

# SECRET CODES: THE HIDDEN CURRICULUM OF SEMANTIC WEB TECHNOLOGIES

RICHARD EDWARDS, UNIVERSITY OF STIRLING

PATRICK CARMICHAEL, LIVERPOOL JOHN MOORES UNIVERSITY

## ABSTRACT

There is a long tradition in education of examination of the hidden curriculum, those elements which are implicit or tacit to the formal goals of education. This article draws upon that tradition to open up for investigation the hidden curriculum and assumptions about students and knowledge that are embedded in the coding undertaken to facilitate learning through information technologies, and emerging 'semantic technologies' in particular. Drawing upon an empirical study of case-based pedagogy in higher education, we examine the ways in which code becomes an actor in both enabling and constraining knowledge, reasoning, representation and students. The article argues that how this occurs, and to what effect, is largely left unexamined and becomes part of the hidden curriculum of electronically mediated learning that can be more explicitly examined by positioning technologies in general, and code in particular, as actors rather than tools. This points to a significant research agenda in technology enhanced learning.

## SECRET CODES: THE HIDDEN CURRICULUM OF SEMANTIC WEB TECHNOLOGIES

### INTRODUCTION

There is a long tradition in educational studies of examination of the hidden curriculum, those elements which are implicit or tacit to the formal goals of education. This article attempts to draw upon that tradition to open up for investigation the hidden curriculum and assumptions about students and knowledge that are embedded in the coding undertaken to facilitate learning through information technologies. In other words, we seek to explore the secrets of the code of information technologies, in particular those associated with recent developments of semantic web technologies for educational purposes. Drawing upon an empirical research study of case-based pedagogy in higher education that is being supported by innovations involving such semantic web technologies, we examine the ways in which code becomes an actor in both enabling and constraining knowledge, reasoning, representation and students. This points to an important agenda for those involved in educational research given the impetus to introduce new technologies into curriculum practices.

Our argument in this article is that while the notion of the hidden curriculum is familiar in discussions of distance and online learning (Anderson 2001), the implications of recent developments in the area of semantic web technologies are poorly understood, while at the same time having the potential for ‘new ways of hiding’ from lecturers and students as they seek to make the curriculum more flexible and interactive. This is located not within the ‘content’ of the curriculum as it is usually understood, but rather within the software and associated practices of computing through which these pedagogical interventions are produced. It is the coding and the linking of data, the applications of technical standards and ways in which decision making and reasoning are articulated in computer code that (along with the hardware and the electronic infrastructures of networks) makes things (like search

engines, web applications, e-assessment systems) perform in particular ways and become actors in the pedagogic process. What hidden codes do is to make it more difficult to determine how and when they are acting, and on what basis. In this article, we are responding to those researchers in information systems who identify technological artefacts as under-theorized (Orlikowski and Iacano 2001) and suggesting one way of doing this in education is through the notion of the hidden curriculum. In addition, in positioning coding as an actor, we are drawing upon the literature of Science and Technology Studies and actor-network theory wherein agency is not the sole preserve of humans and their intentions, but is an effect of particular assemblages of the human and non-human (Latour 2005). This is a strand of theorizing which is of increasing importance in educational research (Fenwick and Edwards 2010). In this way, we seek to intervene in the black box of semantic technologies to explore the work they might be argued to do.

Particularly significant, and yet at the same time largely unrecognized, is the role played by forms of classification and standardization associated with the development of databases, and the ways in which complex knowledge is represented (Lampland and Star 2009). While, as Bowker (2005: 140) argues, ‘you can’t store data without a classification system’, our argument is that, as with much technological innovation elsewhere, in education, how this occurs and with what effects is largely left unexamined and unquestioned. With the passing of time and the incorporation of such data into new assemblages and application, the pre-history of data, selections, application of standards, the application of rules disappears from view and becomes part of the hidden curriculum of electronically mediated learning. With the advent of *semantic web technologies*, which allow data to be shared, aggregated and reused across a linked web of databases and applications (including educational systems), any act of classification, any assumption encapsulated in a rule expressed in the code of a program, or any decision to exclude certain results from the scope of a search may have implications far

beyond its original setting. These implications are indeterminate in nature or impact, and of which the ‘end user’ may be entirely unaware. With the reuse of data across the Internet (itself enabled by the use of the same semantic web technologies) and what Nelson (1965), the pioneer of hypertext systems, described as ‘transclusion’ (the incorporation of all or part of one document or data set into another), there is potential for the propagation and reinforcement of such outcomes. Even now, trying to check the veracity of information in *Wikipedia* is often made difficult because so many other online sources now draw their information ... from *Wikipedia*!

#### THE HIDDEN CURRICULUM

Discussion of the hidden curriculum in education is long standing (for example, Snyder 1971, Apple and King 1983, Margolis 2001). The hidden curriculum refers to the things that are learned by students that are not the intended outcomes of curriculum and pedagogy: an aspect of the enacted rather than prescribed curriculum (Bloomer 1997). In his discussion of distance education, Anderson (2001: 30, emphasis in original) refers to three different senses of the notion of the hidden curriculum:

“(1) a kind of *indoctrination* that attempts to maintain social privilege... (2) the subtle effects of the *setting* in which formal education occurs ... (3) the *unstated* rules for necessary completion of formal education studies”

The notion of the hidden curriculum has been used as part of the critique of educational institutions in reproducing the unequal relations of power in the social order. It is argued that while students may be learning particular subjects and skills at a prescribed level, at an enacted level they learn many other things, including the possibilities they have within the existing opportunity structures of the social order. Institutions, in particular schools, might

officially develop the curriculum to support the learning of all students. However, those things which are selected as part of the formal curriculum provide hidden messages to certain types of students that education and educational advancement is not for them. In this way, it is suggested that individuals become socialized into particular roles in the social order and inequality is reproduced.

