Enhancing Interoperability services in the U-campus Digital Library Project

Morales-Salcedo Raul, Hiroaki Ogata and Yoneo Yano
Department of Information Science and Intelligent Systems, Faculty of Engineering,
The University of Tokushima, Japan
[rauls, ogata, yano]@is.tokushima-u.ac.jp

Abstract

Digital libraries have the potential to not only duplicate many of the services provided by traditional libraries but to extend them. Basic finding aids such as search and browse are common in most of today's digital libraries. But just as a traditional library provides more than a catalog and browseable shelves of books, an effective digital library should offer a wider range of services. Using the traditional library concept of special collections and the concept of virtual spaces, in this paper we propose that explicit creating collections using virtual spaces in the digital library –virtual collections– can benefit both the library's students and teacher's contributions and increase its viability. We first introduce the concept of a virtual collection, outline the costs and benefits for defining such collections, and describe an implementation of collection-level metadata to create virtual collections for learning proposes in a distributed digital library. We conclude by discussing the implications of virtual collections for enhancing interoperability and sharing across digital libraries, such as those that are being developed as part of the Ubiquitous Campus project (U-campus).

1. Introduction

Most of the digital library research and development to date has centered on issues related to the technology and content of digital libraries [10]. This work has focused on issues such as developing effective ways to digitize and store resources, how to efficiently deliver resources over the network, providing ways to search for resources, and how to enable digital libraries to interoperate.

These are fundamental issues to be sure, but to be viable in the long run a digital library must be more than a collection of digital objects that can be efficiently stored and transported. Just as the traditional library evolved to provide services to make its contents more accessible to its users, the effective digital library must develop a range of services to assist its users in finding, sharing, cataloging, understanding, and using its contents. Moreover, in its digital form the library has the potential to not just emulate traditional libraries in the services it provides but to improve and extend them by capitalizing on advantages inherent in the medium.

One important area where the digital library can extend the services it provides beyond that of the traditional library is in integrating and highlighting user contributions. With the exception of especially unique or noteworthy contributions, the traditional library is rarely eager to receive resource contributions outside of its usual channels, as the effort needed to catalog and integrate outside contributions into a physical library is substantial. Digital libraries, on the other hand, are more often willing to receive contributions. It has been demonstrated that a combination of minimal submission data and basic verification procedures can result in high-quality digital library contributions with low rejection rates [6] [12]. Such contributions enhance the value of the digital library by increasing its size and diversity and the process of cataloging and integrating contributed resources into a digital library often requires less effort.

However, the aspects that make digital libraries built from user contributions valuable—diversity of content, potential for large growth—also create potential drawbacks. For example, search and browse facilities enable users to find learning resources based on features such as author, subject, or keywords, but as a digital library grows, finding specific resources of interest among the entire collection can become more
difficult. At the same time, the prominence of a given contributor’s contributions becomes diminished as the library grows.

One way to help users find resources of interest in a digital library while ensuring that contributors receive recognition is to borrow a concept that has long been part of traditional libraries: the special collection. By defining and making available virtual collections we believe the digital library can extend the specific collection model and—at a modest cost—provide benefits to both its users.

2. Collections

Traditional libraries often contain, in addition to their main holdings, special collections. In these settings a special collection is generally defined as a group of related materials that is given some form of special treatment. The special treatment might be due to the rare or delicate nature of the materials (rare books or antique maps, for example), or because the library wants to highlight the materials in some way (collected material of some classes or specific areas).

In contrast to traditional libraries, the special or collections of digital libraries can be much more fluid. Where the holdings of a traditional library are physically constrained to a single space and a single ordering, resources in a digital library can be distributed across many servers, can be owned by different universities or organizations, and can be displayed in many different orderings and arrangements. As suggested in [4], however, even a broad definition of a collection in the context of digital libraries can be ambiguous. It can, for example, be influenced by the point of view of those making the definition. Defining sub-collections can be even more flexible as there are many possible factors that can suggest how sub-collections can be formed. A sub-collection can be defined by including all those resources that share a topic or other significant attribute (the collection of all Japanese, Spanish or English language classes), those contributed by a specific organization (the resource collection of the university of foreign languages), or those used for a specific purpose (all resources used for the online course of languages).

