Journal of Universal Computer Science, vol. 21, no. 2 (2015), 326-364
submitted: 26/4/14, accepted: 5/1/15, appeared: 1/2/15 © J.UCS

A Review of Linked Data Proposals in the Learning Domain

Guillermo Vega-Gorgojo
(Universidad de Valladolid, Valladolid, Spain
Universitetet i Oslo, Oslo, Norway
guiveg@tel.uva.es, guiveg@ifi.uio.no)

Juan I. Asensio-Pérez, Eduardo Gómez-Sánchez
Miguel L. Bote-Lorenzo, Juan A. Muñoz-Cristóbal
(Universidad de Valladolid, Valladolid, Spain
{juaase@tel, edugom@tel, migbot@tel, juanmunoz@gsic}.uva.es)

Adolfo Ruiz-Calleja
(Tallinn University, Tallinn, Estonia
adolfo@tlu.ee)

Abstract: This study critically reviews the recently published scientific literature on Linked Data proposals in the educational field. After systematically searching online bibliographic databases, 33 original works satisfied the scope and quality criteria, and thus were included in this review. Studies were classified with respect to TEL research areas; interoperability, personalization and contextualized learning were the main areas addressed. Many studies have a foundation on learning object and repository research, where Linked Data practices are applied to simplify the integration of educational datasets. As learning institutions are gradually exposing their key datasets as Linked Data, an emergent educational data web is being constituted. A number of the reviewed works consume these data for different purposes, reporting reusability and enrichment benefits. Nevertheless, upcoming proposals should be aware of existing challenges, derived from the Linked Data model, such as the lack of control of data sources or varying degrees of data quality. We also give some recommendations for delivering Linked Data-based proposals in education, including a classification of vocabularies, datasets and technological products. Future research directions include the release of new datasets as Linked Data, federation and interlinking practices to improve the cohesion of the emergent educational Web of Data, generation of learning artifacts, curation and enrichment of educational data, novel educational applications consuming Linked Data, and performance improvements.
Key Words: Computer Uses in Education, Intelligent Web Services and Semantic Web, Libraries/information repositories/publishing
Category: K.3.1, I.2.12, J.8.12

1 Introduction

The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web [Bizer et al., 2009a]. Following these practices, publishers can expose their datasets using the infrastructure of the Web and, more importantly, provide links to other pieces of data in order to facilitate the discovery of further information. Linked Data are published using Semantic Web technologies such as the RDF [Cyganiak et al., 2014] family of standards for data interchange and SPARQL [Prud'hommeaux and Seaborne, 2008] for query. Indeed, the adoption of the Linked Data principles has emerged as the means to reach the Semantic Web vision [Berners-Lee et al., 2001], thus leading to the creation of the Web of Data — a giant, global and interconnected database that improves the way we discover, access, integrate and use data. In stark contrast to the traditional Web of Documents, the Web of Data allows: semantic searches referred to concepts instead of word occurrence [Guha et al., 2003]; retrieval of contextual information by following Linked Data typed links [Bizer et al., 2009a]; and data merging from multiple sources [Allemang and Hendler, 2008, ch. 3].

Since the proposal of the Linked Data principles in 2006 [Berners-Lee, 2006], the Web of Data has exponentially grown, forming a global data space from numerous sources [Heath and Bizer, 2011, ch. 3]. The sort of topics covered is impressive, ranging from geographic locations to books, scientific publications, films, genes, or census information, to name a few. With such a significant volume of Linked Data published on the Web, a new breed of applications that exploits this Web of Data has emerged [Bizer et al., 2009a]. For instance, there are Linked Data browsers such as Tabulator [Berners-Lee et al., 2008] that allow users to navigate the Web of Data, e.g. from the RDF description of Miguel de Cervantes in a dataset, follow the *birthPlace* link to the description of Alcalá de Henares, and from there obtain from other datasets the coordinates of this city and the list of UNESCO's World Heritage Sites in Spain — a difference from the Web of Documents is that here the links are typed, so users, and especially applications, do not have to figure out the nature of a relationship such as *birthPlace*. There are also semantic search engines, e.g. Falcons [Cheng and Qu, 2009], that crawl Linked Data sources and provide different querying capabilities, e.g. searching for concrete items such as people or concept searches for locating classes and properties. Domain-specific applications that exploit Linked Data technologies are also being developed; for example, BBC Programmes and Music [Kobilarov et al., 2009] uses Linked Data for repository integration and augments BBC's own content with additional information from the Web of Data, e.g. MusicBrainz[1], thus dramatically reducing BBC's authoring effort.

In the learning domain, significant research efforts have been traditionally devoted to the application of Semantic Web technologies in areas such as metadata, learning objects, recommendation systems or intelligent tutoring systems — see [Sampson et al., 2004, Devedzic, 2006, Naeve et al., 2006, Bertini et al., 2011, Carmichael and Jordan, 2012, Tiropanis et al., 2012]. Early demonstrators have shown the benefits of semantics for supporting the discovery and delivery of cur-

[1] `http://musicbrainz.org/`

ricular content [Carmichael and Jordan, 2012]. However, raised expectations like learning object reusability or seamless repository integration have not been completely attained, due to issues such as lack of terminology consensus or difficulties to annotate large volumes of learning content [Tiropanis et al., 2012]. Moreover, these initiatives were hindered by the scarcity of semantic data available for reusing and sharing [Shadbolt et al., 2006, Hendler, 2008]. As a result, early semantic developments in the education field had to produce their own datasets, typically incurring in high authoring costs, and leading to a landscape of disconnected data islands. As an example, Ontoolsearch [Vega-Gorgojo et al., 2010] is a semantic search engine of educational tools that employs its own isolated dataset that is hardly updated due to authoring costs.

Nevertheless, the situation has dramatically changed with the adoption of the Linked Data principles and the consolidation of the Semantic Web technology stack, so large volumes of data are readily available. In this regard, a number of educational institutions are releasing some of their key datasets as Linked Data — Linked Universities[2] lists some of these institutions. Furthermore, there are ongoing initiatives to bootstrap the creation of an emergent educational Web of Data: Linked Education[3] collects educational datasets and applications based on Linked Data; Linked Education Cloud[4] is a catalogue of datasets relevant to the learning domain; and LinkedUp Challenge[5] is a competition of educational Linked Data-based applications. Noteworthy, journals and conferences on educational technology are publishing an increasing number of research works in Linked Data and there are some specialized workshops under way, e.g. Linked Learning[6] and Learning Analytics and Linked Data[7].

At this stage, the Linked Data movement promises to significantly improve existing practices of system integration, resource sharing and personalization to support learning [Tiropanis et al., 2012]. Moreover, the emergent educational data web offers outstanding opportunities for data reuse across institutional and application boundaries. As a note of caution, early experiences report that setting up a Linked Data service is not trivial [Hannemann and Kett, 2010]. Furthermore, academics and students expect easy-to-use applications that improve their current practice [Carmichael and Jordan, 2012].

In order to clarify the current landscape of Linked Data proposals in the learning domain, we aim to review the state of the art of existing research

[2] http://linkeduniversities.org
[3] http://linkededucation.org
[4] http://data.linkededucation.org/linkedup/catalog/
[5] http://linkedup-challenge.org/
[6] http://lile2013.wordpress.com/, in conjunction with the 22nd International World Wide Web Conference (WWW2013) — http://www2013.org/
[7] http://lald2012.wordpress.com/, in conjunction with the 2nd International Conference on Learning Analytics and Knowledge (LAK12) — http://lak12.sites.olt.ubc.ca/

works reported in the literature. Therefore, we attempt to provide a coherent picture of existing proposals, as well as to assess the advantages and limitations of Linked Data for the learning realm. Moreover, this review can serve to guide upcoming educational Linked Data projects by distilling a set of recommendations. Further, our final goal is to foresee possible research opportunities in this field. To conduct this research we follow a systematic literature review [Kitchenham and Charters, 2007] which provides a framework for methodically searching the literature, extracting the data and performing the necessary analysis. Note that the focus on Linked Data affordances in the education domain is a unique characteristic of this review. While there are some works that cover the use of Semantic Web technologies in learning such as [Devedzic, 2006], only [Dietze et al., 2013] and [d'Aquin, 2012] explicitly address the topic of Linked Data in the educational field. In this regard, [Dietze et al., 2013] presents a related survey of challenges and approaches for interlinking educational resources, although the main focus relies on Web APIs and interfaces for repository integration. [d'Aquin, 2012] is a report about the opportunities of Linked Data for open and distance learning. These two works are used as input sources in this review.

The rest of this paper is organized as follows: Section 2 discusses the methodology we have followed to perform the review. Section 3 presents a classification of the Linked Data research works that have been selected. Next, section 4 extracts the advantages and limitations of Linked Data found in the domain. Section 5 follows with a set of recommendations for publishing and consuming educational Linked Data. Section 6 concludes with a discussion identifying open research questions. We also include an Appendix A with background information of Linked Data and the Web of Data, while Appendix B summarizes the studies reviewed in this paper.

2 Review method

After outlining Linked Data and the Web of Data, we aim to review the existing research works in the learning domain. More specifically, the goals of this study are fourfold: (1) to identify Linked Data developments in education and provide a coherent classification; (2) to acknowledge the advantages and limitations of Linked Data for the learning realm; (3) to extract recommendations for delivering Linked Data-based proposals in education; and (4) to identify current trends and open research questions in this field.

To carry out this research, we have followed a systematic literature review [Kitchenham and Charters, 2007]. Though this type of reviews requires more effort than traditional ones, the results obtained with a systematic procedure are less likely to be biased. Note that [Kitchenham and Charters, 2007] is primarily aimed at software engineering, although it can be applied in a wider set of contexts, e.g. adaptive and intelligent systems for collaborative learning [Magnisalis et al., 2011]; indeed, this methodology is based on existing guidelines for systematic reviews in medical and social sciences.

When conducting the review, we selected the main electronic databases in Computer Science: IEEE Xplore, ScienceDirect, SciVerse Scopus, ISI Web of Science, ACM Digital Library, SpringerLink and Google Scholar. In addition, we manually searched the proceedings of the following workshops: Linked Learning (2011 and 2012 editions), and Learning Analytics and Linked Data 2012, since they were not indexed in the aforementioned databases, but they are still relevant for the scope of this review. To perform the search, we broke down the question into a learning and a Linked Data facets. After aggregating synonyms, the resulting search string was `("Linked Data" OR "Web of Data" OR "Data Web" OR "Open Data" OR "Linked Open Data") AND (Learning OR eLearning OR e-learning OR education OR educative OR teaching)`. We applied this search to the title, abstract and keywords in most cases, since not all electronic databases support them. In addition, search results were restricted to articles published in journals and conference proceedings from 2006 onwards for additional filtering, given that the Linked Data principles were proposed in 2006. The literature search was undertaken in June 2012, and then repeated in March 2013.

After performing the searches, each candidate study passed through a set of stages until its eventual selection: (1) assess the title and the publication name, discard if not related to engineering or technology; (2) read the abstract, exclude if unrelated to Linked Data and learning; and (3) retrieve the study and critically appraise it, discard in case of out of scope, no credibility, minor contribution or low quality. Then, a data extraction form[8] was filled for each selected study in order to gather evidence for the review goals.

Overall, 763 studies were retrieved, 464 abstracts were considered in stage (2), and 109 were thoroughly reviewed. In the second search we used a script to detect duplicates from the previous iteration, obtaining 34 (out of 109) new candidate studies in stage (3). The first two stages of this workflow were particularly effective, for instance, to filter out unrelated papers with the common expression "*we linked data*". In the third stage, the inclusion/exclusion decision of each study was taken by one of the authors playing the referee role and who was also in charge of filling its corresponding data extraction form. A pilot study was carried out beforehand to refine the data extraction form and to unify criteria among the referees (the authors of this review). In this regard, a candidate study had to make a relevant contribution based on Linked Data for the learning domain. Further, 18 doubtful studies were screened in a panel with all the

[8] Available at `http://www.gsic.uva.es/reviewLinkedDataEducation/ReviewForm.pdf`

Table 1: Publication types and distribution in time of the selected studies.

