Transcendental Metadata: A Collaborative Schema for Electronic Resource Description

Charlene Sorensen (charlene.sorensen@usask.ca)
University of Saskatchewan

Craig Harkema (craig.harkema@usask.ca)
University of Saskatchewan

Karim Tharani (karim.tharani@usask.ca)
University of Saskatchewan

Abstract

Academic libraries are attempting to manage growing collections of diverse electronic resources in a chaotic environment of evolving standards and systems. The transition from a print-dominated resource environment to an electronic one has complicated the decision-making process. Current discourse primarily focuses on meeting patron needs and has distracted researchers from looking at librarian needs. The authors discovered that librarians want a better understanding of the nature, extent, and diversity of electronic resources for decision making, assessment, and accountability. Drawing from the collaborative methods and design philosophies of other disciplines, this paper outlines an approach to leveraging Web 2.0 philosophy and Business Intelligence techniques to address these needs. This approach will serve as a guide for academic librarians to transcend their current practices in order to develop innovative, collaborative, and holistic approaches to the joint stewardship of library electronic resource collections.

Keywords: Academic libraries; Electronic resources; Web 2.0; Business Intelligence; Social metadata; Collaboration

Introduction

Libraries are attempting to manage growing collections of diverse electronic resources in a rapidly changing if not chaotic environment of evolving standards and systems amid growing user expectations. The massive proliferation of information, remote access, and the ability to bring together diverse media types are among the most beneficial characteristics of electronic resources. In addition, the Library 2.0 movement has brought about the use of social web technologies and a new culture of increased online interaction between users and libraries. This movement, as well as the rise of various technological advances such as link resolvers, faceted browsing, and web-scale discovery systems, has allowed academic libraries to improve discoverability and access to their collections. While this user focus is certainly important, it has perhaps distracted libraries from looking closely at the needs of librarians who are struggling to find accurate, appropriate, and timely information in order to improve decision making and assessment pertaining to their electronic collections.

In response to this challenge and to the overall rapidly changing information landscape, the authors propose a process that draws on collaborative methods and design philosophies of other disciplines in order to provide a framework for describing electronic resources in an uncharted future. This approach aims to transcend the constraints imposed by established systems and standards and seeks to provide the right information in the right place to enable librarians to be more effective stewards of electronic resources. Using examples from the University of Saskatchewan Library, this paper addresses several aspects of the problem of electronic resource metadata. First, as a necessary part of understanding the various challenges, librarians’ needs were considered before deciding upon the solution or solutions. To address
this gap in understanding, librarian focus groups were consulted and their needs were analysed and thematically arranged. Second, this needs analysis led to the investigation of a more collaborative approach using Web 2.0 concepts to describe, present, and manage metadata about electronic resources that would benefit University of Saskatchewan librarians. The Web 2.0 philosophy directed efforts toward using a combination of metadata from existing library systems and librarian-generated tagging to produce more meaningful descriptions of the electronic resources. Third, the needs analysis also pointed to the misconception that library systems should act as effective reporting tools. This problem was addressed by using Business Intelligence techniques, an approach identified in the late 1980s by Howard Dresner to describe a data-driven decision-making process (addressed later in this paper) that is able to integrate metadata from disparate library systems into an interactive reporting tool.

Background

The University of Saskatchewan (U of S) is a medical-doctoral institution composed of 13 academic colleges and three interdisciplinary graduate schools. The University of Saskatchewan Library is an Association of Research Libraries member providing access to over 1,100 electronic resources to approximately 18,000 students and 1,000 faculty. Like other academic libraries, the U of S Library is no longer at the point where electronic resources are a small specialized part of the collection. The U of S Library is constantly assessing and enhancing workflows and systems and the current planning for web-scale discovery and electronic resource management systems are signs of these improvements. Despite these developments, or perhaps because of them, the implications of this growing collection are not fully understood and there remains much to consider and assess.

In recognition of this, a U of S Library task force was established in late 2009 to create a project plan for a review of electronic resources. The main goals of the review were to ensure that the U of S Library provides the best resources for its users within budget, to be accountable to the university administration for the library budget, and to respond to potential targeted cancellations dictated by the annual acquisitions budget. The members of the task force soon realized, however, that the information needed to achieve their goals was neither sufficient nor readily available. In other words, information from the catalog and other library systems needed to be consolidated in order for the librarians to assess the collection and make collaborative decisions.

