¹⁷ and the University of Oxford.¹⁸ Outside the UK, several other universities and education institutions are joining the Web of Data, by publishing information of value to students, teachers and researchers with LD. Noticeable efforts include the Linked Open Data at University of Münster, Germany, and the corresponding LODUM initiative,¹⁹ or the Norwegian University of Science and Technology exposing its library data as Linked Open Data. These universities thus leverage LD to facilitate both internal data exchange and make the data easily and programmatically accessible for external consumers. Serving a university's data through a SPARQL endpoint creates a *one-stop shop* for standardized access, and, in combination, creates an interlinked *university graph*. First applications with students as the main user group have been released [61, 96]. In addition, educational resources metadata has been exposed by the mEducator project [95,94]. The problems connected to the heterogeneity of metadata can be addressed by converting the data into a format that allows for implementing the LD principles. Most often this means that the data which is provided as part of RDBMS or in XML format – or, on occasion, in other formats – are converted to RDF. The data model of RDF is a natural choice as it allows for unique identification, interlinking to related data, as well as enrichment and contextualization. Therefore, general-purpose tools (see Section 4) are often used to convert proprietary datasets to RDF. It is common to use DBPedia or other big datasets as “linking hubs”. One of the advantages of this approach is that such datasets are commonly used by other datasets, which automatically leads to a plurality of indirect links. In the case of more specialized applications, it is beneficial if domain-specific datasets or ontologies can be found and linked to. This has been successfully demonstrated by specialized projects such as Linked Life Data²⁰ in the biomedical domain, Organic.Edunet²¹ in organic agriculture and agroecology, and mEducator in medical education [94].

¹⁰<http://openlearn.open.ac.uk/>

¹¹<http://ocw.mit.edu/>

¹²<http://globe-info.org/>

¹³<http://www.europeana.eu/>

¹⁴<http://collection.britishmuseum.org/>

¹⁵<http://data.bnf.fr/>

¹⁶<http://www.ncbi.nlm.nih.gov/pubmed/> and <http://thedatahub.org/dataset/bio2rdf-pubmed>

¹⁷See <http://data.southampton.ac.uk>.

¹⁸See <http://data.ox.ac.uk>.

¹⁹See <http://data.uni-muenster.de>.

²⁰<http://www.linkedlifedata.com>

²¹<http://www.organic-edunet.eu>

A more thorough overview of educational LD is offered by the Linked Education platform,²² whose international team maintains a growing collection of datasets, models, tools, applications, as well as related events and calls.

3.2. *Scientific datasets*

Transparency and reproducibility are core principles for scientific work, yet many disciplines have been struggling to guarantee these beyond the traditional publication of peer-reviewed articles. While most research environments have been undergoing a fundamental change moving into a digital environment, traditional paper publication is still prevalent. Different disciplines have tried to tackle this challenge by setting up repositories for their research data, with limited success: Proprietary APIs and domain-specific data models did not live up to the requirements of a meaningful interlinking of a publication with the data, processes, and software that lead to the presented results. Over the past years, the potential of the LD approach has been realised to achieve this integration. The ultimate goal is a Linked Science approach [53] that takes the traditional idea of Open Science [22] to a new level, where all aspects of a study are available online, referring to each other through shared vocabularies. Ideally, this should enable the development of *executable papers* [54] that automatically gather all data and software for a paper (e.g., packed as a *Research Object* [6,74,97]) and enable the reader to review the whole process at a mouse click. Thinking beyond the use of vocabularies and ontologies to support scientific workflows, the long-term vision is to use them as knowledge artifacts that support the researcher in testing theories and finding new insights [15].

Having all resources relevant for a paper interlinked and available online has obvious advantages for reproducibility and secondary use of often expensive-to-produce datasets. Nonetheless, data, software and processes are useless without the paper that documents the work and serves as entry point and anchor for any related information. In order to fulfill this function in a LD driven research environment, it is vital that bibliographic data is made available according to the LD principles [7]. This requirement, combined with the long-standing tradition and expertise in (meta-)data management the librarians have built over

the decades, leads to the libraries being at the forefront of the LD movement. This is reflected both in the high number of libraries already publishing their catalogues as Linked Open Data (see Table 2) and in a comparably high degree of standardization with agreed-upon vocabularies [31,38,47,84] and the publication of key pieces of the library infrastructure, such as Digital Object Identifiers (DOIs) and thesauri already published as LD [9,45,64]. Beyond a mere conversion of MARC 21 [65] catalogue entries, new approaches towards *semantic publishing* [1,85] are currently being discussed, enabling detailed annotations of article contents, such as hypotheses, and the different ways in which papers are being cited [16,83].²³ The World Wide Web Consortium addresses library data as a central pillar of the Web of Data with the Library Linked Data Incubator Group²⁴ stressing the central role of library LD not only for science and education, but for the Semantic Web as a whole [4]. BibBase has been introduced as a service that supports authors in managing and publishing semantically annotated metadata for their papers [92].

The actual data used and produced during research projects can be divided into static datasets and dynamic, semantically annotated data streams. Static datasets are often hosted on repositories such as The Data Hub²⁵ or DataCite²⁶, with the latter offering persistent DOIs for uploaded datasets to make them citable [11]. These services facilitate data publishing for institutions and individual researchers that lack the resources or expertise to publish the data themselves. Moreover, they also act as central hubs, therefore adopting a role similar to catalogue services that users and developers are likely to visit when looking for datasets. Many self-hosted datasets can also be found on repositories as dumps because of this exposition. The actual datasets found here cover a variety of scientific disciplines, ranging from drug research [49] over cultural heritage [44] to deforestation data from the Amazon rainforest [55,23]. Evidently, some fields have made more progress in the adoption of semantic web technologies than others. The biomedical domain is an example of a field that has a long tradition in us-

²²See <http://linkededucation.org>, co-maintained by the authors of this paper, among others.

²³See the Semantic Publishing and Referencing (SPAR) ontologies for details: <http://purl.org/spar/>.

²⁴See <http://www.w3.org/2005/Incubator/1ld/>

²⁵See <http://thedatahub.org>.

²⁶See <http://datacite.org>.