Publication type	2009	2010	2011	2012	2013	2009–2013
# conference proceedings	1	4	6	8	2	**21**
# journal papers	1		2	6	2	**11**
# reports				1		**1**
TOTAL	**2**	**4**	**8**	**15**	**4**	**33**

referees to reach a decision. After this process, 54 studies were selected. In the cases that multiple papers referred to the same proposal, we opted to choose the most representative one, as suggested in [Kitchenham and Charters, 2007]. Therefore, 33 studies are cited in this review – 12 as a result of the second search iteration (2 from 2011, 6 from 2012, and 4 from 2013). A summary of each of the cited papers is provided in Table 10, in Appendix B.

Table 1 shows the distribution along the time of the selected works, as well as their publication type. Each year the number of studies increases[9], thus reflecting a growing interest of Linked Data in the education domain. As a consequence, this seems a timely moment for a review on this topic.

3 Classification of the selected educational Linked Data proposals

The 33 studies selected from the literature present a captivating picture of disparate approaches that exploit Linked Data in the learning domain, as summarized in Appendix B. This heterogeneity can be attributed to the divergent backgrounds, perspectives and purposes of the research groups involved. To untangle this diversity, we first aim to classify the proposals under an educational point of view. To this end, we have employed the Technology Enhanced Learning (TEL) research areas identified in [Sutherland et al., 2012] — the outcome of a recent and ambitious study for providing a coherent view of the TEL landscape and elaborating a strategic agenda. These areas provide a good coverage of the most relevant research themes in the TEL field, ranging from technological issues, e.g. interoperability, to different types of learning practices. Consequently, referees were asked to classify each assigned study under one or more TEL research areas.

The resulting classification of the proposals is shown in Table 2. It can be seen that they are not uniformly distributed within the TEL research areas. The most referenced area is *Interoperability* (52% of the proposals); this is not surprising since Linked Data and Semantic Web technologies are commonly employed

[9] We expect an increase in 2013, but our last literature search was in March 2013.

Table 2: Classification of proposals with respect to TEL research areas.

TEL research area	Proposals	#
Computer-Supported Collaborative Learning	[Rob11, Tir09]	2
Connection between Formal and Informal Learning	[Sve10, Zab12]	2
Contextualized Learning	[Abe11, dAq12, Die12, Fer11, Lam12, Sve10, Tir09, Wai10, Yu12]	9
Emotional and Motivational Aspects of TEL	–	0
Games Enhanced learning	[Bra12]	1
Improving Practices of Formal Education	[dAq13a, Dav10, Hea12, Jun11, Lam12, Sha09, Tir09, Sve10]	8
Informal Learning	[Alo12, Sve10, Wai10, Zab12]	4
Interoperability	[dAq13a, Dav10, DeV12, Die12, Die13, Fer11, Isa12, Jer13, Lam12, Pir10, RuC12, RuR11, Sch12, Sic11, Sve10, Tir09 , Zab12]	17
Personalization of Learning	[Abe11, Alo12, dAq12, Jun11, Lam12, RuC12, Sia12, Tir09, Wai10, Yoo11, Yu12]	11
Reducing the Digital Divide	–	0
Technology Enhanced Assessment	[Bra12, Fou11, Rey12]	3
Ubiquitous and Mobile Technology and Learning	[Pie12, Rob11]	2
Workplace Learning	[Abe11, Sia12]	2

for system interconnection. As an example, [Fernandez et al., 2011] proposes a federated repository of video lectures by exploiting the Linked Data principles to provide the necessary structure and interlinking. A significant number of the proposals tackle the TEL research areas of *Personalization of Learning* (33%) and *Contextualized Learning* (27%); in these cases, Linked Data provide access to a huge amount of structured data, allowing fined-grained queries to retrieve contextual and customized data — see for instance [Ruiz-Calleja et al., 2012] that proposes a semantic search system of educational tools out of the Web of Data.

A notable number of proposals covers the areas of *Improving Practices of Formal Education* (24%) and *Informal Learning* (12%); they are composed of Linked Data-based applications designed to improve existing practices both in formal, e.g. recommendation lists in Higher Education [Heath et al., 2012], and informal learning, e.g. services for tracking early childhood education and care [Alonso-Roris et al., 2012]. In the remaining TEL areas, the coverage of analyzed proposals is marginal or even non-existent. This may be due to the fact that these areas are more specific than the precedent ones. However, these proposals are especially innovative since they apply Linked Data in non-evident application areas: [Bratsas et al., 2012] and [Foulonneau, 2011] use the Web of Data to generate quizzes and assessment items, respectively; [Siadaty et al., 2012] presents

Table 3: Learning categories and topics covered by the proposals.

Category	Topic	Proposals	#
Educational material	Annotation	[Alo12, DeV12, Die13, Fer11, Jun11, Lam12, RuR11, Yu12]	8
	Ebook	[Rob11]	1
	Learning objects	[dAq12, DeV12, Die13, Jun11, Isa12, Lam12, Sic11, Sve10, Yoo11]	9
	Math lecture notes	[Dav10]	1
	Quizzes	[Bra12, Fou11, Rey12]	3
	Resource lists	[Hea12, Sha09]	2
	Videos	[Fer11, Wai10, Yu12]	3
Systems	Dataset integration	[dAq13, Die12, Fer11, Jer13, RuC12, Yoo11]	5
	Services & tools	[Die12, Die13, RuC12]	3
	Learning management systems	[RuR11]	1
	Libraries	[Sch12]	1
	Personal learning environments	[Jer13, Sia12]	2
	Recommendation systems	[Abe11, Hea12, Pie12, Zab12]	4
	Repositories	[dAq12, dAq13, Die12, Die13, Isa12, Lam12, Pir10, Sic11, Yoo11, Zab12]	10
	Search	[Dav10, Isa12, RuC12, Wai10, Yu12, Zab12]	6
Educational setting	Computer-supported collaborative learning	[Rob11, Tir09]	2
	Early childhood education	[Alo12]	1
	Distance learning	[dAq12, Yu12]	2
	Higher Education	[Tir09]	1
	Mobile learning	[Pie12, Sve10]	2
	Workplace Learning	[Abe11, Sia12]	2
Other	Assessment	[Fou11, Rey12]	2
	Internationalization	[Bra12]	1
	Learning analytics	[dAq13a]	1
	Learning design	[RuR11]	1
	Social Web	[Abe11, Rob11, RuR11, Zab12]	4

a Linked Data-based platform for workplace learning; and [Piedra et al., 2012] proposes a mobile application that suggests courses and learning contents by consuming Linked Data sources, to name a few.

The precedent classification offers a coarse-grained picture of the Linked Data proposals in the educational domain. This view is complemented with a more detailed categorization of the learning topics addressed, as shown in Table 3. This classification emerges from the analysis of the keywords that were suggested by the referees for each assigned study. While referees could freely choose their own tags, we held a work session afterwards to homogenize the keywords and to derive the main categories. One of the most prominent categories is *Educational material*, that accounts for a 64% of the proposals. Indeed, many of these works apply the Linked Data principles to the traditional learning objects research in the education domain. In particular, some proposals deal with specific content types such as videos, quizzes or resource lists. Similar to the learning object literature, many proposals cover the topics of resource annotation and search; an important distinction here is that the Web of Data can be used as a source of annotations, although the automatic annotation of learning content is far from resolved.

67% of the proposals fit in the *Systems* category that encompasses diverse kinds of educational software (see Table 3). Repositories are particularly popular, since a number of works describe the process of exposing datasets as Linked Data, e.g. statistical data about Italian universities [Pirrotta, 2010]. This way, an organization can increase the value of their data, facilitating their consumption by other applications — see examples from the Open University of the UK in [Zablith et al., 2012]. Other works deal with the difficult problem of educational dataset integration. For instance, [Dietze et al., 2012] reports a Linked Data-based approach for integrating resources and services for biomedical education. In addition, a variety of systems use Linked Data to attain new desirable features: [Jeremic et al., 2013] proposes a Personal Learning Environment that is able to integrate and share disperse learning resources in online repositories; [Ruiz-Rube et al., 2011] presents an extension of the LAMS[10] system for describing desired learning resources and afterwards retrieving matching resources from a Linked Data repository; while [Heath et al., 2012] proposes a recommender system of resource lists by harnessing Linked Data about learning resources. Concerning resource search, the use of structured data enables the formulation of detailed queries.

With respect to the *Educational setting* category, there are some proposals targeted to specific settings such as distance or mobile learning. In the majority of the cases there was no reference to any educational context or pedagogy, so only 27% of the proposals were classified into *Educational setting*. This reflects that the Linked Data movement is mainly applied with a technological aim, so new systems, platforms and material are being delivered, while their impact in education practice has seldom been explicitly addressed.

The remaining *Other* category contains a diverse range of topics. Social Web is the most relevant one, explained by the popularity of the Web 2.0 in the education field [Andersen, 2007]; aligned with the objectives of the so-called Social Semantic Web [Breslin et al., 2009], proposals such as [Robinson et al., 2011] aim to gather the best of both worlds by extracting social data and converting them to Linked Data. Another topic is learning analytics and given its emergence and its focus on data processing [Siemens and Long, 2011], we can expect upcoming Linked Data proposals in learning analytics in the near future. The a priori integration of educational datasets for analytics, as well as the a posteriori enrich-

[10] `http://lamsfoundation.org/`

Table 4: Classification of proposals as Linked Data consumers, publishers, or both consumers and publishers.

Type	**Proposals**	#
Linked Data consumer	[Abe11, Alo12, Bra12, dAq13a, DeV12, Fou11, Pie12, Rob11]	8
Linked Data publisher	[Dav10, Hea12, Isa12, Jun11, Pir10, Sha09, Sia12, Sic11, Sve10]	9
Linked Data consumer and publisher	[Die12, Fer11, Jer13, Lam12, Rey12, RuC12, RuR11, Wai10, Yoo11, Yu12, Zab12]	11

ment of analytics results for facilitating their interpretation, are some of the cited synergies between Linked Data and learning analytics [d'Aquin and Jay, 2013].

After presenting the proposals under an educational perspective, we classify them in terms of Linked Data. For this purpose we have employed a simple taxonomy — though commonly employed in the Linked Data literature, e.g. in [Heath and Bizer, 2011] — that distinguishes among Linked Data consumers, publishers, or both consumers and publishers. After excluding secondary studies such as surveys, the resulting classification is presented in Table 4, showing a uniform distribution of the proposals in the aforementioned categories. Linked Data consumers take advantage of the availability of structured educational data for different purposes, especially to generate learning material or metadata annotations. Linked Data publishers contribute with new datasets to the educational Web of Data. The remaining category corresponds to cases that not only consume Linked Data, but also play the publisher role.

All in all, we can observe that Linked Data practices are being applied to address some existing problems of the education domain such as learning object authoring or repository interoperation. Other cases exploit Linked Data to improve the capabilities of learning systems, e.g. improved recommendations or support for fine-grained searches. Besides, educational Linked Data enable the advent of new consuming applications. Therefore, we can summarize the following uses of Linked Data in the learning domain: (1) exposure of educational datasets as Linked Data; (2) integration of disparate repositories; and (3) source of content for applications.

4 Advantages and limitations of Linked Data for the learning domain

After providing an overview of the educational uses of Linked Data proposals, we have extracted the reported benefits and drawbacks in the selected studies, ana-

Table 5: Linked Data advantages for the learning domain.