It should not come as a surprise that the librarians felt they did not have sufficient information to grasp the intricacies and demands of the electronic resources collection. Electronic resources have surpassed print resources in complexity, cost, percentage of the budget, and popularity in academic libraries. The growth in the variety of materials available online combined with changing user expectations is creating new challenges for librarians to describe, organize, and manage these resources. Making significant progress in meeting these challenges will promote collaboration and aid in accountability, decision-making, and assessment, and ultimately help faculty and students use the library in more effective ways.

To date, libraries have had to focus their efforts on dealing with the initial challenges posed by the transition from print to electronic resources: first, the front-end details of making electronic resources easily accessible to library users; and second, managing the business details of back-end staff functions related to acquisition, payment, and licensing that facilitate user access. These two challenges have manifested themselves in two often disparate library systems— the online public access catalog (OPAC) and the electronic resources management system (ERMS). The evolution of the OPAC from what was essentially an online representation of the card catalog to the “next generation” catalog with various social web features and then to the web-scale discovery system represents well how libraries are improving the accessibility of content for users. The proliferation of electronic resources has also led to various ways of managing back-end staff functions. These administrative data can be managed within an ERMS or in a series of interrelated systems that collocate administrative data generally for the purposes of the electronic resources librarian. While ad-
Advances are being made in these two areas, the library literature neither recognizes nor provides solutions for meeting the needs of librarians not directly involved in the acquisition and licensing of electronic resources.

**Needs Identification and Analysis**

In order to fill this gap, the authors looked outside of the library profession for ideas. The “Requirements Engineering” process commonly used by the software development industry seemed like a good starting point. This process involves identifying the users, eliciting and documenting their needs for analysis, and subsequent implementation. The tangible benefit to using this process is that it can be applied iteratively in various situations. It also provides an opportunity to truly understand the needs of librarians in ways that transcend existing administrative processes, systems, and standards.

In order to elicit and document the requirements of the U of S librarians, two focus group discussions were held. Each focus group was composed of librarians at different stages of their careers and with a range of subject expertise. Responses were gathered through note taking and the sessions were audio recorded. In both sessions, the facilitator focused discussion around the following questions:

- What is your definition of electronic resources?
- What are your frustrations around using, managing, teaching, evaluating, and assessing electronic resources?
- What do you wish you had on hand, point of need, to guide decision-making related to electronic resources?
- What are the messages you would share with electronic resource decision makers? What should they know from your perspective?
- What current (top 5) tools, sites, or systems do you use at work? Why do you use them? How could they be improved?

Although these questions generated wide-ranging discussions among the participants, the discussions did not produce straightforward or specific requirements for how to address the U of S librarians’ situation. These sessions, however, did move the conversation forward and generated a series of themes for further analysis.

First of all, the participants felt that they did not have the right information in the right place in order to make decisions. There were, of course, a variety of opinions about what the “right” information actually would be. The uncertainties are understandable because these needs vary from librarian to librarian and institution to institution. The first attempt at addressing the needs resulted in a set of attributes (see Table 1) based on metadata from the U of S Library’s systems.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Source of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliographic number</td>
<td>Library catalog</td>
</tr>
<tr>
<td>Title</td>
<td>Library catalog</td>
</tr>
<tr>
<td>Source</td>
<td>Local spreadsheet</td>
</tr>
<tr>
<td>Source</td>
<td>A locally developed and accepted list of values to indicate the acquisition source of the resource. Options include: locally-created, vendor-subscribed, consortium, open access, one time purchase.</td>
</tr>
<tr>
<td>Pricing model</td>
<td>Local spreadsheet</td>
</tr>
<tr>
<td>Pricing model</td>
<td>An attribute to distinguish between resources that are one-time purchases and those that have maintenance and subscription charges.</td>
</tr>
</tbody>
</table>
## Table 1. Attributes of electronic resources needed for decision making

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Content management system</td>
</tr>
<tr>
<td>Internally created schema to align with colleges and programs at the U of S. As appears in the library’s subject pages.</td>
<td></td>
</tr>
<tr>
<td>Format/type</td>
<td>Library catalog</td>
</tr>
<tr>
<td>A locally developed and accepted list of values. Options include: index, aggregate, document collection, e-book collection, reference.</td>
<td></td>
</tr>
<tr>
<td>Locally hosted (yes/no)</td>
<td>Local spreadsheet</td>
</tr>
<tr>
<td>An attribute to distinguish between resources that are hosted on U of S Library server and those that are externally hosted.</td>
<td></td>
</tr>
<tr>
<td>License digitized (yes/no)</td>
<td>Local database of digitized licenses</td>
</tr>
<tr>
<td>An attribute to identify resources whose license information is available in the local database of digitized licenses.</td>
<td></td>
</tr>
<tr>
<td>Perpetual access (yes/no)</td>
<td>Local spreadsheet</td>
</tr>
<tr>
<td>An attribute to identify resources for which the library has perpetual access rights.</td>
<td></td>
</tr>
<tr>
<td>COUNTER compliant (yes/no)</td>
<td>Local spreadsheet</td>
</tr>
<tr>
<td>An attribute to identify resources that provide usage statistics as outlined by the COUNTER Code of Practice.</td>
<td></td>
</tr>
<tr>
<td>Date acquired</td>
<td>Library catalog</td>
</tr>
<tr>
<td>An attribute to record the date a resource was created or acquired.</td>
<td></td>
</tr>
</tbody>
</table>