For instance, the hidden curriculum has been held to convey the message to working class students that education is not for them, but rather they should take ‘working class jobs’ with limited possibilities for social mobility. Similarly, it is suggested that the hidden curriculum conveys to many women that their primary role in the social order is to be carers. From this perspective, the hidden curriculum is one of the primary educational ways through which social inequality is reproduced. Given that the maintenance of social order was one of the primary motivations for the establishment of many national education systems, it is perhaps unsurprising that the attempts to critique this role have formulated various notions of the hidden curriculum.

The reference to ‘curriculum’ in the term suggests that hidden curriculum is primarily concerned with the forms of knowledge and authoritative discourses made available in educational institutions. When boys mostly (used to) study technical subjects and girls domestic subjects, there was and is clearly a reproduction of established gender roles being learnt through such practices. However, it is also recognized that the hidden curriculum may not simply be about knowledge, but also the forms of teacher-student and student-student interaction that are deemed allowable in the curriculum. The same is the case for the resources used in classrooms, the books, artifacts, furniture and, of course, learning technologies: all may carry hidden as well as explicit messages. This has informed, for instance, the debates about appropriate or authentic content in texts for ‘English as a Second Language’ students

within the context of globalizing processes (Cope and Kalantzis 2000, Snyder 2002). Norms and values are embedded in all of these, which may formally be taken to be universal, but actually differentiate unequally as not all will subscribe to or be encompassed by such universals. Deciphering and the capacity to make one's own meaning is often seen as the redemptive educational strategy in opening up the hidden curriculum to students, wherein code-breaking and code-switching enable the making rather than the simple reproduction of knowledge. However, the extent to which such unveilings truly create transparency or fresh veils is open to debate.

The concept of the hidden curriculum is therefore very broad. It is essentially one that is used to try and explain the inequality that is reproduced through the ostensibly universal provision of a curriculum and some would say democratic practices of education, in particular schooling. It is a way of reproducing social order by providing individuals with messages that they have had their opportunity, but they are only fit for certain purposes, for certain positions in that order. These are often ones that reproduce the types of work and lives that their parents had.

Obviously, the very existence of such a concept and the practices developed to challenge the inequalities in and reproduced by education demonstrates that the hidden curriculum is not as totalizing as it might seem. Nor should we suggest that those practices of education that challenge the hidden curriculum in support of equity do not themselves have a hidden curriculum. In a sense then, there are always unintended outcomes of the curriculum, given that it is always a selection in a particular situation and enactment is always emergent and multiple. Conceptually, for our purposes, the notion opens up for scrutiny those aspects of curriculum and pedagogy which may be overlooked in any focus on the formal curriculum alone as prescribed and described, and teaching and learning outcomes, wherein technologies

are considered as simply tools by which the curriculum is ‘delivered’. In particular, we are interested in the effects of developing standards and coding in the use of information and communication technology on the representation of information and knowledge, the forms of teaching and learning that are possible, and the notion of the student assumed and enacted. For, as Pargman and Palme (2009:181) argue,

What can and what cannot be expressed when it comes to electronic communication is, in the end, determined by the underlying and in many respects invisible infrastructure of standards that enables (and, at the same time, constrains and restricts) such communication.

In this article, we seek to open up what might be considered the merely technical aspects of coding to scrutiny as to its socio-material practices and effects (Bowker 2005, Lampland and Star 2009). In other words, our interests are in the practices of computer coding, and particularly in the context of emerging semantic technologies, the possibilities and constraints placed upon curriculum and pedagogy by the forms of representation that are stabilized within particular standardized computational forms. What norms are built into and excluded by these standards? What sorts of actor is the software, and to what effect? First, however, we need to clarify what it is we are investigating when we write of ‘code’ and of ‘semantic web technologies’ in particular.

#### CODE, STANDARDS AND OPENNESS

Thrift (2005: 240) observes that:

software has grown from a small thicket of mechanical writing to a forest of code covering much of the globe... code runs all manner of everyday devices, from electric toothbrushes to microwave ovens,

from traffic lights to cars, from mobile phones to the most sophisticated computers.

To do this requires enacting information infrastructures and integral to this are the development, adoption and application of standards that enable the organizing and mobility of data. At the same time, software and the practices associated with its development also enact certain notions of the user of technologies (Woolgar 1991), in our case lecturers and students. Thus,

It is not just bits and bytes that get hustled into standard form in order for the technical infrastructure to work. People's discursive and work practices get hustled into standard form as well. Working infrastructures standardize both people and machines. (Bowker 2005: 111-2)

It might therefore be argued that it is not access to information infrastructures or semantic web technologies alone that are transforming learning and teaching, but that the standards and coding associated with them that support or restrict different possibilities dependent upon the alliances and networks of which they are part. In this respect, it is instructive to look at the way in which standards such as LOM, the 'Learning Object Model' (for describing online learning content) or QTI, the 'Question & Test Interoperability Specification' (for online assessment) not only represent particular pedagogical approaches and practices in ways that are widely understood and that support interoperability, but also frame and constrain those practices in particular ways. Adoption of these standards (on the part of an institution, a content producer or the developer of a virtual learning environment) represents a particular positioning, which, again, may have important but indeterminable consequences for teachers and students.



