D. Grant Campbell. 2011. RDA and RDF: A discourse analysis of two standards of resource description. In Smiraglia, Richard P., ed. *Proceedings from North American Symposium on Knowledge Organization*, Vol. 3. Toronto, Canada, pp. 207-16.

D. Grant Campbell (gcampbel@uwo.ca)
University of Western Ontario, London, Ontario Canada

RDA and RDF: A Discourse Analysis of Two Standards of Resource Description

Abstract: The World Wide Web Consortium's Resource Description Framework (RDF) and the library community's new cataloguing standard, Resource Description and Access (RDA), both profess to provide sophisticated and flexible means of describing resources for modern Web environments. But both have attracted scepticism from potential users, who argue that their supposed innovations are overrated. A comparison of the two standards using Michel Foucault's theory of discourse formations suggests that while the two standards differ in their community contexts and their use of intermediaries, they are similar to each other in their commitment to consistent, rigorously-defined entities and relationships; this shared commitment sets them apart from Web 2.0 developments, and offers the potential for fruitful collaboration.

Acknowledgements
Funding for this research was provided by the Social Sciences and Humanities Research Council of Canada.

1. Introduction

This paper addresses the relationship between *Resource Description and Access (RDA)*, the latest standard for bibliographic description in libraries, and the Resource Description Framework (RDF), developed by the World Wide Web Consortium as part of its Semantic Web initiative. Using Michel Foucault's theory of discourse formation as a basis for comparison, and treating resource descriptions as discursive objects, I will argue that RDA and RDF differ substantially in their underlying assumptions about how resource descriptions emerge, and how they are authoritatively defined and accepted into information systems. Beneath these differences, however, lies an underlying consensus on how resources are specified and differentiated. I will argue that this shared commitment to rigorous, conscientiously-applied, machine-readable definitions and relationships sets both library catalogues and Semantic Web standards apart from Web 2.0 systems and tools, particularly those which are frequently (if questionably) classed as emergent systems.

2. RDF and the Semantic Web

Ten years ago, Tim Berners-Lee, who had changed the world by inventing the protocols of the World Wide Web, appeared poised to change it again. Frustrated by the limitations of the Web that he'd helped create, Berners-Lee argued eloquently and persuasively that our new world of static Web pages in HTML was only the beginning. To exploit the Web to its full extent, he argued, we needed to make fundamental enhancements to Web environments. We needed to make Web authoring easier: writing to the Web, he argued, should be like taking out a pen, rather than taking out the lawn mower (Berners-Lee & Fischetti 2000, 159). We needed to exploit the Deep Web: the fund of data that lay unused in structured databases; we needed to move beyond simple keyword searching with our search engines by introducing contextual information that would make hyperlinks, and the data they accessed, more meaningful (Berners-Lee & Fischetti 2000, 180). Above all, we needed to alter the relationship between human information processing and machine information processing, using the machine's speed and efficiency at calculation, together with its capacity to enact inferential logic, to enhance the human mind's ability to draw creative and intuitive conclusions from those calculations (Berners-Lee & Fischetti 2000, 177).

Changing this relationship would involve work: arduous and seemingly unrewarding work, at first. Before consigning important calculations to machines, you need to make data machine-understandable, and not merely machine-readable. Your machine won't draw accurate or effective inferences from data unless you've encoded a multitude of rules and relationships and connections that most of us process automatically. Ambiguities and inconsistencies that cause few problems to human minds can pose insurmountable problems for a Web agent. The Semantic Web, therefore, as Berners-Lee envisioned it, would rest upon a set of standard encoding layers:

- A semantic markup language, to enable the meaningful encoding of data;
- A standard means of describing Web resources: their origin, authorship, content and other bibliographic, technical and administrative metadata;
- A set of ontologies which could reconcile differences in the usage of terms, and establish fundamental relationships of hierarchy and equivalence.

The World Wide Web Consortium has been hard at work since the end of the twentieth century, bringing those standards about. Consortium members have developed and promoted XML and XHTML as the preferred markup languages for a Semantic Web; they maintain the Working Ontology Language (OWL) as the ontology standard; and they have developed both the Resource Description Framework (RDF) and the RDF-Schema as means of encoding metadata for Web resources. The standards are complicated, written as they are in the standard specification language that characterizes the documentation of computer languages and computer programs. Using them forces us to take implicit relationships and to make them explicit, and that process always involves unpacking seemingly simple connections and finding to our dismay that

they're complicated and unwieldy. The Semantic Web will require large amounts of RDF-encoded metadata, and creating that metadata will not be easy.

