Shared Bibliographies as Hypertext

Erik Wilde
ETH Zürich (Swiss Federal Institute of Technology)

TIK Report 224 (May 2005)

Available at http://dret.net/netdret/publications#wil05o

Abstract

The creation, management and dissemination of bibliographic information is a common task for almost all people working in a research environment, and it also is a (often weak) way of knowledge management. Current tools and methods are either centered on the process of document preparation using bibliographic references, or on the aspect of creating annotations and/or relationships describing bibliographic resources. As a result, bibliography management in many cases is still carried out with fairly simple tools and methods, and with little or no support for sharing the information. In the ShaRef project, the areas of document preparation, knowledge management, and information sharing among workgroup members are treated as equally important. As a result, ShaRef enables users to create, manage, and disseminate bibliographic information in a wide variety of use cases.

1 Introduction

In this paper, we present a bibliography management application that combines aspects of document preparation, knowledge management, and hypertext functionality. It is geared towards knowledge workers, which in our case primarily are researchers and affiliated persons (such as project partners and students). The motivation and foundation for the *Shared References (ShaRef)* project is an initial survey [14] covering 1'000 respondents (8'000 questionnaires were distributed) at the Swiss Federal Institute of Technology. The survey asked about the tools, methods, and expectations for working with references, where references were defined as being either bibliographic references or Web bookmarks.

As a part of the survey's results, Figure 1 presents the number of references (bibliographic/bookmarks and combined) per respondent, and shows that there is a significant amount of information managed by many respondents. In general, three quarters of the respondents were maintaining their own bibliographic information (a considerable share of the remaining quarter were university employees in non-research jobs, with no need for bibliography management). Among these, sharing the bibliographic information was almost as widespread as not sharing it (45% vs. 55%), but only a minor fraction of the people sharing their information did so using tools and methods for supporting them (11% of the people sharing their information). Most of the sharing was done informally, being based on simple approaches such as editing a shared file using a revision control system.

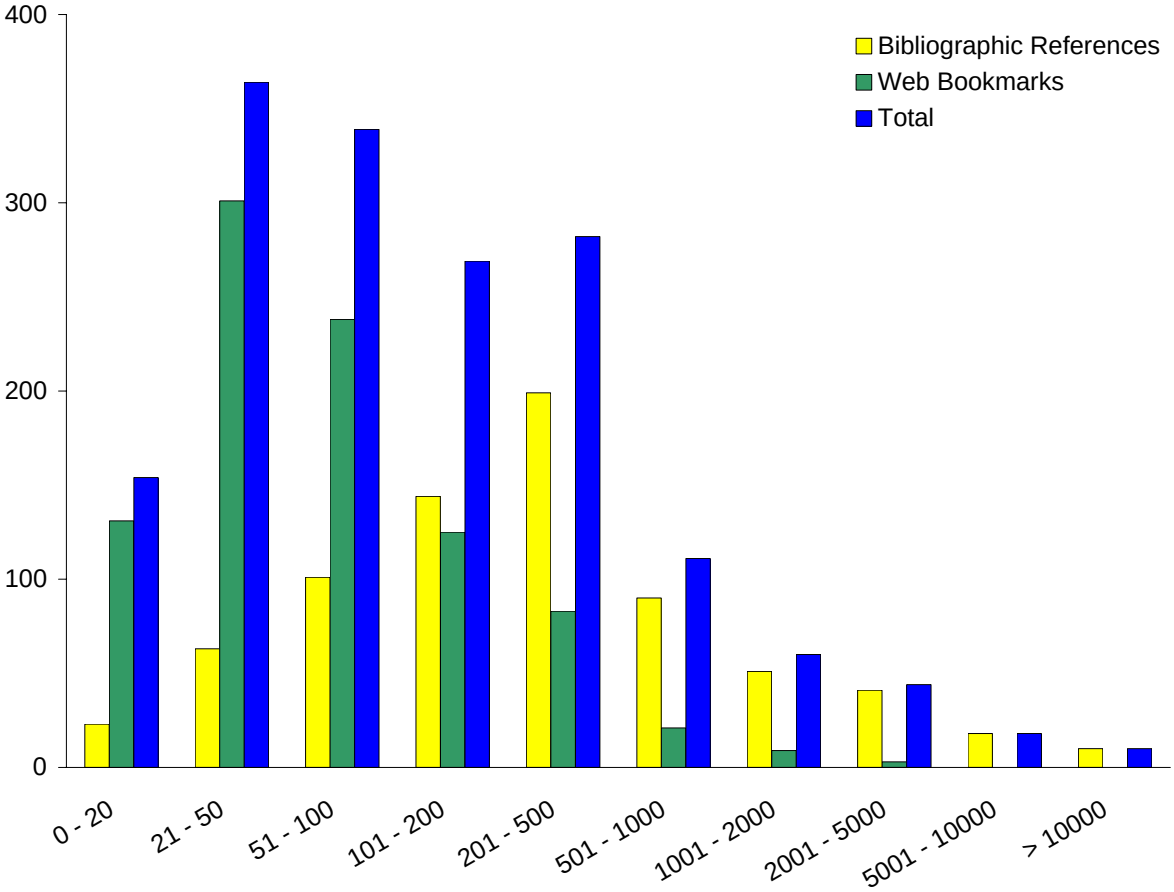

Figure 1: Number of References per User

As a result of the survey and in the context of a university-wide program "to develop and introduce technologies for communication and cooperation independent of time and place", a vision of a tool for better management of references and the sharing of the resulting information was developed [13]. The tool is supposed to be widely usable and thus features a Java client capable of online and offline operation, as well as a Web-based front-end which enables browser-based access to bibliographies. The overall goal is to support individual researchers (by providing them with a powerful tool), research groups (by making information sharing a core idea of the project), and university settings (by making it easy to reuse information, for example as a reading lists for a lecture or seminar).