Yet, if ever there is an area that is ‘black boxed’ or ‘naturalized’ (Bowker and Star 1999, Fenwick and Edwards 2010) not just in everyday life, but in most discussion of technically mediated educational practices, it is the nature, production and performance of computer software. For most educators, computers simply work (or should do), and the rhetoric of hardware, operating systems and software is that of enhanced personal efficiency and the extension of reach and impact. This is accompanied by a powerful rhetoric of institutional change and professional competency which strongly encourages teachers to see technology use as a natural part of their work and that computer use in classroom should be seamless and unremarkable. More generally, the ‘tool’ is the dominant metaphor: office tools, productivity tools, web tools; authoring tools for teachers, ‘Google Web Toolkits’ and ‘Semantic Toolboxes’ for developers. Some tools are branded as ‘apps’ or ‘widgets’, but they are still defined in terms of ‘the job they do (for you)’. The discourse of tools already points away from the differing assemblings through which tools come to be and that which they make possible. Effectively, computers are presented as a form of prosthesis, rather than as part of more complex assemblages within which the software is one element, and which necessarily involve their users in a wide range of socio-material relationships. Computers order, preserve and allow access to resources, but they also promote and preclude certain kinds of social and spatial relationships (Urry 2007), they reconfigure absence-presence (Callon and Law 2004), and they support wider forms of communication than those bounded by the classroom. How the codes they draw upon come to do those things is not usually examined in education and, for us, this is potentially part of the hidden curriculum, or might be said to be part of what Thrift (2004) has termed a ‘technological unconscious’ of contemporary life. This is part of a wider issue that Bowker (2005) also alludes to that what interests those who use computers is not necessarily what interests computer scientists. Nor are the concerns of computer scientists necessarily couched in terms that everyday users of the computer systems would recognize,

even when they themselves as ‘the user’ are represented in the system diagrams that computer scientists produce.

We need to be cautious about over-generalization, here, as, over the years, there have been attempts to engage the ‘users’ of computers in software design as more than potential consumers and to involve them in the processes of coding or at least articulating processes that are then transformed into code. Olsson (2004) reviews a range of ways in which non-computer scientists can be engaged in design, development and evaluation of computer software ranging from those in which they provide requirements as potential consumers to those in which they become fully engaged in development processes, working intensively with technologists on the design of data, code and interfaces. The release of code ‘open source’, as exemplified by the Linux operating system and championed by Raymond (2001), has contributed to increased collaboration and sharing amongst software developers, and offers more transparency as to coding practices. However, it remains primarily of interest to experienced developers: users of software produced using open source may benefit from the increased efficiency and productivity it offers developers of the applications they use rather than directly engaging with the code or gaining any additional insights into the coding practices that generate it.

But these tendencies towards openness and participation coexist with, and even promote increased ‘hiddenness’. Part of this stems from a tendency towards a kind of ‘benevolent concealment’ of complexity on the part of coders, who have to balance demands for openness and transparency with the need to produce software that is easy to use, has intuitive interfaces, and works across operating systems and on different kinds of devices from desktop computers to ‘smartphones’. But more significant still is the fact that coders themselves do not necessarily have to, or wish to, engage with all the complexities of the code that they use. It is

not just that a software application is complex in the way that Law (2002) describes aircraft as being fractional and multiple, but also that they are often, in material terms, multi-authored assemblages of components, modules and shared libraries from different sources and with different creators. Thanks to open source sharing these may have no contact with or knowledge of each other, of those benefiting from their action of sharing. The author of a useful code ‘module’ is unlikely to have much awareness of (or may not even be interested in) how others have used it, just as the beneficiaries may have no particular interest in the motivations of its originator, or of its internal functioning. Developers of open source software applications will characteristically not ‘know’ how every part of the applications they produce actually ‘works’. The use of the term ‘canned scripts’ amongst web programmers in particular articulates this black-boxing perfectly: I don’t need to know how it works, I need to know how to build it in to my own code and I am happy if it just does ‘what it says on the can’.

The rhetoric of openness applies to the content of educational technology applications as well. Friesen (2009), acknowledging the limitations of earlier approaches based on purportedly ‘pedagogically neutral’ ‘learning objects’, describes how ‘open educational resources’ (of which the Massachusetts Institute of Technologies ‘Open Courseware’ initiative is perhaps the best known example) offer teachers and students the opportunity to engage not just as consumers of content, but as authors and editors, although standards still play an important role in ensuring their interoperability and reuse. More recently, the same concern with openness and transparency has accompanied the introduction of semantic technologies through the idea of a ‘linked web of data’ (Bizer *et al.* 2009), which has been adopted by governments (including those in the USA and UK) as a way of offering *some* ‘open data’ for public reuse, and, in educational settings, for ‘authentic’ data to be brought into learning environments. Once again, though, with openness comes concealment, benevolent or

otherwise: requests for streams of ‘open’ government data have to be configured in particular ways and consequently constrained, parameterized and indeed conceptualized as the data providers offer it. The standards that allow the access to take place flatten and package the data according to normative and hegemonic categories: for example, queries for information about people with disabilities derived from UK government data sets, while they are certainly openly available, have to be couched in terms of numbers of new registrations, incidences of health service provision, or benefits claimed.

Standards, then, both those that allow software developers to reuse program code in open source environments, and those that describe various kinds of ‘content’, contribute to an understanding of openness that privileges content reusability and technological interoperability. However, by insisting that things are described and knowledge represented in particular ways, potentially at the expense of a critical and exploratory pedagogical openness, much is hidden from the teacher and learner, including the fact that hiding has taken place. The picture here is therefore one of multiple complexities, and is more than simply concern about the ‘reliability’ of information found online. The issue is not whether content from *Wikipedia* or similar sources should find their way into educational settings, but of multiple hidden translations, some effected by human and some by non-human actants, that are incorporated into educational technology applications through codes, ontologies and metadata (Millerand and Bowker 2009).