Although the specific needs of librarians may vary from institution to institution, Table 1 illustrates the attributes that would help address the U of S librarians’ needs as expressed in the focus groups. These attributes could be considered “formal” metadata in contrast to the “social” tagging side of this approach. Table 1 provides a prioritized list but does not contain every possible attribute to answer all possible questions. The needs analysis phase allowed for the number of attributes to be limited through prioritization for the first iteration of this process. When the U of S Library implements an ERMS, there will be a dramatic increase in the amount of readily available metadata about electronic resources that may provide an opportunity to revisit the needs of the librarians and possibly find more or different pieces of information pertinent to the management of electronic resources. Table 1 is provided in order to demonstrate what may come out of a needs analysis process, but being so specific to the U of S Library, the various items will not be explained in detail. Although these attributes are understandably simplistic, further iterations may lead to more sophisticated descriptive elements. Further discussion around “the right place” clarified that there was a desire for the information to be integrated and available in a single place.

Secondly, the participants expressed the desire to be more “hands on” in two ways. They wanted to be able to find information about the electronic resources themselves instead of making a request through an intermediary, such as someone from the technical services department. The participants knew that some of this information could possibly be obtained from the library catalog, but the querying process was seen as too complicated for someone who used it only occasionally. They also wanted to be able to contribute to the description of the electronic resources. The participants stated that the ability to add tags to the electronic resources would...
help describe the resources in a way that was meaningful to them at this particular institution.

The themes elicited from the consultation sessions, namely the right information in the right place and the do-it-yourself focus, naturally led the authors to investigate the Web 2.0 design philosophy that emphasizes these very concepts. Within this theoretical framework, two practical approaches were chosen in order to fulfill the librarians’ needs. A method was established to create appropriate descriptions of the resources using both existing metadata and librarian-generated tagging. This information was then pulled together in a self-service tool leveraging Business Intelligence techniques (discussed later).

**Web 2.0 Design Philosophy**

In order for a Business Intelligence approach to have fuller effect in our situation, however, it was important to understand and take great advantage of Web 2.0 resources. Web 2.0 is a set of principles and practices that have radically changed the way the online world is experienced. These concepts have manifested themselves in libraries as Library 2.0, which has resulted in increased interaction between patrons and libraries through social web technologies. But these same concepts – user focus, social organization and description, and collaborative tool development – can also be applied within an organization and between different units and branches of a library. Therefore, in order for librarians to create useful tools for themselves, they “can build on their understanding of read/write Web rights with tools that support peer production and participation. The underlying thread in Library 2.0 is trust, and letting go of perfection.” Librarians must trust their capacity to create tools that take advantage of and emphasize the open sharing of databases to extract meaningful metadata, develop smaller applications, and focus on problem solving over sustainability.

While library systems like OPAC and ERMS serve the patron and fulfill operational needs of the library, their usefulness for the non-acquisitions librarian remains limited. These vendor-based systems are developed in a slow and controlled manner, rely on standards that develop slowly, and have evolved from a print-based business model. The commercial organizational models in which these systems are created further entrench the cumbersome nature with which they develop and react. Similar to library systems, existing metadata standards also must make numerous compromises as they attempt to offer a single set of attributes that can be used by all librarians, regardless of expertise, background, or preferences.

By contrast, using Web 2.0 philosophies helps librarians to develop smaller applications that are simple to produce and are not driven by the marketplace, release schedules, or competing systems developers. In addition, the applications’ smaller code bases and more focused target demographics allow applications to be released or revised very quickly. Similarly, having the flexibility to generate local metadata as tags provides the necessary agility to harness the collective intelligence of librarians across the library. This allows librarians to take more risks, experiment with ideas that address niche problems, and focus on problem solving over marketability.

Most librarians understand very well the Web 2.0 environment; they create blogs, engage with Facebook and Twitter and look for ways to enrich their online content by harnessing communities of users. It may be a subtle paradigm shift to consider how these ideas and technologies may be used internally, but librarians are more than capable of integrating new tools and metadata into their existing work flows.