Advantage	**Brief description**	**Proposals**
Data reuse	Lots of Linked Data available for diverse application needs	[Alo12, Bra12, dAq12, dAq13a, Fou11, Hea12, Isa12, Jer13, Lam12, Rey12, RuC12, RuR11, Sch12, Sha09, Sic11, Tir09, Wai10, Yoo11, Zab12]
Enrichment	Use of the Web of Data to semantically enrich and interlink educational data	[Abe11, Alo12, dAq12, Die12, Fer11, Lam12, Rob11, Sve10, Yu12, Zab12]
Accessibility	Uniform and interoperable way to access educational Linked Data	[Bra12, dAq12, Dav10, Die13, Pir10, Sch12, Zab12]
Integration	Merging of educational data is simplified through the use of Linked Data and Semantic Web technologies	[dAq12, DeV12, Die12, Die13, Fer11, Isa12, Jer13, Jun11, Pie12, Pir10, Sch12, Sia12, Sic11, Sve10, Tir09, Wai10]
Discoverability	New data can be obtained by exploiting resource links	[dAq12, DeV12, Sch12, Zab12]
Internationalization	Multilingual datasets offer a simple way for application internationalization	[Bra12]
Accuracy	Linked Data annotations are much more accurate and explicit than other free text-based annotations	[Yu12]

lyzed them and finally synthesized the results in Tables 5 and 6. Beginning with the advantages (Table 5), *Data reuse* is the most referenced one; this fact emphasizes the disruptive nature of the Web of Data as a rich source of multi-domain structured information. This way, educational applications can directly query the Web of Data to satisfy their information needs: [Foulonneau, 2011] retrieves encyclopedic knowledge to generate assessment items; [Ruiz-Calleja et al., 2012] gathers descriptions of educational ICT tools; [Ruiz-Rube et al., 2011] obtains resources related to musical and cultural concepts; and [Waitelonis et al., 2010] gets new terms and concepts to make non-evident recommendations of academic videos, among others.

While Linked Data can be reused for different purposes, a number of studies employ the Web of Data for the *Enrichment* of educational resources. For example, [Robinson et al., 2011] enriches user annotations of ebooks with DBpedia concepts to provide context; [Lama et al., 2012] also makes use of the categories provided by DBpedia to improve the classification of a repository of learning objects. Since Linked Data practices do not impose a fixed set of vocabularies or data sources, educational resources can be easily enriched with unanticipated metadata [Dietze et al., 2013]. Given the high costs involved in metadata creation and maintenance [Currier et al., 2004], this use of Linked Data is especially promising for educational repositories and digital libraries.

As Linked Data offer a standardized and uniform way of consuming data with independence of their content or location, *Accessibility* of educational data is facilitated. In this regard, [Bratsas et al., 2012] remarks that Linked Data provide a structured way for developers to access knowledge, while [Dietze et al., 2013, Pirrotta, 2010, Zablith et al., 2012] indicate that the effort required to access and process educational Linked Data is low. Derived from the ease of data accessibility, a significant number of the studies observe advantages of Linked Data for the *Integration* of educational datasets. For example, [Dietze et al., 2012] achieves a federation of repositories in biomedical education; [Fernandez et al., 2011] integrates video lectures from different sources; and [Sicilia et al., 2011] presents a federation of learning repositories in the domain of organic agriculture. The overall approach for integration consists of exposing data sources as Linked Data, converting legacy datasets to RDF if necessary, and then setting links among datasets. Although there may be mismatches due to different vocabularies, mediation techniques can be used for term reconciliation, e.g. see [Isaac et al., 2012].

Another reported benefit is the *Discoverability* of additional data by exploiting the interconnections of datasets. This way, applications can traverse these links to discover new data at run-time. For example, the Social Study application [Zablith et al., 2012] explores similarities between student profiles and course offerings to recommend potentially interesting courses and study partners; while [De Vocht et al., 2012] proposes an environment to find knowledge about research topics, offering an interface for discovering conferences and researchers based on the learning objects they share. As for the remaining benefits in Table 5, [Bratsas et al., 2012] discusses that multilingual datasets provide a simple way for application *Internationalization*, illustrated with the generation of quizzes for primary education from the Greek DBpedia. Finally, [Yu et al., 2012] tested a video annotation tool with students, finding that Linked Data-based annotations were assessed as much more *Accurate* than free-text ones.

With respect to the limitations (see Table 6), many works report concerns derived from the data sharing model promoted by Linked Data practices. In this sense, it is important to understand that, instead of a unique authoring process with precisely defined rules, the Web of Data is composed of a myriad of datasets with varying degrees of *Data quality*. In spite of the heterogenous authorship of descriptions found in the Web of Data, an evaluation study [Ruiz-Calleja et al., 2012] showed that the quality of educational tools descriptions gathered from Linked Data sources was comparable to the descriptions

Table 6: Linked Data limitations for the learning domain.

Limitation	Brief description	Proposals
Data quality	Dealing with incorrect, incomplete or unexpected information from the Web of Data	[Bra12, dAq12, Fou11, Rey12, Sch12, Sha09, Wai10]
Data control	Lack of control of data sources, some are not maintained or disappear after some time	[RuR11, Sha09, Sch12, Sve10]
Data privacy	Concerns of exposing sensitive information	[Jer13, RuR11, Sve10, Tir09]
Data provenance	Need to know how different sources contributed to a given aggregated dataset	[Sch12, Sha09]
Data licensing	Acknowledge the right to reuse Linked Data	[Jer13, Sha09]
Vocabulary agreement	Need to agree on vocabularies for the educational domain	[Die13, Fer11, Sch12, Yu12]
Information loss	Mapping techniques may imply a loss of information	[RuC12, Sch12, Tir09]
Vocabulary availability	Need of vocabularies for specific domains	[Dav10, Zab12]
Enrichment cost	High costs of enriching and interlinking educational metadata	[Die13, Jer13, Rob11, RuR11, Sch12, Wai10]
Publication cost	High costs of publication of a dataset as Linked Data	[Die13, Fer11, Jun11, Tir09, Wai10, Zab12]
Computational cost	Traversing Linked Data graphs or querying a dataset are computing intensive tasks	[Lam12, Sve10, Tir09, Wai10, Yu12]
Interaction	Improve the navigation and visualization of Linked Data	[dAq12, Sch12]

authored by an expert, although DBpedia and Factforge were the only datasets employed.

Data may be contradictory, stale, or even discontinued due to a lack of *Data control*. While this situation is similar to the Web of Documents, Linked Data are assembled automatically without direct human supervision for discarding low quality content. As a result, Linked Data developers should take cautionary measures when reusing data, e.g. [Ruiz-Rube et al., 2011] suggests the use of tracking measures to discard outdated information. *Data privacy* is also a concern, so Linked Data practices should be applied with care when dealing with sensitive information such as student records — anonymizing data or including a security layer [Ruiz-Rube et al., 2011] are some measures that could be taken.

Another issue of Linked Data is *Data provenance*, i.e. once a number of datasets is combined, how to track the contribution of a piece of data to the aggregated view. This is important for weighting information, e.g. when recommending resources to achieve a particular learning goal [Shabir and Clarke, 2009]. Since data reuse is one of the major advantages of

Linked Data in education (see Table 5), it is important to acknowledge whether permission to reuse a dataset is granted. This way, *Data licensing* with specific terms of use should be promoted, e.g. [Shabir and Clarke, 2009] provides a explicit mechanism for licensing educational resources. Note that data provenance and data licensing issues are not specific to the use of Linked Data, although they introduce some important challenges to the effective use of Linked Data.

Other limitations stem from the publication of educational Linked Data. Lack of *Vocabulary agreement* is sometimes reported as a cause of fragmentation, e.g. [Fernandez et al., 2011], thus requiring mapping mechanisms for interoperation. Such data reconciliation processes may imply some *Information loss*, as in the conversion of bibliographic formats [Schreur, 2012] or educational tool descriptions [Ruiz-Calleja et al., 2012] from diverse sources. Additionally, *Vocabulary availability* is not ensured for every specific domain, so it may be necessary to develop new vocabularies, e.g. mathematical formulae representation [David et al., 2010]. While the use of the Web of Data for enrichment is considered a significant advantage in the educational domain (see Table 5), *Enrichment costs* are not negligible: enrichment typically requires finding suitable Linked Data entities [Jeremic et al., 2013], but fully automatic annotations are still challenging [Dietze et al., 2013, Schreur, 2012]. In addition, *Publication costs* of Linked Data are considered high in a number of works.

Another concern is the *Computational cost* of applications that process Linked Data; for example, [Lama et al., 2012] requires the exploration of large RDF graphs for obtaining relevant DBpedia categories to annotate learning objects, while [Yu et al., 2012] reports high response times of some searches. Finally, some works reflect about providing new ways of *Interaction* with the Web of Data, especially to improve navigation and visualization processes [d'Aquin, 2012].

5 Recommendations for delivering Linked Data proposals in the learning domain

While there are remarkable advantages such as the availability of a huge and vivid educational data web, publication and consumption of Linked Data entail a non-trivial effort. This section aims to help educational Linked Data researchers in upcoming developments. Specifically, we have collected the vocabularies, datasets and technological products that were referenced; we have then analyzed these artifacts; and finally categorized them to provide a coherent picture. In addition, we give some advise for publishing and consuming educational Linked Data, contextualizing general recommendations from the literature, e.g. [Heath and Bizer, 2011], to the learning domain.

Vocabularies are especially important for data publishers, since adhering to a popular one can simplify dataset integration processes or make a dataset automatically processable for applications. Therefore, **it is worth considering existing vocabularies rather than implementing a new one**. Even if there are no existing vocabularies for a specific domain, new developments should provide mappings to other general and related vocabularies (see [Gómez-Pérez et al., 2004] for a discussion on vocabulary development). In our study, we have identified the vocabularies presented in Table 7[11]. As can be seen, there is a broad coverage of topics, ranging from educational specific ones such as academic or learning resources, to horizontal topics, e.g. social web. We encourage prospective educational publishers to **check Table 7 in order to find relevant vocabularies**.

Some vocabularies such as AIISO, Dublin Core, DBpedia ontology, SKOS, or FOAF are widely employed, but the use of others is sparse. In many cases this can be attributed to the specificity of a domain, such as OpenMath for mathematical formulae [David et al., 2010]. Since not every reviewed paper is exhaustive when referring to the employed vocabularies, reported usage should be considered with care. Anyway, **further compatibility and interoperability gains can be achieved if adhering to a popular vocabulary**, as discussed above.

There are competing vocabularies that can be employed for similar purposes. For example, course information can be described with AIISO, Courseware or XCRI. Another example is the description of learning objects, since available choices include Dublin Core, LOM, LRMI, LOCWD, LOCO and SemUNT. **When choosing among different vocabularies, consider the following factors: popularity, simplicity, coherence, comprehensiveness, dataset compatibility, or availability of documentation**, as recommended in [d'Aquin, 2012, Heath and Bizer, 2011, ch. 4]. For instance, Dublin Core is a versatile and simple vocabulary that can be employed for basic annotations of learning material or bibliographic resources; however, other options should be considered for describing additional metadata elements such as learning objectives for learning material or journal information for bibliographic resources. Nevertheless, **vocabulary mappings can be created to reconcile terms if necessary**, e.g. [Ruiz-Calleja et al., 2012] uses mappings to homogenize educational tool descriptions from different sources.

With respect to datasets, Table 8[12] presents those referenced in the selected papers. **Datasets are of special importance for data consuming applications, and upcoming educational Linked Data initiatives should be aware of their presence**. In this regard, many Higher Education institutions are beginning to expose some of their core databases as Linked Data. These typically include course offerings, learning material, publications and university

[11] Available at `http://www.gsic.uva.es/reviewLinkedDataEducation/Vocabularies.pdf`

[12] Available at `http://www.gsic.uva.es/reviewLinkedDataEducation/Datasets.pdf`

Table 7: Vocabularies referenced by the proposals.