These sub-collection examples are instances of collections that cannot be easily replicated in traditional libraries. They are made possible by exploiting advantages the digital environment inherently provides: objects can exist in multiple collections, collections with the same objects grouped in different ways can co-exist, collections can be created dynamically and exist for varying amounts of time. They become virtual collections and as such—in contrast to the traditional library—enable a digital library to provide a limitless number of sub-collections based on a wide range of features.

3. Benefits of Virtual Collections in a Ubiquitous Campus

Although it is common for traditional libraries to create and maintain special collections, many digital libraries do not attempt to provide a similar service. Most digital libraries do create the most basic of virtual collections—the result set dynamically created from a search request or category browsing—but rarely do they explicitly create and promote the sort of virtual collections described above. By providing access to users to this kind of virtual collections using internet, wireless and mobile technologies allows them to interact with the digital library environment for learning activities anywhere at anytime.

A digital library that is available anywhere at anytime containing virtual collections helps its users in several ways. Firstly, it provides permanency; users never lose their work unless it is purposefully deleted. A new user who may be intimidated by a digital library’s search interface or the number of results returned by a query might be better introduced to the digital library through the more easily exploitable partitioned set of resources in a virtual collection. In addition, all learning activities and processes are recorded continuously. A directory of the virtual collections contained by a digital library, as shown in Figure 1, can provide a good introduction and overview of the library’s contents to new or casual users.

Figure 1. Virtual collections available in the U campus digital library project

Associating resources with virtual collections enables those resources to be found more easily, either by browsing the contents of a highlighted virtual
collection (launched from a page such as that in Figure 1) or through standard search and browse interfaces. Figure 2 shows how virtual collections are available from the browse page of the U-campus Digital Library Project, a distributed digital library of learning resources.

By proving accessibility, users have access to the digital library's virtual collections from anywhere. Adding virtual collections to search facilities, such as that of the U-campus digital library of educational resources shown in Figure 3, enables a user to perform a standard search but restrict it to a specific virtual collection, which could provide a more manageable and higher-quality result set than by searching the entire digital library, therefore wherever user are, they can get any information immediately.

Looking at the use of the digital library from a "learning-oriented perspective" [9], other benefits to the user stem from a more productive use of time. In [11] it is suggested that sub-collections can facilitate learning by isolating a group of related content and enabling a user to focus on those resources. Defining virtual collections makes it easier for users to find and work with such groupings of related content, either through a listing of available collections as in Figure 1, or by a "related resources" link based on virtual collection associations and tied to specific resources. Additionally, the virtual collection description might include links to related information outside the digital library, thus guiding users to more materials for their learning.

In most cases those who contribute resources to digital libraries are not directly recognized, yet digital libraries often depend largely on contributions for the content they provide. In a University library for example, teachers usually put supplementary material of their classes in reserve, making these “special collections” available to all students for a certain period of time. It is, therefore, in the best interests of the digital library to find ways to encourage new and repeat contributions. Virtual collections can recognize contributors in several ways. First, they provide an alternative distribution outlet. Users often have collections in which they have invested effort in creating and would like to see used more widely (for example, supplementary multimedia material of a language class at the university). Because a digital library will generally have a much larger base of regular users than contributors, contributing the collection gives the contributor’s resources more exposure.

Virtual collections can not only help improve a set of resources and support their distribution for learning proposes, but can also offer basic infrastructure of services. In some cases, such as with the U-campus Project where resources are quite large (video and audio files), contributing resources enables the contributor to share resources without the overhead of storing and managing them, while retaining an association with them. If a contributor owns a large number of resources, this is a significant benefit itself, and one that has been taken advantage of by several contributors at the U-campus project.

