JISC - SemTech Project Report

July 2009

www.semtech.ecs.soton.ac.uk

Thanassis Tiropanis
Hugh Davis
David Millard
Mark Weal
Su White
Gary Wills

Surveys:
Asma Ounnas
Faith Lawrens
Heather S. Packer
Daniel A. Smith
Learning Societies Lab, ECS

Survey web-site:
Marcus Ramsden

University of Southampton

Acknowledgements

SemTech would like to acknowledge the invaluable support from the JISC CETIS Semantic Technology working group and JISC:

Sheila MacNeill (CETIS)
Lorna Campbell (CETIS)
Phil Barker (CETIS)
Helen Beetham
Simon Buckingham-Shum (Open University, UK)
David Davies (University of Warwick)
Michael Gardner (University Essex)
Tony Linde (University of Leicester)
Wilbert Kraan (CETIS)
Sue Manuel (University of Loughborough)
Lou McGill
Graham Wilson (LT Scotland)
Robin Wylie (LT Scotland)
David Kernohan (JISC)

SemTech is grateful to the members of the community who engaged with its activities and participated in its workshop:

Colin Allison (University of St. Andrews)
Chris Bailey (University of Bristol)
Liliana Cabral (Knowledge Media Institute, Open University)
Patrick Carmichael (University of Cambridge)
Tom Franklin (Franklin Consulting)
David Kay (Sero Consulting)
George Magoulas (London Knowledge Lab, Birkbeck College)
Uma Patel (City University)
Alex Poulovassilis (London Knowledge Lab, Birkbeck College)
John Scott (University of Essex)

Table of contents

1 FOREWORD 4

2 EXECUTIVE SUMMARY 5

3 TERMS OF REFERENCE 7

3.1 SOFT AND HARD SEMANTIC TECHNOLOGIES 7
3.2 LINKED DATA AND SEMANTIC TECHNOLOGIES 7
3.3 TEACHING AND LEARNING RELATED ACTIVITIES IN HE/FE 8
3.4 INSTITUTIONAL REPOSITORIES 9
3.5 THE UNDERPINNING PEDAGOGY OF TECHNOLOGY 9
3.6 THE EVIDENCE OF 'INTENT' FOR SEMANTIC ENRICHMENT 9
3.7 WEB 2.0 AND WEB 3.0 10

4 A SURVEY OF SEMANTIC TOOLS AND SERVICES FOR LEARNING AND TEACHING SUPPORT 11

4.1 SEMANTIC TOOL AND SERVICE CATEGORIES 11
4.2 SEMANTIC TOOLS AND SERVICES VALUE 12
4.2.1 THE LEARNING AND TEACHING PERSPECTIVE 13
4.2.2 THE INSTITUTIONAL PERSPECTIVE 14
4.2.3 HE/FE ACTIVITIES AND ACTORS 16
4.2.4 SURVEYED SEMANTIC TECHNOLOGY VALUE 17

5 SEMANTIC TECHNOLOGY USE IN UK HIGHER EDUCATION 18

6 A ROADMAP OF SEMANTIC TECHNOLOGY ADOPTION 20

6.1 STAGE ZERO: CREATING THE LINKED DATA FIELD ACROSS HE/FE INSTITUTIONS 20
6.2 STAGE ONE: APPLICATIONS ON A LINKED DATA FIELD 21
6.3 STAGE TWO: ONTOLOGY-BASED APPLICATIONS FOR SEARCHING AND MATCHING 22
6.4 STAGE THREE: EMERGENCE OF MORE PEDAGOGY AWARE APPLICATIONS 23

7 CONCLUSIONS 25

ANNEX A: SURVEY OF SEMANTIC TOOLS AND SERVICES FOR LEARNING AND TEACHING 27

ANNEX B: ADOPTION OF SEMANTIC TOOLS AND SERVICES IN THE UK HIGHER EDUCATION 28

1 Foreword

This report presents and discusses the findings of the SemTech (Semantic Technologies for Learning and Teaching) project that was funded by JISC and commenced its activities in September 2008.

SemTech addressed the following questions:

- What are semantic technologies?
- Which tools that make use of semantic technologies are, or could be, relevant to education?
- What is the actual use of semantic tools and services in UK HE and FE?
- What is the value of such tools in the context of learning, teaching and support?
- How do we envisage the adoption of semantic tools in higher education in the future?

Although a definition of semantic technologies was not in the scope of the project it was considered necessary for conducting the survey of semantic tools and services. SemTech distinguishes between (i) soft semantic technologies like topic maps and Web 2.0 applications, which provide lightweight knowledge modelling in formats understood by humans and (ii) hard semantic technologies like RDF, which provide knowledge modelling in formats processable by computers.

The outcomes of an extensive survey of semantic technologies relevant to learning and teaching are documented online (http://semtech-survey.ecs.soton.ac.uk). A total of thirty-six relevant tools and services have been identified by SemTech and the community at the moment of writing. Most of the surveyed tools were not purpose-built for education and they find value in semantic technologies for well-formed metadata or data interoperability and integration. The surveyed tools can be classified to (i) collaborative authoring and annotation, (ii) searching and matching, (iii) repositories, VLEs and authoring tools or (iv) infrastructural technologies for linked data and semantic enrichment.

The existing use and uptake of related tools and services by UK HE institutions was investigated and has been documented online (http://wiki.semtech.ecs.soton.ac.uk/). Repositories that are widely adopted in UK HE provide well-formed metadata using semantic technologies. A potential transition from soft semantic modelling in institutional wikis to hard semantic technologies could be implemented given the wide adoption of wikis and recent examples of *Wikipedia* and *dbpedia.org*. Repositories and collaboration-ware in HE could enable advanced searching and matching in line with HE requirements identified in this report.

Analysis of the findings of this report suggests that building a field of linked open data across UK HE/FE institutions by selectively and securely exposing repositories and institutional data (often data that can already be found on institutions' web pages) can provide significant value and pave the way for pedagogically meaningful applications powered by application-wide or community-wide agreed ontologies in the future. Encouraging institutions to use linked open data technologies and to document successful adoption of semantic technologies is considered of critical importance in this report. HE/FE challenges can be addressed by efficiently linking information across institutions.

SemTech engaged with the JISC CETIS Semantic Technology Working Group and the UK HE community for this report; during a workshop in January 2009 use cases of semantic technologies were constructed from both an educational and an institutional perspective.

2 Executive Summary

A definition of semantic technologies was one of the first challenges that SemTech had to address, although not in the initial scope of the project. However, it was considered necessary to set clear criteria on which technologies would be regarded as semantic for this report. SemTech makes a distinction between what we call *soft* and *hard* semantic technologies and between *linked data* and “traditional“ metadata using *ontologies*.

- We define *soft semantic technologies* as those that let people document certain concepts in formats that are easy to communicate to other people. These concepts can be communicated as part of learning and teaching processes. Examples of soft semantic annotation are folksonomies and topic maps.
- We define *hard semantic technologies* as those that support efficient exchange and processing of semantic data between programs and machines. An example would be *linked data* constructed from RDF statements.
- We distinguish between *linked data* that express the existence of relationships between resources and “traditional” metadata that express such relationships using *ontologies*.

The survey performed in SemTech shows that there is extensive use of soft semantic technologies in HE at the moment. Hard semantic technologies like RDF are initially used in some HE/FE repositories for interoperability. The HE/FE community seems to have identified the benefits of wider adoption of semantic technologies and there is relevant ongoing work.