Academic		
AIISO: Academic Institution Internal Structure Ontology	http://vocab.org/aiiso/schema	[dAq12, dAq13, Pie12, Sha09, Zab12]
Courseware (course descriptions)	http://courseware.rkbexplorer.com/ontologies/courseware	[Zab12]
XCRI: eXchanging Course Related Information	http://www.xcri.co.uk/what-is-xcri-cap.html	[dAq13, Zab12]
JACS: Joint Academic Coding System	https://www.hesa.ac.uk/jacs3	[Sha09]
MLO: Metadata for Learning Opportunities	http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlo_rdfs.xml	[dAq12, dAq13, Zab12]
Semantic Web Italian University Project	http://sw.unime.it/swiup/schema/	[Pir10]
University of Southampton Ontology (university concepts)	http://rdf.ecs.soton.ac.uk/ontology/ecs	[Zab12]
Learning resources		
Dublin Core (basic resource annotation)	http://dublincore.org/documents/dces/	[Abe11, dAq12, dAq13, DeV12, Die13, Fer11, Isa12, Lam12, Pir10, Sch12, Sve10, Yoo11, Yu12, Zab12]
LOCO: Learning Object Context Ontologies	http://jelenajovanovic.net/LOCO-Analyst/loco.html	[Jer13, Sia12]
LOM: Learning Object Metadata	http://kmr.nada.kth.se/static/ims/metadata.html	[Die13, Isa12, Lam12, Sic11, Yoo11]
LRMI: Learning Resource Metadata Initiative	http://www.lrmi.net/the-specification	[Die13]
LOCWD: Linked OpenCourseWare Data	http://purl.org/locwd/schema#	[Pie12]
Ontoolcole (educational tools)	http://www.gsic.uva.es/ontologies/ontoolcoleModel.owl	[RuC12]
OMDoc: Open Mathematical Documents	http://www.mathweb.org/wiki/Index.php/OMDoc	[Dav10]
OpenMath (mathematical objects)	http://www.openmath.org/ontology/	[Dav10]
Organic.Edunet Metadata Application Profile	http://wiki.organic-lingua.eu/Organic.Edunet_Metadata_Application_Profile	[Sic11]
SemUNT (learning objects)	http://semunt.supelec.fr/	[Isa12]
Time and events		
Event ontology	http://motools.sf.net/event/event.html	[dAq12]
LODE: Linking Open Descriptions of Events	http://linkedevents.org/ontology/	[Zab12]
Time Ontology	http://www.w3.org/TR/owl-time/	[dAq13]
Timeline ontology	http://motools.sf.net/timeline/timeline.html	[Yu12]

Table 7: Vocabularies referenced by the proposals (continued).

Cross-domain		
Creative Commons (copyright)	`http://creativecommons.org/ns`	[Pie12, Zab12]
DBpedia ontology (general)	`http://dbpedia.org/Ontology`	[Abe11, Alo12, Bra12, dAq13, Die12, Fou11, Lam12, Pie12, Rey12, RuC12, Sia12, Wai10, Zab12]
DMOZ (general)	`http://rdf.dmoz.org/`	[Fer11]
Open Organisations	`http://purl.org/openorg/`	[dAq13]
OpenCyc (general)	`http://www.cyc.com/platform/opencyc`	[RuC12]
SKOS: Simple Knowledge Organization System	`http://www.w3.org/2009/08/skos-reference/skos.html`	[Abe11, Alo12, dAq12, dAq13, DeV12, Fou11, Isa12, Lam12, Sch12, Sia12, Wai10, Yu12, Zab12]
Umbel (general)	`http://umbel.org/`	[RuC12]
voiD: Vocabulary of Interlinked Datasets	`http://rdfs.org/ns/void#`	[dAq13, Sha09, Sic11]
YAGO (general)	`http://www.mpi-inf.mpg.de/yago-naga/yago`	[Fou11]
Social		
Contact Ontology (contact information)	`http://www.w3.org/2000/10/swap/pim/contact`	[Zab12]
FOAF (social networks)	`http://xmlns.com/foaf/spec/`	[Abe11, dAq12, dAq13, DeV12, Fer11, Fou11, Isa12, Jer13, Lam12, Pie12, RuC12, RuR11, Sch12, Sic11, Sve10, Yu12, Zab12]
FOAF Realm (access control)	`http://www.deri.ie/content/foafrealm-ontology-specification`	[Jer13]
Grapple Core (user modeling)	`http://wis.ewi.tudelft.nl/rdf/grapple-core.owl`	[Abe11]
IntelLEO ontology framework (educational communities)	`http://intelleo.eu/index.php?id=183`	[Sia12]
OPO: Online Presence Ontology	`http://online-presence.net/opo/spec/`	[Jer13]
SIOC: Semantically Interlinked Online Communities	`http://rdfs.org/sioc/spec/`	[DeV12, Jer13, Sia12, Zab12]
SWRC: Semantic Web for Research Communities	`http://ontoware.org/swrc/`	[DeV12]
vCard Ontology (people and organisations)	`http://www.w3.org/TR/vcard-rdf/`	[Zab12]

Table 7: Vocabularies referenced by the proposals (continued).

Libraries		
Bibliographic Ontology	http://purl.org/ontology/bibo/	[dAq12, dAq13, Pie12, Zab12]
CiTO: Citation Typing Ontology	http://purl.org/spar/cito/	[Zab12]
Dewey Decimal	http://dewey.info/	[Yu12]
LCC: Library of Congress Classification	http://id.loc.gov/ontologies/lcc	[Yu12, dAq12]
MulDiCat: Multilingual Dictionary of Cataloguing Terms and Concepts	http://iflastandards.info/ns/muldicat20100901.rdf	[Zab12]
RDA: Resource Description and Access (library cataloging)	http://rdvocab.info/	[Sch12]
Media		
Media RDF Vocabulary	http://purl.org/media	[Zab12]
Music Ontology	http://purl.org/ontology/mo/	[RuR11]
Ontology for Media Resources	http://www.w3.org/TR/mediaont-10	[dAq12, Fer11, Zab12]
Tags		
CommonTag	http://www.commontag.org/Specification	[Sia12]
MOAT: Meaning of a Tag	http://moat-project.org/ns#	[DeV12]
Nice Tag Ontology	http://ns.inria.fr/nicetag/2010/09/09/voc.html	[Fer11, Zab12]
Geography		
Basic Geo (basic geoposition)	http://www.w3.org/2003/01/geo/	[dAq12, Pie12, Sve10, Yu12]
GeoNames (geographical names)	http://www.geonames.org/ontology/	[Pir10, Yu12, Zab12]
Postcode Ontology (UK postcodes)	http://data.ordnancesurvey.co.uk/ontology/postcode/	[Zab12]
Other		
AGROVOC thesaurus (agriculture terms)	http://aims.fao.org/standards/agrovoc/about	[Sic11]
ADMS: Asset Description Metadata Schema	http://www.w3.org/TR/vocab-adms/	[Die13]
Buildings and Rooms Vocabulary	http://vocab.deri.ie/rooms	[Zab12]
CIDOC Conceptual Reference Model (cultural heritage)	http://www.cidoc-crm.org/	[RuR11]
Galen (biomedical)	http://purl.bioontology.org/ontology/GALEN	[Die12]
GoodRelations (e-commerce)	http://purl.org/goodrelations/v1	[Zab12]
LOIUS Ontology (educational statistics)	http://sw.unime.it/loius/dump/type_dimensions.nt	[Pir10]
MeSH (biomedical)	http://www.nlm.nih.gov/mesh/	[Abe11, Die12, Fou11]
Service-finder (services)	http://www.service-finder.eu/ontologies/ServiceCategories	[Die12]
SNOMED (biomedical)	http://www.ihtsdo.org/snomed-ct/	[Die12]
SCOVO: Statistical Core Vocabulary	http://purl.org/NET/scovo#	[Pir10]
Weighted Interest Vocabulary	http://purl.org/ontology/wi/core#	[Abe11]

staff, while some institutions offer information about research groups, university buildings and even vending machines![13] Interestingly, **academic datasets can be exploited for providing innovative educational applications, often in combination with other Linked Data sources**; a number of works in this review exemplify this, e.g. [Yu et al., 2012] can be used to annotate and search educational video resources from the Open University Linked Data portal and from other media sites such as the BBC.

Besides academic institutions, other initiatives expose learning resources as Linked Data, typically associated to specific projects, e.g. mEducator Linked Educational Resources is a dataset of biomedical learning objects; and Organic.Edunet publishes resources for agriculture education. Library and cultural heritage datasets are also employed in some works, since these domains are closely related to the education field; as an example of the usage of a library dataset, [Isaac et al., 2012] gathers bibliographic metadata from FacetedDBLP. In addition, datasets from horizontal domains such as geography, media or social web are exploited for diverse purposes: [Fernandez et al., 2011] integrates video lectures from different sources to provide a uniform access to students; while [Abel et al., 2011] uses social web data to build profiles of user knowledge and interest. A significant number of works reports the usage of DBpedia, especially in data enrichment processes — this is consistent with the popularity of the DBpedia ontology (see Table 7).

Noteworthy, [d'Aquin et al., 2013] studied the state of the educational Linked Data landscape after producing the aggregated dataset Linked Education Cloud (included in Table 8). Their assessment is that the overall educational data network is heterogenous and not very cohesive, due to the use of different competing vocabularies, as discussed above. However, [d'Aquin et al., 2013] reports that, with a reduced effort, the overall connectivity is dramatically improved by creating mappings among vocabularies, mainly DBpedia ontology, Bibliographic Ontology, XCRI and AIISO. Therefore, we encourage educational data publishers **to adapt these mappings for their purposes in order to further increase the interoperability of their exposed datasets**.

To conclude this section, we report in Table 9[14] the technological products referenced in the reviewed studies. They can serve as a guide for solving some demanding tasks in upcoming educational Linked Data developments. When publishing Linked Data, the most basic infrastructure is an RDF triple store and there is an ample number of choices available, e.g. [Pirrotta, 2010] uses OpenLink Virtuoso. **Triple stores allow a high degree of control of the data exposed and can be employed virtually in any situation** [d'Aquin, 2012] — similarly to database management systems. However, **a specialized repos-**

[13] `http://data.southampton.ac.uk/dataset/vending-machines`
[14] Available at `http://www.gsic.uva.es/reviewLinkedDataEducation/TechnologicalProducts.pdf`

Table 8: Datasets referenced by the proposals.

Academic		
Aalto University Linked Data portal	`http://data.aalto.fi/`	[dAq13]
Linked Education Cloud: repository of educational datasets	`http://data.linkededucation.org/linkedup/catalog/`	[dAq13]
University of Münster Linked Data portal (LODUM)	`http://data.uni-muenster.de/`	[dAq12, Die13]
Open University Linked Data portal	`http://data.open.ac.uk/`	[dAq12, dAq13, dAq13a, Die13, Jer13, Yu12, Zab12]
University of Oxford Linked Data portal	`https://data.ox.ac.uk/`	[Die13, Jer13]
University of Southampton Linked Data portal	`http://data.southampton.ac.uk/`	[dAq12, dAq13, Die13, Jer13]
UK educational open data	`http://education.data.gov.uk/`	[dAq13, Zab12]
Learning resources		
mEducator Linked Educational Resources: biomedical	`http://linkededucation.org/meducator`	[dAq13, Die13]
OpenLearn: open educational resources	`http://www.open.edu/openlearn/`	[dAq12, Die12]
Organic.Edunet: organic agriculture resources	`http://organic-edunet.eu`	[dAq13, Die13]
SEEK-KB: educational tools	`http://seek.rkbexplorer.com/`	[RuC12]
Libraries		
Bibliothèque nationale de France	`http://data.bnf.fr/`	[dAq12, Die13, Sch12]
British National Bibliography	`http://bnb.data.bl.uk/`	[dAq12]
CrossRef: scholarly articles (non Linked Data)	`http://www.crossref.org/`	[Hea12]
FacetedDBLP: publications	`http://dblp.l3s.de/`	[dAq13, Isa12]
Library of Congress Subject Headings	`http://id.loc.gov/authorities/subjects.html`	[Sha09]
Linked Periodicals Dataset: journal metadata	discontinued	[Sha09]
NTNU University Library	`http://www.ntnu.edu/ub`	[Die13]
NPG Linked Data Platform: Nature publications	`http://data.nature.com/`	[dAq13]
Online Computer Library Center (OCLC)	`http://www.oclc.org/data.en.html`	[Sch12]
SUDOC French Higher Education library catalogue	`http://www.sudoc.abes.fr/`	[Sch12]
Cultural heritage		
British Museum Semantic Web Collection Online	`http://collection.britishmuseum.org/`	[Die13]
Europeana: European museums' collection	`http://data.europeana.eu/`	[dAq12, Die13, Sch12]
WorldHistory: historic facts (non Linked Data)	`http://www.worldhistory.com/`	[Yu12]
Geographic		
Google Maps: maps (non Linked Data)	`http://maps.google.com`	[Yu12]
Ordnance Survey: UK geographic data	`http://www.ordnancesurvey.co.uk/`	[Zab12]

Table 8: Datasets referenced by the proposals (continued).