Table 1
Selection of educational datasets

Dataset	Source & SPARQL Endpoint	License
mEducator Linked Educational Resources	http://linkededucation.org/meducator http://meducator.open.ac.uk/resourcesrestapi/rest/meducator/sparql	©*
Open University Linked Data	http://data.open.ac.uk http://data.open.ac.uk/query	CC-BY
Achievement Standards Network	http://achievementstandards.org/ http://asn.jesandco.org:8890/sparql	CC
education.data.gov.uk	http://education.data.gov.uk/ http://services.data.gov.uk/education/sparql	©*
Ariadne RDF	http://www.ariadne-eu.org/ http://knowone.csc.kth.se/sparql/ariadne-big	©*
Nature Publishing Group	http://data.nature.com/ http://data.nature.com/sparql	CC-0
SEEK-AT WD	http://www.gsic.uva.es/seek/ http://seek.rkbexplorer.com/sparql/	©*

* No license info given, so we assume that these datasets are *not* under an open license.

Table 2
Selection of library datasets

Dataset	Source & SPARQL Endpoint	License
British National Bibliography	http://bnb.data.bl.uk http://bnb.data.bl.uk/sparql	CC-0
Calames	http://calames.abes.fr –	©
Bibliothèque nationale de France	http://data.bnf.fr –	CC-BY
NTNU special collections	http://api.talis.com/stores/gunnbib-digitalmanuskripter http://api.talis.com/stores/gunnbib-digitalmanuskripter/sparql	PD
Leibniz Information Centre for Economics	http://zbw.eu/beta/ http://zbw.eu/beta/sparql	©*
ECS Southampton EPrints	http://eprints.ecs.soton.ac.uk –	CC-0
DBLP Bibliography	http://dblp.13s.de/d2r/ http://dblp.13s.de/d2r/sparql	©
Hungarian National Library	http://nektar.oszk.hu/wiki/Semantic_web http://setaria.oszk.hu/sparql	©
HBZ NRW (college libraries Northrhine-Westphalia)	http://lobid.org http://lobid.org/sparql	CC-0

* No license info given, so we assume that these datasets are *not* under an open license.

ing ontologies and has already published a significant number of datasets.²⁷

Table 3 gives an overview of the range of such relatively static scientific datasets. They manifest the output of scientific studies and can act as input to new

analyses. In addition, the Semantic Sensor Web [17, 82] is a source of semantically annotated data streams that provide near-realtime sensor observations of phenomena. Services that offer data on the current weather or water conditions, for example, enable reasoning on those data streams [26] and on-the-fly integration with additional data sources [48]. In this case, Semantic Web technologies are not only used to annotate and ex-

²⁷See <http://linkedlifedata.com/sources> for an overview.

change research results, but also part of the scientific analysis toolchain. As the semantic annotation of such services [66] has experienced significant adoption in the Sensor Web community, Linked Sensor Data [17] have been introduced as an application of the LD principles [7] to observation data. Despite the challenges that the often high frequency and large volume of observations pose when serving them as LD [59], the approach has recently gained traction, with first applications demonstrated [75,86,5].

3.3. *Vocabularies and Schemas for Science and Education*

As discussed in previous sections, the eLearning and Technology Enhanced Learning communities have been very prolific in building metadata standards, especially to describe and organise educational resources. Similarly, the scientific community has created formats and vocabularies to exchange information about research. These standards have however rarely been reused in or adapted to the LD world, for a variety of reasons, and a new set of open vocabularies are now emerging to represent different aspects of science and education.

As a shared interest between science and education is the representation of bibliographical references, to support citation, reuse and discovery. Several ontologies exist that can be used to represent references, including the Bibliographic Ontology, BIBO [38], and others like FABIO [84] based on the more sophisticated FRBR model [47]. At the basis of a lot of these efforts is the Dublin Core metadata schema [31] which represents a common ground for the description of resources and documents. Other ontologies have been created that focus on more specific aspects of bibliographical references, such as the CiTO ontology [83].

Another prominent aspect of representing information for science and education is the representation of the communities and institutions in which scientific and educational activities are realised, especially academic organisations. AIIISO (Academic Institution Internal Structure Ontology) [87] for example is a common, simple ontology for the representation of the sub-organisations of an academic institution and of its offerings. The Bolowgna ontology [25] focuses more specifically on the representation of notions related to the Bologna process in Europe. General vocabularies to represent organisations (such as the W3C core organisation ontology [78]), as well as connections within communities (such as FOAF [14] and

SIOC [76,13]) can also be used to represent academic institutions and communities.

Core vocabularies to represent elements related to science and education have of course also emerged. For example, TEACH [58] and Linked Science (LSC) [3] represent two core vocabularies, respectively for the representation of courses and of scientific resources. The Ontology of units of Measure and related concepts (OM) addresses the annotation of research data with the correct units of measure, addressing a requirement for semantic interoperability across many fields that work empirically [79]. Metadata for Learning Opportunities (MLO) [33] and eXchanging Course Related Information (XCRI) [80] are two attempts at providing standard formats for advertising and exchanging information about the course offerings of academic institutions, which had both been adapted to the LD world (through an RDFs version, see <http://linkeduniversities.org/lu/index.php/vocabularies/>).

While a lot of effort was spent on metadata standards for learning resources, there have not been significant initiatives to transfer this effort to LD, besides the Learning Resource Metadata Initiative (LRMI [88]), which attempts to provide a `schema.org` compliant vocabulary for educational resources and the mEducator resource schema [34], which provides an adaptable RDF schema for describing educational resources. Many other vocabularies are used however that are dedicated to the general representation of (online) resources, and applied to the representation of educational content (e.g., the W3C Ontology for Media Resources [63] for the representation of multimedia artefacts). It is important also to notice that significant efforts are required and being spent on making available common topic classifications to annotate these resources, including for example the SKOS representations of the Mathematics Subject Classification [2] or of the UK JACS codes [81]. In many cases, the LD-compliant URIs from the Library of Congress Subject Headings would also be used [64].

