Conceptual Structures for STEM Data: Linked, Open, Rich and Personal

Su White

Web and Internet Science,
ECS, University of Southampton, UK
saw@cs.soton.ac.uk

Abstract. Linked and open data is increasing being used by governments, business and administration. Awareness of the affordances and potential utility of open data is being raised by the emergence of a host of web-based and mobile applications.

Across the educational and research communities applications applying the principles linked data principles have emerged.

Systems developed and used by researchers and academics are most likely to be predominantly in the hands of the early adopters and current developments found in higher education tend to be atomized, yet there is potentially considerable advantage in associating and integrating applications for organisational, educational and administrative.

This paper presents an argument for how we can move from early adopters to early majority, and at the same time presents a roadmap which will outline some of the significant challenges which remain to be addressed.

Keywords: linked data, open data, semantic annotation, higher education, organizational change.

1 Introduction

A strong thread of the use patterns which have accompanied technological advances of computational machines has been their use for data processing. The classic history of computers will inevitably acknowledge the use of Holerith Machines for the US census and the development of LEO to handle administration for Lyons (a company famous in Britain for its chain of corner tea houses). It will refer to the development of COBOL in response to apply computing power to the problems of provisioning the US Navy and the subsequent widespread growth of computer use for all aspects of business administration and record keeping.

Universities, like any other large organisation, made use of computers for administration and like the rest of the business world universities have integrated the use of personal computers into their business processes over the past thirty years. Further transformations in business and personal interactions with and use of computers followed on from the introduction and subsequent refinements of

H.D. Pfeiffer et al. (Eds.): ICCS 2013, LNAI 7735, pp. 1–21, 2013.

the World Wide Web during the 1990s. It is common to refer to three generations of the Web.

- The vanilla web: early implementations, the web as a publishing device—a basic web of documents
- The social web: an enhanced web of documents, the read write web introducing blogs and wikis
- The semantic web: "an extension of the current web in which information is given well-defined meaning" [1],

Following this was much discussion of what was meant by the Semantic Web and how it could be realised. The discussion was between the purists who preferred the path of hard semantics to the more pragmatic approach of soft semantics. Beliefs and attitudes shifted and changed [2]. Five years on from the original publication in 'The Semantic Web revisited' Shadbolt et al [3] asserted

> The Semantic Web we aspire to makes substantial reuse of existing ontologies and data. It's a linked information space in which data is being enriched and added. It lets users engage in the sort of serendipitous reuse and discovery of related information that's been a hallmark of viral Web uptake. We already see an increasing need and a rising obligation for people and organizations to make their data available. This is driven by the imperatives of collaborative science, by commercial incentives such as making product details available, and by regulatory requirements.

Alongside the debate as to the instantiation of the semantic web, Berners-Lee was considering the nature of change which was inherent in the way that the web worked. Evolving the principle that the semantic web was concerned with better enabling computers and people to work in co-operation, he began to refer to the 'two magics'.

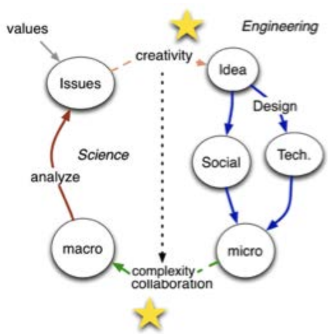

Fig. 1. Berners-Lee's science and engineering approach with magic modified to show complexity and collaboration

'Two magics' incorporates a generative interaction between social activity with the web and technological development. This model goes some way to explaining the way in which use and applications have taken off [4,5]. This was first described by Berners-Lee et al and then subsequently developed into the graphical form presented to the Web conference in 2007.

Alongside the changes related to the engineering of the web, the wider population were developing a conception of, relationship with, and reliance upon the web and its artefacts. Clay Shirky discusses the social web in his 2003 blog, and defines the social web as "software that supports group interaction' [6]. Evidence of the social web and social internet have a long history in discussion forums and Usenet groups. The social web, along with O'Reilly's observations of web 2.0 which can be formally dated from a 2005 blogpost and a 2007 paper [7,8] have become intermingled in the minds of the casual observer. This is of interest because if applications are to move from the early adopters to the early majority then they are more likely to succeed through the use of familiar metaphors (which are understandable and accessible).

In some ways we have been observing sets of memes being adopted by the general public. Just like the recommender systems deployed by Amazon and other commonplace shopping companies, everyday folk understand that applications which behave like applications you know and love are also likely to become applications which you know and love. Everyday folk may not understand the implications of big data, but they can begin to gain an idea of the mechanisms if they participate in citizen science, or see a news item featuring health benefits which have accrued from a massive genome data set collaboration. People are learning: learning from the technology—what it does and how it does it; learning from the people—what they do, how they use the technology. In academia we call it learning from good practice; gradually the concepts of open and linked data are seeping into the everyday consciousness. Alongside this comes some understanding of ontologies and semantic annotations. Concepts like domain models, reasoning and analysis may be more difficult to understand, but awareness is being raised. The educational domain is ready for change. It is being pressured, like so many other businesses, to streamline its processes. Academics are gaining hands-on experience of workplace tools in their research and are becoming ready to generalise these processes across their institutions. It will be interesting to see how these changes affect teaching and administration.

In the rest of the paper that follows, the Background section introduces an analysis of the means by which we can interact with a web of data rather than a web of documents, considering the scope of hard and soft semantics. These ideas form an important conceptual backbone to our understanding of the ways in which STEM data can be harnessed in university education. Through generic examples of implementation of open and linked data in 2.1 it will analyse current approaches and consider how everyday experience of open data shapes expectations and may therefore drive future developments. 2.2 presents a brief account of big data. The specific cases of linked and open data usage in Higher Education have been derived from the work of a number of communities, which are identified in 2.3. Specific educational approaches are examined in subsection 4 and the subsections which follow. The future directions of STEM education and the role played by conceptual structures in that future are then examined. The final section discusses the implications of the previous section in the light of challenges and opportunities in the educational domain and suggests some conclusions.

