McCathieNevile, C. & Méndez, E. (2007). Library Cards for the 21st Century. *Cataloging & Classification Quarterly*, 43 (3-4), pp. 21-45.

DOI: 10.1300/J104v43n03_02

Library Cards for the 21st Century

Charles McCathieNevile

Eva Méndez

SUMMARY

This paper presents several reflections on the traditional card catalogues and RDF (Resource Description Framework), which is "the" standard for creating the Semantic Web. This work grew out of discussion between the authors after the Working Group on Metadata Schemes meeting held at IFLA conference in Buenos Aires (2004). The paper provides an overview of RDF from the perspective of cataloguers, catalogues, and library cards. The central theme of the discussion is re-source description as a discipline that could be based on RDF. RDF is explained as a very simple grammar, using metadata and ontologies for semantic search and access. RDF has the ability to enhance 21st cen-tury libraries and support metadata interoperability in digital libraries, while maintaining the expressive power that was available to librarians when catalogues were physical artefacts.

KEYWORDS

RDF, resource description framework, Semantic Web, vocabularies, catalogues, library cards, bibliographic control, metadata interoperability

INTRODUCTION: REMEMBERING LIBRARY CARDS

The 20th century presented many significant changes, but one central feature remained constant, and would last until the end of, but not beyond, the entire century. In our own university libraries, very far away one from another (University of Melbourne in Australia and University of León in Spain) when we first arrived as students in the 1980s-1990s they still held pride of place, the central object you saw as you entered, taking up the main space of the ground floor. Big, wooden, familiar, box after box held drawer after drawer filled with catalogue cards. Authors, Titles, Subjects, and classification codes served to order the infinite cards of 11.5 × 7.5 cm (4.5 × 3 inches) with a small hole punched at the bottom where a rod was inserted to keep them in order.

Those cards held the secrets of the library. Each one contained information about a book (or perhaps a map, a vinyl record, or even a video-cassette, because they were already modern libraries). The modern librarians would have to make several cards for every book, and then organize them one by one, following alphabetic criteria and different rules. There might be several more that were indexed by subject. A hundred years of history had given them personality, from the hand-written cards in copperplate that we learned to appreciate and decipher in equal measure, to the typewritten cards with their letters not quite straight, and the ultra-modern printed cards that came out of a specific kind of software for printing the cards or even directly from the computerized catalogue, and into the boxes just for completeness in the transition.

As late as the mid-1990s those cards were the best of all catalogues. The computer catalogue was faster, but not yet complete. Many of the important facts were reduced to one giant set of notes at the end of the computer record. A shame, because the cards in the box held rich descriptions, cross referencing each other and pointing to trails of investigation just waiting for the curious.

It is fair to say that the computer system in use at the time was not the best available in the world. And it should be noted that the “librarian” on duty was often simply a student, making some relatively easy money by working in a place they more or less knew, close by, with only the briefest of introductions to the systems they were expected to use. A true librarian could be expected to know how to catalogue a collection, how to ensure that the cross references were all correct and correctly entered, that the work was listed according to the appropriate subjects. And a true librarian might also have gone to the card catalogue looking, checking the information against what was there already.

The catalogue, then, was made to produce reliable results, to display differences between books and cards, to bring together what has to be together, to present meaningful choices and to locate what users want. Modern catalogues, as we move into the 21st century, want to do the same. But in automatic cataloguing and database systems, catalogue data has to be encoded in well-defined ways, which requires a data format. Systems used a variety of formats, sometimes proprietary and not always homogeneous. What was needed for communication between systems was an exchange format; particularly, a machine readable cataloguing format. It had to contain a large number of fields, typically structured into subfields. The content of the fields, however, was governed by the cataloguing rules. The format was and will be a container, and different containers may be used to communicate the same content. Nowadays, these containers could be MARC21, XML (Extensible Markup Language), or even RDF (Resource Description Framework) (Méndez, 2005) in a more semantic Web. We will devote this paper to the value of using RDF for libraries, after the era of catalogue cards has really passed.

In the world of library cards, there were separate physical catalogues for authors and titles, subjects and situation codes. In the world of library OPACs, there were always combined formal and subject access points in the same database if not generally in the same index. However, in the world of WebPACs and the WWW, bibliographic information from the catalogues still largely remains buried within the "hidden Web"–and that, as long as different layers of information remain blended in bibliographic records, the non-librarian world probably is better off without these thousands of identical bibliographic records pointing simply to different items or manifestations (Gradmann, 2005). Furthermore, an important amount of our current library users never have searched in a card catalogue and they have though searched a lot in all kind of search boxes on Internet/Web search engines.

At the end of the last century, all the Librarianship and Information Science (LIS) research had a prospective tone, as if digital information management changed completely and all the librarians waited for a kafkaesque mutation into computer scientists. Now, entering the 21st century we have realised that the user's needs are basically the same, even though users are more autonomous. Information professionals, librarians, and specialised cataloguers are learning to adapt their skills in processing of books to a technological processing of "resources."

Traditional cataloguing blends the main elements: tangible documents, processes (bibliographic description, content analysis–mainly indexing and abstracting), and products, which are fruits of this activity (typically bibliographic records collected as a catalogue). Resource description implies: documents like information objects, the creation–automatic or manual–of information about the attributes of those objects (metadata), making metadata records and metadata repositories (Table 1).

THE WEB: OUR BIGGEST EVER LIBRARY?

With the growth of Internet and, in particular, the success of the World Wide Web, the myth of the "universal library" from Borges has been revived in many scholarly and philosophical discussions. That dream of a library where the whole of human knowledge would be accumulated has been seen to come closer to fruition with the Web. Every once in a while we see a paradigmatic new project or a new birth of the *universal library*.[1] Attempts to create universal libraries, from the mythical Library of Alexandria, failed in previous centuries. In the 21st century knowledge is not perceived as a solid structure any more. The universal library is still a utopia despite the Web. Notwithstanding, the Semantic Web and the technologies used in its construction let us dream again of universal access to knowledge, which in a more modern and technological way we would call something like: "interoperable access to distributed digital knowledge."

