rdfs:frbr– Towards an Implementation Model for Library Catalogs Using Semantic Web Technology

Stefan Gradmann

SUMMARY. The paper sets out from a few basic observations (bibliographic information is still mostly part of the 'hidden Web,' library automation methods still have a low WWW-transparency, and take-up of FRBR has been rather slow) and continues taking a closer look at Semantic Web technology components. This results in a proposal for implementing FRBR as RDF-Schema and of RDF-based library catalogues built on such an approach. The contribution concludes with a discussion of selected strategic benefits resulting from such an approach.

Stefan Gradmann, PhD, is Head, Hamburg University "Virtual Campus Library" Unit, which is part of the computing center and has a mission of providing information management services to the university as a whole, including e-publication services and open access to electronic scientific information resources.

Address correspondence to: Stefan Gradmann, Virtuelle Campusbibliothek Regionales Rechenzentrum der Universität Hamburg, Schlüterstrasse 70, D-20146 Hamburg, Germany (E-mail: stefan.gradmann@rrz.uni-hamburg.de).

[Haworth co-indexing entry note]: "rdfs:frbr–Towards an Implementation Model for Library Catalogs Using Semantic Web Technology." Gradmann, Stefan. Co-published simultaneously in *Cataloging & Classification Quarterly* (The Haworth Information Press, an imprint of The Haworth Press, Inc.) Vol. 39, No. 3/4, 2005, pp. 63-75; and: *Functional Requirements for Bibliographic Records (FRBR): Hype or Cure-All?* (ed: Patrick Le Boeuf) The Haworth Information Press, an imprint of The Haworth Press, Inc., 2005, pp. 63-75. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: docdelivery@haworthpress.com].

http://www.haworthpress.com/web/CCQ

Digital Object Identifier: 10.1300/J104v39n03_05

KEYWORDS. FRBR, Semantic Web, ontologies, RDF-Schema, library automation, deep Web, hidden Web

CONTEXT AND MOTIVATION: WAYS OUT OF THE GOLDEN CAGE . . .

The following contribution was initially motivated by four observations, some of which I dealt with in more detail in an earlier publication (Gradmann 2003) and which originally led me to suggesting Semantic Web technology and the conceptual framework of FRBR as two promising areas for making librarian and generic WWW information services converge or even prepare some sort of integration scenario.

The first of the initial observations that motivate a closer simultaneous look at FRBR and Semantic Web technology is the fact that bibliographic information originated by libraries 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 and thus 'polluting' search engine results with massive amounts of redundant information.[1]

However–and as a result of this–born-digital resources, once they are integrated in library catalogues as part of hybrid library settings, needlessly risk sharing the fate of printed resources and thus of being buried in the 'deep Web' with the rest of bibliographic records.

The second observation (closely related to the first one) is that some aspects of librarian data models together with their technical implementation in most library automation systems have very little potential for WWW-transparency, and this specifically applies to complex entities involving record-linking techniques such as multi-volume and/or continuous publications. This probably is mostly an implementation issue, since heavily linked information structures are a common thing in the WWW–but the way most library automation environments implement these linking structures internally is rather tedious to translate to generic linking concepts in the WWW. This observation probably is valid in a more general sense regarding most library automation applications and the data architecture underlying these, which strangle libraries, creating a structural lack of technical flexibility that adds a lot to the paralyzing potential of the quantitative aspects observed later in this article.

The third observation is that the sheer amount of data that would probably present major problems when migrating to more generic technical environ-

ments prevents most librarians from seriously considering technical and functional alternatives to the current situation. This has led to a rather ridiculous situation (at least speaking from a German context), where relatively insignificant alternatives–such as the potential use of AACR instead of our German RAK-rules–are fervently discussed instead of seriously considering structural alternatives.

These three observations may explain, to some degree at least, why libraries until now have been so reluctant to seriously consider FRBR as a basis for new librarian information architectures–yet still, I believe that they are not sufficient to explain the relatively slow take-up of FRBR (even though the first brave early adopters are now starting to enter the playing field).