Thirty-six semantic tools and services were found to be relevant to education and were surveyed at the moment of writing. Relevance was established on their suitability for teaching and learning activities or by their potential in HE/FE learning and teaching support. The survey was to detail this relevance and establish the value that semantic technologies add to them. The surveyed tools and services either use hard semantic technologies or they use soft semantic technologies but with intent of semantic enrichment. The semantic tools for learning and teaching that we surveyed included:

- Repositories
- Collaborative content creation tools
- Knowledge modelling and annotation tools (e.g. semantic wikis)
- Search tools
- Matching tools (e.g. matching experts to certain keywords)
- Mashups making use of linked data

The way these tools can be employed in an educational context at the moment can be seen to involve teachers, students, administrators and researchers; the lack of tools built or repurposed for learning reflects the lack of specific support for educational uses. However, there is evidence that a number of Web 2.0 tools that support soft semantics are used successfully in HE/FE; this report discusses the unique value that enrichment with hard semantics can introduce to such tools.

We considered the value of semantic technologies to the tools that we surveyed. In eighty-six per cent of cases there was value in using well-formed metadata, in forty-four per cent there was value in easing integration and interoperability costs, and in thirty-three per cent there was a value in improved data analysis and reasoning.

RDF is clearly the prevalent technology for annotation. The use of RDF to link data in ways that do not necessarily require agreement on ontologies makes exposing relationships among data more efficient; more expressive knowledge modelling can then take place in application-specific contexts, potentially over larger linked data sets where RDF concepts are mapped to context-specific ontologies. Some of the tools that seem to make heavier use of knowledge modelling involve matching of people and resources or providing support for argumentation and critical thinking.

The surveyed tools classified as repositories make use of metadata to enable more advanced searching or to expose their data in interoperable and machine processable ways. Semantic wiki tools enable efficient representation of collaboratively authored content, while semantic searching enables contextualised queries. Intelligent recommendation of relevant content, to assist collaborative authoring and Q&A systems, is enabled for tools that match people and resources. The currently exploited value in terms of support of learning and teaching processes appears to be mainly in linking repositories using RDF and in enabling searches across these repositories.

We conclude that the adoption of semantic technologies on a wider scale will be enabled if a sufficiently large volume of linked data is exposed in machine processable and declarative formats like RDF. There are many examples of RDF repositories linked to each other, like *dbpedia.org* and *freebase.com*[1], and many more[2]. Many of the semantic applications we surveyed access RDF repositories or harvest linked data from existing sources using ad-hoc approaches.

Activities to encourage exposure of data by HE/FE institutions in formats like RDF and the linking of data repositories can pave the way for the development of semantic applications to efficiently support learning and teaching; for example matching University courses to student interests or assisting in curriculum alignment across the HE sector. In addition, they can provide a foundation on which more rigorous knowledge modelling can flourish and support innovative applications such as argumentation tools that require more advanced ontologies and reasoning.

The development of semantic applications for teaching and learning for HE/FE over the next five years could be supported in a number of steps:

1. Encouraging the exposure of HE/FE repositories, VLEs, databases and existing Web 2.0 lightweight knowledge models in linked data formats. Enabling the development of learning and teaching applications that make use of linked data across HE/FE institutions; there is significant activity on linked open data to be considered[3].
2. Enabling the deployment of semantic-based searching and matching services to enhance learning. Such applications could support group formation and learning resource recommendation based on linked data. The development of ontologies to which linked data will be matched is anticipated. The specification of patterns of semantic tools and services using linked data could be fostered.
3. Collaborative ontology building and reasoning for pedagogical ends will be more valuable if deployed over a large volume of education related linked data where the value of searching and matching is sufficiently demonstrated. Pedagogy-aware applications making use reasoning to establish learning context and to support argumentation and critical thinking over a large linked data field could be encouraged at this stage.

---

[1] http://blog.dbpedia.org/2008/11/15/dbpedia-is-now-interlinked-with-freebase-links-to-opencyc-updated/

[2] http://linkeddata.org/

[3] http://linkeddata.org/, http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData, http://www.umbel.org/lod_constellation.html

3 Terms of Reference

3.1 Soft and Hard Semantic Technologies

Semantic technologies enable the expression of the meaning of resources such as content, programs and people, and their relationships in machine processable ways. They also provide the mechanisms to draw conclusions (reasoning) based on this meaning; these mechanisms are independent of the meaning itself. Thus semantic technologies can provide for more precise resource discovery and complex queries over various datasets, for example '*provide me a list of the modules in computer science curricula that were introduced after the year 2000 in UK universities'*.

Metadata cannot qualify as semantic technologies if they do not come in formats that are processable by machines alone. Nevertheless, machine processable metadata (e.g. in XML) cannot qualify as semantic technologies, for the purposes of this report, if the schemas they conform to are meaning specific (e.g. XML vocabularies for specific domains like education, chemistry, etc) and therefore cannot be processed using generic inferencing algorithms. Typical relational databases cannot qualify as semantic technologies, since they cannot provide for reasoning, i.e. they cannot infer information that is not explicitly stated.

Metadata expressed in RDF/XML can qualify as semantic technologies since the schema for serialising RDF in XML is not meaning specific (e.g. the RDF schema does not vary per domain like education, chemistry, etc). Tagging uniquely identified resources could qualify as a semantic technology if tagging information becomes available in formats like RDF/XML or N3; tagging could also be in microformats like RDFa[4] (RDF in attributes).

In this report we distinguish between *hard semantic technologies,* which fit the definition above and *soft semantic technologies,* which provide for the expression of meaning for resources in ways that can be understood and processed by people but not by machines. Examples of hard semantic technologies are RDF[5], FOAF[6], SKOS[7] or Triple Stores (large RDF metadata repositories), while examples of soft semantic technologies are traditional tagging tools and topic maps. The simple use of the term semantic technologies in this report implies hard semantic technologies.

3.2 Linked data and semantic technologies

Expressing the meaning of resources and relationships requires unambiguous identification of the concepts to be expressed, i.e. agreement on ontologies, which has often presented a barrier for semantic technology adoption. The case for employing semantic technologies is often based on one or both of the following prerequisites: (i) sufficient *volume* of resources annotated with 'meaning', and, (ii) support for sufficiently complex reasoning based on resource annotations. Applications that use a large volume of resources often have to rely on annotations using a small number of commonly agreed concepts, while applications that feature complex reasoning often rely on a smaller volume of resources annotated using more expressive ontologies. When it comes to resources available in a textual form the performance of a deep linguistics analysis to extract meaning can be employed;

---

[4] http://esw.w3.org/topic/RDFa

[5] Resource Description Framework http://www.w3.org/RDF/

[6] Friend Of A Friend vocabulary http://www.foaf-project.org/

[7] Simple Knowledge Organisation System http://www.w3.org/2004/02/skos/

current implementations include *COGITO®* by *ExpertSystem*[8] and *Nactem*[9].

In recent years we have witnessed the emergence of *linked data,* which involves exposing existing data repositories using semantic technologies like RDF. RDF makes it possible to query and merge information from different sources without requiring detailed knowledge of database schemas or ontologies; although such knowledge can help to interpret the query results.

Linked data provide for applications that establish value first in the volume and scale of annotated resources rather than in the support for complex reasoning. This enables a different approach for the development of semantic applications to what has so far been followed. Linked data enable for a bottom-up approach of exposing data first and then considering ontologies to which linked data can be mapped depending on our intended use of it. In the bottom-up approach the vocabulary behind linked data could come from the database schema from which the data potentially originates, Dublin Core, or other vocabularies. There are a number of requirements on linked data[10] among which is the requirement for dereferencable URIs.

The transition of simple semantic annotations to richer ones may draw attention in the future. It seems that mapping linked data to richer ontologies can be enabled in specific domains where agreement on such ontologies is possible. In addition, mapping linked data to richer but very idiosyncratic local ontologies could provide educationally interesting applications. However there are scarcely any examples of such applications at the time of writing.

3.3 Teaching and learning related activities in HE/FE

There are various classifications of teaching and learning activities for HE/FE and design for learning[11]. Establishing the relevance of the semantic tools and services to these activities can be subjective since there are different views on how such activities need to be performed and on their pedagogical objectives and value.