2 Background

The historical partnership between technology and administrative processes was traced in the introduction. The World Wide Web Consortium (W3C) and strong leadership from Tim Berners-Lee have been strongly influential in emerging standards for the Web. At the same time they have been taking forward the debate on ways in which infrastructure can be used and further developed. Much academic effort has been expended on the semantic web, following the 2001 article by Berners-Lee et al in Scientific American [1]. Widespread use of the Web has confronted users with the reality of shortcomings of the early systems. The early web was criticised as being a library where all the books had been thrown on the floor. As people used the web, understanding of what it could and could not achieve began to surface. The initial implementation was only a small part of the specification envisaged by Tim Berners-Lee and he has continued to work with W3C to realise the broader potential which he wanted to achieve.

Table 1. The Scope of hard and soft semantics

Hard semantics pure	Soft semantics pragmatic
Machine readable	Human readable
Rigorous modelling	Lightweight modelling

The story of the web is a story of engineering; it is a realisation of the difference between a model or proposal and the actual implementation. The web as we experience it has had structure imposed, after the fact. It is overlaid and there are many inconsistencies. As usage has developed in an ad hoc manner, although there are standards, they are many and varied. Early solutions to the combinatorial explosion which will surely follow any attempt to create a rich interlinked hypertext were mostly focused on abandonment of hyperlinking, and resorting to backend databases which serve pages engineered for delivery, but not for interaction. The social web went some way to creating a read-write web of the original conception. The social web has become a place for conversations and discourse. But it is the spam bots which demonstrate the power of machine processing of web pages over individual participation. As discussed in the Introduction, while the general public were becoming accustomed to the social web, the experts were discussing the nuances of the Semantic Web as summarised by Table 1; in particular whether hard semantic solutions were preferable to a more pragmatic approach of soft semantics.

2.1 Linked and Open Data

The UK government has become an enthusiastic supporter of Open Data. Early response to proposals were positive. Plans were initially cancelled with the incoming new Government in 2009, but have been fairly rapidly revived. In the UK, the Open Data Institute formally opened its doors in October 2012. Nigel Shadbolt commented on the ODI blog “Less than a year ago Tim Berners-Lee and I were writing a briefing note for Government outlining the opportunity for an Institute dedicated to realising the economic value of Open Data. Earlier in 2011 the Chancellor in a speech at the Google Zeitgeist stated “Our ambition is to become the world leader in open data. The economic impact of this open data revolution will be profound…”. `http://www.theodi.org/blog/` 1st October 2012.

The Open Data Institute was established with an objective of demonstrating the commercial value of open data. The UK government provides a set of case studies of Open Data on their web site `http://data.gov.uk`. Useful information on the economic impact of open data is available via the LinkedGov wiki which provides an index of a selection of peer reviewed papers. The UK government's 2011 Open Data White Paper [9] identified five agenda items for open data in the UK: i) building a transparent society; ii) enhanced access; iii) building trust; iv) making smarter use of data; v) The future transparent society. Examples of transparency and the benefits of it cover areas of transport, crime and spending. Making data available to citizens and businesses can enable government to more clearly account for their activities and spending, and provide information for feedback loops which can justify or promote changes in behaviours or responses. For government open data is key to understanding the nature of the businesses with which they are engaged. It also enables government to meet requirements of Freedom of Information legislation. The government proposition is that enhanced access leads to increased trust and thus to smarter use of data. There are a number of case studies provided to illustrate these arguments. The government has set itself a standard for information publishing and is promoting the five star scheme of data re-use originally proposed by Tim Berners-Lee. The government had already established a set of Information Principles.

Everyday experience of open data is typically mediated by mobile apps or web sites. Private companies and social collaborations, along with government, have produced and published a large amount of open data. Geographic data has been created in each of these three domains. In the UK the Ordnance Survey publicises its open data, but app developers use a range of different sources. However, people who use an app to display a map and overlay points of interest or to navigate between two locations will at the same time gain some kind of intuitive grasp of what may be possible, even if they do not yet understand the principles of publishing open data. Commonplace experience of apps like Tripit, LinkedIn, Open Street Map and Mendeley each in their different way help users build up an instinctive understanding of what is useful. System designers have learnt that making their API available is good for business and companies can compare the success of open and closed systems and draw their own conclusions

Table 2. Summarizing the semantic enhancements of Shotton et al.

Generic Enhancements	Adding value to the text
Providing access to actionable data:	Semantic annotations for key concepts
making the datasets available	Document summary and study summary
Data Fusion from Other Sources:	Tag tree and tag cloud
enriching the basic journal data	Supporting claims tooltip;
Making information more accessible	Various citation analysis tools
Provenance information	Alternative language abstract

as to the best way forward. In academia, researchers have been experimenting with open data. In 2009 Shotton et al report on an interesting experiment to see how it was possible to semantically enrich a traditional academic paper [10]. The account of this activity identifies a set of enhancements which can be seen in Table 2. The whole activity provided an immensely rich experience which if it could be produced in a replicable automated manner would add considerable value.