As described above, vocabularies and ontologies for LD in the area of science and education are currently gaining momentum and new initiatives are emerging frequently. Specialised vocabularies that can represent precisely the notions related to these areas are needed, that relate to more general vocabularies for online resources, organisations, communities, etc. While these vocabularies are being developed, additional effort is however needed in achieving a widespread adoption of

Table 3
Selection of scientific datasets

Dataset	Source & SPARQL Endpoint	License
Bio2RDF datasets	http://bio2rdf.org http://*.bio2rdf.org/sparql	various *
Italian National Research Council	http://data.cnr.it http://data.cnr.it/sparql	CC-BY-NC-ND
2000 U.S. Census Data	http://rdfabout.com/demo/census http://www.rdfabout.com/sparql	CC-NC
Chronicling America	http://chroniclingamerica.loc.gov -	PD
Drugbank	http://www4.wiwiss.fu-berlin.de/drugbank/ http://www4.wiwiss.fu-berlin.de/drugbank/sparql	©
Linked Life Data	http://linkedlifedata.com http://linkedlifedata.com/sparql	©
European Nature Information System	http://semantic.eea.europa.eu http://semantic.eea.europa.eu/sparql	CC-BY
GeoSpecies Knowledge Base	http://lod.geospecies.org/ http://lsd.taxonconcept.org/sparql	CC-BY-SA
Linked Sensor Data	http://wiki.knoesis.org/index.php/SSW_Datasets http://sonicbanana.cs.wright.edu:8890/sparql	CC-BY
World Bank Linked Data	http://worldbank.270a.info/ http://worldbank.270a.info/sparql	CC-0

* BioRDF is a collection of datasets which have been published under separate licenses.

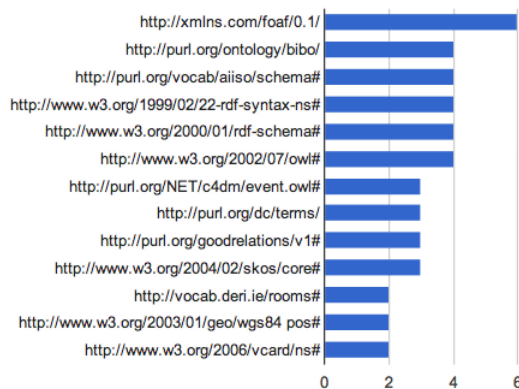


Fig. 1. Vocabularies used in six datasets related to education. Taken from the LUCERO Blog [20].

these common vocabularies. Indeed, as shown in Figure 1, generic vocabularies are still prominently used in educational datasets, and only a few of the prominent specialised ones (BIBO, AIIISO) are really being commonly employed (therefore creating entry points for data exchange and interoperability).

4. Tools and Technologies

For what concerns the publication of LD, in many cases, the same tools and technologies that are being used in other domains would apply similarly in science and education (see e.g., [18]). These include of course triple stores, SPARQL engines and URI delivery mechanisms.

In terms of data processing, the need for specialised tools might appear. Indeed, while generic tools such as Google Refine²⁸ with the RDF extension²⁹, Triplify³⁰ or D2RQ³¹, can be applied on many different types of data, specialised data conversion and processing tools could be more adequate to transform specific formats used in science and education. These include in particular SIMILE RDFizer's converters³² for formats such as MARC records³³ (for library catalogues), OAI-

²⁸<http://code.google.com/p/google-refine/>

²⁹<http://refine.deri.ie/>

³⁰<http://triplify.org/>

³¹<http://d2rq.org/>

³²<http://simile.mit.edu/repository/RDFizers/marcmods2rdf/>

³³<http://simile.mit.edu/repository/RDFizers/marcmods2rdf/>

PMH³⁴ (for open archive repositories) and OCW³⁵ (for MIT OpenCourseWare metadata). Other specialised RDF extractors include Bibtex2RDF³⁶ to convert bibliographical references in the Bibtex format, or the Youtube2RDF³⁷ tool that converts Youtube playlists into RDF using media vocabularies (see [18]). Interactive mapping tools such as Karma [62] support the user in selecting meaningful annotations for their data during the conversion process to RDF.

It is however in developing applications of LD in science and education that certain types of tools and technologies become really crucial. As described in the next section, resource discovery and recommendation are clearly amongst the most crucially needed features both in teaching and research (where resources can be learning material, articles, experts, etc.) Typical recommendation approaches can be employed to realise such features, either by content or through collaborative filtering. However, on a number of examples (including the ones given below), such approaches have been shown to improve (in performance and/or usability) through relying on the structured representation, links and semantics brought by LD (see e.g. [72]).

More sophisticated techniques for resource discovery, as well as for other scenarios in exploiting LD in science and education are also becoming more and more important in these areas. These include for example the use of Natural Language Processing (especially Named Entity Recognition) to analyse and characterise textual resources in terms of LD entities. DBpedia Spotlight [68] has for example been employed in several applications (including DiscOU mentioned below) in order to connect learning resources to DPpedia entities, which in turn provide a hub to all sorts of other datasets on the LD cloud. Similarly, data mining and data analytics techniques are attracting more and more attention in science and education (especially through the emerging Learning Analytics domain [35]), where the traditional “data acquisition bottleneck” is being replaced by an abundance of rich, structured and interconnected data from the LD cloud. To tackle this, common tools such as Gephi³⁸ and R³⁹ have been ex-

tended to handle LD, and the use of such extensions in applications for science and education are starting to emerge.⁴⁰

5. Applications

As mentioned in [18], we can identify 3 main use cases in learner centric applications. These scenarios are especially relevant in open and distance learning, but can be extended to the education domain in general, to research and to science.

Resource delivery and navigation: This includes the possibility to browse through resources in an organised manner and obtain structured data for these resources. In such applications, the structures, classifications and links that are included in the underlying data are used to provide navigational and browsing mechanisms improving the ease of access to relevant resources. Faceted browsing could for example be used, exploiting structured data to provide relevant filters to apply on search results, as in Faceted DBLP [27]. Relevant resources can also be included within more general environments based on links between such environments and available LD. This has been done for example in the course catalogue application of the Open University (see descriptions in [18,19]) and is at the basis of the myExperiment⁴¹ collaborative platform (see [40]). The Glottolog/Langdoc project,⁴² providing a catalogue of the languages of the world, is an impressive showcase of applying SKOS to interlink resources and make them navigable [73]. Similarly, data portals, such as the Linked Science portals for Brazilian Amazon Rainforest data, humanitarian response data from the 2010 earthquake in Haiti, and geographic information science community data on researchers and publications [60] provide exploratory user interfaces that allow non-specialist users to explore the data.⁴³ Initial projects such as Open PHACTS [90] are now trying to go beyond resource delivery and put Semantic Web technologies to use, in this case for drug discovery.