Cross-domain		
DBpedia: general	`http://dbpedia.org/`	[Abe11, Alo12, Bra12, dAq12, Dav10, Die13, Isa12, Jun11, Fou11, Jer13, Lam12, Rob11, RuC12, Sch12, Wai10, Yoo11]
Factforge: general	`http://factforge.net/`	[RuC12]
Freebase: general	`https://www.freebase.com/`	[Alo12]
Wordnet: lexical database for the English language	`http://wordnet.princeton.edu/`	[Alo12]
Media		
BBC programmes information	`http://www.bbc.co.uk/programmes`	[Yu12]
OU podcasts	`http://podcast.open.ac.uk/`	[Fer11]
OU video repository	`http://data.open.ac.uk/context/youtube`	[Yu12]
Videolectures.net: video lectures (non Linked Data)	`http://videolectures.net/`	[Fer11]
Youtube: videos (non Linked Data)	`https://www.youtube.com/`	[Fer11, Yu12]
Yovisto video index	`http://www.yovisto.com/`	[Wai10]
Social		
Facebook: social data (non Linked Data)	`http://facebook.com/`	[Abe11]
Twitter: social data (non Linked Data)	`http://twitter.com/`	[Abe11]
Other		
AGROVOC: agricultural thesaurus of the FAO	`http://aims.fao.org/standards/agrovoc/functionalities/download`	[Alo12]
Best Kids Apps: children applications and games (non Linked Data)	`http://www.bestkidsapps.com/`	[Alo12]
DailyMed: medicaments	`http://www4.wiwiss.fu-berlin.de/dailymed/`	[Alo12]
PBS Kids: children applications and games (non Linked Data)	`http://pbskids.org/`	[Alo12]
Pubmed: medical resources	`http://www.pubmedcentral.nih.gov/`	[Die12]
SmartLink: service descriptions	`http://smartlink.open.ac.uk/`	[Die12]
USDA National Nutrient Database for Standard Reference	`http://ndb.nal.usda.gov/`	[Alo12]

itory with Linked Data support may be more effective in particular cases. For example, Drupal is a well-known content management system that can be configured to expose Linked Data; and ePrints is a document repository that exports RDF according to vocabularies such as the Bibliographic Ontology and Dublin Core — the institutional research repository of the University of Southampton[15] uses ePrints.

Legacy datasets can be converted to RDF using a so-called RDFizer for common formats or a database mapper such as D2R. Custom converters can also be developed for special cases, e.g. the Open University extracts RDF from their podcast platform through RSS feeds [Zablith et al., 2012]. There are also **lookup services to discover vocabularies and datasets**. In this regard, sameAs provides a way of finding co-references in the Web of Data, thus serving to obtain the different names (URIs) of a particular resource, e.g. a publication available in different digital libraries.

Many works in this review aim to improve the discoverability and interlinking of educational data. A number of **semantic extractors can be used to automatically enrich poorly structured data**; for example, [Lama et al., 2012] uses DBpedia Spotlight to obtain DBpedia categories out of descriptions of learning objects for enrichment purposes, while [Jeremic et al., 2013] employs the Ontotext Semantic Platform for annotating lessons and other educational resources. There are also **annotators to manually generate semantic metadata**, e.g. OpenRefine or the annotation tool of Ontotext Semantic Platform.

For other specific tasks, semantic APIs can provide a solution, e.g. some triple stores only provide a SPARQL endpoint, and Pubby can be applied to a SPARQL endpoint to add a proper Linked Data interface with dereferenceable URIs — this usage of Pubby is reported in [Isaac et al., 2012]. Finally, we include a category of other technological products not directly applicable to Linked Data that might be useful for related tasks such as text indexing and search with Lucene — see an example in [Zablith et al., 2012].

6 Discussion and conclusions

In this review we have analyzed the most relevant Linked Data research works in the educational domain that were found in the literature. Many of them are focused on improving the interoperability of systems and tools in the educational field. In this regard, Linked Data practices are crystalizing as a way of achieving the long-desired goal of spreading the exchange and reuse of educational resources. The technology stack promoted by Linked Data facilitates the distribution and sharing of data, as well as data integration processes; this is illustrated with the reviewed works proposing federated repositories of learning

[15] `http://eprints.soton.ac.uk/`

Table 9: Technological products referenced by the proposals.

RDF triple stores		
4Store	`http://4store.org/`	[dAq12]
AllegroGraph	`http://franz.com/agraph/allegrograph/`	[Die13]
ARC RDF Store (discontinued)	`https://github.com/semsol/arc2/wiki`	[Yoo11, Zab12]
Dydra	`http://dydra.com/`	[dAq12]
Fuseki	`http://jena.apache.org/documentation/serving_data/`	[dAq12, dAq13a]
Jena SDB	`https://jena.apache.org/documentation/sdb/`	[Die13, Sia12]
Joseki (discontinued)	`http://sourceforge.net/projects/joseki/`	[Die13]
Mulgara	`http://www.mulgara.org/`	[Die13]
OWLIM	`http://owlim.ontotext.com/`	[Die12, Isa12, RuC12, dAq12]
Sesame openRDF	`http://www.openrdf.org/`	[dAq12, Die13, Isa12, RuC12, Yu12]
OpenLink Virtuoso	`http://virtuoso.openlinksw.com/`	[Die13, Pir10]
Semantic repositories		
Drupal: RDF-compliant content management system	`https://www.drupal.org/`	[Die13]
ePrints: RDF-compliant document repository	`http://www.eprints.org/`	[dAq12]
Fedora: RDF-compliant document repository	`http://fedora-commons.org/`	[dAq12]
RDF converters and mappers		
Bibtex2RDF: converter of Bibtex format to RDF	`http://www.l3s.de/~siberski/bibtex2rdf/`	[dAq12]
D2R: database to RDF mapper	`http://d2rq.org/d2r-server`	[dAq12, Die13]
SIMILE RDFIzer: converter of common data formats to RDF	`http://simile.mit.edu/RDFizers/`	[dAq12]
Triplify: database to RDF mapper	`http://triplify.org`	[dAq12, Die13]
Youtube2RDF: RDF extractor of Youtube videos	`https://code.google.com/p/luceroproject/wiki/Youtube2RDF`	[dAq12]
Lookup services		
DataHub: dataset registry	`http://datahub.io/`	[dAq13]
sameAs: lookup service of co-references in the Web of Data	`http://sameas.org/`	[Sch12]
Sindice: lookup index for Semantic Web documents (discontinued)	`http://sindice.com/`	[Yu12]
Semantic APIs		
Jenabean: persisting Java beans to RDF	`https://code.google.com/p/jenabean/`	[Sia12]
Protégé OWL API: OWL data model manipulation	`http://protege.stanford.edu/`	[Jun11]
Pubby: Linked Data frontend for SPARQL endpoints	`http://wifo5-03.informatik.uni-mannheim.de/pubby/`	[dAq12, Die13, Isa12]
RDF2Go: abstraction over RDF repositories	`http://rdf2go.semweb4j.org/`	[Yu12]
rdfQuery: client-side RDF processing API	`https://code.google.com/p/rdfquery/`	[Yu12]

Table 9: Technological products referenced by the proposals (continued).

Semantic extractors and annotators		
Apache Stanbol: metadata extractor	`https://stanbol.apache.org/`	[Die13]
Bioportal: categorization service of unstructured text to biomedical vocabularies	`http://data.bioontology.org/documentation`	[Die12]
DBpedia Spotlight: semantic annotator tool	`http://spotlight.dbpedia.org/`	[Die12, Die13, Jer13, Lam12]
iServe: annotation service	`http://iserve.kmi.open.ac.uk/`	[Die12]
LIMES: link discovery framework for the Web of Data	`http://aksw.org/Projects/LIMES.html`	[dAq12]
Ontotext Semantic Platform: suite for semantic enrichment, integration and annotation (formerly KIM)	`http://www.ontotext.com/products/ontotext-semantic/`	[Jer13]
OpenCalais: semantic annotator tool	`http://www.opencalais.com/`	[Die13]
OpenRefine: RDF-compliant data cleanup and transformation (formerly Google Refine)	`http://openrefine.org/`	[dAq12]
SILK: link discovery framework for the Web of Data	`http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/`	[Alo12, dAq12, Die13]
TextWise: categorization service of unstructured text to Open Directory Project concepts	`http://www.textwise.com/categorization`	[Fer11]
Zemanta: semantic annotator tool	`http://www.zemanta.com/`	[Yu12]
Other		
Facebook Graph API: retrieving and publishing data in Facebook	`https://developers.facebook.com/docs/graph-api`	[Zab12]
Gephi: network analysis and visualization tool	`http://gephi.org`	[dAq13]
Jersey: RESTful Web Services	`https://jersey.java.net/`	[Sia12]
jQuery: multi-browser Javascript library	`http://jquery.com/`	[RuR11]
jQuery UI: Javascript library for user interfaces	`http://jqueryui.com/`	[Zab12]
Lucene: text search engine	`http://lucene.apache.org/`	[Zab12]
Simple Query Interface: query interface employed in many educational repositories	`http://nm.wu-wien.ac.at/e-learning/interoperability/SQI_V1.0beta_2005_04_13.pdf`	[Sic11]

resources. Moreover, the emergent educational Web of Data offers unprecedented opportunities for data reuse, such as quiz generation, enrichment of educational data or resource recommendation. This way, new ways of contextualized and personalized learning practices can be delivered through Linked Data.

Since publishers are gradually joining this movement, the educational data web is expected to grow and thus potentially become more useful for existing and upcoming applications. Indeed, a number of the reviewed works report data publishing experiences. Particularly, many proposals exploit the Web of Data for educational metadata enrichment and dataset integration, overcoming the scalability and sustainability limitations of prevailing single authoring practices. It can be argued that Web 2.0 approaches have already solved these issues by applying practices such as crowdsourcing for metadata enrichment or so-called "web mashups" for dataset integration. However, free form annotations suffer from problems of imprecision and ambiguity [Mathes, 2004]; and Linked Data-based integration is more generally applicable and flexible, since Linked Data applications feed on an unbounded, global data space, while web mashups are commonly created on a one- to-one basis against a fixed set of data sources [Bizer et al., 2009a][Heath and Bizer, 2011, ch. 1]. Anyway, social and semantic approaches seem more complementary than competitive (see research on Social Semantic Web [Breslin et al., 2009]). Indeed, DBpedia is considered the prototypical dataset of the Web of Data, whereas it is the result of "RDFizing" the collaboratively edited Wikipedia. Further, some works in this review integrate data from the Social Web (Twitter, Facebook, Youtube) with Linked Data sources, thus obtaining much more comprehensive and vibrant applications, such as the video lecture discovery service reported in [Fernandez et al., 2011].

Despite the aforementioned benefits, it is worth reflecting on the challenges ahead. In this regard, the Linked Data model implies a fundamental paradigm shift in which strict control over data cannot be enforced. Therefore, new facets come into play, such as quality assurance of a dataset, privacy, provenance or licensing rights. These aspects have to be considered when consuming or publishing Linked Data, e.g. to avoid poor search results due to the exploitation of a low-quality dataset. Another risk is the fragmentation of the educational data web due to the use of disparate vocabularies — see [d'Aquin et al., 2013]. In this regard, ongoing competition may lead to vocabulary convergence processes and winner-take-all situations, although mediation techniques and vocabulary mappings can mitigate interoperability problems in the short term. Other shortcomings stem from the effort required to publish Linked Data, as well as the computational costs involved in consuming Linked Data. Nevertheless, these challenges are not exclusive of the learning domain and we can expect improvements in the near future, since tools and practices are rapidly maturing.