Nonetheless, Berners-Lee and his colleagues argued persuasively in a 2001 *Scientific American* article that the payoff would be considerable. They described scenarios of the future in which Web agents function as useful and working collaborators in the business of life: prompting the user about urgent appointments, suggesting courses of action at moments of decision-making, performing useful background duties automatically, and making the user aware of useful information in the right place and at the right time (Berners-Lee, Hendler & Lassila 2001). For Berners-Lee and other apologists for the Semantic Web, the Web of the future would be embedded into users' lives: the machine would extend beyond information retrieval and presentation, and integrate itself into the holistic area of information use.

That was the plan at the turn of the century. But something strange happened. Many of Berners-Lee's predictions came true; but they came true without the Semantic Web. The advent of multimedia capabilities within devices of widely varying size and portability, together with the rise of social networking sites and the explosion of e-commerce, have indeed made computing essential to our daily lives, and have indeed turned the Web-enabled computer from a lawn mower into a pen. But these things have happened without resting on a base of RDF- and OWL-encoded metadata. While the Semantic Web standards continue to exist, and while they have proven useful in various contexts, they have not succeeded in bringing about a revolution in Web design.

Seely Brown and Duguid warn us about the perils of prediction: all too often, they suggest, we gleefully anticipate the technical advances with considerable accuracy, but embed them in anachronisms: social and cultural assumptions that lose their cachet as the years pass. It is easier, they argue, to imagine the advent of the jet pack than to predict the advent of feminism or the civil rights movement (31-2). The World Wide Web Consortium accurately predicted that a change was coming. But predicting the arrival of database harvesters was easier than predicting the arrival of Facebook and Twitter.

3. RDA and Cataloguing

While Berners-Lee and the World Wide Web Consortium were positioning themselves for a new revolution, the library cataloguing community was grappling with a far less publicized, and far less glamorous problem: the revision of the *Anglo-American Cataloguing Rules, Second Edition*, the standard for bibliographic description that had prevailed in Europe and North America since 1978. *AACR2R* had proved, and continues to prove, remarkably durable as a descriptive standard, thanks largely to three strengths:

- The distinction between the bibliographic unit and the work, while hardly airtight, provided a highly useful way of simultaneously satisfying the catalogue's objectives of identifying specific editions and collocating works and editions;
- *AACR2R* was conceived as media-neutral: however firm its implicit bias towards monograph books, it was explicitly designed to describe resources of any material or bibliographic type;
- The standard has been regularly and thoroughly revised, to accommodate new materials, such as computer files and remote-access electronic resources.

Despite this durability, it was clear by the mid-1990s that things could not continue this way forever. Web resources were proving to be more than merely a new material type to be embraced. Web environments were altering the nature of other older forms such as the monograph and the serial. They were exhibiting new patterns of authorship and publication.

Above all, Web environments were changing the nature of the library catalogue and its relationship to other search tools, such as databases, search engines and metadata repositories. At first, it was a matter of increased access within the library community: the Z39.50 protocol made it possible to search multiple library catalogues simultaneously. Then, we developed portals that enabled users to search catalogues, search engines, databases and metadata repositories through a common interface, forcing the catalogue to exist alongside tools of different structure and design. And now, as more and more information on the Web is constructed on the fly through dynamic Websites that harvest data from multiple sources, library catalogues will be increasingly isolated, unless their data can be encoded in ways that are congenial to other forms of encoding in use elsewhere. Pressure has been building, therefore, to create bibliographic descriptions that can easily translate into metadata standards and encoding that emerged outside, or partially outside, the library community, such as the Dublin Core and RDF.

In order to meet these challenges, the Joint Steering Committee for the Revision of AACR2 embarked on the creation of an ambitious new standard: one that, after months and years of delay, finally emerged last June as *Resource Description and Access (RDA)*. While the standard continues to support the creation of bibliographic records very like those created by *AACR2*, it has followed a path similar to that of the World Wide Web Consortium. It has attempted to distinguish the various intellectual acts of description from the mechanical act of encoding, and to make distinctions that were formerly available only to human inference available to machine inference as well. To that end, RDA has made the following changes:

First, it is no longer securely embedded in the structure of the International Standard Bibliographic Description (ISBD) and its accompanying punctuation conventions. The ISBD is presented in the Appendices as an optional syntax for the bibliographic record, but the

cataloguing rules themselves no longer adhere strictly to the ISBD areas of description; nor do they mention the familiar ISBD punctuation between data elements.