2 Bibliographies and Hypertext

Bibliographic data, or more generally metadata about resources, is primarily a question of the metadata model and the features implemented on top of this model. ShaRef has a special focus on the interconnected nature of resource metadata, which in fact turns ShaRef into a hypertext application. However, because ShaRef is not primarily a hypertext application, the goal has been to find the optimal balance between the following three issues:

1. *Support for Document Preparation:* The main reason for most researchers to maintain bibliographic information is document preparation. Since different document preparation systems require different bibliography formats (MS Word works well with EndNote, LaTeX works well with BibTeX), the preferred document preparation system often dictates the choice of bibliography software.

2. *Knowledge Management:* In addition to the metadata required for document preparation, additional information about resources may be used to describe the resources and their relationships. Classical examples of resource descriptions are annotated bibliographies. The relationships between resources may be described in relations such as "cites", "updates", "confirms", or "disproves".

3. *Information Sharing:* While document preparation and knowledge management are often carried out individually, there are also many occasions when sharing of the acquired information would enable researchers to work more efficiently, such as when maintaining only one bibliography for a research group, or when publishing parts of a bibliography as reading lists for lectures, summer schools, or the publication list of the own research group.

It is important to note that ShaRef is not meant as a library-scale tool, which means that bibliographies in ShaRef are not meant to have hundreds of thousands of even millions of entries. Instead, it is meant as a personal bibliography management tool with sharing capabilities, so that individuals (with hundreds and maybe thousands of entries) can share and/or combine their information to access bibliographic entries in the magnitude of tens of thousands.

Because of this decision to design ShaRef as a personal tool, one of the most important design decisions was to keep it as simple as possible. Researchers typically want to concentrate on their work rather than the tools they need for managing their work[1], and thus the model and the interface must be mainly self-explanatory and usable in a way which can be easily explained without requiring a lot of training.

Because ShaRef's goals go beyond the capturing of metadata, and include hypertext functionality in addition to the resource metadata model, there must be some "glue" to actually connect the individual resource metadata instances to a hypertext structure. This glue is provided by two concepts, which are not generic hypertext concepts, but application-specific concepts for the area of bibliography hypertext:

- *Keywords:* Keywords are used to classify metadata records in many metadata formats. Some formats treat keywords as free text with predefined semantics (thus enabling search features on this text), while other formats treat keywords as something being taken from a predefined and controlled vocabulary. ShaRef supports both notions, the first one through annotations which may contain any text, and the second through dedicated keyword references, which point to keyword definitions. These keyword definitions are also items within bibliographies, and may be referenced within a bibliography, or across bibliographies (in which case the bibliography containing the keyword definitions effectively serves as a controlled vocabulary).

[1] In the survey, some of the respondents answered questions about their personal bibliography management tools and methods very enthusiastically, obviously being very concerned about this activity. However, the overwhelming majority of respondents regarded bibliography management as a chore, rather than an opportunity to capture some of their knowledge.

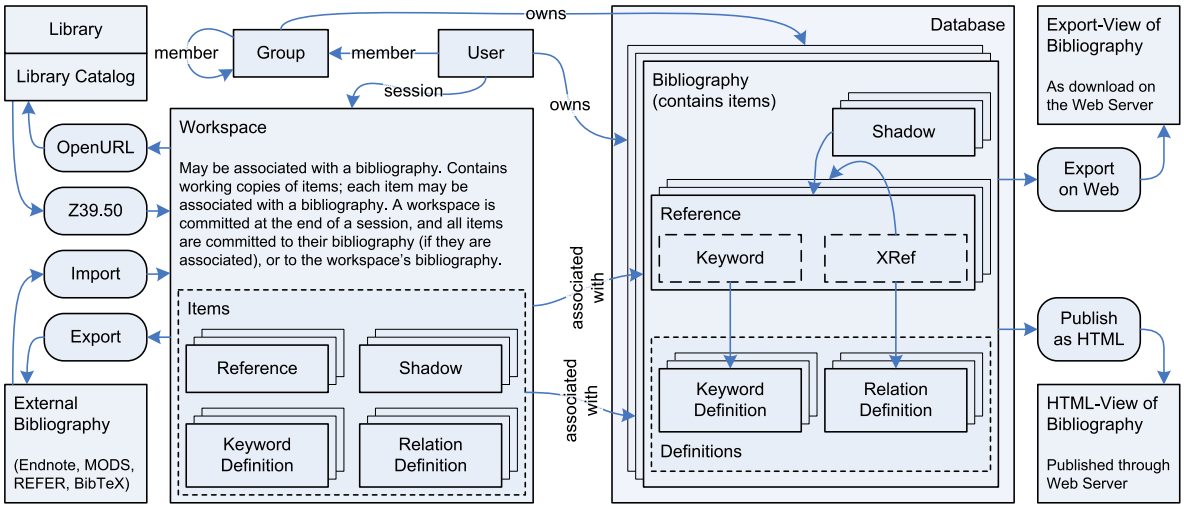

Figure 2: ShaRef Data and Functional Model

- *Relationships:* Whereas a keyword basically associates a reference with a property, a relationship associates two references. We decided to support binary relationships only, because the binary model is easier to understand and can be implemented with a simpler user interface. Relationships turn the individual references into a network of interconnected references. Relationships can be typed (by using defined relationship types, which like keyword definitions are part of a bibliography) or untyped. Relationships may occur on the reference level (in which case they associate the whole reference with another reference), or inside of rich text which is part of a reference. In the latter case, the relationship only associates the rich text fragment with another reference, for example in an annotation stating "this paper strongly suggests that the results reported by [SomeRef] may be misleading."