Finally, if the digital library shares information about resource usage, either directly to its contributors or as is increasingly common, through most recommended, the contributor can gauge the relative demand of his contributions. This is helpful not only to contributors and the users of the digital library, but also “helps new contributors to understand what is considered a good item” [7].

4. Implementing Virtual Collections

The benefits of virtual collections do not come without a price, of course. For a digital library to be able to easily create and remove virtual collections, to associate resources with different virtual collections in a flexible way, and to help users find and use the virtual collections, the library must have a structured approach to representing these collections. Moreover, to make creating such collections practical, this approach should also strive to minimize the costs associated with creating virtual collections.

In the remainder of this paper we describe an approach to implementing virtual collections in personal and group spaces based on our research in creating
learning collections for a digital library in the U-campus project. We first define personal and group spaces in terms of learning activities and then review the current research related to representing collections in digital libraries for learning proposes and describes the costs and benefits of different types of metadata used to represent these collections. We then describe how we used and extend this information to define a collection-level schema for Educative Virtual Collections (EduVC) digital library and discuss practical issues related to implementing the schema in the U-campus project.

4.1. Virtual Spaces

We define personal space as a virtual area that is generated, owned and maintained by users to persistently keep resources objects or references to resources which are relevant to a task or set of tasks the user needs to perform within the learning processes. Personal spaces may thus contain digital documents in multiple media, personal schedules, visualization tools, and user agents. Resources within personal spaces can be pre-assigned according to the user’s role. For example, a research user would have access to research-specific topic materials, visualization tools and interfaces to communicate (video, audio or text based chat) with his/her colleagues. Agents may be available for recommending virtual collections or library materials that are relevant to the research topic and the personal space could be enriched by the agent’s suggestions.

Similarly, we define group spaces as virtual areas in which users can meet to conduct collaborative activities synchronously or asynchronously. These virtual areas are created dynamically by a group leader or facilitator who becomes the owner of the space and defines who the participants will be. Group spaces can be generated automatically when a number of users have been detected to have similar user profiles or interests around a given topic or task. In addition to direct user-to-users video-communication, users should be able to access virtual collection’s materials and make annotations on them for every other group participants to see.

4.2. Metadata

Metadata is a key element of any library, traditional or digital. Metadata is used by libraries to describe and organize item-level resources and by users to search and browse the library. It consists of a set of elements that describes a resource. Collection-level metadata performs a similar function for collections and is used in traditional libraries for discovery across collections. Work on collection-level metadata from several fields including archives, museums, libraries, and the Internet is relevant to the design and implementation of virtual collections for learning proposes. As outlined in [13], each field defines collections differently and has different standards governing collection description. The past few years have seen a movement to create a standard for collection description that is informed by, yet transcends, the fields from which it is derived. There are existing technologies, standards and ongoing initiatives for collection-level metadata for learning proposes. The alliance of remote instructional authoring and distribution networks for Europe [1], Dublin Core initiative [5], IMS global learning consortium [6] and, IEEE learning object metadata working group [15] are the most important initiatives dealing with metadata for computerized learning. Work in UK, USA and MEXICO has had also resulted in the formulation of goals for collection-level metadata and the definition and development of schemas to describe collections.

Based on work with the eLib working group on Collection Level Descriptions, the RIDING Clump Project created a searchable database of collection descriptions to provide information about what was available in its libraries [2]. The purpose of its scheme was to describe any type of collection—physical or virtual (electronic), networked or otherwise. RIDING collection metadata should allow users to discover, locate and access collections, search across multiple collections and allow software to provide services based on user preferences.

The Research Support Libraries Program (RSLP) Collection Description Project developed a model allowing all the projects in its program to describe collections in a consistent, machine readable way [14]. The RSLP builds upon the RIDING goals above by requiring that collection metadata allow the refinement of distributed searching approaches based on the characteristics of collections.

The UDLA Digital Library project in the Universidad de las Americas - Puebla, Mexico [16] is a research and development digital library project focused on provide access to special collections such as antique books, newspapers and historical documents. UDLA project allows to navigate, to visualize and to consult this kind of special collections identified by static metadata.