#### THE SEMANTIC WEB, SEMANTIC TECHNOLOGIES AND THE LINKED WEB OF DATA

The semantic web is conceptualized as ‘an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation ... data on the Web [is] defined and linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications’ (World Wide Web

Consortium, 2001). Early visions of a ‘future web’ stressed the potential for seamlessly integrated systems in which data would be reasoned across by machines in ways that replicate human reasoning. Berners-Lee *et al.* (2001) presented such an example related to personal healthcare management, as do Anderson and Whitelock (2004) and Lytras and Naeve (2006) in relation to potential educational implications of this broad vision of a semantic web.

It is helpful to separate these broad ‘envisionings’ and models of global networked communications from more localized applications and implementations of the raft of ‘semantic web *technologies*’ currently under development. These are guided and framed in part by the broader semantic web project, but also useful in their own right as components of other kinds of web applications and computer programs. These include standardized metadata vocabularies, taxonomies and technical ontologies designed to allow the consistent representation of information; digital repositories and specialized databases (notably semantic triplestores – of which more shortly); data conversion utilities to gather existing data into standard formats; applications to allow users to query these data and display what they discover; and visualization tools to help in the navigation and understanding of the often complex and very large collections of data which are made available.

Over the past decade, the distinction between semantic web and other web technologies has been blurred by two developments. The first of these is the rapid rise of Web 2.0 or Social Web applications and their widespread adoption both by the wider public and, to some extent, in educational settings (see for example, Alexander 2008). The other is the reframing of some semantic web activity as contributing to a ‘Linked Web of Data’ (Bizer *et al.* 2009). Moves to put ‘open data’ from government sources online in order to create a ‘linked data cloud’ have encouraged discussion of opportunities for public engagement and new educational possibilities, although the limitations that current institutional information-sharing practices

place on this development have been recognized (Robinson *et al.* 2009; Ding *et al.* 2010).

While less ambitious in its scope than the broader semantic web vision, the idea of the linked web of data ‘lowers the bar’ to participation and to realization of the benefits of a wide range of semantic web technologies. In educational settings, this has enabled a pragmatic adoption of linked data and semantic web technologies with the potential to enhance existing systems and applications (see for example, Tiropanis *et al.* 2009). However, at the same time, this can mean that their use is even more concealed from the teachers and students who use the educational applications into which they are incorporated: there are no clear signs on web pages that declare ‘Caution: machine-based reasoning at work’. This development may indicate, as Anderson and Whitelock (2004: 6) suggest, there are significant questions about the extent to which the semantic technologies can in principle act in a similar way to humans in dealing with ‘meaning-filled signs’, or whether these are merely pragmatic questions yet to be answered as different kinds of technologies develop.

The complexities and many potential combinations of different semantic web and other technologies have proved a challenge for their developers and advocates of their adoption, who have had to resort to different metaphors and forms of representation. Some of these focus on a single dimension of the problem: the growing ‘linked data cloud’ map (reproduced in Figure 1) concentrates on how different data from different providers might be linked through common taxonomies or ontologies (Cyganiak and Jentzsch 2010).

[FIGURE 1 ABOUT HERE]

Another approach attempts to capture the multiple associations between the elements of a web application, the majority of which would be hidden from the user. A common metaphor and representation since the early days of semantic web development has been that of the ‘layer cake’ with more fundamental technologies and standards underpinning successive layers of

which only the most superficial is typically presented to the ‘end user’. The simplicity of these has been subsequently complicated, with Nowak’s (2010) three-dimensional version (reproduced in Figure 2) offering one solution. Despite the separation of data from protocols, ontologies, rules, logic and presentation, a typical semantic application will contain elements of all of these to differing extents. However, these associations may not be apparent to the user, nor the work they do to enact data in particular forms.

[FIGURE 2 ABOUT HERE]

Underpinning both the broader semantic web vision and many of its associated technologies is a central assumption that knowledge can be expressed as sets of standardized propositions, as it is only on this basis that computer programs can act as agents capable of reasoning in a similar manner to humans (Anderson and Whitelock 2004). The simplest representation of knowledge, to which other, more complex forms can be reduced is the ‘triple’, comprising a subject-predicate-object proposition. The elements of these triples might be URI’s (Uniform Resource Indicators such as web addresses) in which case they are unique and unambiguous, or ‘literals’ such as text strings or numerical values.

This seems to cause few problems, when linking individuals to details like phone numbers, or publications to titles, authors and dates:

“Hotel Pennsylvania”, <<http://xmlns.com/foaf/0.98/Phone>>, “Pennsylvania 6-5000”

“ISBN: 0140620869”, <http://purl.org/dc/1.1/elements#title>>“Alice’s Adventures in Wonderland”

“ISBN: 0140620869”, <<http://purl.org/dc/1.1/elements#creator>>“Lewis Carroll”

“ISBN: 0140620869”, <<http://purl.org/dc/1.1/elements#date>>“1994”

Except, that, of course, the first example presupposes that there is only one ‘Hotel Pennsylvania’ (which is not the case) and that the telephone number, here related to the hotel name by means of the FOAF (Friend of a Friend) standard remains correct (it is not, and it is now 212-736-5000 as things have moved on since the Glenn Miller Orchestra sang about it).