TABLE 1. Traditional Cataloguing Elements vs. Electronic Resources Organization

	TRADITIONAL CATALOGING	ELECTRONIC RESOURCES ORGANIZATION
OBJECT	Books/documents	Resources/DLOs (Document-Like Objects)
PROCESS	Bibliographic Description	Resource based metadata/Descriptive Metadata Creation
	Classification	Subject based metadata*/Specific vocabularies
PRODUCT	Bibliographic Record	Metadata Record
	Catalogue/OPAC	Metadata Repository
	Technical processes	**Technological processes**

*Ahmed, Kal et al. (2001) made this distinction between *resource based metadata* and "subject based metadata." Resource based metadata are those metadata used for cataloguing and identification, used to associate specific properties and their values, and the traditional example is the library catalogue record giving its title, author, publisher, etc. On the other hand, *subject based metadata* are the metadata which represents subjects and their interrelationships and also usually designates specific information resources as belonging to these subjects; this kind of metadata implies specific vocabulary construction and encoding such us ontologies, thesauri, or topic maps.

But while there are many smart librarians, working in libraries great and small, and managing information in ways that a 19th century librarian might not recognize, there is more information being produced. It has been claimed that at the beginning of the 21st century the Web allowed more information to be published in a single year than there had ever been published in human history until that year. Publishing to a large audience is no longer the preserve of the wealthy, the large publishing house, or the highly motivated pamphleteer. Anyone with access to the Internet (somewhere between a quarter and a half or so of the world's population, although it is difficult to measure) can publish persistent documents or "document-like objects," as the post-modern "text" has become known in a digital world.

The number of librarians graduating this year will not exceed the total number of librarians who have ever lived and worked before. Their traditional work is increasingly being done by students or neophyte librarians with relatively little training, or by authors themselves who may have no formal training. The rise of "tagging" or personal tagging to produce "folksonomies" shows a kind of catalogue being constructed in a chaotic way by people whose average training in the use of ontologies and subject thesauri is vanishingly close to zero. To make the most effective use of this information requires being able to determine some kind of order. Folksonomies pursue an ancient technique for organising libraries by natural language description as a modern approach to organising the Web, but rely on the ability to look up tags much faster than in a physical library. WikiPedia,[2] the radical worldwide open encyclopaedia project, can be an enormously valuable resource to a town too far off the beaten track to be visited by an encyclopaedia salesman, but the traditional approach of getting experts to write articles is replaced with a process that allows anyone to add information. A lot of "instant librarians" are indexing in free language their photos on Flickr,[3] their resources on del.icio.us,[4] etc., and a lot of authors are, creating Web sites, digital information, electronic resources, or even encyclopaedias.

On the other hand, information professionals have been discussing about digital libraries, virtual libraries, global libraries, universal libraries, etc., for more than eight years. Because of their virtual nature, professionals have fallen to the temptation of equating the World Wide Web itself to a giant digital library (Brisson, 1999), and metadata, as the technical basis of processing in digital libraries, increased the interest on cataloguing, even for the World Wide Web Consortium (W3) and Semantic Web theory and standards (which also flirt with the universal library ideal).

COMPUTER CATALOGUING AND STANDARDS CHANGE

As Brisson (1999) said, with the Internet and the explosion of Web information, the world has discovered cataloguing. When libraries began to use MARC format for their library catalogues in the late ’60s, they converted the existing records on cards into electronic form for storage and retrieval. Moreover, since the beginning of the computer era, it has been clear that computers were useful for any task based on looking up information in tables, from cracking military codes to tracking student records or the legendary genealogy projects of the Church of the Latter-Day Saints (the “Mormons”). Computer catalogues can do cross-indexing by themselves, looking up any of several fields over thousands or millions of records at lightning speeds. For a large library this is a wonderful thing–in the time it takes to find the first card that has the author you wanted, the computer found all of them.

But this comes with a drawback. For example, the correct way to write one of the authors’ names is “McCathieNevile,” but other people could (and do) write it McCathie-Nevile, MacCathyneville, MacCarthy-Nevile, or even MacKazinevil if you ask to write it to a Spaniard. A person can readily match all of those with a modicum of intelligence, the thing that comes naturally to people. But computers are notoriously not intelligent. It is possible to write rules they can follow which will match all things almost the same, but again they don’t manage to reject the right things, so you are left reading through a large number of records that are close, according to the rules, but obviously not right.

Moreover, with the growth of electronic catalogues, it was no longer possible to write notes on the card, like “Tim usually hides this book in the opposite shelf” easily. In order to add more useful information, cataloguing systems became extremely complex–a complete MARC record is something that most catalogue systems hide from real users, and with good reason–it is a scary thing to behold.

Similarly, computers need standard ways of cataloguing and classifying in ways that people do not, because they cannot make the leap to connect “rabbits” with “bunnies” except where the brand “Playboy” is involved, and they just don’t learn the word “coney” because it is too rare for programmers to run across it. (Or perhaps was until *The Lord of the Rings* taught it to a new geeky audience.) Subject schemes have become extremely complex, in order to describe them in the hierarchies that traditional database and catalogue systems rely on.

Library cards and library records are possibly the most distributed rights management artefacts in the world, and there is a great amount of content that doesn't get examined enough because of licensing restrictions. Attribute-based authentication systems can now open the door to bringing more of the world's recorded knowledge into a common setting without impinging identity or privacy. There is potential for a great improvement in both libraries and Web search systems if we could identify some ways for them to easily talk to each other.

Representing and organizing information in a digital form has a long tradition in libraries, using various rules and standards. The now-venerable MARC format is joined by a large variety of new formats and metadata schemas based on SGML, XML, or even HTML. Standards in libraries are changing, even in the highly structured world of AACR and MARC[5] both in terms of encoding standards like MARC21 and content standards, defined now as RDA (Resource Description and Access) (JSC, 2005-06).