2.2 Big Data

The storage capacity of computer systems and the speed and power of data processing have enabled the collection of big data sets. Big data refers to massive datasets containing many billions of information items. Big data is too large to be analysed by conventional database tools. It comes from many different sources. These include:

i) data gathered by local and national governments as a result of providing services of systematic survey; ii) data gathered by businesses as a result of their interactions with clients and customers; iii) data from the natural world through scientific observations or experimentation; iv) collated and aggregated data sets which are gathered from diverse sources covering similar or identical subject areas; v) historic data sets;

Data sets may be examined to provide evidence and feed into businesses processes for bringing about change. Many large data sets are proprietary—owned and protected from wider use by copyright, privacy or business imperatives. Some large datasets are made available for distributed analysis—the SETI@home project was used to analyse large volumes of astronomical data. Big data can benefit from crowdsourced analysis just as big data sets may be assembled by crowdsourcing.

2.3 Educational Communities Around Linked and Open Data

Developer communities have a crucial role to play in the dissemination and sharing of ideas and helping establish good practice. In educational communities there are broadly four basic types of intersecting communities which are concerned with education specific linked and open data i) formal associations;

ii) institutional initiatives; iii) evangelist practitioners and researchers; iv) loose associations or communities of practice. Community has a very strong role to play in the development of standards in this area. All four types of communities, plus their respective associations of users of linked and open data, are involved in different ways [11]. The examples below are drawn from the UK, but similar development is taking place in many different countries across the world. Probably the most significant of these is the international OpenCourseWare Consortium.

Formal associations like XCRI (`http://www.xcri.co.uk`) can be seen as a combination of bottom up, specifications and demand arising from the community, met strategically with top down input from funding bodies to pursue a common objective. The XCRI initiative is funded by the JISC, the UK agency for technology infrastructure and development in Higher and Further Education. Formal associations produce tangible outcomes. For example, the XCRI initiative has developed and is now working to a standard model of course information.

Institutional initiatives are manifest in a number of ways. Some institutions, such as the UK Open University, pursue open and linked data because there is a strong business case in terms of managing administrative processes and gaining business intelligence. The University of Southampton has a very close link with the development of the semantic web. Tim Berners-Lee holds a chair at Southampton and in an initiative led jointly with Nigel Shadbolt has established the UK Open Data Institute. A number of other high profile institutions have this level of commitment.

Evangelist practitioners and researchers form loose associations irrespective of institutional ambitions. Researchers and application developers in universities often pursue their objectives through passion and academic interest. Often their collaboration is a mixture of face-to-face and online interactions supported by blogging and microblogging which support extended discussions and knowledge sharing.

The linked universities (`http://linkeduniversities.org/`) are a loosely coupled community of practice who work collaboratively on emerging standards. Being in academia, they can have a symbiotic relationship with the funding council through JISC funded standards related work (CETIS) and developers forums such as Dev8D. They have worked to establish a number of agreed vocabularies and to bring together a significant amount of expertise relating to linked data initiatives in the UK and across Europe.

All of these communities add to the common understanding by making visible their discussions and publicising their achievements. Further examples include the informative set of case studies made available via the UK XCRI-CAP web site, while the linked universities describe vocabularies and work in progress and present a collection of relevant publications.

Another significant educational community is that associated with Open Educational Resources. The OER community has two different manifestations. Some institutions have developed repositories which they are making open to share worldwide, while other repositories are shared efforts across institutions, sometimes with disciplinary groupings. Davis et al provide a comprehensive account of the roots of OERs and the experience of community building [12]. In many UK universities OER communities have strong ties with the open and scholarly publications community, and there is evidence in the literature that experience from one field sometimes informs the others. Open educational resources bring together those who take a resource based approach to learning (which survives from many of the early applications of hypertext) and those associated with formal learning design, working from an IMS perspective.

Open Educational Resources: (OERs) explicitly collected or assembled for sharing and reuse. Since 1990s standards have evolved which support and enable publication and re-use e.g., IEEE Learning Object Metadata (LOM) [14,15]. The standards enable resources to be found and provide systematic descriptions. Further standards such as IMS-CP support interoperability. It is possible to transfer sets of identified files and unpack them for use on another server. Further standards such as SCORM RTE and IMS-LD CopperCore [16] can be used to support sequencing and assembling resources for learners. OERs have a long pedigree. The MERLOT project was an international consortium which worked in 1997 to build a platform for open educational resources. In Europe, Rob Koper from the Open University in the Netherlands was influential in early work on learning objects specifying an educational modelling language [13]. Some major players in the OER community are also providers of OpenCourseWare. In 2005 the OpenCourseWare Consortium was established bringing together major international interests committed to open sharing of a range of educational resources presented as discrete courses. The OER and OCW community have been working in the area of modelling and linked and open data, but their activities have been driven from a bottom-up perspective of sharing and achieving interoperability rather than from a top down design approach.

Large-scale learning: "Learning analytics is essential for penetrating the fog that has settled over much of higher education" [17]. Just as business organisations use their big data on customers' behaviour to improve their bottom line, so learning analytics is being harnessed in HE. It is not surprising therefore that just as research into data mining has been influential in many approaches to big data, mathematical methods of examining large data sets in education have emerged as Learning Analytics. Universities and educators are much concerned at the micro level with checking whether learners understand whether learning has taken place, what feedback is needed and how progress may be measured. Data is collected per student and aggregated across cohorts. This data is analysed and reported internally and externally. Topics with which educators are particularly concerned include student achievement (measured through courseworks, assignments, tests and exams). Because institutions are concerned with awarding degrees, data on attainment contributes to progression and retention information which will be discussed and analysed internally and reported externally. Learning analytics brings the approaches of big data to many institutional agendas, particularly those of measuring learning, attainment, progression and retention. From a business process perspective this data may also be relevant to providing evidence for establishing financial cases for educational activities.

Larger institutions such as open universities may already have systematic data collection processes.