Considering that each activity involves a number of actors interacting with information systems and with each other allows us to take a more pragmatic approach and *outline* activities in these terms. We therefore establish the relevance of tools to learning and teaching in the following way:

- We identify the education sector that a semantic tool or service is aimed for (e.g. HE, K-12)
- We identify the actors involved in the use of each semantic tool or service (e.g. students, administrators)
- We identify the individual activities supported (e.g. information gathering)
- We identify the collaborative activities supported (e.g. collaborative content annotation)

The degree to which each semantic tool or service makes use of semantic technologies is established based on:

- Its support of hard and soft semantic technologies for knowledge modelling
- Its support of annotation activities

---

[8] http://www.expertsystem.net

[9] http://nactem/

[10] http://www.w3.org/DesignIssues/LinkedData.html

[11] Beetham, H., Sharpe, R. (Eds) (2007) Rethinking Pedagogy for a Digital Age - Designing and delivering e-learning, Routledge

- The value of hard semantic technologies for the specific tool or service

The relevance of tools and services to HE/FE does not involve only teaching and learning activities but also activities that support the education environment; for example learning repositories, collaboration environments, admissions support.

3.4 Institutional repositories

A significant amount of information is maintained by HE/FE institutions in internal databases, VLEs (Virtual Learning Environments), file systems and internal or external Web pages. Such information may include teaching material, research material, admissions data, course syllabi and learning outcomes.

These types of repositories are often non-interoperable across institutions or even departments due to a number of reasons: database schemas are not known, database servers are placed behind firewalls, files are in different formats and require a user account to be accessed. Data available on Web pages in (X)HTML can be accessible across departments or institutions. Data available in XML, in linked data formats or in (X)HTML enriched with RDFa attributes[12] can be accessible across departments or institutions and can be processed by software.

The use of information across institutional repositories could be relevant to addressing important HE/FE challenges. HE/FE institutions could start exposing such repositories in linked data formats starting with information that is already available on their Web pages (e.g. course syllabi). This could make their information more readily available for mashups and search engines. Search engines like Yahoo! and Google are already starting to make use of exposed information as linked data.

3.5 The underpinning pedagogy of technology

Most of the tools and services surveyed for this report were not purpose-built for education and therefore there was no pedagogical intent in their development. A categorisation of these tools and services based on their functionality classifies them in the following categories:

- Collaborative content authoring and annotation
- Searching and matching
- Repositories, VLEs and Authoring tools
- Infrastructural tools and services that support development in the above three categories

The pedagogical value of tools that enable collaborative authoring, searching, matching, repositories and VLEs has been established in many scientific publications. However, many of these applications lack support for specifying and considering the educational context in which they are used, which limits their learning and teaching value. In this report we focus on pedagogically meaningful features that can be enabled with the enrichment of such services with semantic technologies.

3.6 The evidence of ‘intent’ for semantic enrichment

Some of the tools and services that were surveyed seem to make use of soft semantic technologies at the moment but in most cases there seems to be intent for future transition or adoption of hard

[12] Metadata from RDFa enriched (X)HTML documents can be gleaned using GRDDL tools (http://www.w3.org/2001/sw/grddl-wg/).

semantic technologies. This intent can be established in the following cases:

- When there is evidence of starting to adopt hard semantic technologies (e.g. use of RDF to export data from a repository)
- When there is stated intent of adoption of hard semantic technologies by the tool/service creators (e.g. in interviews or articles)

Often there is awareness of the potential of hard semantic technologies but there are a number of reasons that prevent their adoption. These are discussed in section 4.2 and section 6 of this report.

3.7 Web 2.0 and Web 3.0

Web 2.0 is used to describe 'the social Web', Web technologies enabling more efficient collaboration over the Web chief amongst which is collaborative authoring and collaborative annotation with tagging. In most cases what we define as *soft semantic technologies* are employed in Web 2.0 environments to describe and classify collaboratively authored content.

Recent developments indicate that there is need for a transition to hard semantics in order to extract a more formal specification of Web 2.0 content classification and to enable more precise searching and matching (for example the classification of *Wikipedia* is made available in semantic formats in *dbpedia.org*). This transition from a Web of documents with soft semantics to a Web of linked data[10] or Semantic Web or Web 3.0 will involve the use of ontologies in metadata but following a bottom-up approach rather than a top-down one[13].

---

[13] O'Hara, K. and Hall, W. (2009) Semantic Web. In: Marcia J. Bates; Mary Niles Maack; Miriam Drake (eds), Encyclopedia of Library and Information Science Second Edition, Taylor & Francis. ISBN 978-0-8247-2075-9 (In Press) http://eprints.ecs.soton.ac.uk/17126/1/hall-ohara-elis-semantic-web.pdf

4 A survey of semantic tools and services for learning and teaching support

This section presents the results of a survey of semantic tools and services relevant to education, which involved engagement with JISC, JISC CETIS and the academic community. Details on each surveyed tool or service mentioned in this report are available on the project web site www.semtech.ecs.soton.ac.uk.

4.1 Semantic tool and service categories

The objective was to identify tools and services that make use of semantic technologies or demonstrate *intent* to use semantic technologies as defined in the terms of reference of the previous section.

The survey resulted in the identification of 36 relevant tools and services initially, which can be coarsely classified into four main categories based on their main types of functionality:

- Collaborative authoring and annotation tools
- Searching and matching tools using semantic technologies
- Repositories and VLEs that import/export their data using semantic technologies
- Infrastructural tools and services that enable use of semantic technologies.

Collaborative authoring and annotation tools often build on the concept of wikis (e.g. *AceWiki*, *Cicero*, *Mymory*, *Kiwi*) and they present three main features:

- Collaborative authoring and annotation of resources. Annotation can be on different levels of granularity (e.g. *Mymory* lets users annotate not only whole pages but parts of text too)
- Support for argumentation (e.g. *Cicero*, *Debategraph*, *Compendium*)
- Inline discovery of resources relevant to the collaboration topics (e.g. *AWESOME*)

Three of the tools surveyed in this category have been used in education in the UK. One of them is *PROWE*, which relies on FOAF (Friend Of A Friend) ontologies to enable collaboration in academia. Another tool is Compendium, which enables visualisation of arguments during online discussions. Technologies like WiKis, RDF and FOAF provide for well-formed and interoperable metadata and for basic or more advanced reasoning. Finally, the *AWESOME* project is an example of using semantic wiki and inline semantic searching to assist knowledge discovery in higher education.

The value of hard semantic technologies for collaborative authoring and annotation tools is in being able to describe the relationship between content elements more precisely and in formats that can be efficiently processed by machines. The survey shows that both well-formed metadata and advanced data analysis and reasoning constitute the value of semantic technologies for over half of the surveyed tools in this category. Hard semantics can enable more efficient matching of resources to the context of each collaboration session. The recommendation of matching resources can take place inline and improve the quality of the collaborative activity like in *AWESOME*.

It is expected that the advantages of using hard semantics in collaboration environments will become clearer in the near future as a critical mass of collaboratively authored resources becomes available. We believe it will be inevitable that the knowledge structure of the collaboratively authored resources will have to become available in hard semantic formats thus making it relevant and accessible to additional communities in the way that *dbpedia.org* and *freebase.com* are.

Searching and matching tools that use semantic technologies were identified during the survey: tools and services that support annotation for resources and provide for advanced searching (e.g. *Yahoo! SearchMonkey, Watson, Twine*) and matching (e.g. *ArnetMiner, Twine*). *ArnetMiner* is reported to be used in an education context in China. RDF is the main technology employed to provide semantic support with vocabularies like FOAF and Dublin Core. Basic reasoning is used to provide targeted search results. The availability of repository metadata in hard semantic formats like RDF will enable more efficient searching. *ArnetMiner* extracts metadata in hard semantic formats from existing Web resources. *Yahoo! SearchMonkey* and more recently Google make use of the metadata added on Web resources by the authors to provide more targeted searches to their users and developers; such metadata is embedded in Web pages using *microformats* like RDFa or eRDF[14].