**The Right Information**

The needs analysis of the U of S librarians highlighted the desire for accurate, appropriate, and timely information about the electronic resources collection for decision-making, assessment, and accountability. Descriptions of the electronic resources were established using a relatively small standard set of attributes derived from existing U of S Library systems. These attributes then provided a basis for librarians to generate additional descriptive tags. The key point is that this process is not time- and resource-intensive and that librarians, per-
haps as much as any user group, are well-suited to creating additional metadata around information resources. This approach is further enriched by having two distinct metadata streams merge to meet the librarians' needs. While identifying the attributes in Table 1 met some of the U of S librarians' needs, these librarians also wanted to add their own descriptions to the resources. Tagging and social metadata are Web 2.0 developments that are generally used to exploit the collective intelligence of a group of interested and well-informed users to create richer and deeper metadata. Although not all implementations are successful at improving information retrieval or search functionality, projects such as Flickr and “Steve: The Museum Social Tagging Project” are evidence that social metadata has the potential to greatly enhance and even transform the ways in which metadata is assigned to digital objects. For individual users, these tagging structures provide more control over the information important to them. Collectively, it is possible for user groups to reach a level of consensus over how a digital object should be categorized and described. Guided by the Web 2.0 emphasis on user participation, the authors investigated how social metadata can contribute to the creation of a responsive descriptive framework for librarians given the collaborative and communicative environment it helps to establish.

Social metadata functionality can provide librarians with the opportunity to generate task-specific metadata and build richer metadata around resources that most urgently need it. Some resources may have fewer and less developed tags whereas frequently used or more valuable – however defined – resources may garner more attention. As with most folksonomic and social tagging frameworks, the flaws with this type of organization and description are softened by the amount of flexibility it offers. As explained by Brown and Duguid, combining both formal and social tagging can take advantage of each of their strengths.

While it’s clear that self organization is extraordinarily productive, so too is formal organization. Indeed the two perform an intricate (and dynamic) balancing act, each compensating for the other’s failings. Self-organization overcomes formal organizing’s rigidity. Formal organization keeps at bay self-organization’s tendency to self-destruct. In the U of S Library context, tagging functionality builds upon existing formal metadata and provides increased flexibility in an otherwise rigid structure.

Several areas of description could benefit from the additional metadata generated from these tags, such as subject or discipline, user feedback, and administrative details. The multidisciplinary nature of many electronic resources often adds to the complexity when determining the formal metadata, either diluting the meaning of the attributes or generating a substantial amount of work maintaining the headings. Local practices and cataloging nuances may not be captured if formal subject headings come from the vendor or from a cataloging outsourcing service. Providing librarians with a way to add natural language tags that better define the subject and research areas helps solve these challenges. In addition, librarians’ interactions with patrons provide them with unique and valuable knowledge about how the resource is being used. A tagging function allows them to efficiently share this information while simultaneously incorporating their colleagues’ perspectives into their own understanding of the resource. These tags are made available to all librarians, helping to establish an effective collaborative environment for the improved stewardship of these resources.

Social metadata generally works better with large numbers of participants or users, though this may not be as significant a number as some may think. Golder and Huberman suggest it takes only about 100 users tagging any one item to generate a worthwhile pattern that would assist information-seeking activities. Nevertheless, the number of contributors at an academic library may not approach 100 in total, and even fewer may tag any single item. However, the knowledge and familiarity with the content, along with a more focused end-user (i.e. librarian) community may offset some of the necessity for a large user/contributor base. Social tagging will likely not replace thesaurus-based or algo-
rithmic searching, but it has a role in facilitating scholarly communication that helps to build collaborative communities. Additionally, the use of the formal attributes provides a base from which to build a functional folksonomy.

**In the Right Place**

In addition to identifying the nature and extent of information, the needs analysis of the U of S librarians revealed that they wanted information to be integrated and readily available in a single location and preferably within a self-service model. In many ways, the dilemma of librarians at the U of S is similar to that of managers in commercial organizations who rely on having access to the right information at the right time to make informed decisions to fulfill their responsibilities. One of the ways commercial organizations have successfully dealt with this challenge of informed decision-making is through development of decision support systems, commonly referred to as Business Intelligence (BI). Howard Dresner first introduced the term Business Intelligence in 1989 to describe systems that assist decision makers in understanding the current state of their organization. Since then, BI solutions have been implemented to bring disparate data from existing operational systems together with the purpose of providing strategic insights to help management make better operational and managerial decisions.