Several themes emerge from this survey of requirements. First, it highlights the importance of
establishing standardized collection level metadata schemas that can effectively describe and manage a diverse set of collections and their metadata. Second, it argues that the schemas must support a number of functional library services that enable users to access collections and items, to search for materials, and to comprehend and use them effectively for learning proposes.

4.3. Types of Metadata

One challenge to creating collection-level metadata noted in the literature is the potentially high cost of production. Metadata can be automatically-generated or human-created [10] with the latter clearly imposing more significant costs in terms of human effort and time. In the context of collections, [4] describes two types of roughly corresponding metadata: “inherent” metadata, or information that can be extracted from the resource objects themselves, such as total number of objects or total file-size of the collection; and “contextual” metadata, or metadata which involves human judgment to create, such as a textual description of a collection of resources.

There are significant advantages to utilizing inherent and template-based metadata as much as possible. Because it can be generated automatically or based in general information, metadata has minimal costs associated with its creation and maintenance and can be updated on a regular or automated schedule. In contrast, human-created metadata is time-consuming, error-prone, costly to create, and more likely to be inconsistent. A person assigned to create metadata may only perform this task on an occasional, as-needed basis, and it may be a lower priority task than others for which that person is also responsible. Inconsistencies in metadata assigned to resources can arise due to variations in a given cataloger’s judgment over time and because different catalogers may make varied judgments in cataloging resources.

There are drawbacks to relying solely on inherent and template-based metadata to define virtual collections, however. A risk in complete automation is the loss of many of the benefits of creating virtual collections, and also the application of metadata is mainly limited to content. Our first observation is that such a risk and application of metadata can not describe dynamic objects such as multimedia elements. Metadata can not influence multimedia content itself, because metadata usually contain universal and widely applicable description of objects. Contextual metadata is important because it enables us to give some character and cohesiveness to the virtual collection. Human created metadata is thus vital for articulating the scope, intent, and function of a particular collection, attributes that are likely to make the collection easier to locate. The use of metadata and human judgment in selecting resources to be included in a virtual collection has other benefits. Virtual collections can be described in terms of expected use in addition to being characterized by the terms they actually contain. It can be described in a dynamic way in order to facilitate the I/O behavior of a dynamic element. Resources can be more carefully chosen for inclusion in a virtual collection, with consideration of expected use, resulting in a more concise collection of high-quality resources that is easier to for the user to search, visualize and browse.

However, it is important to recognize, that a collection-level schema that relies heavily on contextual and dynamic metadata is relatively costly to implement and thus less likely to be maintained in the long term. A more viable approach is to define a schema for virtual collections that balances the costs and benefits of each type of metadata. In short, a cost-effective schema should include useful inherent and template-based metadata, supplemented by contextual and dynamic metadata that captures human judgments of a collection’s nature and the selection of criteria for inclusion in the collection. From our point of view, the use of both inherent and contextual metadata schemas requires a new sort of metadata that includes useful, inherent, dynamic and template-based metadata supplemented by human judgments to facilitate the behavior of dynamic elements in digital collections.

5. Virtual Collection in U-campus Project

As stated in the specification of the IEEE’s learning object metadata and according to [15], “a learning object is defined as any entity, digital or non-digital, which can be used, re-used or referenced during technology-supported learning”. Examples of learning objects include multimedia content, instructional content, instructional software and, software tools referenced during technology-supported learning. In a wider sense, learning objects could even include learning objectives, persons, libraries, universities, organizations or events. A learning object is not necessarily a digital object; however the reminder of this article will focus on
learning objects that are stored in the digital library’s virtual collections.

The U-campus digital library that we are developing contains learning items and accepts contributions from anyone, subject to review before being made publicly available. Substantial collections of resources have been contributed to the digital library by a single person or organization. These collections include a group of English learning material that has been digitalized as part of second language learning program, a collection of images and related information of computer supported collaborative learning and, substantial contributions of video from some recorded Japanese TV programs of English.

