Academic Libraries and the Semantic Web: What the Future May Hold for Research-Supporting Library Catalogues

by D. Grant Campbell and Karl V. Fast

Available online 6 August 2004

This paper examines how future metadata capabilities could enable academic libraries to exploit information on the emerging Semantic Web in their library catalogues. Whereas current metadata architectures treat the Web as a simple means of interchanging bibliographic data that have been created by libraries, this paper suggests that academic libraries could use the Semantic Web as a source of rich metadata that can be retrieved and inserted into bibliographic records to enhance the user’s information searches and to expand the role of the academic library catalogue as a research tool rather than a mere locating device.

The paper uses examples from the fields of rare books, literary criticism, and film studies to illustrate how academic libraries, by shifting their emphasis from cataloguing standards to Web information evaluation, can play a new and exciting role in the development of electronic information services.

The rise of the World Wide Web has transformed the nature of academic libraries in many ways. Three effects are of particular importance to us here. First, the Web’s widespread adoption has caused a massive growth in electronic information resources of interest to academic library users. Second, digital library initiatives are springing up everywhere, many of them emanating, if not from academic libraries, from university faculties and departments, making electronic resources available in a wide number of formats, and described according to a wide variety of metadata standards. Last, but by no means least, the accessibility of university library catalogues on the World Wide Web has given the physical resources of academic libraries greater prominence than ever before, as scholars the world over use academic library catalogues to locate and assess information resources that are relevant to their research.

Many academic libraries have responded to the Web in ambitious and proactive ways. First, librarians, cataloguers, and systems designers have created ambitious and useful Web resource guides that are accessible through the library’s Web site. Second, many academic libraries catalogue academically useful Internet resources and incorporate them into their catalogues. Some libraries are participating in digital library projects, and some are now depositing their bibliographic records in metadata archives, so that records of their physical collections are seamlessly available alongside those of digital libraries. The growing number of library metadata applications attests to the commitment of many academic libraries to play a part in the Web information services.
environment rather than existing in proud isolation. The Web, however, continues to grow, both in sheer size and in the range of topics, readers, and activities that it supports. Furthermore, the underlying architecture of the Web is changing, and these changes will present new challenges and new opportunities. The rise of XML, along with the initiatives of the World Wide Web Consortium to create a “Semantic Web,” suggests that a new world of Web-based information lies before us. In that new world, information will be machine understandable, as well as machine readable, enabling intelligent agents to draw sophisticated inferences from the metadata attached to Web-based information. Whether the Semantic Web, as envisioned by the World Wide Web Consortium, lives up to expectations, there is little doubt that the semantic coding of data made possible by XML, and by new markup standards such as the Resource Description Framework, will have a profound impact on the Web and its users. In addition, academic libraries should be thinking now about how this world of semantically coded information will affect their support of academic research.

In this paper, we suggest that the way forward involves the introduction of alternatives. Metadata architectures that underlie many current digital library and Web portal initiatives are based on an assumption that information organizations, whether conventional libraries, digital libraries, or some other information body, create complex descriptions for their own specific purposes and send only simplified reductions of that complexity across the Web in metadata repositories. Such an approach is highly useful to the development of interoperability standards, as witnessed by the success of the Open Archives Initiative, which uses unqualified Dublin Core as its metadata lingua franca for cross-domain resource harvesting. However, while this approach goes a long way towards ensuring smooth and effective delivery of information across the Web, it does not necessarily allow libraries to exploit and contribute to the emerging Semantic Web in a full and exciting way.

The metadata model that we present here proceeds from a basic question: “What resources and functionalities will the Semantic Web eventually offer the academic library, its staff, its catalogue, and its users?” We hope that by speculating on a possible answer to that question, we will also gain some insight into the reverse: “What can academic libraries offer the Semantic Web?” It is important to emphasize from the outset that we are not offering a fully realized, or even partially realized, technical metadata architecture, complete with a comprehensive definition of elements and relationships. Many tools of the Semantic Web are at the prototype stage, and still others have yet to be developed. Rather, we hope to show how the services offered by an academic library catalog can be enhanced, and perhaps even transformed, through the metadata services enabled by the Semantic Web. And while the metadata community is accustomed to thinking of metadata in relation to electronic resources, we suggest that the reach of the Semantic Web extends beyond the purely electronic sphere and that the Web of the future can also have a beneficial effect on the organization and description of those physical resources that form an indispensable part of an academic library’s mandate.