In addition to these two bibliography-specific concepts, ShaRef also supports normal Web-style links to Web resources, which accept any URI and can be used to make references to resources outside of the bibliographic data managed with ShaRef. Web links are unconstrained, the only feature that ShaRef supports are regular link checks, which can be used to report dead links to the owner of a bibliography.

3 Management and Sharing

As illustrated by the project's name, ShaRef is primarily concerned with "References" and "Sharing". In the following sections, these two main issues of ShaRef are discussed, first discussing the basic data and functional model, and then describing some of the sharing and collaboration features of ShaRef. These basic principles are illustrated with an application scenario in Section 5.

3.1 Managing References

The conceptual model presented in Section 2 is implemented through the data and functional model shown in Figure 2. In this figure, the individual components of bibliographies are shown, which are references, keyword and relationship definitions, and *shadows*. Shadows have not been discussed so far and are an important concept for ShaRef's use cases.

For supporting cooperating users and group bibliographies, it is often necessary to replicate information, for example when compiling a bibliography for a seminar from a research group's (much bigger) bibliography. Rather than copying the required references (and thus introducing the problem of how to keep the copies in sync, the classical problem of data denormalization), they can be created as shadows, which essentially are references to references. A shadow may have additional annotations or relationships associated with it, but apart from these additional descriptions it has the same properties as the original reference. If at any point a time the owner of a shadow decides that he wants to change the underlying reference's data, either the original reference must be modified (given the shadow owner has write access to the original reference), or the shadow can be instantiated and then becomes an independent copy of the original reference.

Shadows play an important role in many sharing scenarios, such as the example described in Section 5. They allow the reuse of information without the need to copy it, and they also allow to augment this reused information with additional descriptions. More on shadows can be found in Section 3.2.

The *Workspace* shown in Figure 2 is the metaphor through which users are interacting with the system, it is a temporary storage area in which references from one or more bibliographies are held, and may be worked with. In particular, all modifications of references take place through a workspace, which is synchronized with the actual bibliography through reading and committing references. Apart from working with workspace data through a GUI, data may also be imported and export through filters supporting different bibliographic formats, and through connections to library catalogs.

Even though user interaction takes place through a workspace, bibliography data may also be exported or published on the Web directly. This is achieved by the server side of the system (explained in more detail in Section 4) providing export filters which enable users to directly access a bibliography's contents over the Web. This access can either use an HTML representation of the bibliography (suitable for Web-based browsing), or users may download export versions of a bibliography, so that for example an EndNote or BibTeX version of a bibliography is always available as download and can be used for document preparation using these tools.

3.2 Sharing and Collaboration

While the data and functional model presented in Section 3.1 is sufficient for personal bibliography management, the second focus is on sharing and collaboration, because bibliographic information often could be more productively used when being shared among cooperating people. For supporting sharing and collaboration, the following concepts are important:

- *Group Concept:* Users have an identity and thus can be identified by the system. Groups can be defined which have either users or other groups as group members. A lightweight group concept is implemented, which means that groups can be created, modified, and

deleted by all users, and are the main metaphor for aggregating users and/or groups, and for controlling access to information.

- *Access Control:* Bibliographies are owned by users or groups, and the owners have write access to the bibliography. In addition to the owners, a bibliography may also be made accessible as read-only to other users or groups, or to the general public.

- *Shadows:* Shadows are pointers to references and enable reuse of reference. In particular, shadows are useful for reusing references in new contexts, without losing the association with the original reference. In most cases, users will create shadows of references from authoritative sources, such as a carefully maintained research bibliography, which can serve as a source of references for shadows in personal bibliographies. These personal bibliographies then reuse group information, and may add additional information (new references and/or personal annotations to shadows), thus fostering as much reuse as possible.

- *Messaging:* In addition to working with multiple bibliographies and using ownership and access control mechanisms, messaging is also supported to improve collaboration. Bernheim Brush et al. [3] show that increasing the number of communication channels improves the effectiveness of messaging, so messaging can be either used inside of ShaRef, or it can be redirected to e-mail. Messages can be sent as a means of ad hoc communications, or they notify users of certain events which have been detected by the system, or triggered by other users.

- *Publishing:* While users can always browse bibliographies using ShaRef, publishing features make it possible to provide access to bibliography data through a Web-based interface (an example is shown in Figure 4). This makes it easy to access bibliography data with a Web browser, and thus makes it easy to share data with anybody having Web access.

Apart from these user-oriented concepts, another goal has been to enable the reuse of bibliographic data in other contexts, so that it does not need to be re-created. The export facility can be extended by integrating a user-defined XSLT program, so that any export format can be generated. XSLT has been chosen because it is ideally suited to transform the internal XML data model into any required format (XML or text based), and because it is fairly easy to author for an increasing number of people. The exported data can then be reused in different contexts, for example as a publication list in a Web content management system.