George Siemens is a leading thinker in the area of educational learning analytics and researchers at Athabasca and the UK Open University are prominent and influential in the field of online education termed Connectivism [18]. Siemens provides a personal participative link from Learning Analytics into MOOCs (massively open online courses) having collaborated with Stephen Downes to establish a MOOC in 2008 [19]. MOOCs provide an opportunity to gather learning analytic data, particularly that related to student behaviours. MOOCs provide a context for Open Educational Resources, while learners and participants provide a context for the data collected from a MOOC. Currently MOOCs are provided on a number of different platforms; for example Cousera, Udacity, MITx and EdX. In the US some high profile institutions are running MOOCs, perhaps for their advertising and reputation enhancing potential, or perhaps for their potential to collect learner data which can then feed back into the design of face to face learning activities.

2.4 Learning Approaches

There is a plethora of educational theory to which individual academics may refer, and many institutions will exhibit a broad range of approaches, some of which are explicitly designed within a given educational approach. Some relatively new institutions, especially large open universities may commit to a formal design process across the institution based on an acknowledged set of educational principles, but more often students experience diverse influences from a range of theoretical perspectives.

Figure 2 gives a much simplified representation of some of the key theoretical influences from education and technology which are prevalent in educational approaches in universities at the current time. For learners the network is playing an increasingly important role in learning. This is a reflection of the way in which technology infrastructure has become an intrinsic part of the fabric of everyday life. It also resonates with many educational theories such as constructivism and social constructivism. In Higher Education Laurillard has been highly influential and the conversational model of learning [20] has been credited with widespread impact, certainly amongst UK based educators engaged in technology based learning.

At the same time, work by Siemens has also been influential in proposing a model of connectivism [18] which emphasises the role of the network in shaping and determining the nature of learning and approaches which are relevant and effective. It proposes a model which is particularly relevant to the connected world. It has strong links to social learning theories and stresses the primacy of generative and transformative approaches to learning. Siemens subsequently has been involved with large scale learning activities at the Athabasca University in Canada. Following that work he also identifies specific links between connectivism and learning analytics [21], making connectivism a perspective which is particularly relevant to the scope of this paper.

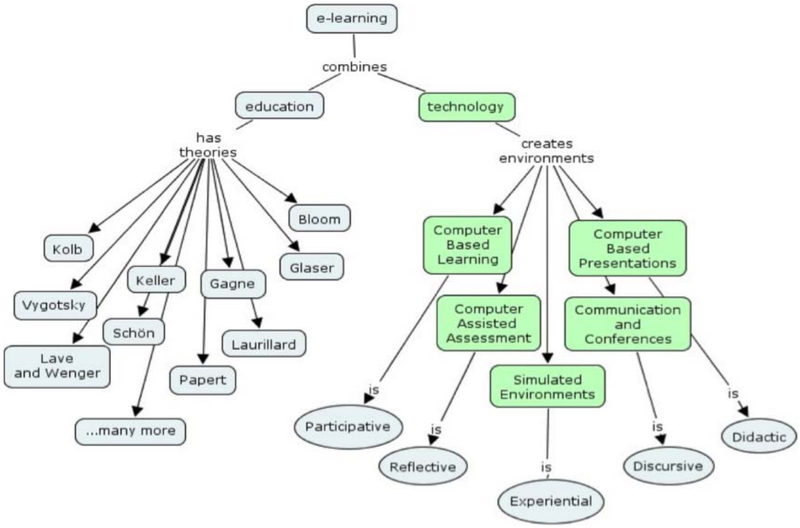

Fig. 2. Educational Approaches—a mass of theories

"Learning analytics currently sits at a crossroads between technical and social learning theory fields. On the one hand, the algorithms that form recommender systems, personalization models, and network analysis require deep technical expertise. The impact of these algorithms, however, is felt in the social system of learning. As a consequence, researchers in learning analytics have devoted significant attention to bridging these gaps and bringing these communities in contact with each other through conversations and conferences" [22].

There are many approaches to educational theory which do not take into account the online and connected world. One recent educator who has been influential in approaches to learning, but who does not deal specifically with technology and learning is Biggs [23]. His models of the student and effective ways of facilitating student learning have gained widespread currency, and there has been a growing interest in understanding what 'what the student does' as a means of modelling and enhancing education. Where Biggs seems to be particularly relevant to the emerging conceptual frameworks in STEM education is via recent interest in Digital Literacies. Much of the discussion within is aligned with Biggs' perspective. Students do things to learn, and can be expected to develop their learning skills whilst undertaking higher level study. Working with students to understand, develop and extend their digital literacies is an increasingly important agenda for Higher Education across the board. For students in STEM areas this will extend to mastering and understanding everyday work tools and to the specific sets of tools which predominate in their chosen specialisms.

Education has a long tradition of working with webs of documents. The role of text in education was a driver for early computer based systems, and in turn influenced models of understanding for approaches to technology based learning which differed from the more strongly industrially influenced approach of computer based training. Hypertext was implemented on mainframes through the Plato. On personal and distributed computers, Apple briefly led the way with their HyperCard system. Hypertext researchers looked at education, and much interest developed in personalisation, customisation and intelligent tutoring systems. Hypertext systems such as microcosm, although developed before the web came into use, used assembled collections of interlinked resources and it was that understanding of texts which informed the developments of early research and educational repositories such as ePrints and EdShare [24]. Different subject areas have tended to privilege different aspects of learning technologies; specifically those which are better suited to their individual disciplinary needs [25]. Independently, academics were programming web pages and producing specialised applications such as simulations, as well as using authentic data sets for structured tutorials.