**Traditional Metadata Architecture**

Early in the development of the Dublin Core Element Set, a difference of opinion emerged concerning the optimal level of complexity that the set should attain. One group, characterized by Weibel as the “minimalists,” argued that the set should remain as simple as possible in order to encourage widespread adoption. The other group, whom Weibel called the “structuralists,” argued that the Dublin Core should support a fair amount of complexity, through the use of qualifiers and schemes to handle complex descriptions and to make the Dublin Core interoperable with the highly structured encoding schemes such as the TEI header and MARC: “Over the years, this goal of ‘simple metadata for coarse-granularity discovery’ came to mix with another goal—that of community and domain-specific resource description and its attendant complexity.”

Ultimately, this tension resolved itself productively in part through the emergence of the Warwick Framework, which presented a modular architecture for metadata design. As the implications of modular metadata packages contained within a containing “wrapper” became clear, two parallel and complementary views of metadata emerged. On the one hand, metadata could be used for object description: Highly complex metadata could be designed and created for purposes of describing information resources in such a way that the end user could make useful and sophisticated decisions regarding the potential usefulness of the resource. The academic library catalogue rests firmly in this view: The Anglo-American Cataloging Rules and the MARC format for information interchange together provide a rich and complex code of information description. More recently, the Library of Congress’s Metadata Object Description Schema (MODS) provides a MARC-compliant XML schema for encoding complex metadata. On the other hand, metadata can be used for resource discovery, particularly discovery across different knowledge and professional domains. In such cases, the academic library, in collaboration with a number of other metadata providers, makes its bibliographic records available in repositories that can be searched and harvested by search agents from numerous different fields. The most prominent current manifestation of this view is the Open Archives Initiative’s Protocol for Metadata Harvesting (OAI-PMH). By using the OAI-PMH protocol, the collection of the University of Illinois at Urbana Champagne can be searched simultaneously with the collections of the Open Video Project, the Perseus Digital Library, and the Michigan State University Digital and Multimedia Center, among many others, using simple Dublin Core metadata elements. In this context, metadata requires interoperability rather than complex description: cross-domain consensus on a few key elements that can facilitate the accurate and productive harvesting of resources from many different sources. As Lagoze suggests, “There remains a compelling need for simple, ‘pidgin’ metadata... Document-centric metadata, where simple string values are associated with a finite set of properties, is most appropriate for generic, cross-domain discovery queries in the Internet commons.”

Wide access, it seems, is inconsistent with deep complexity, at least at present, requiring that a library’s rich and complex descriptions be “dumbed down” to work with other descriptions.

By separating the functions of metadata into these two categories, and incorporating the principle of modular metadata design, an architecture of metadata use has emerged and met with considerable success in projects such as the Gateway to Educational Materials (GEM) (http://www.thegateway.org/) and in various...
The problem with this paradigm is that it presents few options for institutions that are unable, for whatever reason, to create these high-level descriptions. And traditional libraries, despite their best efforts, are often frustrated in their desire to create useful descriptions and to present them in a meaningful set of relationships for the following reasons.

"...traditional libraries, despite their best efforts, are often frustrated in their desire to create useful descriptions..."

Limited Resources
Cash-strapped academic libraries, even when using trained library cataloguers, do not necessarily have the financial resources to provide the fullest or most useful descriptions for their users: “It is no secret to anyone working in an academic library that technical services, and especially cataloging departments, are being cut or eliminated as a way of reducing or shifting library costs.” Despite the presence of bibliographic utilities to facilitate resource sharing, more and more cataloguing tasks are being outsourced, with technical services staff relocated to front-line service. Making the most of limited resources becomes even more challenging for academic libraries that are committed to cataloguing Internet resources: Recent interviews with academic librarians involved with such projects revealed that cataloguing resources are even further stretched by the need to train cataloguers in new procedures and the time and money required to compensate for the initial setup, the development of policies for cataloguing e-journals, communicating

OAI projects such as OAIster (http://oaister.umdl.umich.edu/o/oai ster/) and the UIUC Cultural Heritage Repository (http://nergal.grainger.uiuc.edu/oai/search/). In this architecture, detailed descriptions of a resource are created by the organization that either owns or has recently discovered it: This description frequently involves either a traditional encoding scheme such as MARC or the use of a combination of metadata sets of varying specificity and complexity, specified in an application profile. Out of this complex description, a simplified core description can be extracted for resource discovery.