4 System Architecture

In order to implement the data and functional model presented in the previous section, the design must be based on an open and extensible data model which can be easily used for defining new mappings. To accomplish this, ShaRef is built on top of an XML data model, which is defined by an XML Schema and can be easily extended. Through a mostly declarative mapping process [1], import and export can be defined fairly easily, which is important because ShaRef should be open and easily integrate into existing environments using specific data

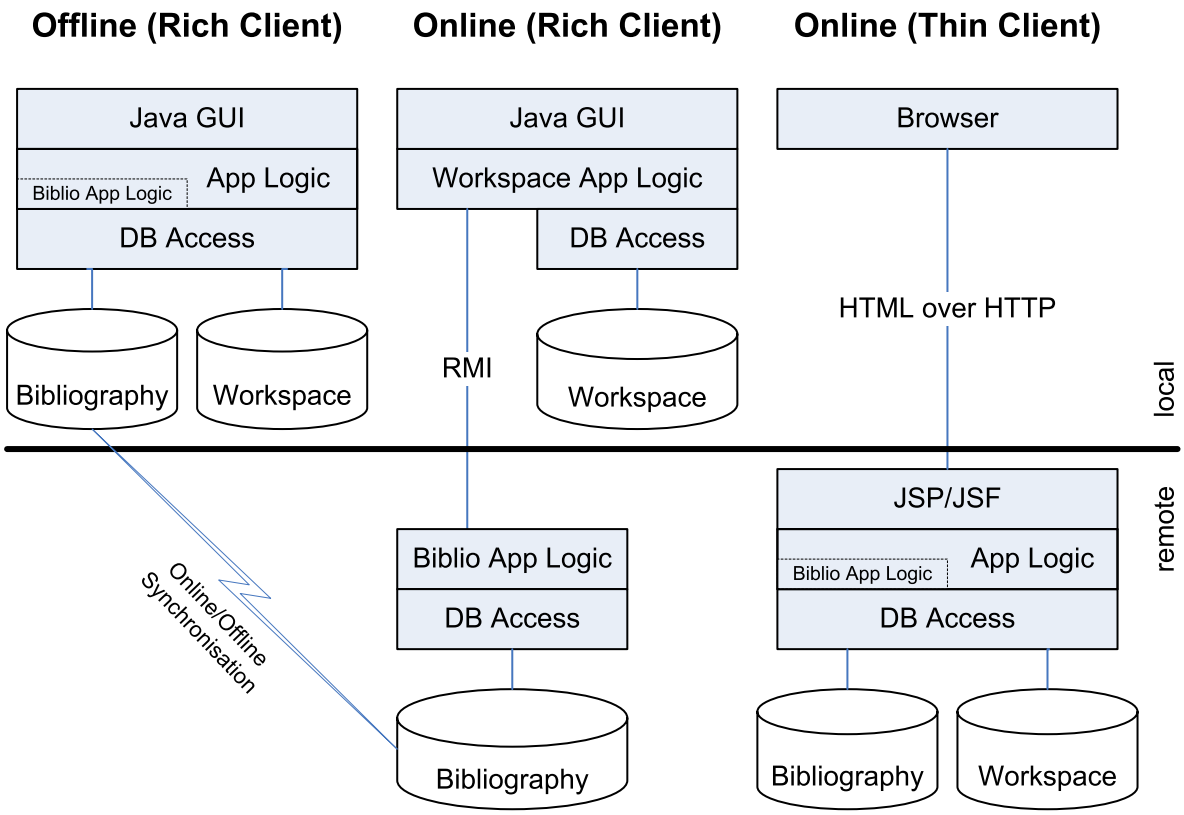

Figure 3: ShaRef System Architecture

models. Mappings for the most wide-spread bibliography formats are available and thus make it easy to use ShaRef with legacy data.

One comment frequently made by the respondents of the initial survey was that any new tool must not lock them in, in the sense that it should be operating system and document preparation system agnostic, and that it should make it easy to export personal data when no longer using the tool. By choosing Java as the implementation platform, and planning for a Web-based interface as well, the system is as independent from any specific operating system as possible. Independence from document preparation systems is achieved by providing bibliography management features and diverse and flexible export methods (through the Java client as well as through the Web interface), so that users can easily import the data into the document preparation workflow of their choice. Finally, by choosing XML as the underlying data model, we can guarantee that exporting native ShaRef data does not lose any information, and can then be transformed into whatever new tool the departing user chooses to use.

Another important requirement voiced by many survey respondents was the ability to work offline. While ShaRef is based on the idea of shared data, and thus needs networking capabilities, it must also be possible to work with ShaRef when being offline, such as from home or on a business trip.

Figure 3 shows the ShaRef system architecture and the different configurations which have been planned for the software components. The Java rich client is a client/server based archi-

tecture implemented using Java RMI, which uses some client-side application logic (workspace management being the most important), as well as some server-side application logic (bibliography management being the most important). However, the bibliography management application logic may also be used on the client side, when working in offline mode (in which case the bibliographies are stored locally). ShaRef uses a pure Java DBMS[2], on the one hand for managing the workspace data (as in-memory database), on the other hand for taking bibliographies offline (in which case a local database is created on the hard drive).

Switching between online and offline operation is a database copy operation from the server to the client, while going from offline to online requires a synchronization process, where only new or modified references are transferred to the online database.

Apart from the Java client, ShaRef also provides a Web-based interface, which can be used for accessing ShaRef bibliographies using a Web browser. This makes it very easy to start working with ShaRef, and also provides an ideal platform for publishing bibliographies, which can be configured for dedicated groups (such as a research group or the participants of a seminar), or world-wide, so that the bibliographic data is available for everybody (Figure 4 shows an example of a ShaRef Web page). The Web-based interface does not offer the full functionality of the Java client, so it can be regarded as a low barrier-to-entry way of using ShaRef, while the Java client is better suited for users wishing to use the full functionality of ShaRef.