³⁴<http://simile.mit.edu/repository/RDFizers/oai2rdf/>

³⁵<http://simile.mit.edu/repository/RDFizers/ocw2rdf/>

³⁶www.l3s.de/~siberski/bibtex2rdf/

³⁷<http://code.google.com/p/lucero-project/wiki/Youtube2RDF>

³⁸<https://gephi.org/>

³⁹<http://www.r-project.org/>

⁴⁰see for example <http://blog.ouseful.info/2011/01/30/open-university-undergraduate-module-map/>

⁴¹<http://www.myexperiment.org/>

⁴²<http://www.glottolog.org/>

⁴³<http://linkedscience.org/data/>

Resource discovery and recommendation: Going a step further with respect to resource delivery, one of the base scenarios in both education and science that can be better supported by the use of LD is resource discovery. Indeed, finding relevant material for learning, teaching or research in the flow of existing resources and publications is becoming a major challenge for both students and scientists. In particular, plenty of work has been done on the automatic generation of educational resources, for instance, assessment items [77,36]. Traditional information retrieval and recommender system approaches can be employed in education, but with the added impact of the common/rich structure brought by LD. A typical example of such an application is Talis Aspire Community Edition⁴⁴ [89], which allows lecturers from UK universities to create and manage online reading lists for the courses they teach. With the help of a common, LD-based metadata representation, the tool offers the possibility to obtain similar, recommended resources to the ones being selected. Similarly, the DiscOU application developed at the Open University⁴⁵ offers recommendations of (open) resources relevant to a TV programme, but in this case based on semantic annotations of the resources using LD entities to measure their relevance and provide meaningful explanations of the proposed results [21]. Additionally, social educational environments such as Metamorphosis+ [52] have adopted LD-mechanisms to integrate, cluster and correlate educational resources and subjects.

In the scientific community, first attempts have been made to utilise LD to collect input data for studies. LinkedDataLens⁴⁶ is a tool that generates input data for network analysis algorithms from LD, and makes the output available as LD again [39,42]. The iExplore tool⁴⁷ [71] implements such a user interface with domain expert users as target group, leveraging the graph structure to assist the researcher in developing new hypotheses. Brennkmeijer et al. [12] propose scientific lenses that make views on scientific linked datasets context aware, so that they can be used

by researchers with different interests or backgrounds.

Personalisation and Social Recommendation: Similarly to resource discovery, the idea of social learning and online collaboration in science is to deliver to users information/resources that are of interest and create personalized learning environments [50]. However, in this case, the idea is to rely on the profile of the user and on his/her interests rather than on similarities with existing resources. This idea is becoming very popular in open and distance learning in particular, where the absence of “classroom-based interactions” needs to be compensated, through technology, by other ways of interacting with fellow learners, teachers and available learning material. Early examples of this type of applications include the Open University’s Course Profile Facebook app⁴⁸ (see [18] for a short description), which allows students to share their “learning journey” and interact with other students on the basis of common interest, courses, etc. Another, particularly innovative approach [51] uses LD as a means to directly interlink unstructured annotations of eBooks in order to improve interoperability and correlation of related learner feedback.

In addition, there are many advantages in employing LD that are not necessarily directly for the purpose of building end-user (student, learner, lecturer, scientist) applications, but that address the general challenges mentioned in Section 2. For example, the Bowlogna ontology [25] has been developed to support information exchange between academic institutions across Europe. Many applications of LD are also targeting support for management and organisation within educational institutions, including for example the monitoring of research communities at the Open University [19] or at the University of Bristol.⁴⁹

6. Conclusions

The LD approach has gained significant traction in science and education. It facilitates the exchange of data within and between scientific communities. Likewise, the OER community leverages LD to enhance annotation and retrievability of learning ma-

⁴⁴<http://community.talisaspire.com/>

⁴⁵<http://discou.info>

⁴⁶<http://linkeddatalens.isi.edu>

⁴⁷<http://knoesis.wright.edu/iExplore/iExplore.html>

⁴⁸<http://apps.facebook.com/courseprofiles/>

⁴⁹<http://researchrevealed.ilrt.bris.ac.uk/>

materials for use in classrooms as well as virtual and distance learning environments. The number of academic institutions publishing LD is constantly growing, as the growing lists of partners at the LinkedUniversities.org and LinkedEducation.org websites show. These hubs, together with blogs,⁵⁰ mailing lists⁵¹ and regular events such as the [Linked Science \[56,57\]](#) and [Linked Learning \[28,29\]](#) workshop series allow the communities to share and discuss best practices, next steps, and collaborations. Beyond the the publication of LD for science and education, standardisation efforts such as VIVO,⁵² euroCRIS,⁵³ or the Learning Resource Metadata Initiative (LRMI) are engaging with practitioners in discussions towards a wider adoption of the LD approach.

Concerning next steps for the LD for science and education community, further internationalisation and interconnection is indeed an important goal with its current centre in Europe, and especially the UK. As the overall landscape is still very fragmented, as highlighted by this special issue, there are significant challenges to be tackled in order to enable LD to live up to its expectations in the academic sector. For example, dereferenceable URIs, shared vocabularies, and a standardized technology stack are only the first building blocks towards the vision of an executable paper that is linked to all relevant data, models, and software. In order to achieve such ambitious aims, tools are required that support the data publishing process, adapted to the specific requirements of the scientific and educational domains. These tools have to be integrated with the practitioners' environments and easy to use, as outlined in [93], so that no specialist knowledge about Semantic Web technologies is required. Comparable to content management systems that allow lay users to administrate websites, data management systems are required that support users in publishing their data online as well as aligning it with existing data sources. Currently, some institutions try to meet the need for such expertise by hiring data scientists; however, this model is unlikely to scale, if we think a few years ahead, when a data portal will probably be a common service to offer for a university, in addition to

its website. With a growing number of datasets being published, provenance, versioning, and archiving becomes more important. Approaches to handle these exist [37,43,67,70], but are hardly applied in practice so far. Likewise, the potential to leverage Semantic Web technology to support the scientific process – besides mere data management and sharing – is still largely untapped [15], for example, when it comes to formalizing hypotheses [41] or breaking down results into nanopublications [69]. Beyond the technology, science and education will have to continue to become more open, with support for the open access movements on a political level.

As Bechhofer et al. [6] note, Linked Data alone is not enough for scientists. It only provides the baseline technology, and there are still plenty of open challenges in terms of tool support, coordination and capacity building, as well as policy development in order to fully exploit the potential for the education and science sectors.