The global information society, where the universal character of the Web claims to offer democratic access to the information, needs more than ever the ideas of internationalisation and interoperability behind the BUC (Bibliographic Universal Control), now as a *Universal Web Control.*

TRADING FREEDOM FOR POWER

Traditional databases have to be constructed in advance, describing the information that you want to add to it before you ever start. If you learn, later, of a better way to classify something, it is a very complex job to add this, and in a distributed system it is likely to be done in different places by different people. The Dewey system, while it provided some common order to libraries everywhere, provides a good example of how hierarchies can become unmanageable, with different people classifying the same resource in slightly different ways leading to it existing in radically different parts of the library. Unfortunately, when you are looking for information on the relationship of Tancred of Antioch to his Uncle Bohemond (or Boamund . . .) of Taranto it is not very interesting to realise that most other people making more general enquiries find the things they are looking for. It is simply a frustration that you have to look in two different places for the same book because two slightly different librarians classified it slightly differently. We are exaggerating at this point, because the catalogue usually makes it fairly fast to find out where the book is in this library. The point is that some people are being inconvenienced and there is no easy remedy. We are thinking of the "new user" that never saw a card catalogue, and has been using Web search engines for 5 or 6 years.

We gave up some of our freedom, the ability to easily scribble on the card and add notes, for the power of a fast search. It is not impossible to add a new field to a database, but it is traditionally very expensive. Making a database understand a new format for a few books is simply not worth the effort, so it does not happen. (Making it work across different catalogues is an enormous amount of work–it is not clear that it can be done except in rare cases.) Is there a way we can get the best of both worlds, without having to run from the computer to the cards and back? . . . Spoiler: YES.

A SEMANTIC WEB: A MORE LIBRARIAN WEB

Let's begin with a description of our suggested solution, and then have a look at how it works. (That way when it gets technical we have already understood the punch line, and we can skip the rest.)

The Semantic Web is built on machine-understandable data, not just machine-readable data (MARC). As Tim Berners Lee (2000) said in his book *Weaving the Web*[6]: If HTML and the Web made all online documents seem to be an enormous single (but multilingual and somewhat schizophrenic) book, the Semantic Web would make all the information seem to be in one enormous data base. The Web needed some structure to constrain mark up in order to achieve standardization and interoperability between metadata schemas, and RDF is a language which allows both machines and human beings to process, organize, and retrieve digital information. The Semantic Web pursues the idea of a *Web Universal Control* (although not formally stated) from mark up structures that can represent ontological knowledge representation.

We want a way of describing things. It should work to quickly look up common things (in the library case, an author, or a title, or a subject), but it should also be reasonable to use it for a small handful of special cases. And in the ideal world we normally only dare visit late at night in our dreams, it should be possible to make our little solution for a handful of things we have locally work everywhere, and work easily with someone else's solution for a similar problem that they developed without talking to us. Just because the Gupapuyngu-speaking folks at the school in Yirrkala and the biological researchers at the Mongolian academy of sciences have no language in common, and have only once ever wanted to mention the same subject, doesn't mean that it should be difficult to discover what that subject was.

So we want to be able to describe things in a way that allows other people to re-use our descriptions when they discover them, adding them at that point to their system. Actually, we want a way for third parties (who have a language in common with the people at each end of this discussion) to describe these descriptions. And of course we want to be able to automatically process the descriptions themselves. This gives us the power of a computer search, the ability to scribble on the virtual library cards, and as an added bonus a way to transfer things across different communities after they are built, without needing direct coordination, and without rebuilding the cataloguing system itself.

We have some pieces in place. We can put anything on the Web, with a URI (Uniform Resource Identifier, a Web address, like http://www.example.com/some/magic/address or similar) and we know people can get to it. That takes care of the "how are we going to share this with people we don't know." We even have engines that crawl around the Web looking at the things there, and letting other people find them. What we need is a way of making descriptions that does what we are looking for. And lo and behold, this is one of the ideas behind the design of the Semantic Web.

Before the Web, discussion about cataloguing was mainly limited to librarians talking about rules, standards, and authority control. In the mid-1990s every Internet community started to develop their own perceptions and their own ways to describe and retrieve electronic resources, through metadata schemas and standards. And it started the different tensions between metadata and cataloguing, between machine understandable data and human understandable data, between MARC and mark up, and so on, generating ink rivers and a lot of bytes of information in the specialised bibliography.[7] Dovey (1999) already talked seven years ago about different schools: the bibliographic control school compared to the structuralist school and the school of structured data. Campbell (2004) recognized that different communities require different granularities of description and he talks about communities of electronic information and a bibliographic organization systems community. The Semantic Web community shared with the "school of bibliographic control" the desire for interoperability and "universal control" of Web resources. In that community, such interoperability and the control it gives are based on formal languages and metadata schemas, both machine and human understandable data.

A question here is why there is such a big gap between library-oriented and other information "worlds" in the perception of universal access to information? Assuming that these essential gaps can be filled, there remain legitimate reasons why a standards-in-common solution poses its own enduring problems, even within the compelling context of interoperability (Howarth, 2005). Lynne Howarth used Stu Weibel's words to point out the need for convergence and cooperation of all these schools, groups, YAMS (*Yet Another Metadata Standard*) and Internet communities: *"The Internet can be thought of as a World-Wide Commons in which many previously-distinct resource description communities are mixed together."*

Several authors (Tillet, 2003; Gradmann, 2005) have envisioned FRBR (Functional Requirements of Bibliographic Records) as the approach to make catalogues more visible and easier to use for any Internet community. Semantic Web technologies and the conceptual framework of FRBR are two promising areas for making librarian and generic WWW information services converge or even prepare some sort of integration scenario (Gradmann, 2005).

In the Semantic Web, universal information retrieval approach is as different as it was in the traditional school of bibliographic control and UNIMARC approach. Tim Berners-Lee and other Semantic Web promoters in and out W3C (such as, James Hendler, Ora Lassila, Eric Miller, Dieter Fensel, or Ivan Herman) are hoping to create a meaningful Web of data. So the computer could learn both about the data and about the information needed to process such data. But the main objective is the same: the global and interoperable Web information processing, which in this case will be founded on RDF, Metadata schemas, and content vocabularies (schemes). Who is guiding the Semantic Web is not the library World, but rather the W3C. The Semantic Web approach needs librarians, because the Semantic Web is, as we said before (Méndez, 2004), a "more librarian Web." Likewise, librarians need the Semantic Web to make their metadata interoperate, not only in MARC-standards domain, but in the entirety of the Web, redefining the traditional strengths and skills of library information organization for the Semantic Web and/or Universal Access era.