In many settings (including the example detailed in the following section), the Java client can be regarded as the console for managing bibliographies and administrative data, such as access rights, and the question if and how bibliographies should be made available on the Web. For the daily read-only access to bibliographies, the Web interface provides a convenient way of browsing bibliographies, using indices (i.e., a set of generated index Web pages) for a better overview of the bibliography data.

Through the *bibconvert* service[3], ShaRef's import and export filters are available in the form of a Web-based conversion service. With this service, users can easily transform their bibliographic data between different formats (including an HTML export option for a Web-based presentation of a bibliography), which for some users is all they want, so they do not need to install the Java software to use the conversion functionality.

5 Working with ShaRef

As an example for ShaRef and in particular ShaRef's hypertext capabilities, one of the system-defined bibliographies uses many of ShaRef's features. It is the bibliography of all IETF RFC documents, a collection of over 4'000 documents which is constantly growing, used by researchers from various disciplines (mainly computer science), and heavily cross-linked through different types of relationships between the RFC documents.

The RFC bibliography is available in our basic ShaRef installation as a centrally maintained, world-readable bibliography, so that every user may use it. Using it means that users can simply browse it, using the hypertext properties to quickly move through the network of relationships between RFCs, or to access the RFC full text from the IETF's Web site. Using the RFC bibliography can also mean to create a shadow of a reference from the RFC bibliography in another bibliography, thus reusing this reference in a new context without losing

[2]The HSQLDB open source software available at http://hsqldb.sourceforge.net/.
[3]Available to the general public at http://dret.net/bibconvert/.

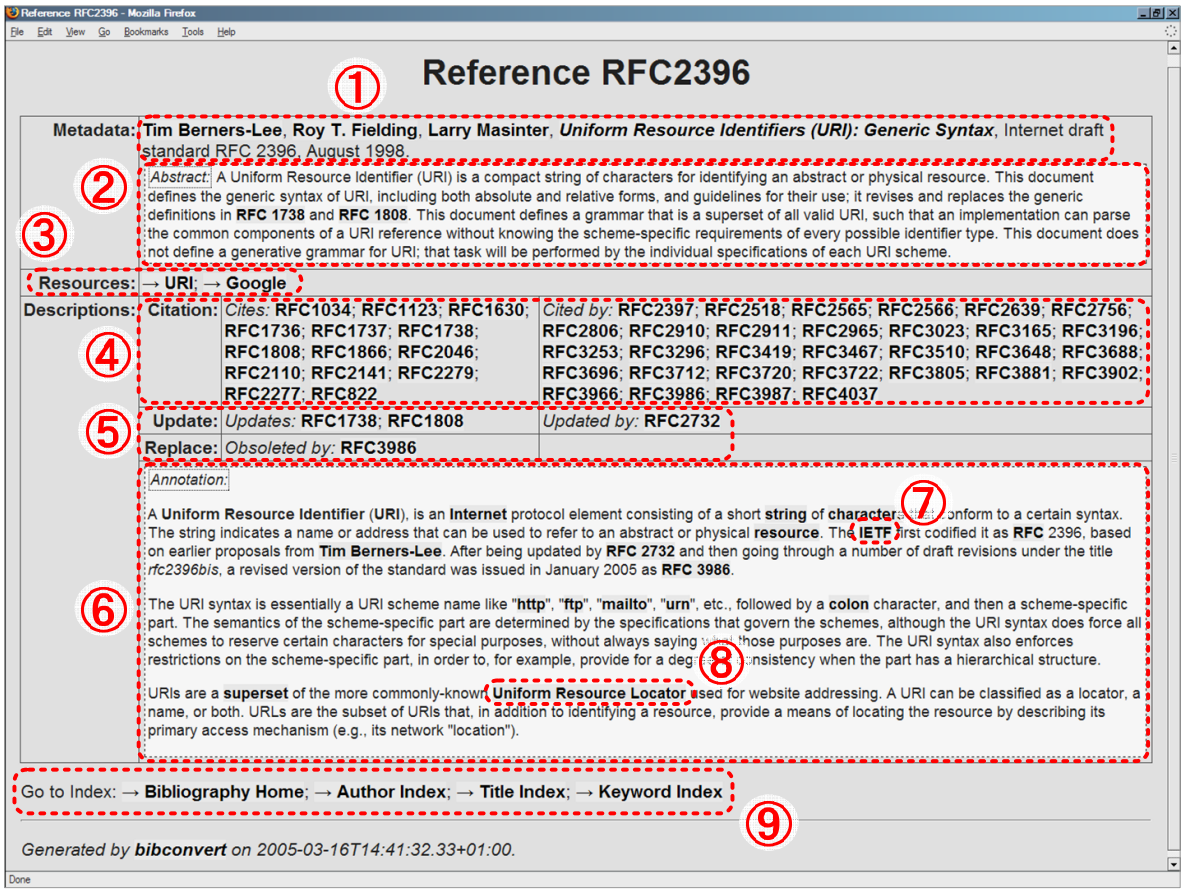

Figure 4: HTML Presentation of a Bibliographic Reference

the connection to the original entry, so that any updates in the central RFC bibliography (such as new RFCs which are updating or obsoleting the shadowed RFC) are also visible for the shadow.

This pattern of a centrally maintained, well-kept bibliography of a certain area, in this case RFC documents, and its reuse in the form of shadows in individual or group bibliographies, is a common usage pattern in ShaRef. The result is that not only bibliographies are hypertexts in the sense that they contain interconnected information, but that the hypertext approach spans across multiple bibliographies, thus forming a bigger hypertext with relationships between different bibliographies.