References

- [1] M.A. Angrosh, Stephen Crane, and Nigel Stanger. Contextual Information Retrieval in Research Articles: Semantic Publishing Tools for the Research Community. *Semantic Web Journal*, accepted.
- [2] Ioannis Antoniou, Charalampos Bratsas, Anastasia Dimou, Patrick Ion, Christoph Lange, and Wolfram Sperber. Mathematics Subject Classification Linked Wiki Version 0.1 Alpha 1. Available from <http://sci-class.math.auth.gr/MSC1W>; last accessed 2012-08-26, 2012.
- [3] Alkyoni Baglatzi, Tomi Kauppinen, and Carsten Keßler. Linked Science Core Vocabulary Specification. Available from <http://linkedscience.org/lsc/ns>; last accessed 2012-08-26, 2011.
- [4] Thomas Baker, Emmanuelle Bermès, Karen Coyle, Gordon Dunsire, Antoine Isaac, Peter Murray, Michael Panzer, Jodi Schneider, Ross Singer, Ed Summers, William Waites, Jeff Young, and Marcia Zeng. Library Linked Data Incubator Group Final Report. W3C Incubator Group Report, available from <http://www.w3.org/2005/Incubator/1ld/XGR-1ld-20111025/>; last accessed 2012-08-26, 2011.
- [5] Payam Barnaghi, Mirko Presser, and Klaus Moessner. Publishing Linked Sensor Data. In *Proceedings of the 3rd International Workshop on Semantic Sensor Networks*, Shanghai, China, 2010.
- [6] Sean Bechhofer, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Darius Michaelides, Stuart Owen, David Newman, Shoaib Sufi, and Carole Goble. Why Linked Data is Not Enough for Scientists. In *e-Science (e-Science), 2010 IEEE Sixth International Conference on*, pages 300–307. IEEE, 2010.

⁵⁰See <http://linkedscience.org>, <http://lucero-project.info/>, and <http://blogs.ecs.soton.ac.uk/data/> for examples.

⁵¹<https://groups.google.com/forum/?fromgroups#!forum/linked-universities>

⁵²<http://vivoweb.org>

⁵³<http://eurocris.org/>

- [7] Tim Berners-Lee. Linked Data – Design Issues. Available online from <http://www.w3.org/DesignIssues/LinkedData.html>; last accessed 2012-04-04, 2009.
- [8] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, May 2001.
- [9] Geoffrey Bilder. Content Negotiation for CrossRef DOIs. Available from http://www.crossref.org/CrossTech/2011/04/content_negotiation_for_crossr.html; last accessed 2012-08-26, April 2011.
- [10] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.
- [11] Jan Brase. DataCite—A global registration agency for research data. In *Cooperation and Promotion of Information Resources in Science and Technology, 2009. COINFO'09. Fourth International Conference on*, pages 257–261. IEEE, 2009.
- [12] Christian Y. A. Breninkmeijer, Chris Evelo, Carole Goble, Alasdair J.G. Gray, Paul Groth, Steve Pettifer, Robert Stevens, Antony Williams, and Egon L. Willighagen. Scientific Lenses over Linked Data: An approach to support task specific views of the data. A vision. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the 2nd International Workshop on Linked Science 2012 – Tackling Big Data (LISC2012), in conjunction with 11th International Semantic Web Conference (ISWC2012)*, Boston, MA, 2012.
- [13] John Breslin, Andreas Harth, Uldis Bojars, and Stefan Decker. Towards Semantically-Interlinked Online Communities. In Asunción Gómez-Pérez and Jérôme Euzenat, editors, *The Semantic Web: Research and Applications*, volume 3532 of *Lecture Notes in Computer Science*, pages 71–83. Springer Berlin/Heidelberg, 2005.
- [14] Dan Brickley and Libby Miller. FOAF Vocabulary Specification 0.98. Available from <http://xmlns.com/foaf/spec/20100809.html>; last accessed 2012-08-26, 2010.
- [15] Boyan Brodaric and Marc Gahegan. Ontology Use for Semantic e-Science. *Semantic Web Journal*, 1:149–153, April 2010.
- [16] Paolo Ciccarese, David Shotton, Silvio Peroni, and Tim Clark. CiTO + SWAN: The Web Semantics of Bibliographic Records, Citations, Evidence and Discourse Relationships. *Semantic Web Journal*, accepted.
- [17] Michael Compton, Payam Barnaghi, Oscar Corcho, Simon Cox, Manfred Hauswirth, Cory Henson, Krzysztof Janowicz, Laurent Lefort, Amit Sheth, and Kerry Taylor. The SSN Ontology of the W3C Semantic Sensor Network Incubator Group. *Journal of Web Semantics*, forthcoming 2012.
- [18] Mathieu d’Aquin. Linked Data for Open and Distance Learning. Available from <http://www.col.org/resources/publications/Pages/detail.aspx?PID=420>; last accessed 2012-08-26, 2012.
- [19] Mathieu d’Aquin. Putting Linked Data to Use in a Large Higher-Education Organisation. In *Interacting with Linked Data workshop (ILD) at Extended Semantic Web Conference, ESWC, 2012*.
- [20] Mathieu d’Aquin. So, what’s in linked datasets for education? Available from the LUCERO blog at <http://lucero-project.info/lb/2012/04/so-whats-in-linked-datasets-for-education/>; last accessed 2012-08-28, 2012.
- [21] Mathieu d’Aquin, Carlo Allocca, and Trevor Collins. DiscOU: A Flexible Discovery Engine for Open Educational Resources Using Semantic Indexing and Relationship Summaries. In *Demo session of the International Semantic Web Conference, 2012*.
- [22] Paul A. David. Understanding the emergence of ‘open science’ institutions: functionalist economics in historical context. *Industrial and Corporate Change*, 13(4):571–589, 2004.
- [23] Giovana Mira de Espindola and Tomi Kauppinen. Semantic Linking of Data for the Brazilian Amazon Rainforest Statistics. In *GeoChange 2010—GIScience for Environmental Change*, Campos do Jordão, São Paulo, Brazil, November 2010.
- [24] Rafael de Santiago and André L.A. Raabe. Architecture for Learning Objects Sharing among Learning Institutions-LOP2P. *IEEE Transactions on Learning Technologies*, April-June:91–95, 2010.
- [25] Gianluca Demartini, Iliya Enchev, Joël Gapany, and Philippe Cudré-Mauroux. The Bowlogna Ontology: Fostering Open Curricula and Agile Knowledge Bases for Europe’s Higher Education Landscape. *Semantic Web Journal*, this issue.
- [26] Anusuriya Devaraju and Tomi Kauppinen. Geo-Processes and Properties Observed by Sensors: Can We Relate Them? In *GeoChange 2010—GIScience for Environmental Change*, Campos do Jordão, São Paulo, Brazil, November 2010.
- [27] Jörg Diederich and Wolf-Tilo Balke. FacetedDBLP – Navigational Access for Digital Libraries. *Bulletin of IEEE Technical Committee on Digital Libraries*, 4(1), 2008.
- [28] Stefan Dietze, Mathieu d’Aquin, and Dragan Gašević, editors. *Proceedings of Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age, in conjunction with the 8th Extended Semantic Web Conference, ESWC2011, Heraklion, Crete, 2012*.
- [29] Stefan Dietze, Mathieu d’Aquin, and Dragan Gašević, editors. *Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data, in conjunction with the World Wide Web Conference 2012 (WWW2012), Lyon, France, 2012*.
- [30] Stefan Dietze, Salvador Sanchez-Alonso, Hannes Ebner, Hong Qing, Daniela Giordano, Ivana Marenzi, and Bernardo Pereira Nunes. Interlinking educational Resources and the Web of Data – a Survey of Challenges and Approaches. *Emerald Program: electronic Library and Information Systems*, 47(1), 2013.
- [31] Dublin Core Metadata Initiative. DCMI Metadata Terms. Available online from <http://dublincore.org/documents/dcmi-terms/>; last accessed 2012-08-02, 2012.
- [32] Erik Duval, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel. Metadata Principles and Practicalities. *D-Lib Magazine*, 8(4), 2013.
- [33] European Committee for Standardization. Metadata for Learning Opportunities (MLO) - Advertising. Available from <ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/WS-LT/CWA15903-00-2008-Dec.pdf>; last accessed 2012-08-26, 2008.
- [34] Mitsopoulou Evangelia, Davide Taibi, Daniela Giordano, Stefan Dietze, Hong Qingand Yu, Panagiotis Bamidis, Charalampos Bratsas, and Luke Woodham. Connecting Medical Educational Resources to the Linked Data Cloud: the mEducator RDF Schema, Store and API. In *Proceedings of Linked Learning 2011, Proceedings of the 1st International Workshop on eLearning Approaches for the Linked Data Age, CEUR-WS, Vol. 717, Heraklion, Greece, 2011*.