This architecture makes two assumptions about how information providers relate to the World Wide Web. First, it assumes that the responsibility for complex description rests with the organization that creates or employs the resource: A digital library that mounts a collection of digitized sheet music, for instance, would have the system capabilities, cataloguing expertise, and understanding of its user base to create resource descriptions, providing authorized versions of composers’ names and maintaining a database of music publishers. Any outside assistance, involving the pooling of descriptions, is confined to other members of the group who are presumably doing the same thing and working towards the same information objectives: “Different communities, each with its own view of the structure of knowledge, are organizing information in databases in ways that seem to them most suited to fit their picture of the world and the nature of their specialized information.”

Second, the Web is assumed in this architecture to be a network of irrigation channels: A space where simplified descriptions can be exchanged and distributed. In this paradigm, the local institution is the place for detail and specificity, and the Web is the place for simplified and efficient transmission (see Figure 1). To use two equally outdated metaphors, the library is the “keeper of the books” while the Internet is the “information highway.”

ACADEMIC LIBRARIES AND THE TRADITIONAL METADATA ARCHITECTURE
This paradigm is an attractive one for academic libraries for at least two reasons. First, the assumption that complexity is created at the source fits with the conventional wisdom that catalogues are best created by those who know their own users and who understand the complexities of information description. Libraries have a long tradition of sharing their cataloguing; indeed, library catalogues would be impossible without the widespread use of bibliographic utilities such as OCLC. These utilities, however, rest on the assumption that only those who have received formal cataloguing training and who create bibliographic records according to standard cataloguing codes are permitted to participate.

Second, this architecture enables libraries to make their resources and collections available across the Web without sacrificing the standards of description that form a fundamental part of the library infrastructure. As long as good, useful, and cost-effective descriptions can be created at the institutional level, the traditional architecture enables those descriptions to be preserved; and in academic environments, where the catalogue serves not as a mere locating device but as a significant research tool, it is all the more important that standards of precision, accuracy, and detail be maintained, even when that system is interacting with other systems that use less exacting descriptive standards.

It is no secret to anyone working in an academic library that technical services, and especially cataloging departments, are being cut or eliminated as a way of reducing or shifting library costs.” Despite the presence of bibliographic utilities to facilitate resource sharing, more and more cataloguing tasks are being outsourced, with technical services staff relocated to front-line service. Making the most of limited resources becomes even more challenging for academic libraries that are committed to cataloguing Internet resources: Recent interviews with academic librarians involved with such projects revealed that cataloguing resources are even further stretched by the need to train cataloguers in new procedures and the time and money required to compensate for the initial setup, the development of policies for cataloguing e-journals, communicating
with various stakeholders, and maintaining links to ephemeral and shifting Web addresses.8

Because of these limited resources, academic libraries often have difficulty maintaining bibliographic and authority records, as well as creating them. Changes in the bibliographic universe over time are difficult to represent in fixed catalogue records and authority files. As cataloguers know from bitter experience, accidents can often happen. The death of an author can result in the unwitting creation of two authorized headings, one with a death date recorded and another without. Multiple authors can be unwittingly collapsed under one heading. Serials frequently change publishers, making the publishing field of their bibliographic records out of date.

**Expertise**

Academic libraries have always been aware of the importance of subject knowledge in cataloguing and frequently try to hire cataloguers who have master’s-level expertise relevant to the library’s collection. However, when a library offers resources in many different knowledge domains, the cataloguing staff may lack the detailed, domain-specific knowledge needed to present the resource in its most advantageous and useful light. The details and relationships prevalent in quantum theory, for instance, may be completely different from those prevalent in Italian Renaissance literature, and the requirements for access points and notes may therefore differ widely.