At the same time commercial interests were promoting systems which has offered to manage the learning process, early systems often worked to extend the book metaphor pacing the learner through their material. It was natural that this work continued through the early years of web technologies, when systems which managed the learning experience Blackboard and WebCT for example were introduced. These systems often integrated with student management systems which the suppliers were also selling. Virtual Learning environments also encapsulate a number of different learning processes, for example tutorial plus questions plus simulations. Typically there are also analytics such as data logging and tracking. They have gained some popularity for their ability to sequence order and organise information and thus drive learning activities. The functionality of VLEs have evolved alongside technology in the wider world, and systems typically now incorporate aspects of the social read-write web, although their objective of being closed systems means that they cannot have the exact same affordances as the wikis, blogs and online discussions which are found on the wild web.

2.5 Semantic Technologies in Higher Education

The discourse and analysis of the potential for semantic technologies in Higher Education has strong links back to the research of the hypertext and adaptive hypertext communities and thus necessarily encompassed contributions from those concerned with AI agendas such as agents and Intelligent Tutoring Systems. The 2009 SemTech Report presents survey findings for semantic technologies in learning and teaching [26]. Things have moved on since, but some of the observations are still relevant. SemTech articulates the need to differentiate between soft and hard semantics in an educational context. Table 3 differentiates 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.

There are many different ways in which semantic led approaches might contribute to educational activities. Table 4 differentiates between those which can

Table 3. Examples of hard and soft semantics in Higher Education

Hard Technologies Using machines to talk to each other	**Soft technologies Helping people to organise knowledge**
managing shared learning content	Tool to link learning and select careers
identify cross-curricular connections	Managing shared learning content
Support for personal learning	Support for personal learning
Search for people (people like me)	Developing reasoning skills and argument
Search for resources	Shared mind-maps/topic-maps

Table 4. Where semantic technologies can contribute to educational processes

Classroom administration	Visible data → dynamic analysis & feedback
Assisting course creation	Aggregate relevant resources & workflow
Aggregate course and module information	Streamline accreditation & quality processes
Learning activities	Assessment, certification, countering/detecting plagiarism
Critical thinking and argumentation support	Learning in the wild
Efficient personal & group knowledge construction	Informal learning
Authentic learning	Self-actuated learning
Group formation	Aggregation, personalisation, customisation

be considered to related to 'classroom' administration and those which might make up a direct component of learning activities.

2.6 Education in STEM Subjects, Some Scenarios

The challenge for educators is to provide the educational opportunities which enable the learner to familiarise themselves with, and then master the necessary knowledge, skills and understandings which can equip them to be competent in their chosen specialism. Can using smart technology help us address these ambitions? How is our model of student learning activities made more complex by these requirements?

At the same time as we address these aims, there is a widespread expectation that learning should be a transformative process, and that the students will be able and ready to make and sustain their contribution to their chosen workplace and career path. It is inevitable that such a trajectory necessitates a mastery which extends beyond academic subject specialism and into the applied discipline in a world where technology is an essential component. Disciplinary differences as investigated by Biglan [27,28] and subsequently Becher [29], play an important role in the nature of teaching in Higher Education. Disciplines determine needs

Table 5. Disciplinary needs of Hard Pure and Hard Applied Subjects [25]

	Curriculum/content	Assessment	Cognitive Purpose
Hard Pure e.g., Natural Sciences	Concepts and principles closely connected Content typically fixed and cumulative Quantitative Teaching and learning activities are focused	Specific and focused exam questions Object tests relying on quantitative nature of knowledge	Logical reasoning Testing of ideas in linear form of augmentation. Reliance on facts, principles and concepts
Hard Applied e.g., Engineering	Concerned with the mastery of the physical environment. Focus is on products and techniques. Knowledge is atomistic and cumulative. Emphasises factual understanding.	Preference for exam questions; especially problem solving.	Logical reasoning. Testing of ideas in linear form of argumentation. Reliance on facts, priniples and concepts

and establish context. The STEM subjects are considered to fall within the Hard Pure and Hard Applied disciplinary space. The context of disciplinary differences in e-learning has been investigated by White and Liccardi whose summary of the disciplinary needs relevant to STEM subjects is presented in Table 5. Alongside the disciplinary needs of learners, it is worth considering the affordances of different types of tools. Based on the observations of the SemTech report, and extrapolating from the experience of enriched publishing reported by Shotton et al. [10]. Table 6 builds on the activity gradient proposed by White and Liccardi, suggesting added value which might result if educational resources were semantically enriched.

The value of ontologies to planning the learning process has been recognised by researchers and educational specialists within some discipline areas [30]. This would come under the classroom administration category suggested in Table 4 which was previously discussed. In the SemTech survey [26] there was evidence of the purposeful use of semantic tools for authentic learning. Experience in the use of repositories and digital collections would suggest that they have strengths for the educator as well as the learner and would indirectly support informal learning, self actuated learning, personalisation and customisation. In the world of MOOCs specialist programs have been established which reflect an industry need; for example the solar power industry `http://solpowerpeople.com/solar-courses/`—which might also be appropriate to provide authentic resources for learners studying on a formal program in a relevant topic. At the time of writing, Coursera, a federation of OpenCourseWare, listed almost 200 courses with a duration of between 4 and 12 weeks. It