As a side note, this approach of using multiple bibliographies for information easily extends beyond references to keyword and relationship definitions (as described in Section 2) as well. This means that a given set of keyword definitions can be used as an individual bibliography, essentially turning this bibliography into a controlled vocabulary for keywords. Any keyword used in bibliographic references then points to this vocabulary, ensuring that references use a consistent set of keyword definitions.

The RFC bibliography is compiled from the RFC index[4] as well as from the RFC full

[4] Available as an XML document at `ftp://ftp.isi.edu/in-notes/rfc-index.xml`.

texts. The index provides the basic information about document titles, authors, status, and publication dates. Additional information available from the index are relationships between RFCs, the index provides information about *updates* and *replacement*. An updated document is affected in the way that it should be read in conjunction with a newer document, whereas a a replaced (or *obsoleted*) document is completely replaced by a newer document and is no longer the relevant document for the given topic.

The RFC full texts are necessary because the XML index does not contain citation information (indicating which RFCs cite other RFCs). Since the RFCs are text documents following strict formatting rules, compiling a citation index from the full texts can be done fairly easily. This citation index enhances the usefulness of the RFC bibliography, because it introduces an additional layer of hypertext information, and thus makes the RFC bibliography easier to navigate. For the example shown here, a third source of information has been used, the Wikipedia, which contains descriptions for some of the technologies defined by the RFCs.

Figure 4 shows the (recently obsoleted) URI specification [2] from the RFC bibliography. This bibliographic reference has all properties that make it an interesting resource from the hypertext point of view. In the HTML presentation of this reference (available through HTML export or the Web publishing feature of ShaRef, as shown in Figure 2), the following information is displayed:

1. *Reference Metadata:* This is a formatted representation of the reference metadata contained in fields such as *author*, *title*, or *date*. This information is used for producing citations for document preparation, it is captured by all tools and formats for bibliography management.

2. *Abstract Text:* As additional metadata, the resource's abstract can be stored, which is particularly useful when using search features for an entire bibliography.

3. *Links to External Resources:* Generated from the metadata and configuration information, links to external information embed the bibliography hypertext into the bigger context of the Web (by referring to known resources, such as online versions of a resource, or linking to search engines for finding online versions of resources) or even the local library.[5]

4. *Citation Index:* This is relationship information which represents document citations. In this example, the citation index is generated from a citation analysis of the RFC full texts.

5. *Other Relationships:* These are relationships as well, but not on the document citation level. In this example, this information was easily obtained from the XML version of the RFC index.

6. *Annotations:* Annotations can be added to further comment on a resource, and in this case the annotation text has been copied from the Wikipedia (using a simple mapping procedure from the Wikipedia's HTML-based text model to ShaRef's rich text model).

[5] This is not shown in the figure, but for books and articles, OpenURL [12] links are generated from the reference metadata and the OpenURL configuration, which provide a very convenient way to access the local library's catalog from within a bibliography hypertext.

Annotations support a structured content model supporting a minimal amount of text structuring and link embedding.

7. *Links to Web Resources:* Annotations may contain links to Web resources, which embed the bibliography hypertext into the bigger context of the Web.

8. *Keywords:* Annotations may contain keyword references, which refer to a keyword definition inside a bibliography. Keywords may be defined within the same bibliography, or they may be reused from a centrally maintained bibliography containing a controlled vocabulary of keywords.

9. *Bibliography Indices:* As backlinks to the starting points for further exploration of the bibliography, links to the bibliography's home page, the title index, the author index, and the keyword index are provided.

This example demonstrates that with the use of appropriate information resources, bibliographies can be modeled as highly interlinked hypertexts. In addition to the linking inside of the bibliography domain (intra- and inter-bibliography), Web links and OpenURLs are pointing to resources outside of the bibliography domain, and thus enable a seamless integration of the bibliography browsing experience with resources available on the Web or from the local library.

While the RFC bibliography has been generated from the data available from the IETF, ShaRef's open data model makes it easily possible to transform data from other sources. Thus, ShaRef can serve as an integration platform, by transforming the bibliography information available in other formats, and then making it accessible (and publish it) through this common format and interface.

6 Discussion

The bibliography model and ShaRef project presented in this paper is an attempt to combine the three areas of bibliographic data for document preparation, knowledge management about bibliographic resources, and sharing and collaboration features for bibliographic information. In Section 6.1, related work is described from these different perspectives. Section 6.2 summarizes the main contributions of this paper. Finally, Section 6.3 describes future plans for the ShaRef project, which is scheduled to run until the end of 2005.

6.1 Related Work

From a principal point of view, the approach described in this paper is applying the principles of hypertexts to the application area of bibliographic references, or, more generally, metadata. Dalgaard [7] discusses this question of how hypertext principles are going to affect text archives, and come to the conclusion that there will be a "paradigm shift [...] from the text in the archive, where the archive is a classified collection of texts, to the archive as a network of texts."

In particular, Dalgaard uses Genette's [8] categories of *intertexts*, *paratexts*, and *metatexts*, and using this classification, the approach taken by ShaRef can be described as follows: Since the concept of intertextuality applies to the text of a resource itself, it is outside the scope of ShaRef, which is not concerned with the actual resources. The metadata kept within

a bibliography can be though of as the paratextual description of the text, however this is only true for the bibliographic metadata (part 2 in Figure 4, which is the information directly inferred from the resource), and not for the more descriptive parts. The descriptive parts, and in particular the annotations, constitute the metatextual information for a reference. (The remaining categories of *hypertexts* and *architexts* also defined by Genette are too abstract to be directly applicable to ShaRef, even though they could be represented through keywords and relationships.)