The case of ‘Alice’s Adventures in Wonderland’ neatly demonstrates that even bibliographical data (represented in ‘Dublin Core’ or ‘DC’ and often a starting point for semantic web applications) is far from easy to reduce to a simple list of triples. The unique identifier used as the ‘subject’ in these triples is the ISBN number of the 1994 Penguin Classics reprint of the 1865 original, of which there have been many editions (each with its own ISBN number), some of which combine this work with ‘Through the Looking Glass’; the title is often shortened to ‘Alice in Wonderland’, sometimes extended to ‘Alice’s Adventures in Wonderland: Down the Rabbit Hole’; Lewis Carroll is of course a pseudonym ... and so on. It would be possible, in the latter example, to construct a set of triples in order to try and express this complexity, and, if these were then aggregated in a semantic database or ‘triplestore’, to construct queries which could then find editions of ‘Alice’s Adventures’ matching particular criteria or to test combinations. For instance, we could ask, ‘Are there any editions of “Alice’s Adventures in Wonderland” illustrated by Mervyn Peake and translated into Spanish?’ or ‘Which abridged editions contain the original illustrations?’ with the term ‘original’ being understood to mean those from the 1865 edition and drawn by John Tenniel, even though this is not explicitly stated in the query. Such a triplestore might be extended to include information about Carroll/Dodgson’s other non-fantasy writings, critical accounts and citations, representations in film and on television, and to trace the associations between these.



In order to step beyond simple pattern matching, additional kinds of data are incorporated into triplestores. *Taxonomies*, like thesauri, provide information about terminology, synonyms, hyponyms and hypernyms, preferred and deprecated terms, and translations. Technical *ontologies* link concepts into an explicit structural representation (ideally based on some degree of consensus). Finally, sets of *rules* explicate how new terms (original and abridged might be examples) can be defined and applied. In a typical semantic web application, the results of this aggregation and searching are often unpredictable. Thus, they might be presented to the user through some kind of data visualization, such as a timeline of ‘Alice-related’ publications, or using ‘faceted browsing’, which allows further filtering, sorting and searching.

Even simple examples such as these raise a number of issues about the assumptions that underpin semantic web applications, for the opportunities they provide to aggregate diverse information from around the Internet:

- There is a great deal of emphasis on ‘identity’ leading to a need to seek explicit ‘disambiguation’ and assign a unique identifier to represent the person, object or concept in question
- While the basic model of simple ‘triples’ can be extended, there is an underlying assumption that all knowledge can be represented in this form, and that queries and reasoning can also be expressed in the same way
- The ability to incorporate definitions, relationships and rules into triplestores means that code, as well as human actors, now performs reasoning. Further, the person(s) constructing the triplestore and incorporating the code, which encapsulates or draws on an ontology or a set of rules, may not be those (or even be in contact with those)

either who compiled these, or those who will ultimately use them through their interactions with a web application.

These issues highlight the complex relationship between standardization and flexibility in the development of these technologies. The preparation of data for reuse in this standardized way can be seen as a means by which systems become more efficient and interoperable (for example, by ‘disambiguating’ concepts, people, places or events); standardization enables the aggregation, linking and mobility of data. It is also a prerequisite for creative reuse, reinterpretation and application of multiple ontologies representing both established and emerging positions. However, there remain questions about what forms those standards take and the work they do in representation in the face of burgeoning complexity, and in the enactment of reasoning, knowledge and knowing.

#### SEMANTIC TECHNOLOGIES AS HIDDEN CURRICULUM: AN EXPLORATION

Our interest in the semantic web and the technologies and standards with which it is associated has developed in the context of ‘Ensemble’, a UK based research and development project to explore their potential to support and enhance teaching and learning with cases in different pedagogical settings in higher education. The initial contention of the research project was that the flexibility of semantic technologies was a good match for the complexities of case-based learning, with the potential of these technologies to link to heterogeneous resources from across the web potentially addressing the common demand that case-based learning ‘bring the real world into the classroom’. Semantic web technologies, combined with the move towards ‘open data’ described above, more specifically addressed the desire on the part of many teachers to bring authentic data into classrooms in order to expose learners to ‘real world’ problems rather than sanitized ‘textbook’ cases. Learning with cases is a common pedagogical model in learning environments in which the subject matter is

complex, controversial or rapidly changing, and in which multiple perspectives and interpretations need to be understood. The affordances of semantic web technologies were seen to provide the technological basis for the development of appropriate flexible tools and associated pedagogies to support learning with cases, in which knowledge is developed, represented, adapted and then transferred. We were particularly interested in the means by which cases from research or from professional practice were transformed and constructed as ‘pedagogical cases’, and, in turn, how these were understood, reconstructed and applied by learners.

The Ensemble project has undertaken research and development activities in parallel, in order to contribute both to the understanding of pedagogical practice in case-based learning across disciplines, and also to the development of semantic technologies and the learning environments into which they are incorporated. The project is working with groups of lecturers and students in three higher education sites in England and across a number of disciplinary settings. It is addressing similarities and differences in the conceptualization, construction and application of cases as an individual and group activity.

What has emerged from this research and development work is that it is only rarely the case that what teachers envisage semantic web technologies ‘doing’ is offering students some kind of unfettered open access to digital resources; nor are disciplinary and professional information spaces oriented towards this kind of access. The vast majority of metadata that are used to describe online data and other resources do not draw solely upon formal taxonomies or ontologies (although there are some examples such as scientific names for plants and animals; the internationally-understood names for chemical compounds; or the Laban Vocabulary for contemporary dance). Similarly, rules that apparently offer ways to define the relationships between terms may not be based on consensual definitions understood

and shared across a disciplinary community given the contestation to which much knowledge is subject, even within the most apparently well-codified settings where international standards and systems of nomenclature are well-established. Despite the enthusiasm for expressing knowledge as technical ontologies and applying rules based on these in order to enable technological progress, these approaches embody only one particular, structural view of knowledge, when actual experience involves a wider range of relations and associations, borrowings, inversions, slipping signifiers, and ideologically loaded and appropriated terminology.