Table 6. Suggested benefits from semantically enriched learning in stem subjects

Teacher led/passive		
Resource	**Conventional use**	**Semantically enriched**
Notes on the web	Teacher: author, publisher. Student: consumer, viewer, use for reference.	Automatically linked to related resources. Dynamic annotation (semantic wiki). Use/integrate with OER.
Tests, questions	Teacher: author, publisher Student: participates/interacts.	Automatically linked. Dynamic and static generation of feedback. Dynamic linking to 'wild' resources. Use/integration of OER
Interactive tutorials. Incorporates learning activities and assumes structure.	Teacher: author, publisher.	Dynamically assembled, dynamically link.
Simulations. Incorporates learning activities and assumes structure.	Teacher: author, publisher. Student: participates/interacts Pathways dynamic/proxy for real world.	Distributed participants, use of authentic data.
'World Ware'	Teacher points to/requires use of authentic tools.	Real world datasets & tools.
Student created artefacts	Student freely utilises authentic tools.	Semantic publication/visibility
Online discussions. Blogs and Wikis.	Students engage is 'social' creation and 'social' learning.	Dynamic interlinking, semantic publication.
Student led/active		

claimed to have more than 1.7 million registered course participants. By far the majority of the subjects were in the stem subject area. High profile open courses have been offered by US Ivy League universities but are also being offered by informal networks of teachers, by individual academics and by small colleges.

The examples above considered the value of semantic enrichment and open and linked data from the perspective of purposeful course design, or as an adjunct to the educational administration. These two perspectives were the main line of analysis identified in the SemTech report. However, there has been considerable discussion across the community which argues for taking a personal learner perspective on educational resources, and to place the use of educational resources within a framework of a Personal Learning Environment. Within the framework of developing campus wide support for learning at the University of Southampton, the personal learning environment has been considered. Semantically rich environments provide ample opportunities for the interlinking and crafting of personal learning resources.

The world is changing and universities must respond to students' needs and expectations in agile and effective ways. Learners enter university with an inevitable diversity of technological familiarity and a mix of naïve and sophisticated approaches to using technology as a part of their learning. Students are using apps and becoming familiar with the potential of linked data. Just as they have learned how to Google for information and to look to Wikipedia as the first source of information, so they are also becoming familiar with technology behaviours which they might reasonably expect to appear in their study environment.

Using online services such as Facebook, Amazon, Delicious, Flickr, YouTube introduces them to a world where artefacts like integration and recommendation are an obvious part of the infrastructure. Familiarisation with these services shapes expectations and also prepares users to be adept at exploiting the affordances for their own reasons. Students develop skills and expectations. Familiarity with these specialised affordances of various common place yet separate applications, may result in students viewing the institutional provision of web sites and virtual learning environments (VLEs) as clunky and out of date. For their part, universities may feel themselves overloaded with the task of providing, maintaining and updating the necessary information needed to inform and educate their students and also to furnish and drive the workflows of their administrative processes.

Many universities are understandably proud of the historic heritage on which University system is based, and the historic roots on which their own institution is established. Yet these same roots and traditions are in some ways likely to be the source of some of the challenges which are faced by the University as an organisation.

Institutions are being changed by external factors. Siemens identifies an altered information cycle brought about by 'participatory technologies'. At the University of Southampton, four fundamental drivers for change were identified. i) support curriculum change and innovation; ii) address student expectations; iii) enable the university to remain credible in its support for learning and teaching—particular to be seen as fluent and innovative in the use of IT; iv) facilitate the adoption of a University-wide educational style. Students may want interconnectivity with external apps, actors in our systems may want to share data from internal info with external apps, but whatever else we certainly need to be able to share and reuse data from and between our internal apps, using a web2.0 approach using the web as a platform, exposing our data and devising services to enable apps to communicate is essential.

An early implementation of educational infrastructure to support teaching was developed in Southampton making use of linked data. We routinely use linked data for info within one part of the University ECS enabling all our info style pages to be generated dynamically. The mix of screen shots collected as a single figure illustrate this approach incorporating a personal page where data is associated to provide information based on real world relationships recorded as linked data. Academics are tutors; tutors have tutees; academics are lecturers; lecturers have teaching allocations, and so on. Assembled sets of relations generate informative and highly functional web pages. This approach creates official home pages and generates module pages for teaching activities. The resources page also accesses the institutional repository and using tags as filters populates the module page resources tab. A wiki is used to add and enter information; there is also a linked HTML web pages on the filestore. Dynamic content I added through retrieving delicious tags associated with course teachers, and a tag cloud index to the delicious data is generated.

We used RDF because it saves time; however, the hand crafted web sites will persist, and some colleagues use paper handouts. Automatically generated pages provide learners with a consistent backbone to which they can refer. Individual differemces will persist. Not everyone uses the EdShare repository, or edits the notes or student wiki.

Even with this proof of concept, the challenge remains, how to port it to the rest of the University? This system was introduced by those whose research is into linked data. Colleagues in Electronics, physic or chemistry might not regard the changes in the same light. Time and again we return to the issue that change is cultural, and individual responses and behaviours are mediated by skills and by available time, and willing priority. Open and linked data can be used by universities for business process management. The University of Southampton established an open data initiative in 2010. Full information can be found at the project's web site `http://data.soton.ac.uk/`.

Benefit can be gained from exposing and sharing the public and private capital of data and information within and across departments and institutions to enable workflows and promote and enable collaboration. Since its inception the project has supported, shown financial returns and won the support of senior administrators and managers who have particularly appreciated the way in which the data can be drawn upon at short notice to provide customised web sites (for example to support a student visit day). Among the achievements made by this initiative i) furnish components of a financial information system; ii) helped address external demands for information provision is a cost effective manner; iii) enable enhanced quality of data relating to room information; iv) increase the efficiency of the on campus catering provision; v) drive a mobile app detailing university and city wide bus services.

3 Discussion

The examples above demonstrate that education has more than one focus: learners, teachers, researchers, administrators; depending on the activity, some people are at the core, whilst some are at the periphery. Establishing a coherent approach to institution-wide change which incorporates technology and introduces new business practices is an ambitious challenge and it would not be surprising if a few challenges were encountered on the way.