Sub-collections resources of the U-campus digital library can be found through various searching and browsing mechanisms. However, for reasons discussed earlier, we felt that creating virtual collections within personal and group spaces to represent the contributed sub-collections would benefit both the contributors and the users of the digital library. Specifically, our primary motivations for developing virtual collections were to highlight the work of authors/creators who contributed a critical mass of materials on a topic, to streamline the creation of item level metadata, to customize learning objects and to provide users with another way of accessing, storing, visualizing and understanding the items available in the digital library within personal and group spaces.

5.1 Defining a Collection-Level Schema

IEEE description schema was chosen by EduVC in U-campus project because the set of elements was universal (it wasn’t created to meet the needs of a specific digital library), yet provided the flexibility for customization, if needed. The IEEE schema was also selected because it is similar to the Dublin Core schema [3]. Dublin Core is a common item-level schema used by many digital libraries, which would facilitate mapping elements and exchanging data. Previous work from the IEEE and the Collection Description Focus [6] resulted in thorough documentation, which facilitated understanding and implementing the schema in a relatively short amount of time. Other projects currently use the IEEE schema to describe large, unrelated, relatively static physical collections in a digital environment. Our contribution is to use and extend the IEEE schema using inherent and contextual static and dynamic metadata to describe “born digital-learning” objects of varying granularities, with varying relationships and at varying stages of collection growth in a digital library.

EduVC formulated requirements used to select IEEE elements, and, more generally, to measure the success of implementing the IEEE schema. Useful collections descriptions can be created dynamically in personal spaces by identifying a subset of elements relevant to users, by ensuring that metadata is complete within a collection description and consistent across collections and by presenting descriptions in an easily understood interface. Low-cost metadata creation can be accomplished by harvesting metadata automatically (template-based metadata), by requiring the collection creator, rather than a cataloger, to describe their collections and by providing an efficient cataloging tool.

Using the IEEE schema, collection descriptions were created for EduVC digital library and its sub-collections. It was important to identify the metadata source (item-level record, collection creator, subject-area reviewer) to track the cost of creating static metadata. By starting with the complete schema we identified elements that aided understanding the collection (especially for learning proposes). This process also identified elements, which were not included in the collection record interface and the collection cataloging tool being developed. The resulting subset of elements met our requirements: collection descriptions could be created and extended with minimal cost while providing sufficient information to aid discovery.

5.2 Implementation of the Collection-Level Schema

The IEEE schema contains lots of elements. Originally, U-campus digital library project implemented a subset of thirteen. After another iteration of testing and design, we extended these thirteen to twenty-one IEEE elements. The element subset records of information about collections were implemented as template-based metadata (catalog id, description, access policies, relationships to other collections, and collection owner contact information). The template-based subset was chosen because the initial cataloging process consistently yielded data for these elements. The subset also matched the types of data reflected in item-level records. This provided users with consistent information between object (item) and collection.
During the initial implementation, some elements were not included in the subset type. Because the IEEE schema has been used to describe physical collections, the developers created a controlled vocabulary to distinguish between collection types. We used the type element during the initial cataloging process, but found that some collections were not often of the same type, so the same vocabulary terms were used repeatedly with no distinction. Also, the terms would have to be explained to collection contributors and users, which could be a barrier to cataloging in personal or group spaces using collections. Recently, however, U-campus digital library project has incorporated the type element into its schema as a means to distinguish between virtual collections created for different spaces and purposes, such as collections organized around a specific personal space contributor and collections containing resources from different personal space contributors (group spaces) intended for a special purpose, such as a test collection.