If we consider one of the applications developed by the Ensemble team, we begin to see some of the ways in which processes of standardization and aggregation can begin to ‘hide’ these complexities in the pursuit of a particular representation of knowledge and the restricted pedagogical discourse that may result. Informed by the accounts of teachers on a postgraduate course in International Journalism, a web application was developed that presents near real-time data about different kinds of natural disasters (such as earthquakes, volcanic eruptions and floods) on an interactive map. Since the intention of the teachers was to challenge students to critically assess how national, regional and global media covered such events, and how political and cultural bias affected this coverage, excerpts of this coverage was also presented through the same web application. A student could, therefore, ‘focus in’ on a particular region to discover how, for example the U.S. media of different political orientations reported and represented natural disasters in China, for example; or how the global media invoked particular tropes and stereotypes in reporting natural disasters in Sub-Saharan Africa (See Figure 3)

[FIGURE 3 ABOUT HERE]

This application makes use of data providers such as the US Geological Service and many of the world's media outlets have made data available according to well-recognized standards for data exchange and reuse. It also draws on the existence of other web services which will, for example, take a latitude and longitude and return the nearest centres of habitation, the national entity within which that point lies, or a brief description of the area. The standardized representation of events (time, latitude, longitude, depth ...), locations (country, name, region, biome) and media coverage (source, title, byline, summary, keywords) allows these to be linked together and presented through a single (albeit complex) interface.

However, despite the appearance of global coverage, immediacy and openness, what the student is actually able to interact with remains limited. Data is not collected or reported in identical fashion across the globe, as, for instance, the coverage of developed countries with earthquake monitoring networks is better than other areas. Not all media outlets, particularly in less economically developed countries, have the kind of digital presence that can be incorporated into this kind of semantic application. And more critically, there is the potential to apply reasoning which might deliberately exclude particular events or media, or which would characterize them as more or less trustworthy, reliable or simply interesting on different grounds: the technical ('data from the network in country X is unreliable'), pedagogical ('the content is not in English and so is inaccessible to our students') or ideological ('Y is a government news agency'). These can be embedded in the code that governs the production of the visualization in such a way that the student user may be entirely unaware of the decisions and choices being made about the limits to their interaction with data and their sources.

In many of the settings in which the Ensemble project is working, a key pedagogical intention on the part of lecturers is to expose students to contested domains and to encourage them to

engage with, and in, multiple interpretations of situations they may have previously seen as unproblematic and fixed. In fact, one of the main affordances of the semantic technologies with which we are working is their very ability to present data organized, described and presented according to radically different ontologies and taxonomies. However, at the same time, it would be possible to maintain the appearance of interactivity while simultaneously selecting data sources and applying sets of rules which enact a particular ideology, disciplinary perspective or personal viewpoint; notions of quality or validity; or ideas about what is or is not pedagogically appropriate and desirable. This hidden curriculum shapes the potential discourses, trajectories for inquiry, and student subjectivities that might emerge from such a learning environment.

Thus far, this presents a view of semantic technologies that might well hide elements of the curriculum, but without doing so in the sense meant by proponents of the idea of a hidden curriculum in terms of reproducing inequality more generally. Where the applications of semantic technologies bring us closer to this latter, conventional understanding of the hidden is that they can also be used to structure, represent, and reason across data that forms part of the content of the curriculum, as well as its role and structure in terms of activities and desired learning outcomes, adapted according to locale, sector and level, not to mention the actual or expected capabilities of the end user students. The same codes and standards that have the potential to open up possibilities, could, therefore equally be used to limit the extent of interactions and the questions that can be explored according to curricular intentions or user models. This convergence of work on modelling knowledge domains, profiling individuals and then ‘personalizing’ their experiences is not really surprising. The vision of the semantic web presented by Berners-Lee *et al.* (2001) was couched in terms of seamless linkage of databases across the web, traversed by semantic agents whose working was concealed from the user who benefitted from such personalized recommendations and service provision.

While the World Wide Web has in some respects become an expression of the ‘collective intelligence’ of its users (Levy 2009), at the same time it also serves as platform for the delivery of highly specific and targeted content.

#### DECODING SEMANTIC TECHNOLOGIES?

The vision of the Semantic Web is based on the capacity for machines to accurately, locate, read, interpret and process data created by hundreds of thousands of different individuals and organizations. (Anderson and Whitelock 2004: 7)

Much debate focuses on the pragmatics of achieving this goal. However, Anderson and Whitelock (2004: 8) argue that ‘questions related to cultural understanding, contextual variations, as well as semantic and ontological underpinnings of information, make the quest for such systems quixotical’. This does not seem to undermine technological attempts to produce the necessary tools, reinforced as it is by the very discourse of ‘tools’ to which we referred above. Our focus, however, has been to point to the hidden curriculum of such work, to the positionings of subjects through the work of coding and the practices and affordances associated with it. There can be an assumption that in adopting a standard like RDF/XML in triples, people are asserting the truth, when it is simply, if powerfully, a way of structuring the assertion, as in our example of ‘Alice in Wonderland’ above. Work is being done here that hides in making possible and giving credibility to particular representations and reasonings.

As an actor-tool, semantic technologies, which ostensibly represent data in an authoritative and technologically naturalized way, have the potential to give credence to information that, if evaluated through alternative methods, would be deemed incorrect and/or reproductive of certain interests and inequalities. The selectivity and evaluation of data sources is therefore part of the hidden curriculum of semantic technologies. It is also the case that the ontologies

or typologies that tend to be drawn upon in the development of applications in different subject areas reflect certain disciplinary, organizational and ideological perspectives. Yet these can remain hidden in the codes and rules of those technologies. One implication is that ‘the world that is explored scientifically becomes more and more closely tied to the world that can be represented by one’s theories and in one’s databases; and this world is ever more readily recognized as the real world’ (Bowker 2005: 152). In other words, the hidden work to produce particular representations of the world is naturalized and lost.