- [35] Rebecca Ferguson. The State of Learning Analytics in 2012: A Review and Future Challenges. Knowledge Media Institute, Technical Report KMI-2012-01, 2012.
- [36] Muriel Foulonneau and Valentin Grouès. Common vs. Expert knowledge: making the Semantic Web an educational model. In *Proceedings of Linked Learning 2012 – The 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France, 2012.
- [37] Daniel Garijo and Yolanda Gil. Augmenting PROV with Plans in P-PLAN: Scientific Processes as Linked Data. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the 2nd International Workshop on Linked Science 2012 – Tackling Big Data (LISC2012)*, in conjunction with 11th International Semantic Web Conference (ISWC2012), Boston, MA, 2012.
- [38] Frédéric Giasson and Bruce D’Arcus. Bibliographic Ontology Specification. Available online from <http://bibliontology.com/specification>; last accessed 2011-12-29, 2009.
- [39] Yolanda Gil and Paul Groth. Linkeddatalens: Linked data as a network of networks. In *Proceedings of the sixth international conference on Knowledge capture, K-CAP ’11*, pages 191–192, New York, NY, USA, 2011. ACM.
- [40] Carole Anne Goble and David Charles De Roure. myExperiment: social networking for workflow-using e-scientists. In *Workshop on Workflows in support of large-scale science*, 2007.
- [41] Bernardo Gonçalves, Fabio Porto, and Ana Maria Moura. On the Semantic Engineering of Scientific Hypotheses as Linked Data. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the 2nd International Workshop on Linked Science 2012 – Tackling Big Data (LISC2012)*, in conjunction with 11th International Semantic Web Conference (ISWC2012), Boston, MA, 2012.
- [42] Paul Groth and Yolanda Gil. Linked Data for Network Science. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the First International Workshop on Linked Science 2011, in conjunction with the International Semantic Web Conference (ISWC2011)*. Bonn, Germany, October 24, 2011.
- [43] Olaf Hartig. Provenance Information in the Web of Data. In *Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009)*, 2009.
- [44] Bernhard Haslhofer, Elaheh Momeni, Manuel Gay, and Rainer Simon. Augmenting Europeana Content with Linked Data Resources. In *Proceedings of the 6th International Conference on Semantic Systems*. ACM, 2010.
- [45] Julia Hauser. Linked Data Service of the German National Library. Available from http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html, July 2012.
- [46] IEEE. IEEE Standard for Learning Object Metadata, IEEE Std 1484.12.1-2002. doi: 10.1109/IEEESTD.2002.94128, 2002.
- [47] IFLA Study Group on the Functional Requirements of Bibliographic Records. Functional Requirements of Bibliographic Records, 1998.
- [48] Krzysztof Janowicz, Sven Schade, Arne Bröring, Carsten Keßler, Patrick Maué, and Christoph Stasch. Semantic Enablement for Spatial Data Infrastructures. *Transactions in GIS*, 14(2):111–129, 2010.
- [49] A. Jentzsch, O. Hassanzadeh, C. Bizer, B. Andersson, and S. Stephens. Enabling Tailored Therapeutics with Linked Data. In *Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009)*, 2009.
- [50] Zoran Jeremić, Jelena Jovanović, and Dragan Gašević. Personal Learning Environments on the Social Semantic Web. *Semantic Web Journal*, this issue.
- [51] Julien Robinson. Johann Stan, Myriam Ribiere. Using linked data to reduce learning latency for e-book readers. In *Proceedings of Linked Learning 2011 - the 1st International Workshop on eLearning Approaches for Linked Data Age*, Heraklion, Greece, 2011.
- [52] Eleni Kaldoudi, Nikolas Dovrolis, and Stefan Dietze. Information Organization on the Internet based on Heterogeneous Social Networks. In *Proceedings of 29th ACM International Conference on Design of Communication (ACM SIGDOC&AZ11)*, Pisa, Italy, 2011.
- [53] Tomi Kauppinen, Alkyoni Baglatzi, and Carsten Keßler. Linked Science: Interconnecting Scientific Assets. In Terence Critchlow and Kerstin Kleese-Van Dam, editors, *Data Intensive Science*. CRC Press, USA, forthcoming 2012.
- [54] Tomi Kauppinen and Giovana Mira de Espindola. Linked Open Science—Communicating, Sharing and Evaluating Data, Methods and Results for Executable Papers. *Proceedings of the International Conference on Computational Science (ICCS 2011)*, *Procedia Computer Science*, 4(0):726–731, 2011.
- [55] Tomi Kauppinen and Giovana Mira de Espindola. Ontology-Based Modeling of Land Change Trajectories in the Brazilian Amazon. In *Geoinformatik 2011—GeoChange*, Münster, Germany, June 2011.
- [56] Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors. *Proceedings of the 1st International Workshop on Linked Science 2011 (LISC2011)*, in conjunction with 10th International Semantic Web Conference (ISWC2011), Bonn, Germany, 2011.
- [57] Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors. *Proceedings of the 2nd International Workshop on Linked Science 2012 – Tackling Big Data (LISC2012)*, in conjunction with 11th International Semantic Web Conference (ISWC2012), Boston, MA, 2012.
- [58] Tomi Kauppinen and Johannes Trame. Teaching Core Vocabulary Specification. Available from <http://linkedscience.org/teach/ns>; last accessed 2012-08-26, 2011.
- [59] Carsten Keßler and Krzysztof Janowicz. Linking Sensor Data—Why, to What, and How? In Kerry Taylor, Arun Ayyagari, and David De Roure, editors, *Proceedings of the 3rd International workshop on Semantic Sensor Networks 2010 (SSN10) in conjunction with the 9th International Semantic Web Conference (ISWC 2010)*, volume 668 of *CEUR-WS*, Shanghai, China, 2010.
- [60] Carsten Keßler, Krzysztof Janowicz, and Tomi Kauppinen. spatial@linkedscience—Exploring the Research Field of GIScience with Linked Data. In Ningchuan Xiao, Mei-Po Kwan, Michael F. Goodchild, and Shashi Shekhar, editors, *Geographic Information Science. 7th International Conference, GIScience 2012, Columbus, OH, USA, September 18–21, 2012, Proceedings*, volume 7478 of *Lecture Notes in Computer Science*, pages 102–117, 2012.
- [61] Carsten Keßler and Tomi Kauppinen. Linked Open Data University of Münster—Infrastructure and Applications. In *Demos*

