Linked Data as a Foundation for the Deployment of Semantic Applications in Higher Education

Thanassis TIROPANIS[1], Hugh C. DAVIS, Dave MILLARD, Mark WEAL and Su WHITE
University of Southampton

Abstract. The value of semantic technologies in the context of learning and teaching has often been associated with the use of reasoning to support learning processes. This paper discusses the value of linked data in addressing data interoperability and integration across higher education institutions and repositories. This value is related to higher education challenges and a proposal on deploying linked data in higher education is presented and discussed.

Keywords. Higher education, semantic technologies, linked data, learning

Introduction

The relevance of semantic technologies to learning and teaching has been examined in a number of different contexts in recent years [2, 4], among which is Higher Education (HE). The development of semantic technologies for learning often required agreement on ontologies, annotation of available resources and reasoning to facilitate learning related processes. The requirement for agreed ontologies has often presented a hurdle in the deployment of semantic technologies for learning on a large scale, involving resources in different administrative domains; on the other hand, the use of expressive ontologies on a smaller scale featured advanced reasoning to match learners and resources. The linked data movement advocates a bottom-up approach to ontology agreement [3] by shifting the focus first to the exposure of data in machine processable formats like RDF before agreeing on ontologies for specific applications.

Semantic tools and services relevant to higher education have been prototyped and, as a recent survey [5] shows, they can address the needs of students, teachers and researchers. However, support for additional kinds of higher education users, such as assessors, admissions teams, programme administrators, do not seem to be supported.

The JISC funded project SemTech[2] (Semantic Technologies for Learning and Teaching) performed a survey of semantic technology adoption in the UK higher education sector to outline a roadmap for the future adoption of semantic technologies. A workshop organised by SemTech in January 2009 identified a number of higher education challenges that semantic technologies were expected to address.

[1] Corresponding Author.
[2] http://www.semtech.ecs.soton.ac.uk

This paper discusses these challenges and argues for the creation of a linked data infrastructure on which semantic technologies relevant to higher education can be deployed. Section 1 discusses the relevance of semantic technologies generally to HE challenges and the extent to which these challenges can be addressed by semantic technologies. Section 2 investigates the relevance of linked data specifically to these HE challenges and issues. Section 3 proposes institutional policies that would help to foster linked data deployment across HE.

1. The Value of Semantic Technologies for HE Support

The HE challenges that could be addressed by semantic technologies, as identified at the SemTech workshop, consisted of a number of institutional challenges as well as challenges related to learning and teaching processes. The institutional challenges are particularly related to UK HE but it is expected they are relevant to HE in other countries too. A brief summary of these challenges includes:

- Visibility of degree programmes and research output of HE institutions
- Curriculum design
- Recruitment and retention of students
- Efficiency of accreditation
- Collaboration across departments and institutions through workflows
- Integration of knowledge capital, cross-curricular initiatives
- Transparency of data held by educational institutions

Semantic technologies are expected to provide for more efficient discovery of degree programmes to match the background and objectives of prospective students; the research output of institutions could be more visible to potential funding bodies. Student retention could be supported by more efficient monitoring of student activity and assessment of their progress. Institutional data is often dispersed across databases and is often not interoperable; semantic technologies could provide for such integration and support workflows and collaboration across departments or even institutions. Requirements for transparency on HE processes could be supported by making institutional data available in open formats, which could also assist in obtaining accreditation for degree programmes by external bodies. Curriculum design could also be supported by establishing how different curricula across HE institutions compare to each other and identify potential gaps that new degree programmes could address.

Apart from the institutional challenges, some of the learning and teaching challenges that were identified include:

- Support for course creation and delivery workflows
- Group formation for learning and teaching activities
- Support for critical thinking and argumentation
- Efficient construction of personal and group knowledge
- Assessment, certification and addressing of plagiarism

Work that has been reported in the literature in recent years and prototypes do address these areas, however, the survey performed by SemTech showed that there are no widely adopted tools and services to address these needs.

1.1. Scale

The lack of widely adopted semantic technologies for learning and teaching and the rarity of semantic applications to address HE challenges raise the following questions:

- What is the value of semantic technologies in the higher education domain?
- Is there a sufficient amount of HE data to perform reasoning on?
- Can HE data be exposed in formats easily mapped to ontologies?

The results of the survey performed by SemTech[3] showed that the value of semantic technologies in HE is primarily in well-formed metadata, secondarily in data interoperability and integration and thirdly in data analysis and reasoning. The surveyed tools and services were found to be collaboration tools (e.g. *Compendium*[4]), searching and matching tools (e.g. *Arnetminer*[5]), repositories and VLEs (e.g. *SKUA*[6]) or infrastructural tools supporting semantic annotation, integration, metadata storage and queries (e.g. *D2R server*[7]). The *potential added value* of semantic technologies per category outlined as follows:

- Collaboration tools could benefit from data integration and reasoning for inline recommendation of resources on other repositories.
- Searching and matching tools benefit from data integration and reasoning on a larger scale.
- Repositories, VLEs and annotation tools could provide additional value by linking to other repositories and by exposing machine processable data.

To reach the *potential added value* of semantic applications in higher education, the scale of availability of higher education data is critical and so is interoperability of metadata across institutions and repositories. Most of the higher education challenges, as identified in the previous paragraphs, rely on data integration on a large scale. For example, when it comes to HE curriculum design or alignment, course information is currently available on the Web pages of HE institutions but often not in machine processable formats; exposing HE curricula as linked data in RDF using SPARQL endpoints would enable relevant searches over different HE institutions, comparisons and analysis for this end. In this example, the power of linked data is in the scale or available data sources rather than reasoning.