While issues of access, reliability and dominant ideologies impact upon any educational practice, our argument is that code, standards and their associated practices take on a heightened role and becomes more significant in the context of semantic web applications in the enacting of certain forms of knowledge, knowing and subjectivity. In enacting the curriculum with the use of semantic technologies therefore, there is a need to clarify the particular possibilities and constraints that the codes and standards afford and explore the positionings of teachers and learners and their effects within these assemblages. This opens an important research agenda for all those involved in developing curriculum with semantic technologies, in particular examining the positionings of students and lecturers as they engage with the new technologies and the specific forms of knowledge, understanding and values that are enacted.



## ACKNOWLEDGEMENTS

This paper draws upon research conducted as part of the research project: ‘Ensemble: Semantic Technologies for the Enhancement of Case Based Learning’, funded by the ESRC/EPSRC’s Teaching and Learning Programme in Technology Enhanced Learning (RES-139-25-0403). Full details of the project and the research team are at the project website: <http://www.ensemble.ac.uk>. An earlier version of the article was presented as a paper at the Annual Conference of the European Association for the Study of Science and Technology, University of Trento, September 2010.

## REFERENCES

- Alexander, B. (2008) Social networking in higher education. In N. Katz (ed.), *The Tower and the Cloud: Higher Education in the Age of Cloud Computing* (pp. 197-202). Washington, DC: Educause.
- Anderson, T. (2001) The hidden curriculum in distance education: an updated view. *Change*, Nov/Dec 2001: 29-35.
- Anderson, T. and Whitelock, D. (2004) The educational semantic web: visioning and practicing the future of education. *Journal of Interactive Media in Education*, 1: 1-15.
- Apple, M. and King, N. (1983) What do schools teach? In H. Giroux, and D. Purpel (eds) *The Hidden Curriculum and Moral Education*, Berkeley, California: McCutchan Publishing Corporation.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The semantic web. *Scientific American*, 284, 5: 34-43.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009) Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5, 3: 1-22.
- Bloomer, M. (1997). *Curriculum Making in Post-16 Education*. London: Routledge.
- Bowker, G. (2005) *Memory Practices in the Sciences*, Cambridge, Mass.: MIT Press.
- Callon, M. and Law, J. (2004) Introduction: absence-presence, circulation and encountering in complex space. *Environment and Planning D: Society and Space*, 22: 3-11.
- Cope, B. and Kalantzis, M. (eds) (2000) *Multiliteracies*. London: Routledge.
- Cyganiak, R. and Jentzsch, A. (2010) Linking Open Data Cloud Diagram. Retrieved from <http://richard.cyganiak.de/2007/10/lod/>

- Devedzic, V. (2006) *Semantic Web and Education*, Dordrecht: Springer.
- Ding, L., Difranzo, D., Graves, A., Michaelis, J., Li, X., McGuinness, D. et al. (2010). Data-gov Wiki: Towards Linking Government Data, Proceedings from AAAI Spring Symposium Series 2010. Online at:  
<http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1154>
- Dolog, P. and Schaefer, M. (2005) Learner modeling on the semantic web. In Proc. of PerSWeb-2005 Workshop, Personalization on the Semantic Web, July, 2005, Edinburgh, UK.
- Fenwick, T. and Edwards, R. (2010) *Actor-network Theory in Education*. London: Routledge.
- Friesen, N. (2009) Open Educational Resources: New Possibilities for Change and Sustainability International Review of Research in Open and Distance Learning 10(5)  
 Online at: <http://www.irrodl.org/index.php/irrodl/article/view/664/1388>
- Orlikowski, W. and Iacano, C. (2001) Desperately seeking the 'IT' in IT research - a call to theorizing the IT artefact, *Information Systems Research* 12, 2: 121-34.
- Lampland, M. and Star, S. (2009) (eds) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Ithaca: Cornell University Press.
- Latour, B. (2005) *Re-assembling the Social: An Introduction ot Actor-network Theory*. Oxford, Oxford University Press.
- Law, J. (2002) *Aircraft Stories: Decentering the Object in Technoscience*. Durham, NC: Duke University Press
- Levy, P. (2009) Au-delà de Google: Les voies de "l'intelligence collective". *Multitudes*, 36: 45-52.

- Margolis, E. (2001) (ed.) *The Hidden Curriculum in Higher Education*. London: Routledge.
- Mika, P. (2007) *Social Networks and the Semantic Web*. Dordrecht: Springer.
- Millerand, F. and Bowker, G. (2009) Metadata standards: trajectories and enactment in the life of an ontology. In M. Lampland and S. Star (eds) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca: Cornell University Press.
- Nelson, T. (1965) A File Structure for the Complex, the Changing and the Indeterminate. Proceedings of the ACM 20th National Conference 1965, pp. 84-100
- Nowak, B. (2010) The Semantic Web Technology Stack. Retrieved from <http://www.bnode.org>
- Olsson, E. (2004) What active users and designers contribute in the design process. *Interacting with Computers*, 16, 377-401.
- Pargman, D. and Palme, J. (2009) ASCII imperialism. In M. Lampland and S. Star (eds) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca: Cornell University Press.
- Raymond, E. (2001) *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopol, CA: O'Reilly
- Robinson, D., Yu, H., Zeller, W., & Felten, E. (2009). Government data and the invisible hand. *Yale Journal of Law and Technology*, 11: 160.
- Snyder, B. (1971) *The Hidden Curriculum*. New York: Knopf.
- Snyder, I. (ed) (2002) *Silicon Literacies*. London: Routledge.