I suppose that, as a fourth observation, there is another hidden and mostly even pre-conscious reflection creating a major and mostly unrecognized barrier for FRBR adoption: the awareness that it is vain to attempt to implement FRBR in the context of existing catalogue data and applications (even though FRBR was conceived with a very traditional entity-relationship model in mind!) without using standard Internet technology at the same time, and thus, just creating another–just slightly more futuristic–librarian ivory tower.

On the other hand, patrons need to use such catalogue data (although OPAC use may tend to decrease) and they need stable, running operations: almost no chance, thus, to suspend operations in order to create a new fundament.

Librarian, thus, may start feeling more than slightly uneasy in their digital librarian cage but are not sure which direction to take–and at the same time the WWW continues to ignore librarian services and instead keeps inventing functional models, and all too often bypasses library services.

However, a whole wealth of information (mainly controlled vocabulary applied to information objects) is buried in library catalogues, and could be very beneficial in the tedious business of building ontology resources. This led me to consider, in the earlier article already mentioned, that:

> Semantic Web technology [. . .] and methods based on Semantic Web ontologies more specifically are likely to make new and productive use of the fine-grained semantic metadata which libraries traditionally have been producing. These could be used for enhancing the taxonomies of Semantic Web ontologies. Assertions based on the use of classifications and indexing schemes could easily be transposed into taxonomy elements that in turn greatly broaden the basis inference rules can be applied to. This results in a much richer taxonomic base for ontological operations and could well generate an ongoing process of library work being fed into Semantic Web ontologies. (Gradmann 2003, 38-39)

And, in the same article, I proposed having a closer look at FRBR as a means to overcome the structural incompatibilities that are the fundamental barrier to cross when attempting to free librarian bibliographic data from its golden catalogue-cage and make it systematically available on the WWW.

However, by that time I suggested a simultaneous look at both areas without actually blending both aspects: FRBR and Semantic Web technology then seemed equally interesting but essentially distinct domains in which librarian and WWW information services could be made to interact productively within innovative paradigms of information modelling.

The invitation to write a contribution for this volume suggesting I should take up some arguments from my earlier work made me think again: in this contribution I, therefore, am going to propose a much more integrated perspective on FRBR and Semantic Web technology, but before doing so, it may be appropriate to explain–at least tentatively–what 'Semantic Web Technology' is all about!

. . . INTO THE SEMANTIC WEB

Some have it that the 'Semantic Web' is impossible to define in simple terms, because beneath the ever-changing semantics of this buzzword linger the old dreams and illusions of Artificial Intelligence (AI) that are simply given a new disguise.

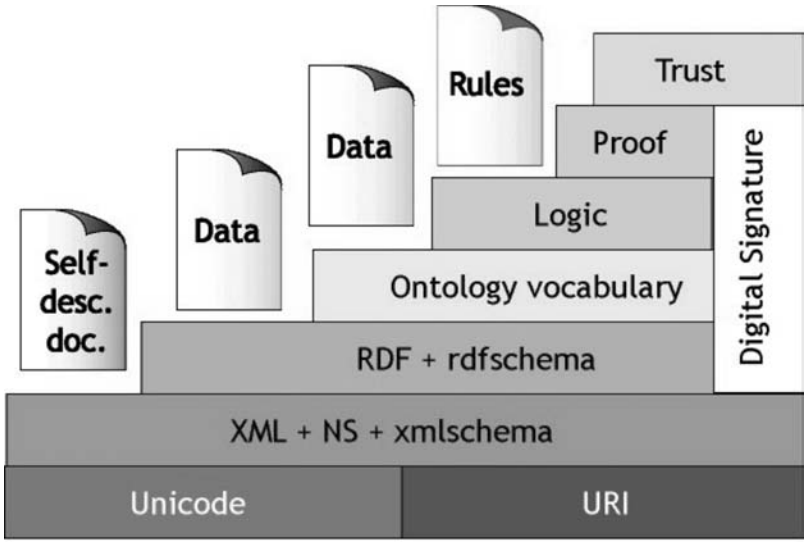
While there is some truth in this assertion (and introductions such as the excellent volume by Davies et al. (2003) certainly do not deny it), Semantic Web technology seems to have learned some lessons from the rise and (temporary) decline of AI in that it combines the visionary attitude and goals of AI (which all too often have been close to a true hype) with the robust pragmatism that is at the very roots of the WWW.