Twelve different repositories and repository related tools that can import or export data in RDF were surveyed. Some of them are global repositories (*Project Gutenberg*, *Sweet Tools*, *freebase.com*, *MyExperiment*) while others can add value to institutional data (e.g. *CIP*). Most of the surveyed repositories make use of RDF and SPARQL[15] endpoints (which can be queried and expose matching RDF statements) to enable interoperability and integration of data from a number of data sources. Specifically, 10 of the 12 surveyed tools in this category rely on RDF for interoperability. Semantic tools that do not appear to use RDF, namely *Autology* and the *MIT Course Picker* do not seem to have intended to address interoperability across repositories.

A number of the surveyed tools and services provide a semantic technology infrastructure to support collaboration, searching/matching and repository services. These infrastructural tools and services enable:

- Exposing relational databases in RDF via SPARQL endpoints (e.g. *D2R Server*, *TALIS*)
- Integration of data from diverse sources (*mSpace*)
- Authoring and/or hosting metadata or RDF data (e.g. *RKBexplorer*, *Konduit*)
- Embedding RDF into Web pages in microformats like RDFa[16].
- Gleaning or scraping metadata from Web resources using *GRDDL*[17] or scrapers like *Solvent*[18].

Some of these infrastructural services are already used by repository, searching/matching or collaboration tools and services. For example, *D2R Server* is already used by *Project Gutenberg* and the *DBLP* repositories to expose parts of their relational databases as RDF via SPARQL endpoints.

4.2 Semantic tools and services value

The emergence of Web 2.0 applications supports collaboration over the Web by means of collaborative content creation on a large scale. The implications of these new services for education have been identified[19].

---

[14] http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml

[15] SPARQL Query Language for RDF, W3C Recommendation, 15 Jan 2008 http://www.w3.org/TR/rdf-sparql-query/

[16] http://rdfa.info/wiki/Tools

[17] http://www.w3.org/2004/01/rdxh/spec

[18] http://simile.mit.edu/wiki/Solvent

[19] Anderson, P. (2007) What is Web 2.0? Ideas, technologies and implications for education, JISC Technology & Standards Watch, Feb. 2007 http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf

Machine processable and interoperable semantic descriptions of people and resources can enhance collaboration with inline recommendations and modelling of collaboratively modelled knowledge. Contextualised searching and matching with semantic technologies (e.g. *ArnetMiner*), can provide for more efficient resource discovery (e.g. *Yahoo! SearchMonkey*), access to data over a large number of repositories and support for critical thinking and argumentation by means of reasoning. In addition, the statements on which conclusions were drawn and the process of reasoning can be traced and validated, which is different to the ad-hoc way recommender systems are often employed. Finally, modelling resources and concepts related to the life of individuals such as their calendars, commitments, abilities and location could provide for more efficient adaptation of resources and interactions for learning and teaching. For example, pointing higher education students of art to museums or galleries that happen to be in their vicinity to enhance their learning experience requires interoperability among a large number of institutional and general repositories.

The most significant barrier to the adoption of semantics so far seems to have been the requirement for agreement on knowledge modelling concepts (ontologies) and the need to annotate and use data across different administrative domains (such as different University information systems or different HE/FE institutions). This challenge is addressed by the emerging linked data movement which follows a bottom-up approach of exposing data first, before seeking agreement on ontologies.

Tools that will enable the selective and secure exposure of existing HE/FE institutional repositories in semantic, interoperable formats presented another challenge. The standardisation of SPARQL (Query Language for RDF) on 15 January 2008 and infrastructural tools like *D2R* can enable the partial exposure of relational databases as linked data and address this challenge. Such tools enable the instantiation of *SPARQL endpoints*; servers that can be queried to provide linked data in RDF format.

4.2.1 The learning and teaching perspective

During the SemTech workshop in London in January 2009 the teaching and learning challenges that semantic tools and services can address in the Web 2.0 era in the UK HE/FE were outlined as follows:

- Enabling access to teaching and learning material across institutions to improve the quality of learning in UK HE/FE. Semantic technologies could provide for contextualised resource discovery based on the field of study, type of teaching and learning activity, learning theory or pedagogical framework.
- Assisting the workflow of course creation, delivery and revision by recommending relevant content and people for specific tasks. Surveyed tools that seem to address these challenges to a certain extent are the ones provided by the EU *LUISA project*, *ArnetMiner*, *Yahoo! SearchMonkey*, *Watson* and *SELF*.
- Assisting students by recommending resources that match the topics of their assignments and people that may be able to support them. Tools that seem to support these activities include *ArnetMiner* and *SELF*.
  Group formation for collaborative work. None of the surveyed tools seems to support such activity but such tools are under development[20].
- Support for critical thinking and argumentation by visualising arguments and linking relevant discussions. Tools and services like *Debategraph*, *Compendium* and *Cicero* provide support for visualisation of arguments during online discussions. *Debategraph* does not appear to be

[20] Ounnas, A., Davis, H. and Millard, D. (2008) A Framework for Semantic Group Formation. In: The 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008), July 1st- July 5th, 2008, Santander, Cantabria, Spain.

using semantic Web standards but its creators are reported to be examining this possibility[21]. In addition, activities on schemas for the exchange of deliberations among Open University's *Cohere & Compendium*, *Debategraph* and *MIT's Deliberatorium*[22] are already underway.

- Efficient support for cross-curricular activities in emerging areas by matching people and resources. None of the surveyed tools addresses this requirement; potentially *ArnetMiner* is most relevant.
- More efficient personalised knowledge construction to assist parties involved in learning and teaching. Tools that are based on topic maps can provide support in this direction by means of soft semantics. *TM4L* utilises topic maps to build ontology-driven repositories. Similarly, *MyPlan* builds a network of access to the knowledge capital of higher education and further education institutions.
- More efficient support of contextualised group knowledge construction. A large number of semantic wikis and additional collaborative authoring tools can enable group knowledge construction, for example: *AceWiki*, *Kiwi*, *Mymory*, *Revyu* and *Swim*. A number of tools provide for collaborative knowledge building in specific contexts, for example SWIM enables collaborative knowledge construction in mathematics.

From a learning and teaching perspective, the potential of each category of the surveyed tools and services can be summarised as follows:

Collaborative content authoring and annotation

- Representation of shared knowledge with precision
- Inline recommendation of related content and people for collaborative activities
- Documentation and support of collaboration workflows on a larger or smaller scale
- Support for argumentation and visualisation of arguments and related resources to enable critical thinking

Searching and matching

- Contextualised queries and searches
- Searches across repositories
- More efficient Question and Answer systems and knowledge bases
- Matching people for collaborative activities

Repositories, VLEs and authoring tools

- Semantic annotation of content to support more precise knowledge construction
- Use of hard semantic technologies to enable interoperability and integration of repositories across institutions

Infrastructural technologies

- Large repositories for efficient storage and search of data from different sources

4.2.2 The institutional perspective

Apart from the challenges in learning and teaching there are well-identified challenges in HE that can be addressed using semantic tools and services. Some of these challenges were identified at the SemTech workshop in which about ten UK higher education institutions and HE/FE related

[21] http://interviews.liveinterviewsonline.com/content/interview/detail/1525/

[22] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1099082

organisations were represented:

- Student retention with more effective student support, access to resources and student progress monitoring. Institutions would like to know of early signs of student disengagement with their studies, which can be established by combining information in different data sources. Analysis on student interests, modules taken and modules attended could further support student retention. Some of the surveyed tools that could be of relevance include *ArnetMiner* for matching experts and *the* VLEs proposed by the *EU LUISA Project.*
- Information in UK HE/FE institution seems to be fragmented and in formats that makes it often inaccessible. Discovery of relevant information over a large number of sources needs to be supported. Information that is publicly available on the institutions' Web pages is not available in machine processable formats making it difficult to compare programmes of study, syllabuses or research angles. HE institutions could attract more international students (or students on the ERASMUS programme) were their programmes available in machine processable formats for comparison. Institutions adopting the XCRI[23] vocabulary could easily expose the specification of their courses in semantic formats like RDF to enable efficient searching and comparisons. Transformation of other institutional information to hard semantic formats using tools like the *TALIS* or the *Virtuoso* ones could be relevant to this challenge; notably, *TALIS Aspire* provides for integration of resources within a HE context.
- There is a challenge for transparency and validation of information sources. Data need to be available from the right sources to the right people. Tools like the *D2R Server* and the *RKBExplorer* could enable the selective export of information in accessible formats at accessible locations addressing this requirement to a small extent.
- Course information and material available in VLEs and information on programmes and courses available in different institutional databases could be integrated using semantic Web standards enabling more efficient curriculum, programme or module design.
- The expertise and research angle in Universities can attract students and industrial funding. Keywords based search is not efficient to establish expertise in finely defined research areas. Semantic technologies could make this expertise accessible in machine processable formats.
- While teaching and research collaboration within institutions within departments is supported and collaboration across departments can under certain circumstances be enabled, collaboration across institutions cannot be supported since the relevant information systems of universities are not interoperable. Large repositories, to which information can be efficiently stored, searched and managed, like the *RKBExplorer,* relate to these requirements.
- The quality of institutional data can often vary since institutional databases can be outdated or incomplete. Semantic technologies could assist in placing institutional data in the right context for interpretation. The *reification*[24] supported by semantic Web standards could be relevant for this end. In addition, data from disparate sources could be normalised in RDF.
- There is lack of a framework to enable each institution to state the IPR of the resources they are eager to expose in order to attract funding and students. Large repositories can address issues to a small extent by time-stamping contributions but provide little support for stating IPR. Semantic technologies could address such issues to a more satisfactory extent but none of the surveyed tools seems to do this at the time of writing.
- UK HE/FE institutions spend significant resources to support their case for accreditation by professional bodies, for research assessment or for similar activities. Exposing institutional data in machine processable formats using semantic technologies like RDF could assist institutions in integrating information across departments and accreditation bodies in accessing

[23] http://www.xcri.org/

[24] http://www.w3.org/TR/rdf-mt/#ReifAndCont

and reviewing this information.

One can observe that many of the above challenges can be addressed by making data that is *already public* on institutional web pages available in machine processable formats. Surveyed tools and services can address these challenges in the following ways:

Collaborative content authoring and annotation

- Representation of the shared knowledge capital of higher education institutions in ways that can be accessed by different faculties, schools, other institutions and the public
- Documentation and support of collaboration workflows within and across higher education and further education institutions

Searching and matching

- Searches across repositories different departments or institutions
- Efficient Question and Answer systems or knowledge bases for learning and teaching support
- Exposing institutions' expertise to attract funding and students
- Combination of information scattered within institutions to enable better monitoring of student progress and recommendations based on declarative statements that can be validated

Repositories, VLEs and authoring tools

- Interoperability across repositories within or across institutions
- Semantic enrichment of classifications in repositories to enable more efficient resource discovery and interoperability
- Integration across repositories to enable visibility of the knowledge capital of institutions and to attract funding and talent

Infrastructural technologies

- Exposing part of organisational data to partners
- Formatting data in interoperable machine-processable formats
- Integration of data from different sources
- Large repositories for efficient aggregation of data scattered on different platforms

4.2.3 HE/FE activities and actors

The majority of the surveyed semantic tools and services were not purpose-built for use in an educational context. This reflects on their potential uses in HE/FE, which seem to be generic, e.g. *computer mediated discussion, collaborative content creation/annotation, information gathering/handling* and *publishing*. Additional activities such as *team building, computer mediated experimentation, role-play based activities, case based learning* or *simulations* cannot be fully supported by existing tools although they relate to the expectations from semantic technologies from the institutional and learning and teaching perspective.

The HE/FE types of users (actors or roles) that could be involved in the use of the surveyed tools and services include mainly *teachers*, *students* and *researchers*. There is no support for more finely identified roles such as *assessors/examiners, university administrators, system administrators, programme/module co-ordinators* or *admissions team*. This can be justified by the fact that most of the surveyed tools are not purpose-build for education and that there are hardly any semantic tools and services to address institutional perspective at the time of writing.

4.2.4 Surveyed semantic technology value

There is value in the use of semantic technology and semantic Web standards for over 4 in 5 of the surveyed tools and services in *well-formed metadata*. For almost half of the surveyed tools there is additional value in *data integration and interoperability* using semantic technologies. Value in *data analysis and reasoning* appears to be the case for just under 2 in 5 of the surveyed tools.

The surveyed services for which semantic technologies add value in well-formed metadata but not in data integration and interoperability are mainly repositories or infrastructural tools that aim to facilitate resource annotation, e.g. *DBLP, EPrints, Project Gutenberg, Talis, D2R Server.* There are however tools that could benefit from data integration with other relevant resources should they become accessible, e.g. *Cicero, Debategraph, Konduit, PROWE.*

Most of the collaboration tools could benefit from data analysis and reasoning for inline recommendation of people or resources in collaborative learning activities. Searching and matching tools for learning and teaching could benefit from data analysis and reasoning and the recent developments in *Yahoo! SearchMonkey* and in *Google*[25] seem to second this claim. Repositories, VLEs and annotation tools could potentially be linked using semantic Web standards and address requirements on collaborative curriculum and module design as identified from the institutional perspective. Given well-formedness of metadata and interoperability, reasoning could provide for more pedagogically meaningful analysis and combination of data and resources.

A list of teaching and learning related activities that semantic technologies could bring potentially unique value to includes:

- Group formation for group learning activities where the learners have backgrounds and objectives that can be aligned to pedagogical ends.
- Activities that enable critical thinking by providing learners with access to relevant resources across repositories and matching them with learners supporting similar or opposite arguments.
- Support of cross curricular learning and teaching activities in emerging areas by matching teachers to new programmes and modules.
- Identification of the context in which learning and teaching activities take place and assistance in constructive alignment with intended learning outcomes and assessment when designing modules or courses.

[25] http://www.tgdaily.com/content/view/41845/113/

5 Semantic technology use in UK higher education

There is evidence that a number of the surveyed tools and services that employ hard semantic technologies are used by higher education institutions. At the same time, there appears to be a significant number of tools, widely adopted by institutions, which make use of soft semantic technologies. In these cases, requirements for interoperability, integration with additional data sources and reasoning might serve as powerful incentives for transition to hard semantic technologies.

Fourteen UK universities appear to have adopted the use of wikis to support learning and teaching on an institutional level. Wikis provide for community agreed lightweight knowledge modelling in addition to collaborative content creation. Current activities on the development of semantic wikis (e.g. *Semantic Media Wiki, Kiwi*, *Samizdat*) indicate awareness of the advantages of adding meaning to the relationships among wiki pages or sections of wiki pages (e.g. *Mymory*). In addition, the support for reasoning with hard semantic technologies can provide for additional collaborative activities such as argumentation, where the relevance between arguments can be precisely identified and used for navigation and visualisation of discussions (e.g. *Cicero*).

Past projects have established the value of knowledge modelling in wiki environments and have been adopted by higher education institutions, for example the JISC funded *PROWE* wiki used by the University of Leicester, which uses FOAF to enable metadata interoperability instead of more complex metadata schemas[26]. Similarly, current work on the exchange of deliberations among Open University's *Cohere & Compendium*, *Debategraph* and *MIT's Deliberatorium* may require RDF or other semantic technologies.