On the more concrete level, related work can be identified in the three main areas that are important for the work presented here, bibliographic metadata, knowledge management in the context of bibliographic data, and sharing and collaboration features.

Bibliographic metadata has been studied for a long time, and today's formats can be broadly categorized by either being tailored for the complex cataloging requirements of libraries, or on the other hand being more geared towards non-librarians, who should be able to work with bibliographic data, too. The classical bibliography formats used for many libraries are different versions and variations of the basic *Machine Readable Cataloguing (MARC)* format, with MARC 21 [9] being the latest version. On the non-library side, the formats implicitly defined by applications such as EndNote and BibTeX have become popular. The *Metadata Object Description Schema (MODS)* [10] is an attempt to create an XML-based schema which is a subset of MARC 21 and is suitable for use in non-library applications. Currently, using a bibliography metadata format is not so much a question of the underlying model, but a question of the available tools and future plans for the bibliographic data. Thus, we have defined our own schema after inspecting the most popular formats, and use mapping tables for importing and exporting other formats.

Canós and Mena [5] have created the *BibShare* system, which is compatible both with MS Office products and LaTeX. It uses a client/server-architecture similar to ShaRef, but does not provide online/offline support. Furthermore, BibShare lacks the import/export support of ShaRef and also does not provide Web publishing features, which makes it a less open approach in comparison to ShaRef. Finally, the data and functional model of BibShare is better suited to describe individual references, which means that the hypertext view of reference information is much more limited than in ShaRef.

The *bibliographic* project of the open source *OpenOffice* initiative is another attempt to define and implement a bibliography management tool. The work is in its early stages, but is seems as if the data model is going to be based on MODS. It remains to be seen how the rather complex model of MODS can be implemented in an application that is more geared towards the casual bibliography user than the advanced knowledge worker, with the latter being probably more willing to adapt to a more complicated but also more powerful data model.

While most tools and formats for bibliographic information have no or only weak support for distribution and/or collaboration, the *Reference Manager* product is different in the sense that it is explicitly designed for workgroups. The product is based on a client/server-model and supports Web publishing features. The area where the product offers no or little functionality is the knowledge management and hypertext facet of bibliographies, because references are modeled as single entities with no relationships or structured annotations.

As an example of a richer end more powerful model of knowledge management in the context of bibliographic references, the *ClaiMaker* system created by Uren et al. [11] is an interesting approach. ClaiMaker provides semantically well-defined types of links between resources, and then provides features to infer information from a network of resource metadata

connected by semantically typed links. ShaRef does not define any relationship types, instead it enables users to define their own types. From the ShaRef point of view, it would therefore be easy to define the ClaiMaker classification in a shared bibliography and then use it as a controlled vocabulary of link types, but this would still not enable users to use the kind of reasoning features that a system with predefined semantics such as ClaiMaker can provide.

On the knowledge management and collaboration side of the spectrum, Wu et al. [15] present a system for document co-organization. Using this system, users individually categorize their documents into hierarchies, and the system then computes a "common view" of the document organization, using the categorization of all users. The goal of the underlying research is to better understand how people organize documents into structures, use other users' structures, and what consensus structure is emerging from this process of individual and collaborative structuring .

Cadiz et al. [4] present an interesting study about the usage of annotations. Even though their study focuses on annotations to resources rather than annotations as part of resource metadata, it is one of the few studies to actually present results of document annotation utilization. Document annotation features today are available in many settings, there are several Web annotation services (some of them supporting sharing), and increasingly document management systems are also supporting collaborative annotations (for example Adobe's Acrobat or the Microsoft Office 2000 suite used in the study). An interesting result of the study is that the notification mechanism was something to improve in following versions, because it was an important factor in improving collaboration and not quite as powerful as users would have wished.

Carr et al. [6] describe the *Dynamic Review Journal (DRJ)* and its utilization in a specific application area. While DRJ is more about an integrated publishing process, taking into account not only the text of articles, but experimentation data as well, it demonstrates the ongoing efforts towards a more integrated process in the context of document management. In the same sense as DRJ can be regarded as an attempt towards integrating the information that is relevant to writing a scientific publication, ShaRef can be regarded as an attempt to integrate the information that is relevant for working with bibliographic references. Thus, approaches such as DRJ and ShaRef could complement each other.

6.2 Contributions

The work presented in this paper has several contributions to the field of bibliographic reference management, and in particular with regard to modeling this application area from the hypertext point of view.

- *Hub for Bibliographic Data:* By defining its own data format and mappings for the most popular bibliographic formats, ShaRef can serve as a hub for bibliographic data. By using the Web-based *bibconvert* service, users can take advantage of this functionality without having to install any software.

- *Bibliographies as Hypertexts:* The data model not only covers bibliographic metadata, but also supports hypertext features. By modeling and presenting bibliographic information as a highly interlinked set of metadata resources, users can easily navigate complex structures of interconnected resources.

- *Knowledge Management for Bibliographies:* References can be described by keywords and relationships, and this can either be done privately or informally, or controlled vocabularies can be used. In both cases, the keyword and relationship features can be used to capture knowledge by using semantically well-defined concepts.
- *Sharing and Collaboration:* Publishing features, messaging, and the ability to create shadows and inter-bibliography links make collaboration and sharing of bibliographic information easy.