Second, the standard is structured upon the FRBR paradigm established by IFLA's Working Group on the Functional Requirements of Bibliographic Records, which was in turn founded on exhaustive work modeling *AACR2R* in entity-relationship format to establish, in database terms, the fundamental data structures. This paradigm establishes four primary entities, linked in a series of one-to-many relationships: a Work is an abstract artistic or intellectual construct, which has one or more Expressions in linguistic or symbolic notation; each of these Expressions has one or more material Manifestations, each of which can be duplicated into one or more physical Items. By using this structure, RDA is able to provide rules for encoding within the bibliographic record a set of primary relationships between work, expression, manifestation and item.

4. Change or No Change?

Both RDA and RDF have had their share of detractors. While criticism has focused on many different aspects of the two standards—unnecessary complexity, ambiguity, obscurity, irrelevance—in most cases the scepticism cases has a disillusioned feel to it. For Clay Shirky, RDF is a way of making syllogisms:

> The Semantic Web is a machine for creating syllogisms. … [It] specifies ways of exposing these kinds of assertions on the Web, so that third parties can combine them to discover things that are true but not specified directly. This is the promise of the Semantic Web -- it will improve all the areas of your life where you currently use syllogisms. Which is to say, almost nowhere. (2003)

RDA, on the other hand, was ironically hailed by Karen Coyle and Diane Hillman as "Cataloguing Rules for the 20th Century." The JSC's ambitious goal to reinvent cataloguing for electronic environments, they claimed, was ultimately defeated by the need to create backward compatibility for *AACR2* records, together with a lack of community support (2007).

For Shirky and for Coyle and Hillman, the revolutionary pretensions of RDF and RDA are suspect. What professes to be a radical change turns out to miss the key point: that the Web is stranger and more radical than originally envisioned, and that you can't do justice to it while preserving your legacy systems.

We have, then, two efforts to create standards of information resource description. Both profess to be significant changes. Both constitute responses to the new realities of the World Wide Web. Both have attracted scepticism on that score. To what extent do they embody a similar response to Web environments? And how significant are they in today's Web culture?

These are not idle questions, because they go to the heart of one of the most vexing problems in information environments, which are rife with the rhetoric of brave new worlds, populated with brilliant new technologies. Theorists and practitioners alike must navigate between utopian and dystopian predictions of the future, on the one hand, and cynical assurances on the other that the more things change, the more they stay the same. As Adam Gopnik pointed out in a recent essay in *The New Yorker*, the world of Internet prediction is divided between the "Never-Betters," the "Better-Nevers," and the "Ever-wases." (2011) Specialists in information organization—both theoretically- or practically-oriented—must continually develop their vocabulary for understanding current and future changes in information systems at both the technological and the conceptual level. How do we evaluate information systems, when some voices claim that they are changing the world, while others claim that they are changing nothing at all?

5. Foucault and Discursive Formations

The remainder of this paper attempts to answer these questions by treating both standards—RDA and RDF—as discursive constructions in the field of knowledge organization. Neither standard professes to be merely a change in technology, although technological change is certainly a factor. RDF seeks to create a common means of doing what most people do anyway in various ways: making statements about Web resources. By drawing on pre-existing standards like XML and the Dublin Core, its designers hope to enhance the variety of ways that such statements can be used and reused. RDA seeks to create a new procedure for describing and providing access to information resources. By distinguishing the cataloguing standard from the ISBD and the MARC standard, its designers hope to enhance the number of contexts in which bibliographic descriptions can be used and reused. Both standards rely on a conceptual framework: the RDF triple of subject, predicate and object, and the FRBR paradigm that underlies RDA.

The work of Michel Foucault—particularly *The Archaeology of Knowledge* (1969) —provides a promising theoretical context for assessing these changes. To begin with, Foucault's comparison of history with the history of ideas raises contrasting images of change which resemble the vexed visions of change confronting information studies. Historians, he argues, have traditionally seen history as a set of linked events that move in slow and stable succession (3), whereas the history of ideas has turned to discontinuities and interruptions in these long processes (4). In his attempts to articulate the nature of profound change in such confusing conditions, Foucault isolates three key conditions which govern the formation of discursive objects:

- Surfaces of emergence: in what social, institutional, administrative conditions are such discursive objects identified and analysed?
- Authorities of delimitation: what individuals or corporate bodies possess the authority to designate and define these discursive objects?