In its simplest form, the architecture of a BI solution has three components: one or more source systems that provide the relevant information to be integrated; a front-end tool for users to visualize and interact with the integrated data; and a data repository to house the integrated data. Typically, the relevant data from the source systems are transferred to the central data repository where the necessary data are extracted, transformed, and loaded to support managerial decision-making. The front-end interactive reporting tool provides the necessary means, with varying degree of sophistication, for users to organize, format, and visualize data in support of their information needs. Data visualization is an essential element of BI. It not only lets users represent their selected data in the form of charts, maps, and other graphical representations, it also empowers users to visually interact with data directly to instil a culture of informed and evidence-based decision making.

The level of sophistication of BI architecture in a given organization is, of course, proportional to factors such as available budget, resources, and the perceived need and importance of BI in the organization. Guided by the Web 2.0 design philosophy, the authors here propose a very simple and pragmatic architecture. In order to extract the necessary information to support the attributes identified in Table 1, relevant U of S Library systems were identified as the source systems. These included the catalog and the Drupal-based content management systems used by the U of S librarians to manage subject pages and the electronic resources A-Z list. According to Gartner’s Magic Quadrant for BI Platforms, a highly influential and credible assessment report on BI vendors and their products, there are twenty vendors offering several BI products including front-end interactive reporting tools. The authors chose Tableau software, which operates on a self-service model where U of S librarians can drill up and down, drag and drop, and filter as well as visualize data in ways that best fit their needs. Since Tableau is capable of working with spreadsheets, the data from source systems were extracted, transformed, and loaded into a spreadsheet to act as a data repository.
Simply implementing a successful idea from the commercial sector in a library setting, however, is easier said than done. There are fundamental differences in how academic libraries operate and implementation of any new tool must be adapted carefully to ensure success. This is where the Web 2.0 philosophy can play an important role in keeping efforts and resources in check while delivering incremental value to the users.

The success of BI in academic libraries will also depend on two crucial factors. First, librarians must understand and appreciate the different purpose and functionality of existing U of S Library systems such as the catalog and ERMS that support the U of S Library operations and the decision support systems such as BI solutions that are put in place specifically to gather and provide insights necessary for better decision-making. Second, and perhaps more importantly, librarians must transcend the constraints imposed by current operational systems in seeking the information they need to make informed decisions. For example, academic libraries must recognize that implementing an ERMS may achieve operational efficiency in managing electronic resources, but having an ERMS in place will not automatically provide integrated, summarized, and historic information to support decision-making. From a BI perspective, an ERMS, like the library catalog, will serve as one of the source systems for extracting necessary information that may be combined with information from other sources to meet the needs of the librarians.

Conclusion

The shared anxiety in the field of information studies centers on the challenges librarians face in managing electronic resources into the 21st century. This is understandable given how the transition from a print-dominated resource environment to an electronic one has complicated decision-making processes, especially data gathering. Librarians must therefore find ways to collaborate and transcend common practice and to focus on what is really needed to solve electronic resource related problems.

The emergence of Web 2.0 design philosophies has provided methods for improving the ways librarians interact with collections of electronic resources. Adopting these philosophies builds a collaborative foundation on which to discuss possible solutions with both technical and non-technical staff. Web 2.0 approaches also highlight the importance of finding ways to bring together information about electronic resources from a variety of sources. The agile and iterative nature of smaller applications will allow librarians to adapt quickly to evolving needs within the changing information landscape. Combining librarian-generated tags with formal metadata from existing library systems, as depicted in Figure 1, shows that popular and emerging trends like social metadata are not just reserved
for library patrons but can be beneficial for librarians with specific needs in the context of collaborative management of electronic resources. In addition, decision making can be improved by providing librarians with an avenue to query integrated data about electronic resources from disparate library systems within a single self-service environment – something that is no longer outside the domain of librarians’ expertise.

There is considerable value in providing a venue for librarians to share their needs. The information created and gathered through this process need not result in the creation of a perfect solution, but if captured and presented effectively, it can lead to a more collaborative, informed, and successful decision-making environment. Taking advantage of the collective intelligence of librarians in this way can help libraries move forward in the management of electronic resources. This paper has outlined an approach that will serve as a guide for similar initiatives and will encourage academic librarians to transcend their current practices in order to develop innovative, collaborative, and holistic approaches to the joint stewardship of library electronic resource collections.

Endnotes

1 For the purposes of this paper, electronic resources are defined as anything digital available remotely through an electronic device. However, e-journals and e-books are not dealt with individually but rather at the collection level.


11 Dr. Jeff Smith (Digital Research Centre, University of Saskatchewan), in discussion with Craig Harkema, October 20, 2010.