As a wrap-up, this review can help educational researchers to envision prospective Linked Data applications in the learning field. Towards this goal, we have provided a classification of existing works and we have discussed the main advantages and limitations that may be considered to assess the feasibility of upcoming proposals. Further, developers can take profit of our guidelines for implementing Linked Data applications in education; in this sense, the reported usage of vocabularies, datasets and technological products constitutes an important asset. To conclude this paper, we summarize below the most promising research directions that were suggested in the reviewed papers:

Exposure of educational data. The emergent educational Web of

Data is mainly the result of pioneering initiatives from some universities, governments and research projects. [Dietze et al., 2012, Siadaty et al., 2012, Tiropanis et al., 2009, Zablith et al., 2012] discuss that more educational institutions should join this movement. Since many of these already maintain high-quality datasets within their institutional realms, they could be offered as Linked Data with a relatively low effort. Privacy concerns should be addressed in this process, such as anonymizing or removing sensitive information before opening [Tiropanis et al., 2009]. As a result, the exposure of these legacy repositories as Linked Data can lead to an exponential growth of the educational data cloud. Although offering Linked Data is not a goal per se, we expect that the publishing organizations themselves will improve their visibility and will drive further research and innovation activities related to data reuse and scrutiny.

New Linked Data-based educational applications. The availability of structured information from the Web of Data can be exploited for different uses in learning, as exemplified by the works reported in this review. For example, new discovery services can be delivered for academic publishing [Schreur, 2012], while [Tiropanis et al., 2009] proposes the use of Linked Data for improving the visibility of course offerings, recommendation of educational material or expert matching. Overall, [d'Aquin, 2012] reflects on the potential of Linked Data to openly access, repurpose and remix many sources of information independently from their origin and primary intent.

Curation and enrichment of educational data. Despite the quality concern of some sources of the Web of Data, there are opportunities for leveraging educational datasets through Linked Data. In this sense, works such as [Lama et al., 2012] automatically annotates learning objects with DBpedia concepts, while [Dietze et al., 2012, Zablith et al., 2012] encourage further investigation to enable efficient, accurate and dynamic enrichment of educational data. In addition, [Bratsas et al., 2012] proposes to provide feedback to dataset publishers for data curation.

Federation and interlinking of educational data. Following traditional research on learning objects, Linked Data practices have been successfully applied to federate educational material, e.g. Organic.Edunet [Sicilia et al., 2011] for learning in the agricultural domain. Nevertheless, effective repository interoperation is still challenging at a Web scale, especially due to vocabulary fragmentation and limited interlinking [Dietze et al., 2013]. In this regard, [Zablith et al., 2012] proposes to develop a common classification scheme for sharing educational resources as Linked Data. Moreover, [Schreur, 2012] suggests the use of crowdsourcing and co-reference services, e.g. sameAs, for providing name aliases to the Web of Data. Significantly, [d'Aquin et al., 2013] demonstrates that educational dataset interconnection can be greatly improved through the use of vocabulary mappings.

Generation of learning artifacts. Authoring reusable learning objects remains an elusive goal [Boyle, 2003], despite the work devoted on creating standards such as IEEE LOM[16] and SCORM[17]. However, the educational data web can be repurposed to generate new learning artifacts, such as quizzes [Bratsas et al., 2012, Rey et al., 2012] or educational ICT tool descriptions [Ruiz-Calleja et al., 2012]. As a result, we can expect new innovative usages of the Web of Data to create different types of learning artifacts.

Improving performance. Some Linked Data applications can be computing intensive, especially querying large datasets [Tiropanis et al., 2009] or graph traversing [Lama et al., 2012]. There are opportunities, though, to improve performance such as the use of cloud computing, parallelization, caching, pre-indexing or query preparation beforehand.

Acknowledgements

This research has been partially funded by the Spanish Ministry of Economy and Competitiveness through the EEE project (TIN2011-28308-C03-02) and the European Commission through the BYTE (FP7 GA 619551) and METIS (531262-LLP-2012-ES-KA3-KA3MP) projects.

References

[Abel et al., 2011] Abel, F., Celik, I., Hauff, C., Hollink, L., and Houben, G.-J. (2011). U-sem: Semantic enrichment, user modeling and mining of usage data on the social web. In *Proceedings of the 1st International Workshop on Usage Analysis and the Web of Data (USEWOD2011) in the 20th International World Wide Web Conference (WWW2011)*, Hyderabad, India. Reviewed proposal labeled as [Abe11] in the text tables.

[Alexander et al., 2009] Alexander, K., Cyganiak, R., Hausenblas, M., and Zhao, J. (2009). Describing linked datasets. In *Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009) in the 18th International World Wide Web Conference (WWW2009)*, Madrid, Spain.

[Allemang and Hendler, 2008] Allemang, D. and Hendler, J. A. (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL.* Morgan Kaufmann, Burlington, MA, USA.

[Alonso-Roris et al., 2012] Alonso-Roris, V. M., Miguez-Perez, R., Santos-Gago, J. M., and Alvarez-Sabucedo, L. (2012). A semantic enrichment experience in the early childhood context. In *Frontiers in Education Conference (FIE), 2012*, pages 1 –6. Reviewed proposal labeled as [Alo12] in the text tables.

[Andersen, 2007] Andersen, P. (2007). What is web 2.0?: ideas, technologies and implications for education. Technical report, JISC, Bristol, UK. URL: `http://21stcenturywalton.pbworks.com/f/What+is+Web+2.0.pdf`, last visited September 2014.

[16] `http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html`
[17] `http://scorm.com/`

[Berners-Lee, 2006] Berners-Lee, T. (2006). Linked Data – Design Issues. URL: `http://www.w3.org/DesignIssues/LinkedData.html`, revised 18-06-2009, last visited September 2014.

[Berners-Lee et al., 2005] Berners-Lee, T., Fielding, R., and Masinter, L. (2005). Uniform resource identifier (URI): Generic syntax. Standard RFC 3986, The Internet Engineering Task Force (IETF).

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American*, 284(5):34–43.

[Berners-Lee et al., 2008] Berners-Lee, T., Hollenbach, J., Lu, K., Presbrey, J., and Schraefel, M. (2008). Tabulator redux: Browsing and writing linked data. In *Proceedings of the 1st Workshop on Linked Data on the Web (LDOW2008) in the 17th International World Wide Web Conference (WWW2008)*, Beijing, China.

[Bertini et al., 2011] Bertini, M., Devedzic, V., Gasevic, D., and Torniai, C. (2011). Guest editorial: Semantic technologies for multimedia-enhanced learning environments. *Interactive Learning Environments*, 19(1):1–4.

[Bizer et al., 2009a] Bizer, C., Heath, T., and Berners-Lee, T. (2009a). Linked data – the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS), Special Issue on Linked Data*, 5(3):1–22.

[Bizer et al., 2009b] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009b). DBpedia – a crystallization point for the web of data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165.

[Boyle, 2003] Boyle, T. (2003). Design principles for authoring dynamic, reusable learning objects. *Australian Journal of Educational Technology*, 19(1):46–58.

[Bratsas et al., 2012] Bratsas, C., Chrysou, D. E., Eftychiadou, E., Kontokostas, D., Bamidis, P., and Antoniou, I. (2012). Semantic web game based learning: An i18n approach with greek dbpedia. In *Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France. Reviewed proposal labeled as [Bra12] in the text tables.

[Breslin et al., 2009] Breslin, J. G., Passant, A., and Decker, S. (2009). *The Social Semantic Web*. Springer.

[Brickley and Guha, 2014] Brickley, D. and Guha, R. (2014). RDF Schema 1.1. Recommendation, W3C. URL: `http://www.w3.org/TR/rdf-schema/`, last visited September 2014.

[Carmichael and Jordan, 2012] Carmichael, P. and Jordan, K. (2012). Semantic web technologies for education - time for a ‘turn to practice’? *Technology, Pedagogy and Education*, 21(2):153–169.

[Cheng and Qu, 2009] Cheng, G. and Qu, Y. (2009). Searching linked objects with falcons: Approach, implementation and evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS), Special Issue on Linked Data*, 5(3):49–70.

[Currier et al., 2004] Currier, S., Barton, J., O’Beirne, R., and Ryan, B. (2004). Quality assurance for digital learning object repositories: issues for the metadata creation process. *Research in Learning Technology*, 12(1).

[Cyganiak and Jentzsch, 2011] Cyganiak, R. and Jentzsch, A. (2011). The Linking Open Data cloud diagram. URL: `http://lod-cloud.net/`, last visited June 2013.

[Cyganiak et al., 2014] Cyganiak, R., Wood, D., and Lanthaler, M. (2014). RDF 1.1 Concepts and Abstract Syntax. Recommendation, W3C. URL: `http://www.w3.org/TR/rdf-concepts/`, last visited September 2014.

[d’Aquin, 2012] d’Aquin, M. (2012). Linked data for open and distance learning. Technical report, Commonwealth of Learning. Reviewed proposal labeled as [dAq12] in the text tables.

[d’Aquin et al., 2013] d’Aquin, M., Adamou, A., and Dietze, S. (2013). Assessing the educational linked data landscape. In *Proceedings of the ACM Web Science 2013 (WebSci’13)*, Paris, France. Reviewed proposal labeled as [dAq13] in the text tables.

[d'Aquin and Jay, 2013] d'Aquin, M. and Jay, N. (2013). Interpreting data mining results with linked data for learning analytics: motivation, case study and direction. In *Proceedings of the Third Conference on Learning Analytics and Knowledge (LAK 2013)*, Leuven, Belgium. Reviewed proposal labeled as [dAq13a] in the text tables.

[David et al., 2010] David, C., Kohlhase, M., Lange, C., Rabe, F., Zhiltsov, N., and Zholudev, V. (2010). Publishing math lecture notes as linked data. In *Proceedings of the 7th Extended Semantic Web Conference (ESWC 2010)*, volume 6089 LNCS, pages 370–375, Heraklion, Greece. Reviewed proposal labeled as [Dav10] in the text tables.

[De Vocht et al., 2012] De Vocht, L., Van Deursen, D., Mannens, E., and Van de Walle, R. (2012). A semantic approach to cross - disciplinary research collaboration. *International Journal of Emerging Technologies in Learning*, 7(2):22–30. Reviewed proposal labeled as [DeV12] in the text tables.

[Devedzic, 2006] Devedzic, V. (2006). *Semantic Web and Education*. Springer, New York, NY, USA.

[Dietze et al., 2013] Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H., Giordano, D., Marenzi, I., and Nunes, B. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches. *Emerald Program: electronic Library and Information Systems*, 47(1):60–91. Reviewed proposal labeled as [Die13] in the text tables.

[Dietze et al., 2012] Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., and Taibi, D. (2012). Linked education: interlinking educational resources and the web of data. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC '12)*, pages 366–371, Trento, Italy. Reviewed proposal labeled as [Die12] in the text tables.

[Fernandez et al., 2011] Fernandez, M., D'Aquin, M., and Motta, E. (2011). Linking data across universities: An integrated video lectures dataset. In *Proceedings of the 10th International Semantic Web Conference (ISWC 2011)*, volume 7032 LNCS, pages 49–64, Bonn, Germany. Reviewed proposal labeled as [Fer11] in the text tables.

[Fielding et al., 1999] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee, T. (1999). Hypertext Transfer Protocol – HTTP/1.1. Standard RFC 2616, The Internet Engineering Task Force (IETF).

[Foulonneau, 2011] Foulonneau, M. (2011). Generating educational assessment items from linked open data: The case of dbpedia. In *Proceedings of the 1st International Workshop on eLearning Approaches for Linked Data Age (Linked Learning 2011), 8th Extended Semantic Web Conference (ESWC2011)*, volume 7117 LNCS, pages 16–27, Heraklion, Greece. Reviewed proposal labeled as [Fou11] in the text tables.

[Gómez-Pérez et al., 2004] Gómez-Pérez, A., Fernández-López, M., and Corcho, O. (2004). *Ontological Engineering*. Springer, London, UK.

[Guha et al., 2003] Guha, R., McCook, R., and Miller, E. (2003). Semantic search. In *Proceedings of the Twelfth International World Wide Web Conference (WWW2003)*, Budapest, Hungary.

[Hannemann and Kett, 2010] Hannemann, J. and Kett, J. (2010). Linked data for libraries. In *Proceedings of the 76th IFLA World Library and Information Congress*, Gothenburg, Sweden.