WHAT THE SEMANTIC WEB LOOKS LIKE: AN EASY WAY TO SHOW RDF

The Semantic Web is an idea that it is possible to extract information from the Web at large. The language used most often by the inventors of the Web to do this is called the Resource Description Framework (RDF, 2006), which even sounds like the boxes of cards, or like "Resource Description and Access," the newest version of Anglo-American Cataloguing Rules.[8]

The RDF suite of specifications[9] consist of six W3C recommendations (RDF, 2006), and they are perhaps the most powerful and most important standard if the Web is to achieve its full potential (Ahmed et al., 2001), that grew out of a requirement to apply descriptions to information resources. It is intended to allow the computer processing of distributed information on the Web. We will devote this part of the paper to explain the RDF model and syntax and RDF schema from the perspective of library cataloguing or "resource description."

In a simplest level RDF is an XML-based language to describe resources, which underpins the Semantic Web paradigm. A resource in RDF stands for either electronic resources, like files (an (X)HTML Web site, for example) or concepts (like RDF/OWL representations of concepts in a thesaurus or Knowledge Organization System) or even a person who has an URI (for example, an e-mail in a FOAF–Friend Of A Friend–description). An RDF resource is basically "anything that has identity." Another easy way to explain it is Jul's (2003) definition: *A resource is any item we wish to describe–one can think of it as what you get when you click on a URL.*

Writing RDF (RDF Model and Syntax)

RDF is based on three-part statements of the form "**something** has **some relation** to **some other thing**." People who like formal language call these three parts respectively: subject, predicate, and object of "the triple," or statement. And it (normally) uses a URI (such as a Web address) to identify each member of the triple (although the object can also be "plain text," a so-called *Literal*).

Subject-Predicate-Object, the three elements that compound a statement or RDF triple, could be also understood as Entity-Relation-Entity in a relational database paradigm or Class-Property-Value in the object-oriented landscape. Let's explain what these elements of an RDF statement means:

- A Subject [in our example (Figure 1): http://www.bartleby.com/173] is the resource that is being described by the ensuing predicate and object, and the URI (URL) stands for a unique concept.

- Predicate [in our example (Figure 1): dc:title; dc:language and dc:creator] is a relation between the subject and the object or a property type referred to the resource which will have a value or object. In RDF, as we will explain below (Table 2), we would define a unique URI for any predicate.
- An Object is either the resource referred to by the predicate or a literal value [in our example (Figure 1), the values: Albert Einstein, en, and Relativity: The Special and General Theory].

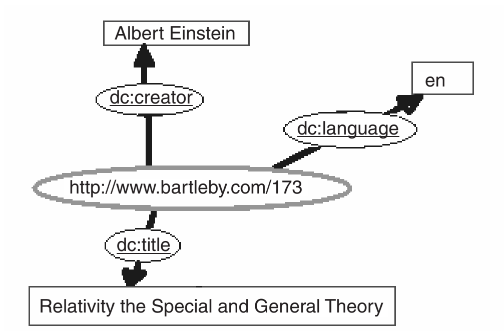
See the code below to identify all these elements → Understanding the following code isn't necessary for understanding this article–all the actual RDF examples here are for completeness, and are accompanied by English explanations of the relevant bits.

```
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
xmlns:dc='http://purl.org/dc/elements/1.1/'>
<rdf:Description rdf:about='http://www.bartleby.com/173'>
 <dc:creator>Albert Einstein</dc:creator>
 <dc:language>en</dc:language>
 <dc:title>Relativity: The Special and General Theory</dc:title>
</rdf:Description>
</rdf:RDF>
```

This code means: There is a thing which we identify with the URI http://www.bartleby.com/173, and which has some properties and values: The creator is "Albert Einstein," the language is "en" (English, from an ISO standard vocabulary for langauges), and the title is "Relativity: The Special and General Theory." So far, there is nothing in this that doesn't seem like a normal library record. The terms we have used to describe our resource are all pretty common, and they come from a standard vocabulary meant for library-like cataloguing called Dublin Core.

One of the useful things about information is that the more of it you have (so long as it is consistent–but we'll ignore that problem for the minute), the more you can do with it. RDF has a simpler approach to information than most systems. If you give an RDF processor two sets of statements about the same thing (a thing identified by the same URI), it just treats them as one collection of statements. Unlike many of its predecessors, the idea behind RDF is not cataloguing according to a predefined scheme in quite the way we are used to. In order to provide useful descriptions, we do indeed need to have some common terms. But in-

FIGURE 1. Graphical Representation of a Basic RDF Statement (RDF Model)



In this and following figures in this paper, the oval (○) represents an URI, a resource; The square (□) represents a value or literal, and the circle in the arrow (-○→) a property or association (a predicate).

stead of building a database, and then finding out afterwards what fits into the database, the Semantic Web is a collection of descriptions, not very different from a library catalogue except in the scope of material. Computers can understand such a collection, present it so humans can also read the information, and by following simple well-defined rules computers can infer further semantic information.

A practical upshot of the way this is done is that we can use any kind of description we like. Let's start, like any decent library card did, with a bit more about the creator, since names can be shared by so many people:

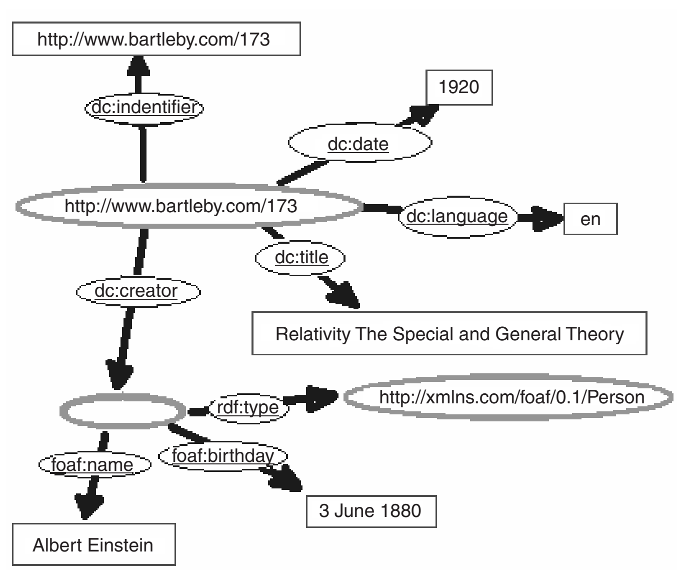
```
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
 xmlns:dc='http://purl.org/dc/elements/1.1/'
 xmlns:foaf='http://xmlns.com/foaf/0.1/'>
<rdf:Description rdf:about='http://www.bartleby.com/173'>
 <dc:creator>
 <foaf:Person>
 <foaf:name>Albert Einstein</foaf:name>
 <foaf:birthday>14 March 1879</foaf:birthday>
 </foaf:Person>
 </dc:creator>
 <dc:language>en</dc:language>
 <dc:title>Relativity
 The Special and General Theory</dc:title>
 <dc:date>1920</dc:date>
 <dc:identifier>http://www.bartleby.com/173</dc:identifier>
</rdf:Description>
</rdf:RDF>
```