In addition, the rise of the World Wide Web, together with the growth of interdisciplinary subject areas, has brought to light the existence of complex knowledge domains that exist “under the radar” of traditional library organization schemes. Comic books, a form which is traditionally overlooked in many cataloguing policies, may well have enormous research significance for a media studies program: What is more, many of these comic books have subtle differences and nuances that would place a heavy demand on the descriptive record. Similarly, various branches of cinema and television have bodies of enthusiasts that keep careful track of details and relationships that might well be overlooked by a cataloguer. While media theory scholars are expecting sophisticated access to comic book narratives, film theory majors may want similar detail in the description of science fiction movies, and women’s studies programs may expect guides to episodes of Buffy the Vampire Slayer that emphasize the construction of gender roles. Academic libraries have never had the resources to represent such information in their catalogues and traditionally rely on external reference sources to provide such guidance.

Finally, the emphasis that academic libraries place on professional cataloguing competence leads them, naturally enough, to trust and consult trained and skilled cataloguers to produce reliable and useful records. But cataloguing expertise is not the same as subject expertise, and no cataloguers can be expected to have an equally passionate interest in every subject. Yet, the Web is full of monuments to the achievements of enthusiasts: the Internet Movie Database (http://www.imdb.com), the Open Directory Project which forms the basis of the Google Directory (http://www.dmoz.org), the Freshmeat open source software archive (http://www.freshmeat.net), and many more.

**Bibliographic Relationships**

Rebecca Green has argued that “the intuitive understanding that humans bring to relationships is not shared by computational devices. At the same time, the expression and manipulation of relationships is perhaps our best hope for infusing higher quality into our retrieval systems.” This statement attests to the importance of infusing human concepts of relevance and relationship into electronic information systems. However, even if libraries had the time and expertise to describe materials within their specific knowledge domains, traditional library catalogues and descriptive codes are able to describe bibliographic relationships only in a limited way. Distinctions between one work and another are often made only by the choice of main entry, a method that originated in paper card systems and translates uneasily to an electronic format. To be sure, some of these problems may be alleviated when cataloguing codes eventually incorporate the findings of the IFLA Report on the Fundamental Requirements of Bibliographic Records (FRBR); but, like any universal scheme, FRBR will have limited ability to represent unique and domain-specific relationships. Similarly, the Library of Congress Subject Headings, the standard controlled vocabulary for library materials, are repeatedly criticized for their inability to represent the finer nuances of subject relationships and the false relationships they accidentally create.10 The term, “male prostitutes,” for example, is linked as a narrower term to “prostitutes,” which in turn is linked as a narrower term for “women,” creating the nonsensical suggestion that male prostitutes are women.

Despite these problems, libraries also have resources and traditions that can be employed and redefined in a networked context. In particular, they have a long-lasting tradition of shared cataloguing, and with the leadership of OCLC’s Con- nexion service and the Library of Congress’s MARC-XML Schema, they are making rapid progress in the migration into XML compliant forms of MARC coding. The question remains: Can the World Wide Web in general, and the emerging Semantic Web in particular, provide resources beyond that of mere metadata transfer that could help academic libraries solve some of these persistent problems?

**RDF and the Semantic Web**

The architecture of the Semantic Web, in conjunction with the current architecture of the World Wide Web, will bring two sets of assets: one technical, and the other social. Technically, the Semantic Web will boast, first of all, a global, standardized addressing system in the Uniform Resource Identifier (URI): the Resource Description Framework accepts as a resource any entity that has a URI.11 Anything, however tangible or intangible, can be given a Web presence: a physical object, a virtual object, an author, a concept, or a relationship. By widening our conceptual understanding of the URI beyond the traditional “Web location,” the Semantic Web vastly increases the number of entities that can be represented and subjected to machine-understandable inference processes.

---

"The architecture of the Semantic Web, in conjunction with the current architecture of the World Wide Web, will bring two sets of assets: one technical, and the other social."