Table 7. Southampton data sets available in autumn 2012

Apps using our data	Links to DisabledGo	Public Phonebook
Buildings and Places	Access Information	Published Accounts
Catering	Local Amenities	Services
Common Learning	Open Data Catalogue	Southampton Bus-routes
Spaces	Open Days July	Southampton Bus-stops
Extra Information	2011Organisation	Southampton Jargon
ECS EPrints Link set	Payments 2010-11 to	Dictionary
EPrints Repository	2011-01	Student Statistics
Easting/Northing	Photographs of	Students Union Events
EdShare	University of	Teaching Room Features
EdShare Video	Southampton Things	Transport Linkset
ECS EPrints Repository	Press Contacts	University of
Events Diary	Information	Southampton Profile
Facilities and Equipment	Programmes (2010-2011	Document
Food Hygiene Ratings	session)	Vending Machines
International Links	Programmes (2011-2012	WiFi
International Links	session)	iSolutions Workstation
DBPedia Data	Programmes (2012-2013	Clusters
JACS Codes	session)	

Surveying the use of Semantic Technologies in Education in 2009, Tiropannis et al observed of the the challenges faced by Higher Education that most could be addressed by querying across institutional repositories (databases, web pages, VLEs). Significant learning and teaching challenges can be addressed by accessing resources across departments, schools, institutions. The emergence of linked data fields across related repositories (seen in Table 7) will enable new applications relevant to identified HE challenges. They consider that the initial value of semantic technology will be in scale rather than reasoning and suggest that institutions will benefit from adopting a bottom-up approach starting from linked data which can be related to (layers of) ontologies later in the context of specific applications. The SemTech perspective quoted here focused on specific implications for learning and teaching, but from the material covered in this paper it would appear that a broader perspective would repay investigation.

This paper has looked at the roots of technology innovation which were created by the web and its technical developments incorporating linked and open data. It has considered a range of different technical innovations which can be found in the business and commercial domains and considered how they relate to educational domains through two mechanisms i) establishing patterns of use and user expectations through familiarity and perceived user benefit; ii) creating organisational gains either in terms of improved efficiency and effectiveness or through directly reduced costs. Furthermore, these changes have the potential to create indirect savings by streamlining processes and gathering valuable business intelligence which can help in strategic planning and direction.

Whilst much of the paper has discussed educational innovations, it may well be that for educational institutions the real gains which can be made are in the area of organisational efficiencies. Core business functions in educational institutions have much in common with commercial and business organisations, albeit there are some very particular constraints found in educational contexts because of the cyclical nature of the business, and the uneven tempo of the academic year. There has been some discussion of formal teaching related affordances which might be available to educators using linked and open data. However, semantic technologies do offer learners the means to independently craft and fine tune their own personal learning environments. The discussion and examples throughout this paper have referred to university education, but the case was made in the opening sections that this work is equally applicable to workplace learning at higher levels. There are particular constraints which apply to workplace learning which differentiate it from university learning.

The affordances of semantic technologies, open and linked data introduce potential for flexibility, dynamism and automation which may be particularly beneficial for those who are studying in a workplace context. Streamlining the ways in which we can assemble and inter-link content offers a considerable gain. This benefit will be as relevant to the work-based learner, topping up expertise or undertaking professional development. The models established in OpenCourseWare combined with the potential for personal learning environments appear to be particularly fruitful areas for future development. It also seems likely that academics, becoming familiar with technologies in their research activities, will find ways of introducing datasets and research practice into learning as authentic activities. The specific digital literacies of each learner within any given STEM area will be closely related to the authentic tools which are routinely used by practitioners (in the workplace or the research lab) associated with that specialism. Time savings may also accrue from a data based approach to gathering summary information about study programmes for accreditation. The curriculum is one area where the effort of building ontologies is beneficial. Institutions expend significant effort trying to gain broad-brush pictures across modules and programmes; work on knowledge modelling in this area could be fruitful.

The potential impact of widespread use of linked data in Higher Education is immense. Everyday understanding of the power derived by placing raw data in the public domain is growing. It promises to transform education, interconnecting administrative data, enriching and embellishing teaching resources while providing tools and resources for learners and researchers alike. Currently, semantic technologies are more widely and systematically used in research and administration than they are in teaching in higher education.

Having discussed the broad challenges and potential of greater use of data in an educational context, it might be constructive to suggest a way forward.

- Experience at Southampton has placed great value in purposefully constructing teams which incorporate a range of organisational perspectives. Existing literature on change processes identifies the need for champions and patrons.

Champions pursue agendas at a local level, while patrons support and visibly promote change at a strategic and trans-institutional level.

- Some of the simple demonstrators which helped disseminate the potential of using data in applications were developed speculatively on low budgets by student interns. Some apps were developed independently by students, mirroring the wider experience of making APIs available for those who will gain most to invest in.
- Many of our data clients were surprised by the simplicity of the changes they needed to make in order to publish their data and accrue additional benefit. A participant who saves money is a great advert and a willing advertiser of your hard work.
- Once the data has been published and example apps developed, new clients are more able to imagine what they want and what they might gain. This can then enable effective collaboration and co-creation of further apps which will in turn accelerate or refine future developments.
- Borrowing from business practices can be fruitful. Hackathons, BarCamps, UnConferences and Competitions can be ways of finding and pairing developers with clients and producing proof of concept apps in short timeframes. The energy created by these types of events will also sustain more measured developments.
- There is a wider understanding of what might be achieved by crowdsourcing: sharing the task of collecting data, or refining and correcting datasets.
- Publishing data is a wonderful way to distribute quality control tasks. Users can spot and correct published data. We found this worked particularly well in the case of our teaching room database, which was previously maintained on a PC and updated on an annual basis, often preserving mis-information from year to year.
- Shared initiatives lead to understanding and sharing organisational objectives
- Being a semantic squirrel may be rewarding. If there is an opportunity to collect data, the cost of storage will be small. If a means or motivation to analyse and use it emerges in the future, half of the job will already be done.
- Engaging in cool projects at your institution will make your techies happy, and give them things which they can go and brag about at developer events.
- Joint projects like the open data activities suggested here provide an opportunity to develop a local community of practice which will in turn enrich organisational knowledge