Here, we have used some terms that were created for describing actual people. We may not know how many people called Albert Einstein wrote something called "Relativity: The Special and General Theory" in 1920, but in case there are a few at least we know that the one we mean was born in 1879.

Importantly, we have here just one random example of annotating the card with something new and different. We could have used any defined term, from vocabularies about describing books, or the kind of words used in text, or detailed taxonomies of literary style or subjects. We could have mixed all of these and other things together. That's what real RDF does in the wild: enabling us to put metadata from multiple vocabularies (and even from multiple sources) in the same record (Figure 2).

Just as any collection of information in RDF can be merged, it can also be decomposed into the three-part statements we mentioned. The technical way of encoding information described according to different "cataloguing schemes" (or vocabularies, or ontologies) is with XML namespace. In the example above, we are merging three vocabularies

FIGURE 2. RDF Statements with Different Vocabularies (Dublin Core and Friend Of A Friend)



In this figure, the oval (○) represents an URI, a resource; The square (□) represents a value or literal, and the circle in the arrow (-○→) a property or association (a predicate).

(RDF's own basic vocabulary, Dublin Core, and FOAF), using three namespace declarations to define which terms in the description are associated with which vocabulary:

```
xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
xmlns:dc='http://purl.org/dc/elements/1.1/'
xmlns:foaf='http://xmlns.com/foaf/0.1/'
```

As an alternative to looking at pictures trying to represent the code for human consumption, we can list the statements that the code makes in a table. Thus, the individual statements represented in Figure 2 are shown in Table 2.

Those who have been following the code and the pictures in detail might have already noticed that we managed to have an **object** in the collection (a circle in the pictures) that is not labelled with a URI. That is a little modelling trick to be able to stack up lots of statements–similar to the way we use "which" or "that" in English, to avoid repeating the

TABLE 2. Definition of the Statements in Our Example (Properties in Triples)

Subject	***Predicate***	***Object***
http://www.bartleby.com/173	http://purl.org/dc/elements/1.1/title	"Relativity The Special and General Theory"
http://www.bartleby.com/173	http://purl.org/dc/elements/1.1/creator	"Albert Einstein"
http://www.bartleby.com/173	http://purl.org/dc/elements/1.1/language	"en"
http://www.bartleby.com/173	http://purl.org/dc/elements/1.1/date	"1920"
http://www.bartleby.com/173	http://purl.org/dc/elements/1.1/identifier	"http://www.bartleby.com/173"
http://www.bartleby.com/173	http://purl.org/dc/elements/1.1/creator	Something (let's call it X here)
The thing called X	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
The same X	http://xmlns.com/foaf/0.1/name	"Albert Einstein"
The same X	http://xmlns.com/foaf/0.1/birthdate	"14 March 1879"

name of something. Except here we can also use it to completely avoid naming something.

"The person whose name is Albert Einstein, and whose birthday is 14 March 1879" seems to be pretty clearly identifying something. But it has no URI of its own (in the picture, one oval shape–in green–is unlabelled to reflect this). On the other hand, we could compare a couple of structures and decide whether they mean the same thing or not. With the FOAF[10] vocabulary that we have used in this example, it is not assumed that having the same name and birthday are enough to say that two people are the same. But it does have a term for a mailbox (a personal e-mail address). In its definition the vocabulary does say, using some Semantic Web glue, that any two people who have the same mailbox are considered to be the same person. This gives us a simple rule for inferring more information. Any two sets of statements about people, where each person has the same e-mail address, are all about the same person.

So we know that we can gather up descriptions to build a collection of information. And we can use any kind of description that has been defined for the Semantic Web. So how are these terms defined? What is the process for getting new terms or whole new vocabularies?

Defining New Vocabulary (RDF Schema)

The Semantic Web uses URIs to identify terms, and to build descriptions. Moreover, several of the vocabularies built in the Semantic Web are about describing terms. So a complete definition of a term is usually made using the basic RDF vocabulary (sometimes people might use OWL, which is a vocabulary that provides for more detailed descriptions).[11] In our cards metaphor, if you want to define a new term, you make a new card for it and put it in the catalogue. Then, when you are looking through the catalogue and do not recognise a term, you simply look it up in the same catalogue. With RDF you can add new properties, new elements, or even new vocabularies as easily as it was done in the card catalogue era, but much easier to use, since the lookup is almost instantaneous.

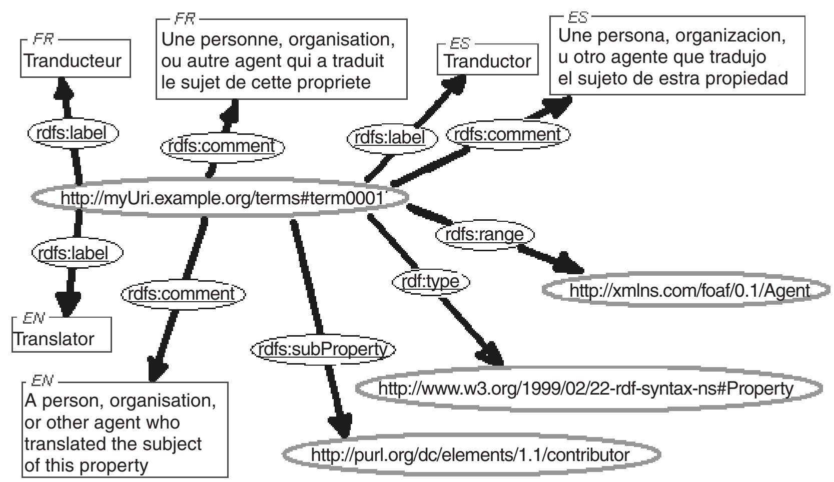
When you name your elements, they will have to be unique within the schema but they do not have to be globally unique. This is the job of the namespaces we mentioned in the section before. With RDF Vocabulary Language (RDF/S) we could create properties, subsets of properties and classes, and even "seeAlso" additional information like cross-references in the card boxes (Figure 3).