The area of bibliographic information, knowledge management, hypertext, and collaboration so far has attracted only little research. In the digital libraries community, personal bibliographies are more or less out of scope, while in the knowledge management community, the application of bibliographic metadata is often considered as being too limited. However, seeing that working with resources and metadata about these resources is one of the few commonalities shared by virtually all researchers, regardless of discipline and organization, we believe that further research in this area and the provision of better tools could benefit a large share of the scientific community.

6.3 Future Work

While the architecture described in Section 4 is mainly geared towards human users, accessing ShaRef through some kind of UI (the Java GUI or a Web-based GUI), an additional goal of ShaRef is to make it accessible as a service through an API, in particular as a Web Service. The system architecture shown in Figure 3 is the starting point for defining an API which can then be used by other applications. A typical example for a use case for such a scenario is a *Content Management System (CMS)*, which needs the data stored inside ShaRef (for example, the publication list of a research group), but needs to reuse it for its own publishing features.

Currently, it is possible to manually export data for a CMS by adding an appropriate output filter (for example, we have implemented such a filter for the *Silva* CMS used at our university), but exporting the data and importing it into the CMS is a manual process, especially cumbersome when frequent updates are required. Integrating ShaRef through an API, so that the CMS could fetch the data whenever required, would allow better integration with other tools and services.

ShaRef is currently in its beta stage, which means that the *bibconvert* service (which basically is a standalone version of the import/export filters) is in production, while the Java GUI and the database part are in their testing stage. We have already scheduled some presentations at our university to present the tool to students and researchers, and we look forward to collecting and presenting usage data about the client as soon as the first version of the client (scheduled for mid 2005) has been released.

7 Conclusions

The intersection of bibliography management, knowledge management, and collaboration presented in this paper to our knowledge is a unique approach. To a large extent, this is based on the hypertext-like properties of the underlying data model, which enable users to

capture the complexity and connectivity which is inherent to any significant set of scientific literature.

Bibliography management can be considered as the one activity which brings together almost all researchers, and for many researchers it is the closest activity similar to any formal kind of knowledge management. By giving researchers better tools for bibliography management, extending them in the areas where many researchers are likely to see an improvement in their workflow and information management, we hope to convince researchers that bibliography management can be improved, and that they, their research group, and the interested public can benefit from these improvements.

References

[1] Sai Anand and Erik Wilde. Mapping XML Instances. In *Poster Proceedings of the Fourteenth International World Wide Web Conference*, pages 888–889, Chiba, Japan, May 2005. ACM Press.

[2] Tim Berners-Lee, Roy T. Fielding, and Larry Masinter. Uniform Resource Identifiers (URI): Generic Syntax. Internet draft standard RFC 2396, August 1998.

[3] A. J. Bernheim Brush, David Bargeron, Jonathan Grudin, and Anoop Gupta. Notification for Shared Annotation of Digital Documents. In *Proceedings of the SIGCHI Conference on Human Factors and Computing Systems*, pages 89–96, Minneapolis, Minnesota, April 2002. ACM Press.

[4] J. J. Cadiz, Anoop Gupta, and Jonathan Grudin. Using Web Annotations for Asynchronous Collaboration Around Documents. In David G. Durand, editor, *Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work*, pages 309–318, Philadelphia, Pennsylvania, December 2000. ACM Press.

[5] José H. Canós and Eduardo Mena. BibShare: An Interoperable System to Access and Maintain Bibliographic References. In *III Jornadas de Trabajo DOLMEN*, Madrid, Spain, November 2002.

[6] Leslie A. Carr, Timothy Miles-Board, Gary Wills, Guillermo Power, Christopher Bailey, Wendy Hall, and Simon Grange. Extending the Role of the Digital Library: Computer Support for Creating Articles. In *Proceedings of the 15th ACM Conference on Hypertext and Hypermedia*, pages 12–21, Santa Cruz, California, August 2004. ACM Press.

[7] Rune Dalgaard. Hypertext and the Scholarly Archive: Intertexts, Paratexts and Metatexts at Work. In *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, pages 175–184, Århus, Denmark, August 2001. ACM Press.

[8] Gérard Genette. *Paratexts: Thresholds of Interpretation*. Cambridge University Press, Cambridge, UK, March 1997.

[9] Library of Congress, Network Development and MARC Standards Office. MARC 21 Concise Format for Bibliographic Data, October 2003.

[10] Library of Congress, Network Development and MARC Standards Office. Metadata Object Description Schema (MODS), December 2003.

[11] Victoria Uren, Simon Buckingham Shum, Gangmin Li, John Domingue, and Enrico Motta. Scholarly Publishing and Argument in Hyperspace. In *Proceedings of the Twelfth International World Wide Web Conference*, pages 244–250, Budapest, Hungary, May 2003. ACM Press.

[12] Herbert Van de Sompel and Oren Beit-Arie. Open Linking in the Scholarly Information Environment Using the OpenURL Framework. *D-Lib Magazine*, 7(3), March 2001.

[13] Erik Wilde. References as Knowledge Management. *Issues in Science & Technology Librarianship*, (41), Fall 2004.

[14] Erik Wilde. Usage and Management of Collections of References. Technical Report TIK-Report No. 194, Computer Engineering and Networks Laboratory, Swiss Federal Institute of Technology, Zürich, Switzerland, June 2004.

[15] Harris Wu, Michael D. Gordon, and Kurt DeMaagd. Document Co-Organization in an Online Knowledge Community. In Elizabeth Dykstra-Erickson and Manfred Tscheligi, editors, *Extended abstracts of the 2004 Conference on Human Factors in Computing Systems*, pages 1211–1214, Vienna, Austria, April 2004. ACM Press.