Vocabulary for Type, which classifies collections by curatorial environment, content or policy, EduVC created a new vocabulary more appropriate to its resources. The cost of creating collection-level metadata can be reduced by automatically populating template-based fields in the collection description. In EduVC, "manually-entered" metadata is provided by the collection creator via a cataloging tool or by a subject-area reviewer when the collection-level metadata is examined. "Automatically populated" template-based metadata is derived from querying specific fields of item-level resources within the collection. In our case, collection descriptions are completely comprised of contextual metadata that is manually entered either at the item or collection level. Currently three fields in the subset can be automatically populated with template-based metadata from the item-level description. For collection description to be cost-effective the cost of item-level metadata creation must be minimized and more fields in the collection description must be dynamically populated. Implementing a cataloging tool within personal and group spaces in a usable interface for collection contributors is another way to reduce the cost of creating collection-level metadata. The prototyped collection cataloging tool shown in Figure 5, metadata for other fields will be supplied in drop-down menus with standardized vocabulary or text boxes that can be modified and extended by collection contributors or reviewers dynamically. This will ensure consistency in collection description.

Also, a well-designed interface with clear instructions should minimize the cost of metadata creation in terms of a contributor's time. For example, when a collection record is rendered in XML, the elements retain their IEEE attributes; however, field names were changed on the interface (IEEE attribute "Concept" becomes "Keyword"; "Super Collection" becomes "Collection is Part Of"). EduVC hopes to pass the majority of the cataloging costs on to its collection contributors as a trade-off for having the collection publicized and also incur some cost through the involvement of the subject-area reviewer as they error-check metadata and recommend changes.

One aspect of metadata creation that U-campus digital library contributors and EduVC subject-area reviewers share is identifying the relationships between collections and expressing them through the relational fields (Contains Sub-collections, Collection is Part Of, Related Collections). These relationships can be applied to collections of varying sizes and granularity, as in Figure 4, which shows the relational fields of EduVC and the U-campus digital library project.

Figure 4. Collection relationship

As collection-level metadata becomes widely used, we believe the relational attributes will be essential not only for discovering resources within personal and group spaces repositories, but also across digital libraries. However, the larger and more distributed the digital libraries become, more difficult will be for users to find valuable resources and the (often small) collections they need to represent in their personal or group spaces. By explicitly representing not only a wealth of virtual collections, but also the relationships among them, regardless of their physical location or collection-level metadata schema we need to improve the navigability of digital library.

6. Conclusion

The collection-level metadata schema that we have developed and extended has started to be tested in our U-campus digital library project enabled us to define virtual collections that benefit both library users and collection-providers in several ways. But in a broad
sense, the most important beneficiary may be the U-campus digital library itself.

Virtual collections encourage us to see a digital repository not as unitary structure, but as virtual and modular representation of learning objects that comprise many sets of resources, some small and others large, some separate and others overlapping, some stable and others transient, some defined by the library managers and others extended dynamically by library users. We think this is a compelling perspective. In fact, large scale digital libraries are increasingly adopting just such a modular structure.

There are a number of reasons this perspective could be attractive to other projects. In the first place, as we have noted, it is often costly to create metadata. Object level metadata is the most costly of all, since it describes the “atomic” digital learning objects in a collection. However, it is often unnecessary to incur this cost: for example, when all members of a set of learning objects are similar, item descriptions are redundant. In such cases, collection-level descriptions will be more cost-effective than object-level metadata. However, a given repository may have some distinct objects, as well as sets of similar components and services. This means that descriptions of resources in a collection should neither be fixed at a low level of granularity (object-level metadata) nor at a high-level (complete collections), but must change as needed. In other words, a cost-effective way of describing a collection will require the flexibility of virtual collection metadata schemas within personal and group spaces such as the one we have presented here.

The EduVC metadata-based framework also addresses the customization of learning objects within personal and group spaces. Having explained the extensions and challenges of metadata, we describe our implementation in U-campus digital library project. Technical details concerning the transaction of collection-level descriptions among federated repositories will also need to be worked out, if metadata is going to be shared across a distributed personal and group spaces in the U-campus digital library at low cost. Fortunately, many of the protocols that have been tested for learning object-level metadata should also apply straightforwardly to collection-level schemas as well. For example, the Metadata harvesting protocol [8] enables personal and group space’s collection to provide easily their metadata to services providers. By agreeing on a standard collection-level metadata schema it should be as simple for virtual repositories to exchange collection information as it now is for them to share learning object records.

7. References