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American 284, 34–43 (2001)
2. Shadbolt, N.R., Gibbins, N., Glaser, H., Harris, S., Schraefel, m.c.: CS AKTive Space or how we stopped worrying and learned to love the Semantic Web. IEEE Intelligent Systems 19, 41–47 (2004)

3. Shadbolt, N., Berners-Lee, T., Hall, W.: The Semantic Web Revisited. IEEE Intelligent Systems 21, 96–101 (2006)
4. Berners-Lee, T., Weitzner, D.J., Hall, W., O'Hara, K., Shadbolt, N., Hendler, J.: A Framework for Web Science. Foundations and Trends in Web Science 1, 1–130 (2006)
5. Berners-Lee, T.: The process of designing things in a very large space: Keynote Presentation. In: WWW 2007 (2007)
6. Shirky, C.: A group is its own worst enemy - A speech at ETech. Clay Shirky's Writings About the Internet: Economics & Culture, Media & Community, Open Source (April 2003), http://www.shirky.com/writings/group_enemy.html
7. O'Reilly, T.: What Is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software (2005), http://oreilly.com/web2/archive/what-is-web-20.html (last accessed June 2010)
8. O'Reilly, T.: What is Web 2.0: Design patterns and business models for the next generation of software. Communications & Strategies 1, 17–37 (2007)
9. Cabinet Office Open Data White Paper Unleashing the Potential Unleashing the Potential Open Data White Paper 52, London (2011), http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential
10. Shotton, D., Portwin, K., Klyne, G., Miles, A.: Adventures in semantic publishing: exemplar semantic enhancements of a research article. PLoS Computational Biology 5, e1000361 (2009)
11. Wilson, S.: Community-driven Specifications: xCrI, sword, and Leap2a. International Journal of IT Standards and Standardization Research 8, 74–86 (2010)
12. Davis, H.C., Carr, L.A., Hey, J.M.N., Howard, Y., Millard, D.E., Morris, D., White, S.: Bootstrapping a Culture of Sharing to Facilitate Open Educational Resources. IEEE Transactions on Learning Technologies 3, 96–109 (2010)
13. Koper, R.: Modelling units of study from a pedagogical perspective the pedagogical meta-model behind EML (2001)
14. IEEE Learning Standards Committee (LTSC) IEEE P1484.12 Learning Object Metadata Working Group; WG12, http://ltsc.ieee.org/wg12/
15. Campbell, L.M.: UK Learning Object Metadata Core Working Draft Version 0.3 1204 (2004)
16. Tattersall, C.: Comparing Educational Modelling Languages on a CaseStudy: An Approach using IMS Learning Design. In: Sixth International Conference on Advanced Learning Technologies, pp. 1154–1155 (2006)
17. Long, P., Siemens, G.: Penetrating the Fog: Analytics in Learning and Education. Educause Review 46, 31–40 (2011)
18. Siemens, G.: Connectivism: A Learning Theory for the Digital Age. International Journal of Instructional Technology and Distance Learning 2, 1–8 (2005)
19. Siemens, G., Downes, S.: Connectivism and connective knowledge: Course delivered at University of Manitoba (September-November 2008), http://ltc.umanitoba.ca/connectivism/
20. Laurillard, D.: Rethinking University Teaching: a Framework for the Effective Use of Educational Technology. Routledge, London (1993)
21. Siemens, G.: Learning Analytics: Envisioning a Research Discipline and a Domain of Practice. In: Second International Conference on Learning Analytics and Knowledge (LAK 2012), pp. 4–8 (2012)
22. Siemens, G., Gasavic, D.: Learning and Knowledge Analytics. Journal of Educational Technology & Society 15, 1–2 (2012)

23. Biggs, J.: Teaching for quality learning at university: what the student does. Open University Press in Association with The Society for Research into Higher Education, Buckingham (1999)
24. Hall, W., Davis, H.C., Hutchings, G.: Rethinking Hypermedia the Microcosm Approach. Kluwer, Boston (1996)
25. White, S., Liccardi, I.: Harnessing Insight into Disciplinary Differences to Refine e-learning Design. In: 36th Annual Frontiers in Education Conference, pp. 5–10 (2006), doi:10.1109/FIE.2006.322553
26. Tiropanis, T., Davis, H., Millard, D., Weal, M., White, S., Wills, G.: JISC - SemTech Project Report 28, Bristol (2009)
27. Biglan, A.: The characteristics of subject matter in different academic areas. Journal of Applied Psychology 57, 195–203 (1973)
28. Biglan, A.: Relationships between subject matter characteristics and the structure and output of university departments. Journal of Applied Psychology 57, 204–213 (1973)
29. Becher, T.: The Significance of Disciplinary Differences. Studies In Higher Education 19, 151 (1994)
30. Cassel, L.N., Davies, G., LeBlanc, R., Snyder, L., Topi, H.: Using a Computing Ontology as a Foundation for Curriculum Development. In: SW-EL 2008 Conjunction with ITS 2008 (2008)