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<rdf:Description
rdf:about="http://myuri.example.org/terms#term0001">
 <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-
ns#Property"/>
 <rdfs:label xml:lang="fr">Traducteur</rdfs:label>
 <rdfs:label xml:lang="en">Translator</rdfs:label>
 <rdfs:label xml:lang="es">Traductor</rdfs:lbel>
 <rdfs:comment xml:lang="fr">Une personne, organisation, ou autre
 agent qui a traduit le sujet de cette propriete</rdfs:comment>
 <rdfs:comment xml:lang="en">A person, organisation, or other
 agent who translated the subject of this property</rdfs:comment>
 <rdfs:comment xml:lang="es">Una persona, organización, u otro
agente que tradujo el sujeto de esta propiedad</rdfs:comment>
 <rdfs:subProperty
rdf:resource="http://purl.org/dc/elements/1.1/contributor"/>
 <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Agent" />
</rdf:Description>
</rdf:RDF>
```

Here we have a resource, identified by the URI http://myUri.example.org/terms#term0001. We have said that it has the type of an RDF Property. Using the RDF Vocabulary Description Language (RDF/S, 2004) *label* and *comment* terms we have provided human-readable explanations of our property in three languages (Spanish, English, and French). This is important. In general discussion of tags we have often give them names based on the text inside the pointy brackets–the actual element name we use for the computer. But that is really just the last bit of a URI, a Web address, and may or may not be meaningful to a human reader.

In principle, it seems to make sense that we choose names we can remember, and it is not a bad idea. However, it is fairly clear that the names that are easy to remember in Arabic are not as easy to remember in Japanese, and vice versa. So these informal “tag names” are formally defined with real, per-language labels that people can read. At the same time, there is a URI that doesn’t change, because computers don’t speak human languages anyway and don’t make a stylistic difference between “real” words and random strings of text. In this way, we can share a URI across languages, and allow people to search for it in a real language.

FIGURE 3. Creating New Properties with RDF/S (*Translator* in Spanish, English and French)



In this figure, the oval (○) represents an URI, a resource; The square (□) represents a value or literal, and the circle in the arrow (-○→) a property or association (a predicate).

We have also declared that it is an RDF subProperty of the Dublin Core contributor property. That is, if something has a translator X, then it has a Dublin Core contributor X. While not all contributors are translators, all translators are contributors. We will see below a concrete example of what this means. Finally, we declared a range for the property. This is a claim that all objects of declarations that use the property have a given type–in this case they are FOAF Agents.

In other words, we have related the URI defined to some English, Spanish, and French text that can be used to present the information to people. We have also related it to other terms in other vocabularies. It allows us to go to the catalogue with some basic rules, and find not just the information we want, but the definitions of new things added to the cards, which we can relate in various ways to definitions that are already known.

RDF/S can be used to make statements defining and describing in a formal (machine-understandable) way, application-specific and existing vocabularies (for example, DCMES, Dublin Core Metadata Element Set)[12] or for creating new vocabularies or new elements of an existing one. Jul (2003) also recognizes this advantage of RDF from

MARC. Several different namespaces would be needed for a standard library bibliographic record or for a resource description that uses library rules and tools. This author also points out that things that may take several years to change in cataloguing rules or in MARC format (like introduce a new element, code, or rule), in RDF could take seconds, where the owner of the namespace might introduce pretty quickly as we did in our example.

Scribbling on the Cards

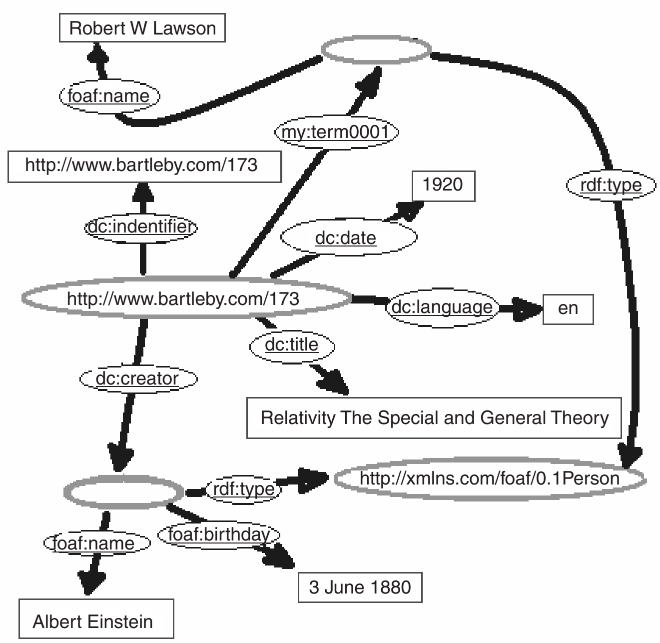
In library cards we could scribble any kind of information, about a book or a library material quickly. We can do the same in RDF, enriching more and more our description capabilities and more precise resources description. Let's include now a new term in our example. We can use our new term (my:term0001), just like any other:

```
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
 xmlns:dc='http://purl.org/dc/elements/1.1/'
 xmlns:foaf='http://xmlns.com/foaf/0.1/'
 xmlns:my='http://myuri.example.org/terms#'>
<rdf:Description rdf:about='http://www.bartleby.com/173'>
 <dc:creator>
 <foaf:Person>
 <foaf:name>Albert Einstein</foaf:name>
 <foaf:birtdate>14 March 1879</foaf:birthdate>
 </foaf:Person>
 </dc:creator>
 <my:term0001>
 <foaf:Person>
 <foaf:name>Robert W Lawson</foaf:name>
 </foaf:Person>
 </my:term0001>
 <dc:language>en</dc:language>
 <dc:title>Relativity
 The Special and General Theory</dc:title>
 <dc:date
rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1920</dc:dat
e>
 <dc:identifier>http://www.bartleby.com/173</dc:identifier>
</rdf:Description>
</rdf:RDF>
```

In library cards we could scribble any kind of information, about a book or a library material quickly. We can do the same in RDF, enriching more and more our description capabilities and more precise resources description. Let's include now a new term in our example. We can use our new term (my:term0001), just like any other (Figure 4).

As we noted above, our translator property is defined as a subProperty of the Dublin Core contributor property. This means that if we give this example to an RDF/S (Vocabulary Description Language) aware processor, it will make the inference that the thing we are describing has a Dublin Core contributor property whose value is a FOAF Person with the FOAF name "Robert W Lawson." This is useful when we want to search a large collection of records, especially if they have been encoded at different times, for different purposes, but with vocabularies that are related to others. In fact it is often possible to take several collections of RDF, describe some relations between properties, and do useful searching across the collections.

FIGURE 4. Graphic Representation of a New Term (my:term0001) in Our Example



In this figure, the oval (○) represents an URI, a resource; The square (□) represents a value or literal, and the circle in the arrow (-○→) a property or association (a predicate).

CONCLUSION

There are a number of additional RDF components that we have not addressed in this article, but which can be used to make it easier to tie the various vocabularies to each other. OWL[12] also allows for various constraints to be specified, such as the fact that a person should only have one birthday. In our definition of a translator above, we could have declared a domain for the property–that is, specified a type that all subjects have, if we could think of a suitable definition of that type (should it be a person? or an organisation? or some combination of those?). Or we can wait for someone else to make the relevant RDF (for example, if Dublin Core proposed a translator term that would be a subProperty of contributor in its RDF/S for Metadata Terms). When that property is described, they, we, or anyone, can declare the two as equivalent by using a term from OWL.

We have not discussed some of the deeper mechanics of how to use RDF in this work for practical reasons, but there is a wealth of literature and tools available already (RDF, 2006). The power of RDF comes in part from its simplicity, but it is quite easy to make nonsensical statements because the original definition of a term did not consider the size of the catalogue that is available.

W3C has developed a series of RDF standards over the last three years (RDF, 2006). RDF has been integrated into a variety of applications from library catalogues and directories to syndication of news and content to personal collections of music, photos, and events, using existing schemas (like DC)[13] or creating new ones. When plans of Universal Bibliographic Control (UBC) were first presented, the standards were not there to support full implementation. Despite the Paris Principles (1961) there was no international standard which could ensure the UBC, and later MARC format would get it. Today, RDF has the potential to be the backbone for the Semantic Web, and help us achieve "Universal Web Control."

It is important to remember that RDF is just a technology, and it is the Semantic Web that is a way of making the Web more useful. We have seen that we can all scribble on the cards in the RDF catalogue, but the main challenge remains the same–how do we ensure that as much useful information as possible gets on there, and as little unreliable or incorrect information is added? On the global scale of the Web, this is in fact a complex problem, but not just a technological one.

Librarianship, including cataloguing and classification, have evolved over the centuries, and continue to demonstrate innovative. This new dimension is based on information architecture, structured knowledge representation, and on the Semantic Web, where standards (old and new) are in continuous evolution. This continuous development means, on a technical level, standards will keep changing and librarians and cataloguers will keep adapting to this evolution. Against Mark Twain's idea: *I'm all for progress; it's change I don't like*, librarians have learned to adapt their professional competencies and their attitude to the Web and the Semantic Web requirements to achieve metadata interoperability and perhaps the UWC too.

NOTES

1. Some examples of this are:

- The European project in the late twentieth century called *Bibliotheca Universalis.* One of the pilot projects of the GIS (Global Information Society) arose from the 1995's G7 Activities about Information Society. In some way, *Bibliotheca Universalis* could be considered the basis for the current TEL (The European Library) project: http://www.theeuropeanlibrary.org.
- Traditional examples of digital global libraries, born with the vocation of being "universal." We are thinking on projects such us: Project Gutenberg <http://www.gutenberg.org> or Cervantes Virtual Library in Spain <http://www.cervantesvirtual.com>.
- The Universal Library from the University of Carnegie Mellon <http://tera-3.ul.cs.cmu.edu/>. The last news appeared in *Library Journal* in April 2006, which considers the OCA (Open Content Alliance) plan against Google's books digitisation project "The Birth of the Universal Library" (Bengtson 2006).

2. Wikipedia: http://www.wikipedia.org.
3. Flickr: http://www.flickr.com.
4. Del.icio.us: http://del.icio.us.
5. See MARC 21 XML Schema: http://www.loc.gov/standards/marcxml.
6. We must point out that we are using the Spanish version of the book and these Berners-Lee's remarks and reflections are in p. 171 in the cited version (Berners-Lee, 2000) and we made a free translation from it, which might not coincide with his original words.
7. We only need to check the last 4-5 years of *Cataloging & Classification Quarterly* <http://catalogingandclassificationquarterly.com/ccqissue.html> (ISSN: 0163-9374) and *Journal of Internet Cataloging* <http://www.haworthpressinc.com/store/product. asp?sku=J141> (ISSN: 1091-1367) to show several examples of all these tensions and scientific discussion.
8. For more information about RDA, please see JSC (2005-06), or http://www.rdaonline.org.

9. The W3C Public Access Specifications for W3C are understood as *de facto* standards. When they achieved the level of Recommendation they are ready to be applied for industry and software. RDF specifications are Recommendations since February 2004 (See: RDF, 2006).

10. FOAF (Friend Of A Friend), See FOAF Project site at: http://www.foaf-project.org.

11. OWL stands for Web Ontology Language and it is the formal language created by W3C Semantic Web activity for representing and encoding ontologies. OWL builds on RDF and RDF Schema and adds more vocabulary for describing properties and classes: among others, relations between classes (e.g., disjointness), cardinality (e.g., "exactly one"), equality, richer typing of properties, characteristics of properties (e.g., symmetry), and enumerated classes. To explain OWL goes further the objectives of this article. For more information about OWL, please see the official W3C site at: http://www.w3.org/2004/OWL.

12. DCMES (Dublin Core Metadata Element Set), i.e., the fifteen elements which are recognised as an ISO standard (ISO 15836), has its own RDF Schema defined in: http://dublincore.org/2003/03/24/dces# where all the elements and its semantics values are described as RDF properties. For example, the element "title" in RDF/S looks like this:

```
<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/title">

<rdfs:label xml:lang="en-US">Title</rdfs:label>

<rdfs:comment xml:lang="en-US">A name given to the
resource.</rdfs:comment>

<dc:description xml:lang="en-US">Typically, a Title will be a name
by which the resource is formally known.</dc:description>

<rdfs:isDefinedBy rdf:resource="http://purl.org/dc/elements/1.1/" />

<dcterms:issued>1999-07-02</dcterms:issued>

<dcterms:modified>2002-10-04</dcterms:modified>

<dc:type
rdf:resource="http://dublincore.org/usage/documents/principles/#el
ement" />

<dcterms:hasVersion
rdf:resource="http://dublincore.org/usage/terms/history/#title-
004" />

</rdf:Property>
```

13. To check some of these (more or less) well know schemas, see SchemaWeb at: http://www.schemaweb.info.

REFERENCES

Ahmed, Kal et al. (2001). *Professional XML Meta Data*. Birmingham: Wrox Press, 2001.

Bengtson, Jonathan B. (2006). The Birth of the Universal Library. *Library Journal*, April 15, 2006. Available online at: http://www.libraryjournal.com/article/CA6322017.html.

Berners Lee, Tim, James Hendler, Ora Lassila (2001). The Semantic Web. *Scientific American*, vol. 284, n. 5, p. 34-43. Available online at: http://www.sciahttp://www.haworthpress.com/Store/E-Text/ViewLibraryEText.asp?s=J104&m=0v=39m.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21.

RDF (2006). Ivan Herman, Ralph Swick and Dan Brickley, (eds.). World Wide Web Consortium, rev. 19 July 1006. Available online at: http://www.w3.org/RDF.

Tillett, Barbara B. (2003). AACR2 and Metadata: Library Opportunities in the Global Semantic Web. *Cataloging & Classification Quarterly*, vol. 36, n. 3/4, p.101-119.

Berners-Lee, Tim (2000). *Tejiendo la red: el inventor del World Wide Web nos descubre su origen*. Madrid: Siglo veintiuno, 2000. [Spanish version of Berners-Lee's book: *Weaving the Web*. San Francisco: Harper, 1999].

Brisson, Roger (1999). The World Discovers Cataloging: A Conceptual Introduction to Digital Libraries, Metadata and the Implications for Library Administrations. *Journal of Internet Cataloging*, vol. 1, n. 4, p. 3-30.

Campbell, Grant D. (2004). The metadata-bibliographic organization nexus. In *International Yearbook of Library and Information Management 2003-2004: Metadata applications and Management*. Maryland: Scarecrow, 2004, p. 185-203.

Daconta, Michael C., Leo J. Obrst, Kevin T. Smith (eds.) (2003). *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Indianapolis: Wiley, 2003.

Dovey, Matthew J. (1999) "Stuff" about "Stuff"–the Differing Meanings of "Metadata." *Vine (Theme issue: Metadata. Part 1)*, 1999, n. 116, p. 6-13.

Gradmann, Stefan (2005). rdfs:frbr–Towards an Implementation Model for Library Catalogs Using Semantic Web Technology. *Cataloging & Classification Quarterly*, 2005, vol. 39, n. 3/4, p. 63-75.

Hjelm, Johan (2001). *Creating the Semantic Web with RDF: Professional Developer's Guide*. New York, etc.: Wiley Computer Publishing, John Wiley & Sons.

Howarth, Lynne C. (2005). Metadata and Bibliographic Control: Soul-Mates or Two Solitudes? *Cataloging & Classification Quarterly*, 2005, vol. 40, n. 3/4, p. 37-55.

Jones, Wayne, Judith R. Ahronheim, and Josephine Crawford (eds.) (2002). *Cataloging the Web: Metadata, AACR, and MARC 21*. Lanham: Scarecrow Press, 2002.

JSC (2005-06). Joint Steering Committee for Revision of Anglo-American Cataloguing Rules. *RDA: Resource Description and Access*. Library and Archives Canada, Firs issued July 2005, rev. 3 august 2006. Available online at: http://www.collectionscanada.ca/jsc/docs/5rda-prospectusrev2.pdf.

Jul, Erik (2003). MARC and Mark up. *Cataloging & Classification Quarterly*, vol. 36, n. 3/4, p.141-153.

McCathieNevile, Charles (2005). RDF for XMLers. *Upgrade: The European Journal for Informatics Professional*, February 2005, vol. VI, n. 1, p. 39-44.

Méndez, Eva (2004). La Web semántica, una web 'más bibliotecaria.' *Boletín de la SEDIC*, 2004, n. 4. Available online at: http://www.sedic.es/p_boletinclip41_confirma.htm.

Méndez Rodríguez, Eva Mª (2005). Libraries and Information Systems Need XML/RDF . . . but Do They Know It? *Upgrade: The European Journal for Informatics Professional*, February 2005, vol. VI, n. 1, p. 51-56.

RDF (2004). World Wide Web Consortium. RDF/XML Syntax Specification (Revised): W3C Recommendation 10 February 2004. Dave Beckett, ed. W3C, 2004. Available online at: http://www.w3.org/TR/rdf-syntax-grammar.

RDF/S (2004). World Wide Web Consortium. RDF Vocabulary Description Language 1.0: RDF Schema: W3C Recommendation 10 February 2004. Dan Brickley and R. V. Guha, eds. W3C, 2004. Available online at: http://www.w3.org/TR/rdf-schema.

RDF (2006). Ivan Herman, Ralph Swick and Dan Brickley, (eds.). World Wide Web Consortium, rev. 19 July 1006. Available online at: http://www.w3.org/RDF.

Tillett, Barbara B. (2003). AACR2 and Metadata: Library Opportunities in the Global Semantic Web. *Cataloging & Classification Quarterly*, vol. 36, n. 3/4, p.101-119.