A useful definition of 'Semantic Web' is given in (Palmer 2001):

> The Semantic Web is a mesh of information linked up in such a way as to be easily processable by machines, on a global scale. You can think of it as being an efficient way of representing data on the World Wide Web, or as a globally linked database.

The Semantic Web can thus be thought of as a technological infrastructure on top of the http transport layer, which implements syntactical constructs in such a way as to enable operations on the semantics of WWW-resources. This infrastructure is clearly layered as illustrated by the very inventor of the term, Tim Berners-Lee (Berners-Lee 2000) (see Figure 1).

FIGURE 1. Tim Berners-Lee's Vision of the Semantic Web



Thus, the picture behind both the architecture and the development sequence proposed by Berners-Lee for building such an infrastructure is a layered approach with each layer acting as a fundament for the following one. The two most fundamental layers of identification/encoding and of XML/XMLschema today are increasingly stable, with standards endorsed by W3C and more and more operational applications being available.

The next two layers create the actual basis for scientific activity on the WWW truly using WWW technology and not just simply using the Web as a transport layer for traditional information objects: the two layers of RDF/rdfschema-based syntactic modelling and of ontology vocabulary building currently are the ones that receive the most attention in terms of development activity–and we will return to them since they also have tremendous potential concerning the main topic of the present contribution.

The top layers of 'logic,' 'proof,' and 'trust' remain very abstract and academic for the time being and probably will only be tackled seriously and massively once a stable basis is established, made up of sufficiently comprehensive rdfschema syntactic constructs and sufficiently rich ontologies.

Coming back to the RDF and ontology levels, I am now quoting from Davies et al. (2003) (which is a good in-depth introduction to this specific area!) in order to first give a clearer idea of what RDF actually is:

> The resource description framework (RDF) is a recent W3C recommendation designed to standardize the definition and use of meta-data descriptions of web-based resources. However, RDF is equally well suited to representing data.
>
> The basic building block in RDF is an object-attribute-value triple, commonly written as A(O,V). That is, an object O has an attribute A with value V. (Davies et al. 2003, 13)

Palmer (2001) gives a more concrete idea of what this actually means:

> A triple can simply be described as three URIs. A language which utilises three URIs in such a way is called RDF [. . .].
>
> Once information is in RDF form, it becomes easy to process it, since RDF is a generic format, which already has many parsers. [. . .] Let's take a quick look at an example of XML RDF right now:

```
<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/0.1/foaf/" >
    <rdf:Description rdf:about="">
        <dc:creator rdf:parseType="Resource">
            <foaf:name>Sean B. Palmer</foaf:name>
        </dc:creator>
        <dc:title>The Semantic Web: An Introduction</dc:title>
    </rdf:Description>
</rdf:RDF>
```

> This piece of RDF basically says that this article has the title "The Semantic Web: An Introduction," and was written by someone whose name is "Sean B. Palmer." Here are the triples that this RDF produces:

```
<> <http://purl.org/dc/elements/1.1/creator> _:x0 .
this <http://purl.org/dc/elements/1.1/title> "The Semantic Web: An
Introduction" .
_:x0 <http://xmlns.com/0.1/foaf/name> "Sean B. Palmer" .
```

RDF schema, in turn, is built on top of RDF and

> . . . takes a step further into a richer representation formalism and introduces basic ontological modelling primitives into the Web. With RDFS, we can talk about classes, subclasses, subproperties, domain and range

> descriptions of properties, and so forth in a Web-based context. (Davies et al. 2003, 14-15)

And a further extension of this formalization method, DAML+OIL delivers yet more in-depth properties and classes as well as–most importantly–some simple terms for creating inferences.[2] In the meantime, DAML+OIL has developed into OWL, the Web Ontology Language, which has been an official W3C recommendation since February 2004.[3]

RDFS and DAML+OIL/OWL, in turn, are the fundamental tools for building Semantic Web ontologies.[4] A sufficiently precise definition of an ontology in the AI-Semantic Web use of the term is the following one:

> A specification of a representational vocabulary for a shared domain of discourse–definitions of classes, relations, functions, and other objects–is called an ontology. (Gruber 1993)

And probably, from a librarian perspective, ontologies are the first thing that comes to one's mind when thinking about possible fields of convergence. Intuitively one might feel that there is a lot of similarity comparing Semantic Web ontologies and librarian thesauri or other techniques of controlled vocabulary and classification.