In addition, the Semantic Web will offer machine-understandable metadata, enabling agents to draw inferences from
that metadata to select and reassemble information in new ways. It will also bring a system of namespaces and ontologies to ensure some measure of interoperability between different metadata standards: “The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work and learn together.”

Socially, the current World Wide Web has demonstrated significant strengths. The ease with which information can be created and mounted on the Web has opened up the information universe far beyond the confines of the traditional library. The result has been the rapid rise of collaborative information resources. Trekkers create elaborate guides to the canonical and noncanonical contributions to the Star Trek franchise; scholars and teachers create complex guides to Internet resources for colleagues and students in their subject areas; the Wikipedia (http://www.wikipedia.org) and the “Earth Edition of the Hitchiker’s Guide to the Galaxy” (http://www.h2g2.com/) both attest to the massive content that can be created through the dedication of enthusiasts. These resources are not created according to library standards, and as Lynch reminds us, the issue of trust is paramount. But while some of this information is undoubtedly suspect, much of it also reflects the benefits of passion and enthusiasm.

If we look at the Semantic Web merely as a cheap mechanism for exchanging metadata between approved metadata providers, we are shutting ourselves out from its potential richness. A Web that boasts a wealth of information, semantically coded and with a global addressing system, could be a source of cataloguing data in and of itself.

**What Might the New Academic Library Catalog Look Like?**

**Basic Framework**

This scenario rests on a vision of the Web as a semantically rich “sea” of information, much of it encoded in RDF, so that it is both human and machine understandable. Assuming that this is the case, the new academic library catalogue might well be structured around three primary entities: a local description identified by a URI, at least one RDF-encoded information resource on the Web, and an intelligent agent.

**Local Description**

The owning institution of a resource would create a skeletal description consisting of the following:

- a URI, which would serve as a unique identifier and address of the resource; and

- whatever information the institution prefers to generate locally, such as call numbers and other information relating to local access.

The principle at work here is the same as the principle behind cooperative cataloguing: Direct description is limited to only those elements that cannot be reproduced from other sources, thereby saving time and expense.

**RDF-Encoded Information**

The professional expertise formerly spent in locating or creating bibliographic descriptions would be reallocated towards locating RDF-encoded information resources that could place this particular resource in a context that is meaningful to an academic research environment, according to its bibliographic nature or its subject content.

The bibliographic resource might be an XML-encoded bibliographic record from a respected source; it could also be an online bibliography created by a subject expert, containing an abstract; it could also, in the case of rare books, be an online descriptive bibliography, giving detailed descriptions of collation, pagination, and binding. In the case of nonbook materials, the resource could be an online database such as the CD Database, the International Movie Database (IMDB), or an art gallery’s online catalogue. The subject resource could be any online resource from a trusted source, which places the particular work in a context of other works according to subject content. In either case, the Web-based resource must meet two requirements:

- it must be encoded in machine-understandable semantic tagging that enables

![Figure 2](link)

**Figure 2**

**Skeletal Bibliographic Record**


- BIBLIOGRAPHIC INFORMATION
- AUTHOR INFORMATION
- WORK INFORMATION
- RELATED WORKS

![Figure 3](link)

**Figure 3**

**Bibliographic Information Imported from the Web**

From the Library of Congress

Bibliographic Record:


xvi, 624 p. (20 no. in 19), [38] leaves of plates: ill ; 23 cm.

From “Charles Dickens: An Annotated Bibliography”:

First edition, 20 numbers (in 19) issued monthly, from March, 1852 - September, 1853. In illustrated blue paper covers; each number contains two plates except No. 9 and 10, which contain one and three respectively. A printed notice in No. 9 offers an explanation.
the relationships between the different data elements to be read by an agent; and

- it must emanate from a source that the cataloging institution considers trustworthy, reliable, and useful.

**Software Tools**

The owning institution would need a suite of software tools which enable the staff to interact with the information on the Semantic Web. Many of these tools would come from agent technology: programs that “can track user interests, explore the Web proactively, learn through interacting with the user, [and] provide personalized data and services.” These tools would exploit the Web’s semantic metadata to enable the cataloger to make intelligent selections: finding appropriate information, mounting appropriate information, and extracting appropriate information from other sources into the bibliographic record. They would also be used to permanently “refresh” the record so that the information would be freshly harvested whenever a record was accessed rather than remaining fixed and subject to obsolescence.