- Grids of specification: how are these objects defined, classified, grouped into relationships, and differentiated from each other? (41-42)

When we analyze each of these conditions for both RDA and RDF, we might some intriguing differences and still more intriguing similarities.

6. Surfaces of Emergence

The data structures of RDA originate from a set of specific administrative and social conditions. RDA conforms to the International Statement on Cataloguing Principles of the International Federation of Library Associations (IFLA) (RDA 0.4). It uses IFLA's Report on the Functional Requirements of Bibliographic Records as its paradigm. And its stated purpose is to enable users to find, identify, select and obtain resources that they need, and to find, identify, clarify and understand key entities such as persons, corporate bodies and concepts (0.0). RDA, then, emerges from an international professional community of librarians. It rests on the conceptual foundations of the *Anglo-American Cataloguing Rules* (0.2), and its ties to IFLA place it within the overall objective of Universal Bibliographic Control, in which libraries "make bibliographic data on all publications issued in all countries universally and promptly available in a form that is internationally acceptable" (Chan 1994, 44).

RDF emerges from a different community: from the community of computer programmers and Web designers who are committed to the formation and maintenance of universal standards of computing and data interchange. The Framework is expressed through a variety of document genres that are typically used in the Web standards community: specifications of RDF Syntax and RDF Semantics, and a Primer for explaining the concepts to users of the standard. These standards resemble RDA less than they resemble MARC manuals: they are less concerned with guiding humans in their interpretive decisions than they are with ensuring that these human decisions are accurately encoded. RDF emerges, not as a parallel or alternative to library resource description, but rather as a parallel and alternative to MARC coding: the ability to transfer information between machines without loss of meaning:

> RDF is intended for situations in which this information needs to be processed by applications, rather than being only displayed to people. RDF provides a common framework for expressing this information so it can be exchanged between applications without loss of meaning. Since it is a common framework, application designers can leverage the availability of common RDF parsers and processing tools. (Manola & Miller 2004).

The two standards, therefore, emerge from very different communities: from the library community on the one hand, and from the Web standards community on the other. Thus, one

standard speaks to library cataloguers making interpretive decisions; the other speaks to Web designers and computer programmers who are making encoding decisions.

7. Authorities of Delimitation

RDA and RDF share a common purpose: to make it easier for people to find resources they want, in part by taking the time and effort to resolve ambiguities inherent in a messy world of information. RDA, however, addresses this challenge by a set of rules that enable the intermediary, in the form of the cataloguer, to recognize and clarify important relationships in the bibliographic universe: relationships that the end user may not immediately recognize, but which could become necessary. Such concerns include clarifying the relationship between different entities and the primary relationships between works, expressions, manifestations and items. In such cases, the library, or other official information intermediary, takes upon itself the task of differentiating between two authors with the same name, between the original and revised editions of an expression, or between the English and the French translations of a work.

With RDF, the power to define entities and discriminate between them is, to some degree, delegated to specific user communities, which are represented on the Semantic Web through namespaces. Rather than dictate a specific vocabulary, RDF enables specific user communities to specify their own vocabularies, both for entities and attributes. Furthermore, the data can theoretically be used by a multitude of communities, for varying purposes: "The ability to exchange information between different applications means that the information may be made available to applications other than those for which it was originally created" (Manola & Miller 2004). The RDF triple, therefore, can fit an indefinite variety of uses, as defined by an indefinite number of user communities.

8. Grids of Specification

RDF and RDA, then, part company on the first two conditions that Foucault establishes. When we get to grids of specification, however—how the two standards define, classify and group concepts—the wide differences pale beside an underlying similarity. To be sure, RDF provides support for a wide range of different ontologies and descriptive standards, and in so doing fosters a diversity that looks very different from the uniformity produced by a cataloguing standard like RDA. When we compare them both to developments in the world of the Web 2.0, however, a more important contrast emerges.

The Web 2.0 represents, in many ways, a simplified version of the Semantic Web; indeed, the co-authors of a 2007 article suggested that phenomena such as user tagging were steps forward on the way to the implementation of Semantic Web technologies on a wide scale (Feigenbaum,

et al. 2007). Predictions for the future of the Web 2.0, however, suggest that it is evolving in a very different way.