Expert matching as provided by the *ArnetMiner* service appears used by the University of Tsinghua in China[27]. The JISC funded *AWESOME* project provides software that combines semantic wiki and pedagogy-aware inline recommendations to empower academic writing and is used by a number of schools at the University of Leeds, University Coventry and University of Bangor[28].

Repositories are the most widely adopted type of infrastructure by the UK HE/FE institutions. Over forty universities are reported to employ repositories in the UK to publish their research results, conference and journal articles, presentations or course material. Most of these repositories provide access to scholarly work rather than good quality teaching and learning material; the latter seems to be maintained mainly in internal institutional Web pages or VLEs. Repositories like *EdShare*[29] are working on filling this gap.

The most common repository platforms in use are *DSpace* and *EPrints,* both of which reported to be adding semantic extensions to enable interoperability with other repositories and information export in additional machine processable formats. The *SIMILE* project is looking at tools that will enable metadata interoperability with *DSpace* as its target user. A number of tools identified as *RDFizers*[30]

[26] http://www.prowe.ac.uk//documents/ALISSonPROWE080207.doc

[27] Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. 2008. ArnetMiner: extraction and mining of academic social networks. In Proceeding of the 14th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining (Las Vegas, Nevada, USA, August 24 - 27, 2008). KDD '08. ACM, New York, NY, 990-998. DOI= http://doi.acm.org/10.1145/1401890.1402008

[28] http://awesome.leeds.ac.uk

[29] http://www.edshare.soton.ac.uk/

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

also aim to support RDF-enabled interoperability. *EPrints* can export its metadata in a number of machine processable formats including RDF.

Apart from repositories, the School of ECS at the University of Southampton provides information on a number of entities like its people, roles, interests, courses, seminars or presentations in RDF format[31]; this RDF is available on the Web alongside the XHTML versions and is also updated regularly on *RKBExplorer*. Apart from the UK, the DERI institute in Ireland also provides RDF on the Web along with its XHTML resources.

There are a large number of infrastructural technologies that support semantic tools and services in UK institutions. Very often it is hard to determine which ones are used in each case, for example a number *RDFizers* can be used in suits of tools to enable export of repository metadata to RDF but it is often hard to identify each one separately.

There is evidence that a number of infrastructural tools are used by UK higher education institutions to expose data from relational databases via SPARQL endpoints (for example the University of Southampton and the University of Oxford) but for the moment this concerns research projects. Nevertheless, D2R is also used for *Project Gutenberg* and the *DBLP* library, which is of high relevance to education institutions engaging in research activities or supporting student access to resources.

A trend for education-targeted mashups[32] has been reported. Examples are *LazyLibrary*[33], *Find the Landmark*[34] and *CampusExplorer*[35]; the latter enables searching for courses across 6000 US universities. Technologies to enable export of organisational information for inclusion to education related mashups might drive further development of infrastructural tools to address transition to metadata using semantic technologies.

The ENSEMBLE project, jointly funded by ESRC/EPSRC, is investigating the potential of the semantic Web in rapidly evolving fields where case based learning is the pedagogical approach of choice[36].

---

31 http://id.ecs.soton.ac.uk/docs/

32 http://emergingtechnologies.becta.org.uk/upload-dir/downloads/page_documents/research/technews/nov08.pdf

33 http://lazylibrary.com/

34 http://landmark.mapsgame.com/

35 http://www.campusexplorer.com/

36 http://www.ensemble.ac.uk

6 A roadmap of semantic technology adoption

Based on (i) the survey of tools and services (ii) the survey of adoption of these tools in the UK higher education sector and (iii) the challenges that semantic tools and services can support as highlighted in the section on 'Semantic technologies value', this report outlines a roadmap for semantic technology adoption in the coming years.

This roadmap foresees three major stages in technology adoption in this period, which to a large extent involve transition from soft semantic technologies to hard semantic technologies, selective and secure exposure of institutional repositories, linked data applications across institutional repositories, collaborative knowledge modelling and support for pedagogy aware applications for learning and teaching.

Each stage involves challenges and drives for adoption of semantic technologies or transition from soft to hard semantic technology use as illustrated in Figure 1. In addition, each stage presents higher education and further education institutions with value; this value makes the case of semantic technology adoption in each step of the roadmap.

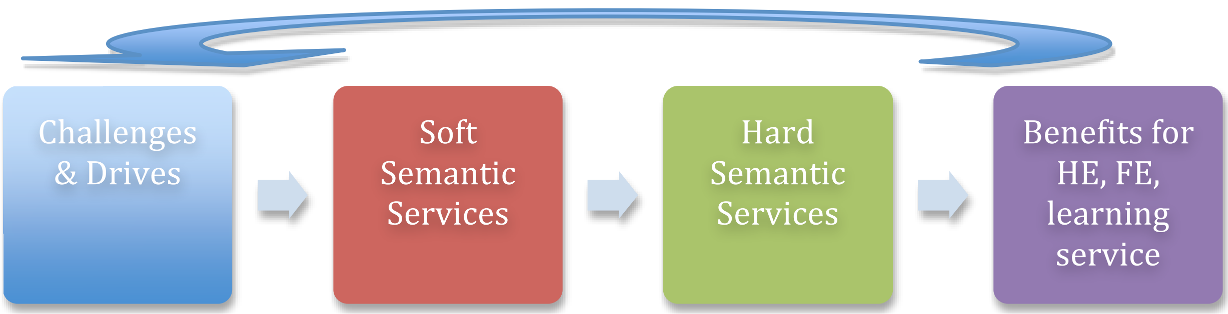

Figure 1: Stage of roadmap for the development and adoption of semantic services for learning

It is likely that the stages outlined in the following paragraphs will present overlaps, particularly between stage one and stage two and between stage two and stage three. The stages of the roadmap as discussed below are also illustrated in Figure 2.

6.1 Stage zero: creating the linked data field across HE/FE institutions

This stage represents the present situation in which a number of repositories relevant to HE/FE are isolated either due to institutional policies or because they are available in formats that are not interoperable. Our survey shows that there is value to be gained by letting institutions have access to external repositories and by sharing their data with them; this value will initially materialise in more powerful cross-repository search utilities. We expect this to apply to both repositories of scholarly work and repositories of teaching and learning material.

Data already exposed by institutions on their public Web pages could also be exposed as linked data to address institutional challenges. Technologies that enable embedding metadata into existing Web page content in microformats (such as RDFa) are growing in popularity and are starting to be used by search engines like Yahoo! and Google to offer more targeted searching. Tools using GRDDL can glean semantics from resources in (X)HTML and microformats. Infrastructural semantic tools like *D2R Server* will ease the exposure of relational database information in interoperable formats like RDF via

SPARQL endpoints. At the same time, large RDF repositories (like *RKBExplorer*) will optimise the storage and management of such metadata. This sets the scene for stage one of semantic technology adoption.

6.2 Stage one: applications on a linked data field

Period: starting now and intensifying for the next two to three years.

Challenges and drives: As university repositories will be able to expose their metadata in interoperable formats like RDF/RDFa and given that support for advanced searching based on metadata is already implemented by search engines like *Yahoo! SearchMonkey* and *Google* it will be in the interests of all parties to successfully enable interoperability among repositories. Agreement on ontologies is elusive when the objective is integration on a large scale. The main objective during this stage, is expected to involve establishing links among resources in educational, institutional and global repositories creating a *linked data field for HE/FE*. Institutions could be encouraged to consider exposing additional information (to what they already make publicly available on their Web pages) as linked data to address additional challenges from the institutional or the learning and teaching perspective. In addition to scholarly work they could also be encouraged to expose teaching and learning material, curricula and syllabi as linked data.

Benefits for HE/FE institutions: Learners and teachers will be able to efficiently search across various repositories. Learning and teaching will be better supported with utilities that enable targeted searching on *authoritative* teaching and learning material across institutional repositories. Prospective students and module designers will be able to make comparisons of curricula if such information is exposed in linked data formats.