It would be important that data exposed in linked data formats that can easily mapped to potentially more expressive ontologies for the development of specific semantic applications that employ advanced reasoning.

1.2. Reasoning

Certain applications establish the value of semantic technologies in advanced data analysis and reasoning and do not require large-scale data interoperability from the start.

---

[3] http://semtech-survey.ecs.soton.ac.uk
[4] http://www.aktors.org/technologies/compendium/
[5] http://www.arnetminer.org/
[6] http://www.myskua.org/
[7] http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/

Such applications include *Debategraph*[8] and *Cicero*[9] and can support learning based on critical thinking and argumentation. Nevertheless, even these applications could feature added value given the availability of additional resources in semantic formats. For example, argumentation tools could enable the discovery of resources to second certain arguments or to link to other argumentation data on additional platforms. Similarly, tools that rely on deep linguistic analysis of resources with textual descriptions to perform reasoning (e.g. *COGITO®* by *ExpertSystem* [10]) could benefit from interoperability with additional resources in additional repositories.

2. Exposing HE Linked Data

A significant amount of information is already exposed by institutions on their public Web pages. This information could also be exposed in RDF as linked data and help address institutional challenges such as exposing institutions' expertise and making their curricula and syllabi available for semantic technology enabled matching to prospective student interests. Infrastructural tools like *D2R server*, *Talis*[11] or *Virtuoso*[12] could provide for exposing data in relational databases as RDF via SPARQL endpoints. The availability of large RDF repositories (e.g. *RKBExplorer*) could also host linked data from a number of institutions and provide for optimised storage and searches [1].

Learning and teaching resources currently available in VLEs and internal repositories could expose their metadata via plugins; such extensions are already featured by repositories for publications such as *EPrints*[13] or *DSpace*[14].

Agreement on common URIs for RDF across institutions and repositories would be desirable but not required. URIs could be HE institution specific, VLE specific, standard (e.g. *DublinCore*) or agreed community-wide. A high degree of reusing URIs will make the mapping of linked data to higher ontologies more efficient.

2.1. Issues

Exposing linked data in HE can provide significant value in addressing HE challenges and in supporting learning and teaching activities. At the same time, there are certain challenges that need to adequately discussed and addressed.

It seems that by exposing information already publicly available as Web pages can support applications with valuable features (e.g. exposing degree program information, research output, expertise, and accreditation related information). Technologies like GRDDL[15] could support transition from HTML to linked data. However, additional data to address student retention like course evaluation data would potentially have to be available to selected parties. Other data would have to be sufficiently anonymized before exposed to any third party in order to protect personal information.

---

[8] http://debategraph.org/
[9] http://cicero.uni-koblenz.de/wiki/index.php/Main_Page
[10] http://www.expertsystem.net/
[11] http://www.talis.com/platform
[12] http://virtuoso.openlinksw.com/
[13] http://www.eprints.org/
[14] http://www.dspace.org/
[15] http://www.w3.org/TR/grddl/

The performance of even lightweight reasoning over resources dispersed across repositories is another issue to be considered. RDF can be stored and queried via SPARQL endpoints at each institution, or could be stored in larger RDF repositories that support optimised queries. Certain information may be required to remain within institutions (e.g. information of a sensitive nature or information frequently updated) while other information could be stored in large RDF repositories.

The cost of exposing linked data is another issue for consideration. Despite the availability of even free or open source *RDFizers*[16] there are additional cost parameters to consider. The existence of additional barriers due to institutional or government policies deserves further investigation.

Potentially novel teaching and learning activities enabled by linked data need to be identified and properly documented to enhance our understanding on the pedagogical potential of semantic tools and services over linked data.

3. Ways Forward

The potential of a deployed linked data field across the higher education sector has been argued in the previous sections together with the challenges that need to discussed, understood and addressed. It seems that there is significant value to be obtained by exposing information currently publicly available as HTML and that there is additional value in exposing data currently available in internal databases.

Taking this forward requires institutional policies on exposing linked data in a way potentially similar to when policies were established for exposing institutional information in HTML. Case studies of applications that could address HE challenges need to be conducted to identify more precisely what information each institution should consider exposing. Successful cases of using semantic technologies in a pedagogically meaningful way need to be documented and become available across institutions.

The cost of exposing linked data and of maintaining triple stores and SPARQL endpoints needs to be investigated. At the same time, the deployment of education related triple stores that will host metadata from institutions that are not able to support their own RDF repositories could be discussed. Best practises for exposing institutional data securely and selectively can be documented and studied by HE institutions.

References

[1] Harris, S. and Gibbins, N., 3store: Efficient Bulk RDF Storage. *In Proceedings 1st International Workshop on Practical and Scalable Semantic Web Systems*, Sanibel Island, Florida, USA, 2003.

[2] Johnson, L., Levine, A., & Smith, R., *The 2009 Horizon Report*, Austin, Texas: The New Media Consortium, 2009.

[3] O'Hara, K. and Hall, W., Semantic Web, *In: Marcia J. Bates; Mary Niles Maack; Miriam Drake (eds), Encyclopedia of Library and Information Science Second Edition,* Taylor & Francis, 2009.

[4] Ohler, J., The Semantic Web in Education – What happens when the read-write web gets smart enough to help us organize and evaluate the information it provides?, *EDUCAUSE Quarterly Magazine*, 31 (4), 2008.

[5] Tiropanis, T., Davis, H., Millard, D., Weal, M., *Semantic Technologies for Learning and Teaching in the Web 2.0 era-A survey*, 1st Web Science Conference, Athens, 2009, http://journal.webscience.org/166/

[16] http://simile.mit.edu/wiki/RDFizers