But even though there is some truth in such an assertion (but which would have to be discussed at length in order to avoid misunderstandings based on over-simplification!), this will not be the main direction to follow when now making a more radical proposal for further blending Semantic Web and librarian information methodology and technology.

HOW TO GET THERE: A PROPOSAL!

The proposal I actually wish to make seems somewhat in the air: both the Medlane-XOBIS project[5] and LibDB[6] already aim at combining librarian data structures and XML-based technology. And the developer of LibDB, one of the few platforms that are largely inspired by FRBR concepts, even mentions the prospect of "going 'all RDF'" at some point in the introduction to the LibDB database schema (Iff 2003)–still, to my knowledge, no one to date has actually taken up the original suggestion T. Berners-Lee made in the presentation quoted earlier, the last slide of which is devoted to the "Killer App for the Sweb." Not surprisingly, the first item listed there as an early adopters' community are ontology implementors. However, the second item on the list is a

surprise to quite some extent: “Catalogs on the Web” is suggested there as the second potential killer application for Semantic Web technology–and the proposal I wish to make is to take up this suggestion.

The proposal is to rethink the technical platforms for librarian metadata implementation in terms of Semantic Web technology and to do so using FRBR as a kind of pivot concept. In that sense, the proposal is not to view FRBR as a kind of ontology to be expressed in RDF, but rather to consider it a kind of specific meta-ontology in the field of librarian information objects, which would have to be expressed using RDFschema (or OWL) as a consequence and which, in turn, would be a suitable basis for catalogue implementation using RDF.[7]

The following RDF fragment[8] outlines the kinds of classes and properties that would have to be modelled in such an approach in order to model FRBR group 1 entities with some selected attributes:

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE rdf:RDF [
    <!ENTITY a 'file:/F:/apps/Kaon/ontologies/frbr#'>
    <!ENTITY kaon 'http://kaon.semanticweb.org/2001/11/kaon-lexical#'>
    <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
    <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>]>

<rdfs:Class rdf:ID="work">
    <rdfs:label xml:lang="en">work</rdfs:label>
</rdfs:Class>
<rdfs:Class rdf:ID="expression">
    <rdfs:label xml:lang="en">expression</rdfs:label>
    <rdfs:subClassOf rdf:resource="#work"/>
</rdfs:Class>
<rdfs:Class rdf:ID="manifestation">
    <rdfs:label xml:lang="en">manifestation</rdfs:label>
    <rdfs:subClassOf rdf:resource="#expression"/>
</rdfs:Class>
<rdfs:Class rdf:ID="item">
    <rdfs:label xml:lang="en">item</rdfs:label>
    <rdfs:subClassOf rdf:resource="#manifestation"/>
</rdfs:Class>
<rdf:Property rdf:ID="language">
    <rdfs:label xml:lang="en">language</rdfs:label>
    <rdfs:domain rdf:resource="#expression"/>
</rdf:Property>
<rdf:Property rdf:ID="title">
```

```
    <rdfs:label xml:lang="en">title</rdfs:label>
    <rdfs:domain rdf:resource="#work"/>
    <rdfs:domain rdf:resource="#expression"/>
    <rdfs:domain rdf:resource="#manifestation"/>
    <rdfs:range rdf:resource="#expression"/>
    <rdfs:range rdf:resource="#manifestation"/>
</rdf:Property>
<rdf:Property rdf:ID="edition">
    <rdfs:label xml:lang="en">edition</rdfs:label>
    <rdfs:domain rdf:resource="#manifestation"/>
</rdf:Property>