“...information would be freshly harvested whenever a record was accessed, rather than remaining fixed and subject to obsolescence.”

**The Basic Process**

With these three entities in place, the process we envision works something like this. First, the bibliographic record, as constructed by the cataloger and viewed by the user, would consist of a skeletal framework, which could be filled in with information in a variety of categories, gathered by the agent from the RDF-encoded resources on the Web. Such categories could include the following:

- the work’s genre: fiction, drama, biography, etc.;
- the work’s historical context;
- the work’s intellectual context;
- the author’s life and oeuvre; and
- bibliographic features, such as the presence of illustrations, indexes, or bibliographies.

The surrogate then would be recreated whenever it is accessed: Changes in relevant information, as encoded on the Web, would appear in the record. In addition, an RDF framework might well exist that could place the work being described within a set of relationships tailored specifically to that topic rather than the generalized relationships permitted by the Anglo-American Cataloguing Rules.

In order to function effectively, this cataloging system would depend on three specific conditions:

1. **Universal use of URIs:** Every resource would need to have a unique identifier that could be expressed in Web terms; this would involve mapping specific editions of works to unique URIs so that there was a Web-identifiable identifier for the edition of the work in question.

2. **Collaboration with faculty expertise:** Because this framework would often be used in interdisciplinary contexts, the library would need to work closely with faculty researchers in various areas in order to identify Web resources of acceptable authenticity, rigor, and reliability.

3. **The existence of significant useful Web sources in RDF format:** Finally, the framework depends on the exis-
tence of Web resources related to library materials. These Web resources must be encoded with semantic metadata, enabling agents to extract from them appropriately.

It is important to note here that academic library catalogues could themselves form part of the RDF environment because the information in these catalogues could be captured and reformulated in nontraditional ways. Bibliographic and authority records from academic libraries, for instance, could be encoded in RDF Dublin Core or MODS format, enabling them to be searched by agents. The cross-references in an authority record could be used to list alternate versions of the author’s name; author and subject authority files, if assigned an appropriate URI, could be used to give a Web presence to people and ideas, which could be used by other agents for other purposes as well.

**EXAMPLE: BLEAK HOUSE**

In this first example, we present a scenario in which a library specializing in rare materials devoted to British literature draws on information sources in two established and respected academic fields: descriptive bibliography and literary studies. One such item is an 1853 edition of Charles Dickens’s *Bleak House*. Literary scholars have been quick to use both semantic markup and the World Wide Web to make electronic resources available; therefore, there exist large numbers of Web resources that, if encoded in RDF metadata, could be used to extract material for a bibliographic surrogate. These include the Victorian Web (http://victorianweb.org), the Voice of the Shuttle (http://vos.ucsb.edu/), and the Victorian Database (http://www.victoriandatabase.com/).

**Skeletal Record**

The material in the skeletal record would consist of a unique identifier for this particular edition, recognized as such both by the local system and within the RDF community from which the information will be extracted. Additional information could include access notes, as well as a local call number (see Figure 2).

**Bibliographic Information**

The bibliographic information could be drawn from an XML version of an existing bibliographic record, as well as from...
Biographical Information

Biographical information is readily available on the Victorian Web. A succinct biography could easily be found, as well as the listing of the standard biographies (see Figure 4).

Work Information

Work information would vary depending on the user base of the catalogue. In the case of literary works such as Bleak House, current Web resources offer introductory information on certain basic categories that are commonly needed in the study of literary works: summaries of the social context, discussions of genre and literary technique, and introductions to common themes. Also available on the Web are resources relating to the material circumstances of production, which would be of great interest to higher level literary scholars (see Figure 5).

Related Works

In the case of related works, academic sites such as the Victorian Web offer far more detailed and sophisticated sets of relationships than are traditionally offered in cataloguing codes. Extracting RDF data from these sites would enable the user to view lists of other works by the same author. In addition, the user might well be able to see listings for works that influenced the work in question, as well as contemporary works that are related in some way (see Figure 6).