Enabling institutional policies: Exposing publicly available information as linked data would be a significant step at stage one. Teaching and learning material can also become available as linked data. Institutions may also start considering which information currently available in internal or external repositories or in repositories is relevant to addressing student retention, employability and other HE/FE challenges.

Enabling technologies: Tools like *RDFizers*, *TALIS*, *Virtuoso* or *Collibra* can enable semantic enrichment, export and integration of metadata. In addition, the knowledge models collaboratively established with Web 2.0 tools like Wikis will need to be exported to interoperable formats; such a trend is already observed with large repositories like Wikipedia, the metadata of which is available in *dbpedia.org* and *freebase.com*. The prompt availability of such tools will be decisive for the transition from soft semantic data in Web 2.0 systems to interoperable hard semantic data. Examples of triple stores that could provide efficient storage and management of RDF include *RKBExplorer*; at the moment *RKBExplorer* focuses on academics and their research output rather than teaching practices but this could be extended.

Supporting deployment: Encouraging the exposure of institutional and educational repositories in RDF/RDFa will contribute significantly to establishing a field of linked educational data. 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 considered. The exposure of education related knowledge models currently available in soft semantic formats like wikis to hard semantic formats could also be fostered. Best practises for exposing institutional data securely and selectively can be documented and considered by institutions.

6.3 Stage two: ontology-based applications for searching and matching

Period: starting two to three years from now.

Challenges and drives: The availability of a linked data field, where metadata is available in interoperable machine processable formats will enable the development of more efficient applications that can perform more advanced data analysis and matching among people and resources. A significant volume of HE/FE related linked data in combination with the emergence of applications that will employ this data to address certain teaching, learning or institutional challenges could initiate a *network effect* that will at this stage, which will enable further exposure of linked data and further development of advanced semantic applications. Advanced reasoning will require the development of ontologies on one hand and efficient mapping of information from the linked data field to those ontologies on the other. Global agreement on ontologies is not required for the development of such tools but, where available, it can prove powerful. Experimentation with rich but idiosyncratic ontologies for education could also be considered.

Benefits for HE/FE institutions: Tools that enrich the repertory of technology enhanced learning activities like critical thinking and argumentation will become possible; more targeted and efficient resource discovery and matching will further support learning, teaching and research work. Ontologies and reasoning on a large volume of linked data will empower higher education institutions with more insightful analysis to attract students and research funding and will enhance their research potential. The research results and expertise of institutions will be even more visible to prospective students and the industry. Efficient matching tools will enable collaboration among departments in different universities, while research and teaching involving cross-disciplinary areas will be better supported.

Enabling institutional policies: Publishing methodologies on how data exposed by HE/FE institutions can be used to address institutional challenges. Exposing learning and teaching material and additional information related to the HE/FE challenges as linked data. Developing relevant applications for internal use by institutions (e.g. morning student progress) or external use by interested parties (e.g. institutional expertise, courses, etc). Publishing methodologies on how data exposed by institutions can be employed for pedagogical ends. Publishing educational patterns on how emerging tools are used for learning in HE/FE.

Enabling technologies: Infrastructural technologies that will enable collaborative ontology building and automated or semi-automated mapping of linked data to these ontologies will be very relevant. In addition, the use of technologies that enable efficient reasoning across linked data sources could be critical at this stage. The caching of linked data in large repositories (e.g. triple stores) will provide for more efficient queries and reasoning. It is also expected that large triple stores might introduce facilities for mapping linked data to more expressive ontologies.

Supporting deployment: Given a sufficiently large volume of HE/FE related linked data is available in repositories and large triple stores as part of stage one, the deployment of searching and matching applications to enhance learning and teaching could be encouraged at this stage. Specifically, group formation and contextualised searching services could be deployed and available for integration with other applications. Repositories to publish methodologies of using linked data to address institutional or pedagogical challenges and education patterns could support developments significantly.

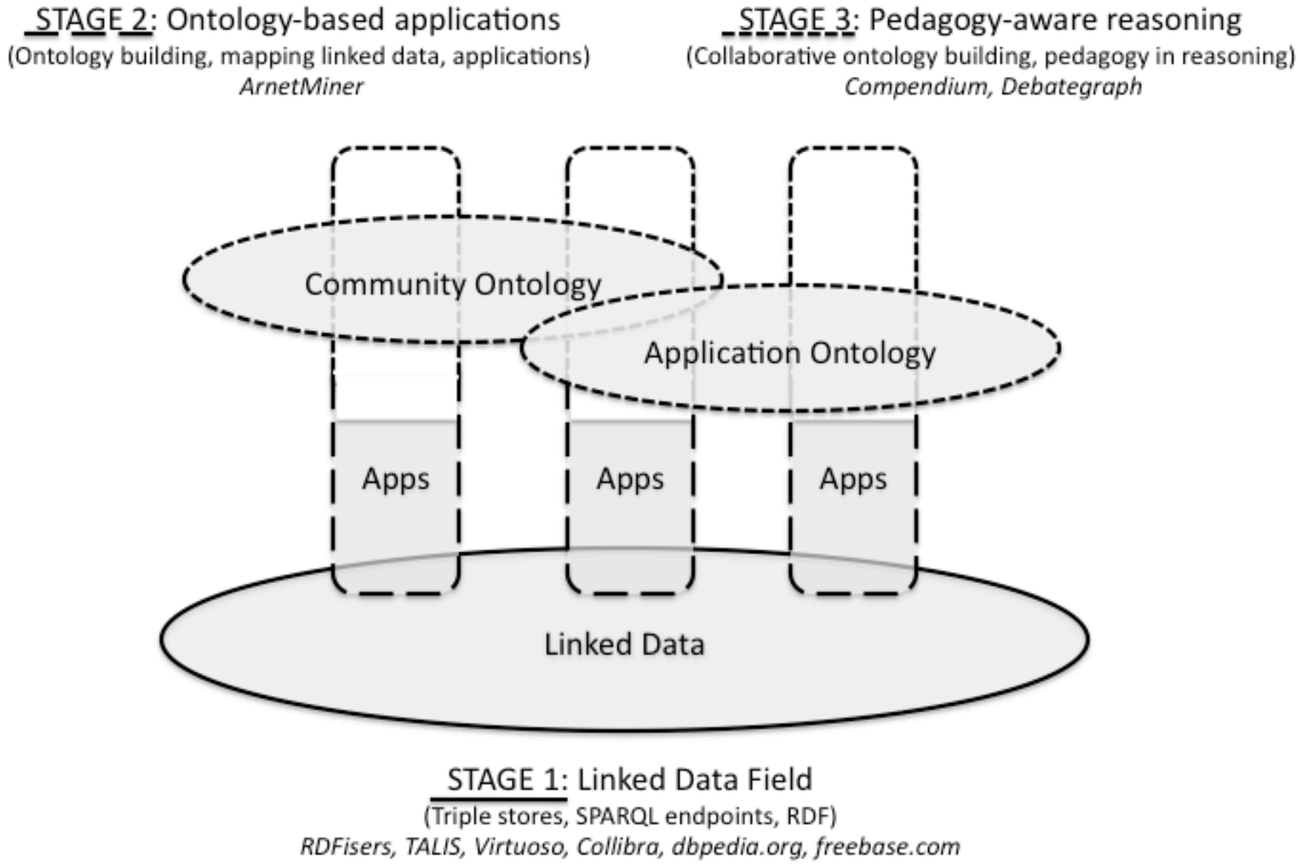

Figure 2: Roadmap of adoption of semantic technologies in education

6.4 Stage three: emergence of more pedagogy aware applications

Period: starting four to five years from now.