Tim O'Reilly and John Battelle argued at a recent Web 2.0 conference that the Web 2.0 is turning into the "Web Squared": a new dimension of Web use that works on the notion of the information shadow: "everything and everyone in the world casts an "information shadow," an aura of data which, when captured and processed intelligently, offers extraordinary opportunity and mind bending implications" (2009). Processing and utilizing this information shadow will involve processes that systems learn: processes very different from those that intermediaries "teach" to their systems. Speech, voice and image recognition programs will evolve to enable systems to recognize entities, not through their unique identifiers, but by assembling multitudes of data from different sources and coherently reconciling it and deriving from it:

> A bottle of wine on your supermarket shelf (or any other object) needn't have an RFID tag to join the "Internet of Things," it simply needs you to take a picture of its label. Your mobile phone, image recognition, search, and the sentient web will do the rest. We don't have to wait until each item in the supermarket has a unique machine-readable ID. Instead, we can make do with bar codes, tags on photos, and other "hacks" that are simply ways of brute-forcing identity out of reality. (O'Reilly & Batelle 2009)

Batelle and O'Reilly are quick to point out that some of this "hodgepodge of sensor data" will include rigorously-defined, consistent and reliable data such as GPS coordinates, while others are vaguely-defined embodiments of an information "shadow."

If we compare these projected developments of the "Web Squared" to the grids of specification enabled by RDA and RDF, the differences between RDA and RDF pale beside an overwhelming similarity. Both standards stand in opposition to the hodgepodge approach, and rest firmly in the camp of those communities that value, and depend on, rigorously-defined and conscientiously-delineated entities. Both standards are committed to teaching processes rather than learning processes: preserving data that systems will take as authoritative and accurate.

9. Conclusion

It would be naïve, therefore, to suggest that the Semantic Web will revolutionize the world as did the first incarnation of the World Wide Web. But it would be equally naïve to write it off as a failure simply because it is unlikely to become ubiquitous. The visions of an embedded Web infrastructure that Berners-Lee expressed in 2000 have largely come true, but without the defined Semantic Web infrastructure. Nonetheless, the shared commitment of standards like RDA and RDF to precise and rigorous information description suggests that libraries and the World Wide

Web Consortium have much to offer each other. If the Web of the future is to rest in part on precise and reliable data, then libraries may well find the Semantic Web to be the most congenial context in which to preserve their cataloguing practices in the twenty-first century.

References

Berners-Lee, Tim, Hendler, James, and Lassila, Ora. 2001. The semantic web. *Scientific American*. May 17. Available http://www.scientificamerican.com/article.cfm?id=the-semantic-web.

Berners-Lee, Tim, with Fischetti, Mark. 2000. *Weaving the Web: the original design and ultimate destiny of the World Wide Web.* New York: HarperBusiness.

Brown, John Seely, and Duguid, Paul. 2000. *The social life of information*. Boston: Harvard Business Press.

Chan, Lois Mai. 1994. *Cataloging and classification: an introduction*. 2nd ed. New York: McGraw-Hill.

Coyle, Karen, and Hillman, Diane. 2007. Resource Description and Access (RDA): cataloguing rules for the 20th century. *DLib magazine* 13.1/2: January/February. Available http://www.dlib.org/dlib/january07/coyle/01coyle.html.

Feigenbaum, Lee, Tonya Hongsermeier, Eric Neumann & Susie Stephens. 2007. The Semantic Web in action. *Scientific American* November 19.

Foucault, Michel. 1971. *The archaeology of knowledge and the discourse on language*. Trans. A.M. Seridan Smith. New York: Pantheon.

Gopnik, Adam. 2011. The information: how the Internet gets inside us. *New Yorker* February 14. Available http://www.newyorker.com/arts/critics/atlarge/2011/02/14/110214crat_atlarge_gopnik

Joint Steering Committee for Development of RDA. 2010. *RDA: resource description and access*. American Library Association.

Manola, Frank, and Miller, Eric, ed. 2004. *RDF Primer*. World Wide Web Consortium. Available http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

O'Reilly, Tim, asnd Battelle, John. 2009. *Web squared: Web 2.0 five years on*. A white paper produced for the 2009 Web 2.0 Summit. O'Reilly Media. Available http://assets.en.oreilly.com/1/event/28/web2009_websquared-whitepaper.pdf

Shirky, Clay. 2003. The Semantic Web, syllogism and worldview. *Clay Shirky's writings about the Internet*. November 7. Available http://www.shirky.com/writings/semantic_syllogism.html