Other Domains

The Bleak House example provides some indication of the resources that could be pulled in to describe a document that is firmly entrenched in an established academic discipline. Queries in academic libraries, however, are notoriously prone to ambiguity. As we move into newer and more interdisciplinary domains, the amount of Web information available is equally impressive. However, the library’s role becomes more complex, as it negotiates the line between relying on the energy of enthusiasts and maintaining the library’s specific mandates.

In the area of film, for instance, databases and Web guides abound, which provide a good deal more information about a movie than any cataloguer could possibly put into a bibliographic record. The library, however, would want to choose which resources offer the most reliable and useful data. At the conservative end, a library cataloguing a copy of Buster Keaton’s 1927 movie, The General, might well draw on the American Film Institute’s Catalog of Silent Films for authoritative bibliographic information (see Figure 7).

At the opposite end of the spectrum, the library might choose to draw on the writing skills and enthusiasm of Jim Petersen, a film buff whose movie Web site, Jeeem’ s CinePad, boasts enthusiastic reviews from the New York Times and is clearly written with the desire to celebrate movies by providing as much information as possible: “I’ve been writing about movies for more than half my life now, since 1977 . . . so it’s particularly exhilarating for me—at long last—to be able to write, edit, photograph, design, and publish all by myself!” Somewhere in the middle between these two extremes lies the massive Internet Movie Database (IMDB), which was established through volunteer effort, and which aims to “provide useful and up to date movie information freely available online across as many platforms as possible.” Like Petersen’s site, the IMDB contains a lot of enthusiastic opinion; unlike Petersen’s site, the IMDB assembles opinions from a variety of quarters and therefore provides an element of user feedback that the library might find useful (see Figure 8).

Figure 8

Record for the General from the Internet Movie Database
Music databases, online bookstores, listerv archives, e-text collections, and government Web sites all offer information that could, in a Semantic Web context, be extracted and inserted into a bibliographic surrogate. In these cases, the cataloguer’s task would shift from transcribing bibliographic data to finding bibliographic and related data and scrutinizing that data for usefulness and authenticity.

**IMPLICATIONS FOR THE ACADEMIC LIBRARY**

Obviously, this vision of the future would involve significant changes in the way libraries catalogue their materials. However, these changes are consistent with the traditions of sharing cataloguing expertise that initially gave rise to online bibliographic utilities, cataloguing consortia, and international standard bibliographic descriptions. They are also consistent with library trends in employing cataloguers with specific subject expertise. The primary change would involve the way in which other library activities—particularly collection development and reference experience—would come to infiltrate the cataloguing process. Rather than confining all bibliographic data within a MARC/AACR structure, the cataloguer in an academic library would spend significant time in the following activities:

- locating RDF-encoded information on specific subjects, scrutinizing its reliability, and assessing its usefulness in meeting cataloguing objectives;
- selecting RDF resources for the specific item being catalogued; and
- participating in markup projects within a specific knowledge domain, thereby promoting the growth of open-access domain-specific metadata.

In so doing, the cataloguer would act, as always, as an information intermediary, using a combination of subject knowledge and information expertise to facilitate the growth of semantically encoded metadata that will rest upon a broad base of expertise and enthusiasm.

**CONCLUSION**

We are not presenting this possible future as a desirable replacement for the traditional architecture that underlies many digital library projects. Clearly, the need for the interoperable transfer of metadata in metadata harvesting programs is deeply important. This proposed model does try, however, to remind all of us that the Web can be more than a vehicle for the transfer of home-grown complex descriptions. With its wide adoption and its support of widely divergent information growth in many official and non-official areas, the Web is a source of richness rather than a mere vehicle for transporting richness from one place to the other. And if the Semantic Web comes into being, we will have practical means of utilizing this richness that comes from giving free rein to the enthusiasm and expertise of people from different areas. Academic libraries stand to benefit from this rich “sea” of information and to play an important part in the growth of it.

**Acknowledgment:** The authors gratefully acknowledge the contribution of Robert Kowaluk, whose unpublished work on the cataloguing of comic books provided a significant catalyst for our thoughts in this area.

**NOTES AND REFERENCES**