- of the *Extended Semantic Web Conference 2012 (ESWC2012)*, Heraklion, Crete, Greece, May 2012.
- [62] Craig A. Knoblock, Pedro Szekely, Jose Luis Ambite, Shubham Gupta, Aman Goel, Maria Muslea, Kristina Lerman, and Parag Mallick. Interactively Mapping Data Sources into the Semantic Web. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the First International Workshop on Linked Science 2011, in conjunction with the International Semantic Web Conference (ISWC2011)*. Bonn, Germany, October 24, 2011.
- [63] WonSuk Lee, Werner Bailer, Tobias Bürger, Pierre-Antoine Champin, Jean-Pierre Evain, Véronique Malaisé, Thierry Michel, Felix Sasaki, Joakim Söderberg, Florian Stegmaier, and John Strassner. Ontology for media resources 1.0. Available from <http://www.w3.org/TR/mediaont-10/>; last accessed 2012-08-26, 2012.
- [64] Library of Congress. Subject Headings. Available from <http://id.loc.gov/authorities/subjects.html>; last accessed 2012-08-26.
- [65] Library of Congress. MARC 21 Format for Bibliographic Data. Available online from url: <http://www.loc.gov/marc/bibliographic/ecbdhome.html>; last accessed 2012-08-13, 1999–2012.
- [66] Patrick Maué, Henry Michels, and Marcell Roth. Injecting semantic annotations into (geospatial) Web service descriptions. *Semantic Web Journal*, accepted.
- [67] James P. McCusker, Timothy Lebo, Li Ding, Cynthia Chang, Paulo Pinheiro da Silva, and Deborah L. McGuinness. Where did you hear that? Information and the Sources They Come From. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the First International Workshop on Linked Science 2011, in conjunction with the International Semantic Web Conference (ISWC2011)*. Bonn, Germany, October 24, 2011.
- [68] Pablo N. Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. DBpedia Spotlight: Shedding Light on the Web of Documents. In *International Conference on Semantic Systems (I-Semantics)*, 2011.
- [69] Barend Mons, Herman van Haagen, Christine Chichester, Peter-Bram 't Hoen, Johan T den Dunnen, Gertjan van Ommen, Erik van Mulligen, Bharat Singh, Rob Hoof, Marco Roos, Joel Hammond, Bruce Kiesel, Belinda Giardine, Jan Velterop, Paul Groth, and Schultes. The value of data. *Nature Genetics*, 43:281–283, 2011.
- [70] Johnson Mwebaze, Danny Boxhoorn, and Edwin Valentijn. Supporting Scientific Collaboration Through Class-Based Object Versioning. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the First International Workshop on Linked Science 2011, in conjunction with the International Semantic Web Conference (ISWC2011)*. Bonn, Germany, October 24, 2011.
- [71] Vinh Nguyen, Olivier Bodenreider, Todd Minning, and Amit Sheth. The Knowledge-driven Exploration of Integrated Biomedical Knowledge Sources Facilitates the Generation of New Hypotheses. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the First International Workshop on Linked Science 2011, in conjunction with the International Semantic Web Conference (ISWC2011)*. Bonn, Germany, October 24, 2011.
- [72] Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, and Markus Zanker. Linked Open Data to support Content-based Recommender Systems. In *International Conference on Semantic Systems (I-Semantics)*, 2012.
- [73] Sebastian Nordhoff and Harald Hammarström. Glotlog/Langdoc: Defining Dialects, Languages, and Language Families as Collections of Resources. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the First International Workshop on Linked Science 2011, in conjunction with the International Semantic Web Conference (ISWC2011)*. Bonn, Germany, October 24, 2011.
- [74] Kevin Page, Raúl Palma, Piotr Ho, Graham Klyne, Stian Soiland-Reyes, Don Cruickshank, Rafael González Cabero, Esteban García Cuesta, David De Roure, Jun Zhao, and José Manuel Gómez-Pérez. From workflows to Research Objects: an architecture for preserving the semantics of science. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the 2nd International Workshop on Linked Science 2012 – Tackling Big Data (LISC2012), in conjunction with 11th International Semantic Web Conference (ISWC2012)*, Boston, MA, 2012.
- [75] Kevin R. Page, David C. De Roure, Kirk Martinez, Jason D. Sadler, and Oles Y. Kit. Linked Sensor Data: RESTfully serving RDF and GML. In *Semantic Sensor Networks 2009 (SSN09)*, October 2009.
- [76] Alexandre Passant, Uldis Bojars, John Breslin, and Stefan Decker. The SIOC Project: Semantically-Interlinked Online Communities, from Humans to Machines. In Julian Padget, Alexander Artikis, Wamberto Vasconcelos, Kostas Stathis, Viviane da Silva, Eric Matson, and Axel Polleres, editors, *Coordination, Organizations, Institutions and Norms in Agent Systems V*, volume 6069 of *Lecture Notes in Computer Science*, pages 179–194. Springer Berlin/Heidelberg, 2010. 10.1007/978-3-642-14962-7_12.
- [77] Guillermo Alvaro Rey, Irene Celino, Mariana Damova, Danica Damljanovic, Ning Li, Panos Alexopoulos, and Vladan Devedzic. Semi-Automatic Generation of Quizzes and Learning Artifacts from Linked Data. In *Proceedings of Linked Learning 2012 - The 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France, 2012.
- [78] Dave Reynolds. An organization ontology. Available from <http://www.w3.org/TR/vocab-org/>; last accessed 2012-08-26, 2012.
- [79] Hajo Rijgersberg, Mark van Assem, and Jan Top. Ontology of Units of Measure and Related Concepts. *Semantic Web Journal*, this issue.
- [80] Ben Ryan, Mark Stubbs, and Scott Wilson. XCRI CAP 1.1. Available from http://www.xcri.org/wiki/index.php/XCRI_CAP_1.1; last accessed 2012-08-26, 2009.
- [81] Nadeem Shabir. JACS (Joint Academic Coding System) Data Set. Available from <http://jacs.dataincubator.org/>; last accessed 2012-08-26, 2012.
- [82] Amit Sheth, Cory Henson, and Satya S. Sahoo. Semantic Sensor Web. *Internet Computing, IEEE*, 12(4):78–83, 2008.
- [83] David Shotton. CiTO, the Citation Typing Ontology. *Journal of Biomedical Semantics*, 1(Suppl. 1):S6, 2010.
- [84] David Shotton and Silvio Peroni. FaBio, the FRBR-aligned Bibliographic Ontology, 2012.
- [85] David Shotton, Katie Portwin, Graham Klyne, and Alistair Miles. Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Comput Biol*, 5(4), 2009.