</rdf:RDF>
```

And a graph visualizing a more complete model covering all three entity groups might look as on Figure 2.

These examples are by no means proposed as first steps of the serious work to be done–they should simply illustrate the kind of work that would have to be done if such a proposal was adopted.

Expressing FRBR in an RDFS model would then allow for implementing catalogues using RDF and for integrating Semantic Web ontologies in such a framework in various fields.

I will not work out this proposal in detail here but simply wish to conclude pointing out some of the enormous benefits that could result from such an approach.

. . . AND THE BENEFITS OF DOING SO

In briefly discussing the benefits of the proposed approach I will not list these exhaustively but rather concentrate on a few prominent examples.

A. Most evidently, an rdfs:frbr based implementation model for catalogues on the Web would effectively solve the problem of catalogue records being buried in the ‘hidden Web’ and would make all objects, instances, attributes, and relations of information objects modelled in catalogues WWW-transparent.
B. In doing so, and intelligently making use of the FRBR layer-model, it would achieve this goal of WWW transparency without automatically drowning the Internet with heavily redundant cataloguing elements, since, on such a basis, layered integration scenarios can be conceived,

exposing (for instance) only 'work' (and maybe 'expression') elements to the WWW, while offering links for anyone who would wish to drill further down to the manifestation/item levels.

C. Inference-based functional models could then be built on this technical basis, generating completely new services for metadata retrieval and also simplifying and automating much of routine cataloguing and indexing work. Generating–for instance–proposals for classification attributes using inference rules may well help a lot in everyday library work. A rule such as, "If a work by a given author has a given classification element associated to it and if the publication years of another work by an author with the same name are adjacent, the same classification element is likely to apply to this item," would probably yield useful and time-sparing classification proposals for newly catalogued items. There are almost no limits for imagination in this respect!

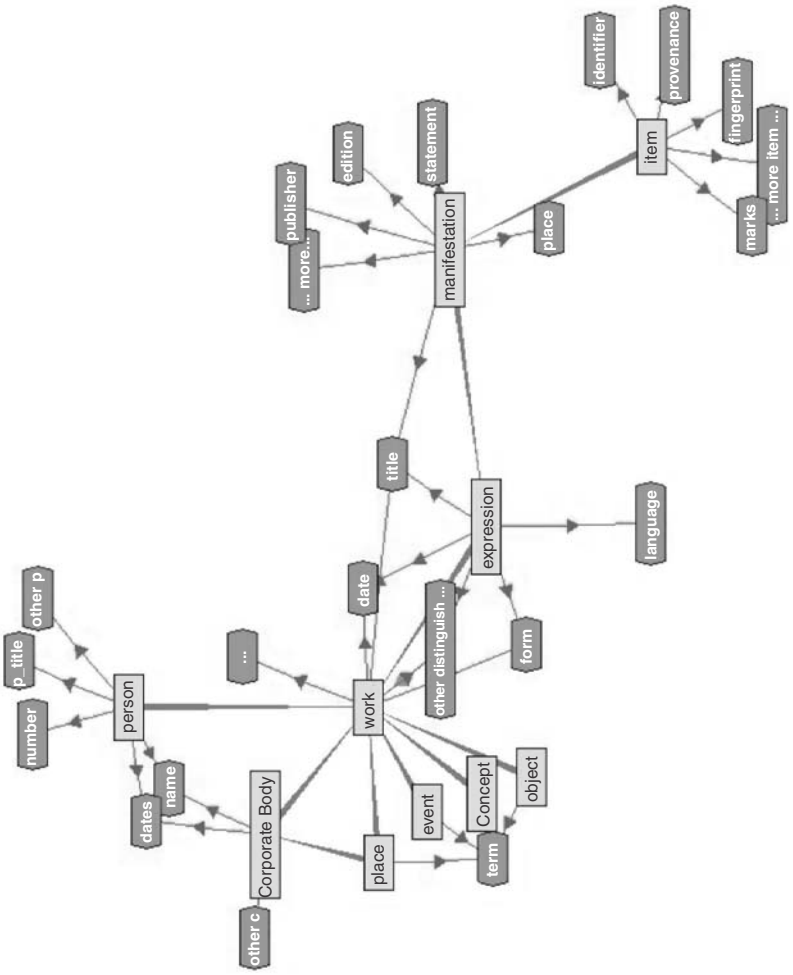
D. More generally, an rdfs:frbr-based methodology is likely to create more systematic junction scenarios between instances as conceived by libraries and as modelled in current or future information architectures in the WWW, and would avoid libraries and the wealth of information they have to offer being wrapped away again in their cataloguing golden cage.

E. As already mentioned above, such integrated architectures would allow for integrating Semantic Web ontologies as successors of librarian models for terminology management and classification in librarian information environments. At the same time, the ontology community has a lot to gain in return from such an approach, as I already pointed out in the passage from my earlier contribution which is quoted in the first part of this paper.

F. Thus–as a side effect!–such an integrated approach would also create grounds for an integrated, WWW-transparent global model for librarian metadata successfully transgressing the divide which separates formal and subject metadata.