[Heath and Bizer, 2011] Heath, T. and Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool.

[Heath et al., 2012] Heath, T., Singer, R., Shabir, N., Clarke, C., and Leavesley, J. (2012). Assembling and applying an education graph based on learning resources in universities. In *Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France. Reviewed proposal labeled as [Hea12] in the text tables.

[Hendler, 2008] Hendler, J. (2008). Web 3.0: Chicken farms on the semantic web. *Computer*, 41(1):106–108.

[Horrocks, 2008] Horrocks, I. (2008). Ontologies and the semantic web. *Communications of the ACM*, 51(12):58–67.

[Hyland et al., 2013] Hyland, B., Atemezing, G., Pendleton, M., and Srivastava, B. (2013). Linked Data Glossary. W3C working group note, W3C. URL: `http://www.w3.org/TR/ld-glossary/`, last visited September 2014.

[Isaac et al., 2012] Isaac, Y., Bourda, Y., and Grandbastien, M. (2012). Semunit – french unt and linked data. In *Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France. Reviewed proposal labeled as [Isa12] in the text tables.

[Jeremic et al., 2013] Jeremic, Z., Jovanovic, J., and Gasevic, D. (2013). Personal learning environments on social semantic web. *Semantic Web*, 4(1):23–51. Reviewed proposal labeled as [Jer13] in the text tables.

[Kitchenham and Charters, 2007] Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, Keele University (UK).

[Kobilarov et al., 2009] Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M., Bizer, C., and Lee, R. (2009). Media meets Semantic Web – how the BBC uses DBpedia and Linked Data to make connections. In *Proceedings of the 6th European Semantic Web Conference (ESWC2009)*, volume 5554 LNCS, pages 723–737, Heraklion, Greece.

[Lama et al., 2012] Lama, M., Vidal, J., Otero-García, E., Bugarín, A., and Barro, S. (2012). Semantic linking of learning object repositories to dbpedia. *Educational Technology and Society*, 15(4):47–61. Reviewed proposal labeled as [Lam12] in the text tables.

[Magnisalis et al., 2011] Magnisalis, I., Demetriadis, S., and Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. *IEEE Transactions onLearning Technologies*, 4(1):5–20.

[Mathes, 2004] Mathes, A. (2004). Folksonomies-cooperative classification and communication through shared metadata. *Computer Mediated Communication*, 47(10):1–13.

[McGuinness and van Harmelen, 2004] McGuinness, D. and van Harmelen, F. (2004). OWL Web Ontology Language Overview. Recommendation, W3C. URL: `http://www.w3.org/TR/owl-features/`, last visited September 2014.

[Naeve et al., 2006] Naeve, A., Lytras, M., Nejdl, W., Balacheff, N., and Hardin, J. (2006). Advances of the semantic web for e-learning: expanding learning frontiers. *British Journal of Educational Technology*, 37(3):321–330.

[Piedra et al., 2012] Piedra, N., Chicaiza, J., Lopez, J., Tovar, E., and Martinez-Bonastre, O. (2012). Combining linked data and mobiles devices to improve access to ocw. In *Proceedings of the IEEE Global Engineering Education Conference (EDUCON 2012)*, pages 1 –7, Marrakesh, Morocco. Reviewed proposal labeled as [Pie12] in the text tables.

[Pirrotta, 2010] Pirrotta, G. (2010). Linking italian university statistics. In *Proceedings of the 6th International Conference on Semantic Systems (I-Semantics '10)*, Graz, Austria. Reviewed proposal labeled as [Pir10] in the text tables.

[Prud'hommeaux and Seaborne, 2008] Prud'hommeaux, E. and Seaborne, A. (2008). SPARQL Query Language for RDF. Recommendation, W3C. URL: `http://www.w3.org/TR/rdf-sparql-query/`, last visited September 2014.

[Rey et al., 2012] Rey, G. A., Celino, I., Damova, M., Damljanovic, D., Li, N., Alexopoulos, P., and Devedzic, V. (2012). Semi-automatic generation of quizzes and learning artifacts from linked data. In *Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France. Reviewed proposal labeled as [Rey12] in the text tables.

[Robinson et al., 2011] Robinson, J., Stan, J., and Ribière, M. (2011). Using linked data to reduce learning latency for e-book readers. In *Proceedings of the 1st International Workshop on eLearning Approaches for Linked Data Age (Linked Learning*

2011), 8th Extended Semantic Web Conference (ESWC2011), volume 7117 LNCS, pages 28–34, Heraklion, Greece. Reviewed proposal labeled as [Rob11] in the text tables.

[Ruiz-Calleja et al., 2012] Ruiz-Calleja, A., Vega-Gorgojo, G., Asensio-Pérez, J. I., Bote-Lorenzo, M. L., Gómez-Sánchez, E., and Alario-Hoyos, C. (2012). A linked data approach for the discovery of educational ict tools in the web of data. *Computers & Education*, 59(3):952 – 962. Reviewed proposal labeled as [RuC12] in the text tables.

[Ruiz-Rube et al., 2011] Ruiz-Rube, I., Cornejo, C., and Dodero, J. (2011). Accessing learning resources described in semantically enriched weblogs. *International Journal of Metadata, Semantics and Ontologies*, 6(3-4):175–184. Reviewed proposal labeled as [RuR11] in the text tables.

[Sampson et al., 2004] Sampson, D. G., Lytras, M. D., Wagner, G., and Diaz, P. (2004). Ontologies and the semantic web for e-learning. *Educational Technology & Society*, 7(4):26–28.

[Schreur, 2012] Schreur, P. E. (2012). The academy unbound – Linked Data as revolution. *Library Resources & Technical Services*, 56(4):227–237. Reviewed proposal labeled as [Sch12] in the text tables.

[Shabir and Clarke, 2009] Shabir, N. and Clarke, C. (2009). Using linked data as a basis for a learning resource recommendation system. In *Proceedings of the 1st International Workshop on Semantic Web Applications for Learning and Teaching Support in Higher Education (SemHE'09)*, Nice, France. Reviewed proposal labeled as [Sha09] in the text tables.

[Shadbolt et al., 2006] Shadbolt, N., Hall, W., and Berners-Lee, T. (2006). The semantic web revisited. *Intelligent Systems, IEEE*, 21(3):96–101.

[Siadaty et al., 2012] Siadaty, M., Jovanovic, J., Gasevic, D., Milikic, N., Jeremic, Z., Ali, L., Giljanovic, A., and Hatala, M. (2012). Semantic web and linked learning to support workplace learning. In *Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012 at WWW-2012)*, Lyon, France. Reviewed proposal labeled as [Sia12] in the text tables.

[Sicilia et al., 2011] Sicilia, M. A., Ebner, H., Sánchez-Alonso, S., Álvarez, F., Abián, A., and García-Barriocanal, E. (2011). Navigating learning resources through linked data: a preliminary report on the re-design of organic.edunet. In *Proceedings of the 1st International Workshop on eLearning Approaches for Linked Data Age (Linked Learning 2011), 8th Extended Semantic Web Conference (ESWC2011)*, volume 7117 LNCS, Heraklion, Greece. Reviewed proposal labeled as [Sic11] in the text tables.

[Siemens and Long, 2011] Siemens, G. and Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, 46(5):30–32.

[Sutherland et al., 2012] Sutherland, R., Eagle, S., and Joubert, M. (2012). A vision and strategy fortechnology enhanced learning: Report from the stellar network of excellence. Technical report, STELLAR. URL: `http://www.teleurope.eu/pg/file/read/152343/a-vision-and-strategy-for-technology-enhanced-learning-report-from-the-stellar-network-of-excellence`, last visited September 2014.

[Tiropanis et al., 2009] Tiropanis, T., Davis, H., Millard, D., and Weal, M. (2009). Semantic technologies for learning and teaching in the web 2.0 era. *Intelligent Systems, IEEE*, 24(6):49 –53. Reviewed proposal labeled as [Tir09] in the text tables.

[Tiropanis et al., 2012] Tiropanis, T., Millard, D., and Davis, H. C. (2012). Guest editorial: Special section on semantic technologies for learning and teaching support in higher education. *IEEE Transactions on Learning Technologies*, 5(2):102 –103.

[Vega-Gorgojo et al., 2010] Vega-Gorgojo, G., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Gómez-Sánchez, E., Dimitriadis, Y. A., and Jorrín-Abellán, I. M. (2010). Semantic search of tools for collaborative learning with the ontoolsearch system. *Computers & Education*, 54(4):835–848.

[Waitelonis et al., 2010] Waitelonis, J., Sack, H., Hercher, J., and Kramer, Z. (2010). Semantically enabled exploratory video search. In *Proceedings of the 3rd Interna-*

tional Semantic Search Workshop (SEMSEARCH '10), pages 1–8, Raleigh, NC, USA. Reviewed proposal labeled as [Wai10] in the text tables.

[Yu et al., 2012] Yu, H. Q., Pedrinaci, C., Dietze, S., and Domingue, J. (2012). Using linked data to annotate and search educational video resources for supporting distance learning. *Learning Technologies, IEEE Transactions on*, 5(2):130 –142. Reviewed proposal labeled as [Yu12] in the text tables.

[Zablith et al., 2012] Zablith, F., Fernandez, M., and Rowe, M. (2012). Production and consumption of university linked data. *Interactive Learning Environments*. In press. Reviewed proposal labeled as [Zab12] in the text tables.

A Linked Data and the Web of Data

The idea of Linked Data is to publish and interlink structured data on the Web [Heath and Bizer, 2011, ch. 2]. Such data have to be machine-readable, their meaning has to be explicitly defined and they have to be linked to other external datasets in order to create a single global data space [Bizer et al., 2009a] — the Web of Data. The so-called Linked Data principles [Berners-Lee, 2006] provide a set of guidelines for publishing data, namely: (1) use URIs (Uniform Resource Identifiers) to name things; (2) use HTTP (HyperText Transfer Protocol) URIs in order to look up those names; (3) provide useful information in response to a URI lookup, using the standards (RDF, SPARQL); and (4) include links to other URIs so that more things can be discovered.

The resulting Web of Data is thus based on the Web architecture, using the Web as an open decentralized platform for interconnecting data offered by different providers. Therefore, Linked Data relies on two fundamental Web technologies: URIs [Berners-Lee et al., 2005] for identifying any entity in the world, from a Web document to a real object; and the application protocol HTTP [Fielding et al., 1999] for transferring representations of those entities across the network. In the Linked Data context, anything that has a URI is a resource, while the process of looking up a URI to get a description is called dereferencing — see [Hyland et al., 2013] for a glossary of terms commonly employed in Linked Data.

When publishing Linked Data on the Web, resources are commonly represented using the Resource Description Framework (RDF) [Cyganiak et al., 2014]. RDF provides a data model conceived to be simple, flexible and tailored towards the Web: anybody can make assertions about anything with RDF, and any two resources may be linked with an RDF triple, thus allowing the creation of a rich Web of Data through this process. More specifically, RDF allows the expression of triples composed of subject, predicate and object, e.g. `MiguelDeCervantes` (subject) `isAuthorOf` (predicate) `DonQuixote` (object). The subject and the predicate of a triple are both URIs, each one identifying a resource, while the object may also be a URI in case of a resource or a literal such as a string or a date. Note that instead of using URIs such as `http://classicwriters.org/MiguelDeCervantes` in our examples, we just write `MiguelDeCervantes` for the sake of readability.

While RDF provides a way to represent data, the RDF Vocabulary Description Language (RDFS) [Brickley and Guha, 2014] and the Web Ontology Language (OWL) [McGuinness and van Harmelen, 2004] are modeling languages for creating vocabularies — often called ontologies [Horrocks, 2008]. A vocabulary defines the terminology of a domain of interest, especially its classes and properties. For instance, a bibliographic vocabulary may define classes `Writer` and `Book`, as well as an `isAuthorOf` property. Structured annotations can then be added with vocabularies such as the triple `DonQuixote rdf:type Book` that uses the important property `rdf:type`[18] for capturing the class-instance relationship and the previously defined `Book` class. Although anybody can create their own vocabulary and use it to publish Linked Data, a best practice is to reuse existing vocabularies if available for the domain of interest [Bizer et al., 2009a, Heath and Bizer, 2011, ch. 4]. For example, the former two triples can be expressed as `DonQuixote rdf:type bibo:Book` and `DonQuixote dc:creator MiguelDeCervantes` using the Bibliographic Ontology and Dublin Core, respectively (see Table 7). This way, data described with well-known vocabularies are more likely to be consumed by other applications without further pre-processing. Note that vocabularies can be themselves connected through mappings or extended as desired.