Challenges and drives: During the first two stages of the roadmap it is expected that the development non-pedagogically purposed applications and the exposure of institutional repositories as linked data will provide HE/FE institutions with significant value. During the stage two of the roadmap, some education purposed applications making use of linked educational data are expected to emerge. In addition, the publication of methodologies on how advanced reasoning on linked data could enable pedagogy related applications is envisaged in stage two. However, efficient support for pedagogically meaningful activities is anticipated in this third stage when both linked data and efficient ways of mapping those data to ontologies are in place.

Benefits for HE/FE institutions: Education purposed tools and services would assist addressing learning and teaching requirements and supporting pedagogy aware activities. Ontology repositories and mapping of collaboratively developed and more expressive ontologies to each other will further support interdisciplinary work and better access to the knowledge capital of higher education and further education institutions. Modelling the underpinning pedagogy and the context of tools and services will encourage teaching innovation across the UK HE/FE sector. Pedagogic innovation and practise may become more transferable across departments, disciplines or institutions. Development needs by staff and students will be more efficiently identified and addressed.

Enabling institutional policies: Encouraging the use of linked data to support and publish the outcome of interdisciplinary activities or activities across departments. Adoption or development of pedagogy-aware tools making use of reasoning over linked data to support critical thinking and argument building could encourage further development of pedagogically meaningful applications. The deployment or use of linked repositories of *arguments* and visualisation tools would add value to applications across institutions.

Enabling technologies: Intuitive knowledge modelling and visualisation tools that make use of lightweight metadata or more expressive ontologies will enable the development of HE/FE purposed tools and services. Ontology repositories and recommenders of ontologies or concepts to support annotation will also be important to this end. Tools and services that establish relationships among higher-level ontologies would enable more efficient reuse of more expressive metadata that might have become available during stage two. Tools like *Compendium* and *Debategraph,* which enable critical thinking could become even more significant at this stage given the availability of a linked data field with semantically annotated argumentation resources. Technologies like RDF can provide for *reification* of statements (e.g. semantic description of *who* a statement is attributed to).

Supporting deployment: There are benefits in encouraging the development of semantic tools and services with advanced reasoning capabilities at this stage of the roadmap instead of earlier stages. One of the benefits is that if these tools are deployed to use a well-established field of linked HE/FE data they will be more likely to work over a larger number of repositories. Another benefit is that there is a better chance to extend and maintain pedagogically aware tools that rely on well-established community-wide ontologies, which are anticipated to emerge in stage two in the roadmap. Even if such ontologies were agreed at the earlier stages it would have been unlikely that a sufficient amount of metadata in those ontologies would have been available.

7 Conclusions

The Semantic Web or Web 3.0 vision has inspired research with significant output and there is an emerging consensus that some form of Semantic Web is an inevitable development of existing technologies[37]. In this roadmap we examined the stages that will lead to this adoption based on survey and analysis and we proposed ways in which adoption could be further supported. The widespread adoption of semantic aware applications for education is placed in a horizon of four to five years[38] but we believe that the last stage of the roadmap will just commence during this period.

We anticipate the emergence of a linked data field and the population of this field with interoperable semantic data before agreement on ontologies and advanced pedagogically meaningful applications will become available; this is a key conclusion from this report. Advanced reasoning applications that will not be built on a well-established linked data field may have to rely on a limited number of repositories or on ad-hoc mappings of unstructured data to ontologies, which may be limiting.

By evolving and disseminating deeper understandings of the potential for semantic technologies across the education sector JISC can work with its stakeholders to establish and develop good practice which enhances the value for money of its investment in innovation and infrastructure in this area. This could include a JISC triple-store to which projects can contribute, targeted keynotes in the area of semantic technologies and a semantic focus within development events.

The following paragraphs outline research and development activities that JISC could encourage in order to foster developments during the stages of the roadmap.

- Stage one (intensifying the next two to three years)
    - Tools and incentives for HE/FE institutions to expose information on their curricula, research outcomes, learning and teaching material in interoperable semantic formats to build a linked data field across HE/FE starting from what is already exposed.
    - Enhancement of additional institutional repositories (online services, databases, Web resources, VLEs) with metadata in linked data formats.
    - Tools and services that will assist selective and secure exposure of institutional data as linked data.
    - Identification, fostering and sharing best practice on the kind of linked data (institutional data, learning and teaching resources) that can be used to address learning and teaching or HE/FE institutional challenges.
    - Continuation of the work of the Semantic Technologies working group to document additional use cases of linked data/semantic technologies to support learning and teaching.
    - *Relevant JISC activities: OER, XCRI[23], Curriculum Design and Delivery, LLL (Life Long Learning, AMG (Auto Metadata Generation), Rapid Innovation, LTIG (Learning and Teaching Innovation Grant), Digitisation programmes, JORUM, OLNET.ORG*

- Stage two (starting two to three years from now)
    - Documentation and demonstrators on integrating advanced searching and matching functionality using linked data into the workflows for learning and teaching support in

---

[37] http://connect.educause.edu/Library/EDUCAUSE+Quarterly/TheSemanticWebinEducation/47675

[38] http://wp.nmc.org/horizon2009/chapters/semantic-aware-apps/

    HE/FE.
  - Encouraging the exposure of additional institutional information as linked data to address further the established HE/FE challenges.
  - Development of utilities that will assist documenting the relationships among linked data in more expressive formats using collaborative ontology editing, visualisation and mapping tools across HE/FE institutions.
  - Development of additional use cases building on the practice of authentic data use in specific research domains (bioinformatics, etc).
  - Development of utilities that will combine linked data with geo-location resources for learning and teaching purposes.
  - Continuation of the work of the Semantic Technologies working group to identify additional types of data (institutional data, learning and teaching resources) that needs to be exposed to address further institutional and learning and teaching challenges.
  - *Relevant JISC activities: Rapid Innovation, LTIG (Learning and Teaching Innovation Grant), OER (Open Educational Resources).*
- Stage three (starting four to five years from now)
  - Enhanced support for documentation and access to pedagogic innovation across disciplines, across HE/FE departments and institutions using linked data/semantic technologies.
  - Development of tools to support collaborative semantic enrichment of linked data and ontology mapping.
  - Development of ontology powered visualisation and reasoning tools to pedagogical ends such as critical thinking, argumentation support and group formation on a larger scale (e.g. across institutions, repositories, etc).
  - Development of tools and services with more powerful reasoning support in a learning and teaching context.
  - *JISC to continue to include semantic web/linked data agendas in activities such as working groups, special interest groups, developers' networks.*

This is a recommendation of prioritising the concerns that JISC can seek to address via its work in the next few years. SemTech has undertaken a number of initiatives to foster developments in this direction by liaising with JISC, the CETIS Working Group on Semantic Technologies, the UK HE sector and the international community interested in this area. Specifically, the following activities were scheduled:

- 1st SemTech workshop organised at JISC premises, London, 19 January 2009.
- A paper on "Semantic Technologies for Learning and Teaching in the Web 2.0 era - A survey" was presented at the 1st Web Science conference, Athens, 18-20 March 2009, http://journal.webscience.org/166/
- A workshop on "Semantic Technologies in Education – exploring the practitioners' perspective" and short paper on "A roadmap for semantic technology adoption in UK higher education" accepted to the ALT-C conference, Manchester, 8-10 September 2009.
- A workshop on "SemHE-09: Semantic Web applications for learning and teaching support in higher education" accepted to the ECTEL 2009 conference, Nice, 28-29 September 2009 http://www.semhe.org/ - this workshop organised in collaboration with the ESRC/EPSRC project ENSEMBLE (http://www.ensemble.ac.uk/).

Annex A: Survey of Semantic tools and services for learning and teaching

The content of this annex is online at: http://semtech-survey.ecs.soton.ac.uk/report/technologies.html

Annex B: Adoption of Semantic Tools and Services in the UK higher education

The content of this annex is online at:
http://wiki.semtech.ecs.soton.ac.uk/index.php/Survey_of_Semantic_Technology_use_across_UK_universities