- [86] Christoph Stasch, Sven Schade, Alejandro Llaves, Krzysztof Janowicz, and Arne Bröring. Aggregating Linked Sensor Data. In Kerry Taylor, Arun Ayyagari, David De Roure, editor, *The 4th International Workshop on Semantic Sensor Networks 2011 (SSN11) in conjunction with the 10th International Semantic Web Conference (ISWC 2011)*, Bonn, Germany, 2011.
- [87] Rob Styles and Nadeem Shabir. Academic Institution Internal Structure Ontology (AIISO). Available from <http://purl.org/vocab/aiiso/schema#>; last accessed 2012-08-26, 2008.
- [88] The Learning Resource Metadata Initiative (LRMI). LRMI Specification version 1.0. Available from <http://www.lrmi.net/the-specification>; last accessed 2012-08-26, 2012.
- [89] Nadeem Shabir Chris Clarke Justin Leavesley Tom Heath, Ross Singer. Assembling and applying an education graph based on learning resources in universities. In *Proceedings of Linked Learning 2012 – The 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France, 2012.
- [90] Antony J. Williams, Lee Harland, Paul Groth, Stephen Pettifer, Christine Chichester, Egon L. Willighagen, Chris T. Evelo, Niklas Blomberg, Gerhard Ecker, Carole Goble, and Barend Mons. Open PHACTS: semantic interoperability for drug discovery. *Drug Discovery Today*, in press.
- [91] World Wide Web Consortium. SPARQL Query Language for RDF. Available online from <http://www.w3.org/TR/rdf-sparql-query/>; last accessed 2012-08-02, 2008.
- [92] Reynold S. Xin, Oktie Hassanzadeh, Christian Fritz, Shirin Sohrabi, and Renée J. Miller. Publishing Bibliographic Data on the Semantic Web using BibBase. *Semantic Web Journal*, this issue.
- [93] Varun Ratnakar Yolanda Gil and Paul Hanson. Organic Data Publishing: A Novel Approach to Scientific Data Sharing. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the 2nd International Workshop on Linked Science 2012 – Tackling Big Data (LISC2012), in conjunction with 11th International Semantic Web Conference (ISWC2012)*, Boston, MA, 2012.
- [94] Hong Qing Yu, Stefan Dietze, Ning Li, Carlos Pedrinaci, Davide Taibi, Nikolas Dovrolls, Teodor Stefanut, Eleni Kaldoudi, and John Domingue. A linked data-driven & service-oriented architecture for sharing educational resources. In *1st International Workshop on eLearning Approaches for Linked Data Age (Linked Learning 2011), 8th Extended Semantic Web Conference (ESWC2011)*, 2011. Workshop proceedings published as CEUR Vol 717.
- [95] Hong Qing Yu, Stefan Dietze, Ning Li, Carlos Pedrinaci, Davide Taibi, Nikolas Dovrolls, Teodor Stefanut, Eleni Kaldoudi, and John Domingue. A linked data-driven & service-oriented architecture for sharing educational resources. In *Proceedings of Linked Learning 2011 – the 1st International Workshop on eLearning Approaches for Linked Data Age*, Heraklion, Greece, 2011.
- [96] Fouad Zablith, Miriam Fernandez, and Matthew Rowe. The OU Linked Open Data: Production and Consumption. In *Proceedings of Linked Learning Workshop, Extended Semantic Web Conference 2011. Heraklion, Crete*, 2011.
- [97] Jun Zhao, Graham Klyne, Piotr Holubowicz, Raul Palma, Stian Soiland-Reyes, Kristina Hettne, Jose Enrique Ruiz, Marco Roos, Kevin Page, Jose Manuel Gomez-Perez, David De Roure, and Carole Goble. RO-Manager: A Tool for Creating and Manipulating Research Objects to Support Reproducibility and Reuse in Sciences. In Tomi Kauppinen, Line C. Pouchard, and Carsten Keßler, editors, *Proceedings of the 2nd International Workshop on Linked Science 2012 – Tackling Big Data (LISC2012), in conjunction with 11th International Semantic Web Conference (ISWC2012)*, Boston, MA, 2012.