G. And finally: an rdfs:frbr-based implementation methodology for catalogues that are still the heart of every library automation system could substantially raise the level of platform- and vendor-independency in the library software market, which still suffers from all too many proprietary and vendor-dependent technologies which restrict the choice of librarians and as a niche market can offer only relatively expensive solutions.

FIGURE 2. FRBR Expressed in a Draft RDFS Model



This list of benefits is selective. Still, it should be sufficiently convincing to make librarians and technicians adopt this proposal and work together to actually implement it.

A heavy and complex agenda will certainly result from actually tackling such a task, but the effort is well invested! I suggest that IFLA and other bodies quickly investigate ways of setting up such an agenda for specification and implementation of rdfs:frbr, and try to do so joining forces with W3C and other relevant communities right from the start!

NOTES

1. More on this issue in Gradmann (2003).
2. More details on DAML+OIL to be found in Davies et al. (2003); a useful introduction is the "DAML+OIL walkthru" at <http://www.daml.org/2001/03/daml+oil-walkthru>.
3. More on OWL at <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.
4. And I will not go into the Byzantine discussions around the actual foundations of this term which certainly could have been better chosen in order to avoid the massive fuzz and misunderstandings generated by the word 'ontology.'
5. More at <http://laneweb.stanford.edu:2380/wiki/medlane/schema>.
6. More at <http://www.disobey.com/noos/LibDB/>.
7. The author is aware of the fact that the relation between RDF, relational database technology, and the possible use of XML-databases in this context remains to be clarified!
8. The examples given here were generated using the KAON tool suite available at <http://kaon.semanticweb.org/>.

WORKS CITED

Berners-Lee, Tim. 2000. *Semantic Web on XML*. Paper presented at XML2000, Washington DC. Available online at <http://www.w3.org/2000/Talks/1206-xml2k-tbl>.

Davies, Jahn, Dieter Fensel, and Frank van Harmelen, eds. 2003. *Towards the Semantic Web: Ontology-Driven Knowledge Management.* Chichester: Wiley.

Gradmann, Stefan. 2003. How Digital Will Libraries Ever Be? Musings on the Limits of a Popular Metaphor. In *Informacné Správanie a Digitálne Knižnice = Information Behaviour in Digital Libraries*. Bratislava: 27-40. Also available online at <http://www.elt.sk/ibdl//media/zbornik.zip>.

Gruber, Tom R. 1993. *A Translation Approach to Portable Ontology Specifications*. Available online at <http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html>.

Iff, Morbus. 2003. *LibDB: Database Schema*. Available online at <http://www.disobey.com/noos/LibDB/?DatabaseSchema>.

Palmer, Sean B. 2001. *The Semantic Web: An Introduction*. Available online at <http://infomesh.net/2001/swintro/>.

SELECTED BIBLIOGRAPHY FOR FURTHER READING

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American* 2001(5).

Hemenway, Kevin. 2002. *The Semantic Web: 1-2-3*. Available online at <http://disobey.com/detergent/2002/sw123/>.

Le Boeuf, Patrick. 2003. *Brave New FRBR World*. Paper presented at the IFLA Meeting of Experts for an International Cataloguing Code, Frankfurt, 2003. Available online at <http://www.ddb.de/news/pdf/papers_leboeuf.pdf>.

Swartz, Aaron. *The Semantic Web in Breadth*. Available online at <http://logicerror.com/semanticWeb-long>.

Swartz, Aaron, and James Hendler. 2001. The Semantic Web: A Network of Content for the Digital City. In *Proceedings Second Annual Digital Cities Workshop, Kyoto, Japan, October, 2001*. Available online at <http://blogspace.com/rdf/SwartzHendler>.