- Thrift, N. (2004) Remembering the technological unconscious by foregrounding knowledges of position. *Environment and Planning D: Society and Space*, 22: 175-90.
- Thrift, N. (2005) Beyond mediation: three new material registers and their consequences. In D. Miller (ed.) *Materiality*, Durham, NC: Duke University Press.
- Tiropanis, T., Davis, H., Millard, D., Weal, M., White, S., & Wills, G. (2009). SemTech Project Report. Southampton: University of Southampton/JISC. Online at: <http://www.jisc.ac.uk/media/documents/projects/semtech-report.pdf>
- Urry, J. (2007) *Mobilities*. Cambridge: Polity Press.
- Woolgar, S. (1991) Configuring the user: the case of usability trials. In J. Law (ed) *The Sociology of Monsters*, London: Routledge.

Figure 1: The Linking Open Data Cloud Diagram



Figure 2: The Semantic Web Layer Cake (3-Dimensional Version by Benjamin Novak). Note the less complex 'linked web of data' alongside the more complex assemblage of the 'semantic web toolkit'

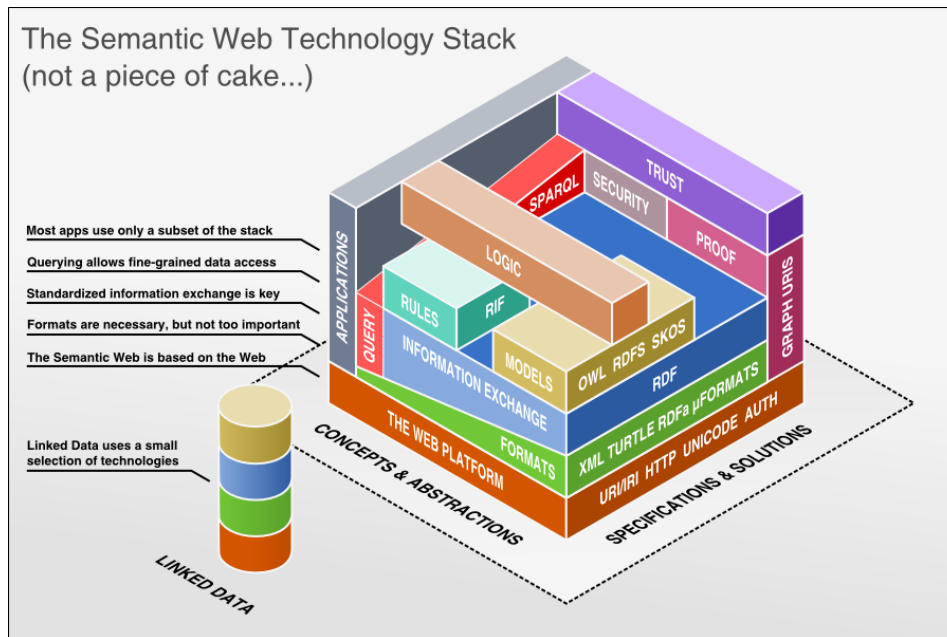


Figure 3: Global Events, Global Media: data from the US Geological Survey plots earthquakes onto a map; clicking a location presents linked data and media coverage

### Global Events, Global Media: Exploring the coverage of seismic events in the world's media

#### Recent Seismic Events

These data are derived from the web service of recent seismic events produced by the [USGS](#). The full source data file is available online [here](#).

Click the link next to an event to see it highlighted in the map and view other data about it.

#### Largest Recent (90 Days)

2010-02-27 Magnitude:8.5 [Show Details](#)  
 2010-02-27 Magnitude:8.3 [Show Details](#)  
 2010-01-03 Magnitude:7.1 [Show Details](#)  
 2010-01-12 Magnitude:7.0 [Show Details](#)  
 2010-02-26 Magnitude:7.0 [Show Details](#)  
 2010-02-18 Magnitude:6.9 [Show Details](#)  
 2010-02-27 Magnitude:6.9 [Show Details](#)  
 2010-01-05 Magnitude:6.8 [Show Details](#)  
 2010-01-05 Magnitude:6.8 [Show Details](#)  
 2010-01-03 Magnitude:6.6 [Show Details](#)  
 2010-01-10 Magnitude:6.5 [Show Details](#)  
 2010-02-07 Magnitude:6.3 [Show Details](#)  
 2010-01-17 Magnitude:6.3 [Show Details](#)  
 2010-02-27 Magnitude:6.2 [Show Details](#)  
 2010-02-15 Magnitude:6.2 [Show Details](#)

#### Most Recent

2010-02-27 Magnitude:5.1 [Show Details](#)  
 2010-02-27 Magnitude:6.2 [Show Details](#)  
 2010-02-27 Magnitude:6.0 [Show Details](#)  
 2010-02-27 Magnitude:5.5 [Show Details](#)  
 2010-02-27 Magnitude:4.8 [Show Details](#)  
 2010-02-27 Magnitude:5.4 [Show Details](#)  
 2010-02-27 Magnitude:5.6 [Show Details](#)  
 2010-02-27 Magnitude:6.0 [Show Details](#)  
 2010-02-27 Magnitude:6.9 [Show Details](#)  
 2010-02-27 Magnitude:8.3 [Show Details](#)  
 2010-02-27 Magnitude:5.6 [Show Details](#)  
 2010-02-27 Magnitude:8.5 [Show Details](#)  
 2010-02-27 Magnitude:6.1 [Show Details](#)  
 2010-02-27 Magnitude:5.2 [Show Details](#)  
 2010-02-27 Magnitude:5.3 [Show Details](#)

#### World Map Display



[Reset Map](#)

#### Location

Location	Date and Time	Magnitude (MWS)	Depth (km)
-35.988 lat, -72.782 long	2010-02-27 at 06:34:09.7	8.5	9

#### Locality

##### Geonames Service Information

Country: Chile  
 Near to: Puyay, Bio-Bio  
 Locality:

#### News Coverage

[Search News Stories](#)