Publishing Linked Data on the Web involves a series of steps [Bizer et al., 2009a, Heath and Bizer, 2011, ch. 4], each of which maps onto one or two of the Linked Data principles: (1) A data provider has to assign URIs to their resources using a Web domain of their own and prepare RDF representations for lookup. For example, a publisher who owns the `http://classicwriters.org` domain can expose the resources `http://classicwriters.org/MiguelDeCervantes` and `http://classicwriters.org/DonQuixote`; when someone looks up one of these URIs, a set of RDF triples referring to the resource involved (see the examples above) should be returned in response. (2) Resources should be described with appropriate vocabularies for the domain, reusing terms from well-known vocabularies when available, e.g. Dublin Core and the Bibliographic Ontology. (3) It is very important to provide RDF links to related resources in other datasets in order to create an interconnected data web. For example, the triple `MiguelDe Cervantes owl:sameAs dbpedia:Miguel_de_Cervantes_Saavedra` indicates that the subject and the objet are URI aliases and they refer to the same entity, so additional information can be retrieved by looking up the URI `dbpedia:Miguel_de_Cervantes_Saavedra`. (4) Finally, published data should be themselves described with metadata to aid

[18] Here `rdf` is an abbreviation of the RDF namespace, namely, `http://www.w3.org/1999/02/22-rdf-syntax-ns#`

consumers in the assessment of a dataset, e.g. using the Vocabulary of Interlinked Datasets (voiD) [Alexander et al., 2009]. Once a dataset is ready, an RDF triple store is commonly employed for publishing Linked Data, although other choices may be considered [Heath and Bizer, 2011, ch. 5]. Besides dereferenceable URIs, RDF triple stores typically provide a SPARQL endpoint, a popular and convenient mechanism for querying a dataset using the SPARQL query language [Prud'hommeaux and Seaborne, 2008].

With all these elements into place, the adoption of the Linked Data principles has been very successful as demonstrated by the exponential growth of the Web of Data, containing billions of triples [Heath and Bizer, 2011, ch. 3]. Bootstrapping the Web of Data can be mainly attributed to the W3C Linking Open Data (LOD) project[19] which is a grassroots community effort devoted to publishing existing datasets available under open licenses according to the Linked Data principles. [Cyganiak and Jentzsch, 2011] periodically reported the state of the LOD cloud — it is interesting to compare its growth along the years. In addition, they provide a classification of the LOD cloud content: datasets are cataloged in cross-domain, geographic, government, media, libraries, life science and user content categories. One of the most prominent examples is DBpedia [Bizer et al., 2009b], a cross-domain dataset that is automatically constructed from Wikipedia. Due to the breadth of topic coverage, DBpedia URIs are commonly referenced in other datasets; as a result, DBpedia is the main hub of the Web of Data and plays a very important role for dataset interlinking.

Although publishing Linked Data entails a non-trivial effort, it provides a generic and flexible mechanism for accessing, discovering and integrating data from different sources. Specifically, RDF offers a single, unifying data model to represent data; the HTTP protocol offers a standardized form of accessing data; resource discovery is enabled by the use of URIs as global identifiers and typed links for interconnecting resources in different sources; and shared vocabularies ease the integration of data from different datasets [Heath and Bizer, 2011, ch. 2]. Publishers can make their transition to Linked Data incrementally, as described in the five-star rating scheme [Berners-Lee, 2006]. Each of these steps facilitates the consumption of data by third-parties, beginning with the exposure of data under an open license on the Web, to using open structured formats, such as RDF, and including links to other sources in order to provide context.

B Summary of reviewed proposals

The reviewed proposals are summarized in Table 10. This appendix is also available at `http://www.gsic.uva.es/reviewLinkedDataEducation/LinkedDataProposals.pdf`.

[19] `http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData`

Table 10: Summary of Linked Data proposals in the learning domain.

Proposal	**Research problem**	**Contributions**	**Contribution type**
[Abe11]	How to use information from the social web, publicly available but unstructured, categorize it, and build user profiles	A conceptual architecture for extracting user profiles from the (unstructured) information publicly available about them on social sites, such as those for blogging or bookmark annotation	Vocabulary Application
[Alo12]	How to enrich the data of an Early Childhood Education and Care platform	The authors employ Linked Open Data repositories to enrich their platform dataset by using a crawling agent. They also exploit pre-selected Web APIs and websites to extract new RDF triples	Dataset Application
[Bra12]	Can the knowledge available in the Web of Data be used for educational purposes?	A web game that generates quizzes based on DBpedia data	Application
[dAq12]	How to adopt Linked Data in distance learning and which are the tools, technologies, and processes to publish and use Linked Data	Description of the process for publishing Linked Data. Review of vocabularies and tools for publishing Linked Data in learning	Study
[dAq13]	Which is the current state of Linked Data for education. In particular, which datasets are available, what common practices are used and how datasets are interconnected	Assessment of the educational Linked Data landscape: available datasets, employed vocabularies, most popular terms and dataset interconnection analysis	Study
[dAq13a]	How to enrich the results of educational data mining with Linked Data so that the analysis of such results is facilitated	An approach to explore and interpret the results of educational data mining using open Linked Data sources	Technique
[Dav10]	How to search existing math documentation (in LaTeX) using semantics	Generation of a Linked Data repository of mathematics lectures by natural language processing of LaTeX documents	Vocabulary Dataset Technique
[DeV12]	How to connect and interlink existing research and educational data in a coherent way	A process for aligning, transforming and presenting various resources of research information to students, researchers and anyone involved with knowledge intensive tasks	Technique

Table 10: Summary of Linked Data proposals in the learning domain (continued).

Proposal	Research problem	Contributions	Contribution type
[Die12]	How to integrate existing TEL repositories to interlink resources in biomedical education	An architecture for federating TEL repositories An infrastructure for educational data and services integration The Metamorphosis+ application for accessing biomedical data	Dataset Application Architecture
[Die13]	What are the challenges and approaches for interlinking educational repositories in the Web of Data	Survey of approaches and challenges for educational repository interlinking in the Web of Data	Survey
[Fer11]	How to interlink video lectures across different institutions and how to categorize them automatically and homogenously	A new dataset that integrates videos from three different sites A technique for automatic categorization of video lectures	Dataset Technique
[Fou11]	Can Linked Open Data be used to generate formative assessment items?	Results of tests about the quality of data in DBpedia for automatically creating assessment items	Technique
[Hea12]	How to link the data generated by different deployments of Talis Aspire RLMS and how to generate recommendations with such linked data	The approach followed to link the data generated in different TALIS Aspire deployments and the techniques used to generate recommendations	Technique
[Isa12]	How to improve the search capabilities of a federation of educational repositories in Higher Education and how to automatically link with Linked Data resources	An ontology of learning resources inspired in LOM The architecture of a federated metadata repository Two sample services for document and expert search	Vocabulary Application Architecture
[Jer13]	How to support the personalization and meaningful integration of distributed learning data/systems	A set of principles for developing Personal Learning Environments (PLEs) that make use of Social Semantic Web and Linked Data ideas. A PLE called DEPTHS with such features for the learning of software engineering	Application
[Jun11]	How to support end users for authoring Linked Data without RDF knowledge	An annotation tool for authoring Linked Data and searching resources. This proposal is employed by teachers to annotate educational resources	Application

Table 10: Summary of Linked Data proposals in the learning domain (continued).

Proposal	**Research problem**	**Contributions**	**Contribution type**
[Lam12]	How to automatically classify learning objects with meaningful semantic metadata	A method that automatically annotates a learning object with metadata from the Web of Data (DBpedia concepts) and the LOM schema. The authors also present the corresponding learning object repository and a search front-end for it	Dataset Application Algorithm
[Pie12]	How to use Linked Data to make sensible mobile applications for students enrolled in OpenCourseWare	A mobile application that consumes Linked Data sources (1) to retrieve learning objects offered by institutions closer to the user's geoposition, (2) to recommend courses, and (3) to present information from the user's social network	Application
[Pir10]	The authors aim at opening statistical data about Italian universities	A vocabulary for describing statistical data about universities An architecture for a data repository of statistical data	Vocabulary Architecture Dataset
[Rey12]	How to semiautomatically generate educational artifacts from the Web of Data	An architecture of how to use the Web of Data to semiautomatically generate quizzes and other learning artifacts	Architecture
[Rob11]	Allow readers of the same ebook to share their annotations and to enrich them using Linked Data sources	A social application that allows sharing annotations of ebooks that are automatically related to DBpedia concepts	Application
[RuC12]	Can existing Linked Data sources, not necessarily related to the learning domain, be used to search educational tools?	A dataset of educational tools that is automatically constructed from the Web of Data An infrastructure that crawls the Web of Data to gather educational tool descriptions and allows searching	Dataset Application
[RuR11]	How to decouple Learning Management Systems (LMSs) from learning resources using a semantic layer and an annotation system	A system that comprises a new module for the LAMS LMS and a semantic annotation tool embedded into WordPress	Application
[Sch12]	Can Linked Data make a big impact on academic publishing?	Discussion of why Linked Data may shake up the academic world of information creation and exchange, identifying advantages, limitations and opportunities	Position

Table 10: Summary of Linked Data proposals in the learning domain (continued).

[Sha09]	How to deal with the four limitations of Linked Data within the context of Talis Aspire RLMS: sustainability of data sources, provenance, licensing and data reliability	A set of techniques for addressing the four issues that the authors found to hinder the use of resource lists in the Talis Aspire RLMS	Technique
[Sia12]	Can we support workplace and self-regulated learning by providing a PLE that integrates external data and tools and allows semantic annotation and interlinking of data	A system for workplace learning that integrates the different tools that employees often interact with. Semantic technologies and Linked Data allow employees to easily document, integrate and retrieve their work no matter the tool in which the contribution was created	Application
[Sic11]	How to redesign a federation of learning repositories in the domain of organic agriculture in order to enable the search and navigation of resources exposed as Linked Data	A redesign of a federation of learning object repositories that improves the search and navegation of resources by means of Linked Data	Application
[Sve10]	How to enrich emerging learning objects with contextual characteristics in a machine interoperable and interpretable manner	An approach to publish learning objects as Linked Data, annotated with contextual information	Technique
[Tir09]	Which is the potential of semantic technologies in education, and how mature are the technologies already in use	Categorization of the semantic technologies already in use in higher education.	Survey
[Wai10]	How to enrich video metadata with the Web of Data in order to support semantically enabled exploratory search of academic videos	A frequency-based heuristic for ranking entity properties and relationships of Linked Open Data resources A prototype modification of the academic video search facility "yovisto" to support exploratory search	Application Technique
[Yoo11]	How to obtain learning objects from external repositories based on learners characteristics	A Linked Data-based approach to retrieve learning objects from internal and external repositories that considers student characteristics	Application

Table 10: Summary of Linked Data proposals in the learning domain (continued).

[Yu12]	How to provide semantic annotations of educational video resources and related to the Web of Data in order to improve the discoverability and reusability of such video resources	The "Video Annotation Ontology" for annotations of instants or periods of a video The web application "Annomation" to add Linked Data annotations to instants or durations on the video timeline The web application "SugarTube" for searching and exploring educational videos	Vocabulary Application
[Zab12]	The paper tackles the problem of conceiving the process to produce Linked Data from already existing data in other formats at OU-UK and developing the technological support for that process	A process defined to generate Linked Data from already existing data at OU-UK, the technological support developed to support it and three sample applications that consume the data generated by such process (an application to search OU experts in a given area, an application that recommends study partners, an application that groups material with podcasts and courses)	Dataset Application