Invited paper

Modelling emerging collaboration functions in Library 2.0

András Micsik
MTA SZTAKI
Department of Distributed Systems
micsik@sztaki.hu
http://dsd.sztaki.hu

Abstract

Digital libraries offer a practical place for collaboration and socializing with others, like traditional libraries did for hundreds of years. In this paper we examine how collaboration is used in current digital libraries, and what are the trends for new library services. Based on this, an enhancement of current digital library modelling is suggested, in order to help us to better understand and to better support the large variety of possible collaborative functions in digital libraries.

Keywords

Collaborative work, groupware, digital library reference model, Web 2.0, Library 2.0, social networking, social software

Introduction

Libraries always had social and cultural roles, and worked as a dedicated place for learning, research and discussion. The Internet era gave birth to digital libraries (DL), and digital libraries can also be a 'virtual place' for social and collaborative functions. However, most of these social functions are taking place outside of current DLs, as they were quite slow in adapting to the new social culture of the Internet.

In this paper, we investigate the possibilities for collaboration in digital libraries, and seek for a generalized model of these collaborations. We hope, that generic models and architectures provide a help for the development of better collaborative support in DL systems. The desired collaborative functionality of a DL can be quite different for each user group, yet we attempt to give an overview of such functionality in section 2. Section 3 describes the concept of Library 2.0, which closely corresponds to our goal of providing better collaboration support. Section 4 starts from DELOS DL Reference Model and suggests its enhancement with a new type of collaboration modelling.

Collaborative functionality in digital libraries

Collaborative functions can be grouped according to the typical cycle of information processing. Commonly, we identify the following steps in the cycle: first, necessary data is *collected*, next the collected information is *interpreted* and *analyzed*, then some new information is *synthesized*, which can be seen as the result of the cycle. The result is then prepared to be *presented* and finally it is *published*, so that the result reaches the desired audience.

Regarding the first phase, data collection can use a wide variety of sources nowadays. On-line search engines, on-line lexicons, or even video repositories help nowadays people wishing to find and learn new facts. In this

phase, users collect information objects and sources of information (such as digital libraries). They are helped by query engines, user interfaces for browsing and searching. Some additional artifacts appear here as intermediate results: queries, query answers (or results), links (references) to other information objects, annotations to information objects and experts. These all can play important roles in collaborative functions.

For example, in repositories supporting complex queries, the construction of a query can be a collaborative action. Query expressions can be shared with others, edited together, and re-used later in other queries (Morris 2008). Using WikiAnswers[1], people can post their questions publicly and collect answers from others. Virtual reference desks are getting more common for all kinds of libraries (Kwon 2007). Users can ask for information or live help using dedicated tools such as chat rooms, e-mails, Skype, etc. In some cases, the user with a question is routed to the proper expert, who can provide the answer. This means that the communication channel is re-routed to the expert. In other cases, the question is distributed through internal channels until it receives a satisfactory answer.

It is more and more common to share the results of individual queries and investigations. With CiteULike[2], one can easily share publication lists, while using faviki[3], information objects can be annotated, categorized.

The second phase of information processing is about the interpretation and analysis of collected data. This also includes the organization, categorization of collected data, and annotations, tags can be used again. Popular tools such as Picasa allow us to easily maintain a tagged photo collection locally, and publish parts of it. Nowadays everything can be tagged and social tagging/bookmarking together with folksonomies help us to keep our things in order (Farooq 2007).

More abstractly, the artifacts used in this phase are: annotations (tags) to information objects, personalized collections. Both of these types are often shared with others. An annotation can take very valuable forms such as a summarization, for which automatic tools are not really available yet. Annotations may include discussion threads as well. Exchanging opinions is essential in this phase, the feedback and a wider viewpoint provided by others can help us to form our own interpretation of the available information.

Ordering and annotating our data enables us to synthesize something new in the third phase. A new information object (or more) is created, for which a wide variety of tools can be used. The synthesis can also be a shared act, aided by joint authoring tools, such as Google Docs or Wiki sites. Abstractly, we have various human roles (author, reviewer, etc.) in this phase, and versions of new information objects. The result can not only be a document, but also new media. Here, the borders between the synthesizing and presenting phase are obscure, usually the results can be presented or published as they are. However, the presentation phase can enrich the result by embedding or linking external resources, such as illustrations.

[1] http://wiki.answers.com/
[2] http://www.citeulike.org/
[3] http://www.faviki.com/

There are many ways to publish new information objects: videos, audios, presentations or scientific papers can be easily shared with services like YouTube, SlideShare or CiteSeer. Within these services the users engage into a specific society, where authorship is only one aspect, but comments, connections and awareness of others are other very important dimensions. Blogging is a more personal way of publishing, but blogs are also heavily cross-referenced with each other. Wikis are sometimes less popular because of the proprietary (although simple) markup language used, but still the best solution for the collaborative maintenance of information. Podcasts or webinars involve the audio and video formats into the publication possibilities. Abstractly, the information objects are published into digital collections in this phase, where users can not only access, but annotate and re-use them. In the last phases complex specialized information objects appear in the list of used artifacts as the results of presentation and publication. Some examples are: blog or micro-blog entries, slide shows, wiki pages, annotated videos, etc. These objects are often segmented, but more importantly, they are interlinked with each other, by means of reference or embedding.

Previously we saw that collaborative functionality can take many different forms. Sometimes very similar functionalities have special names (e.g. tagging and annotating). However we can find some common characteristics among these. For example, the *asynchronous or synchronous* modes of communication are one of these characteristics. The difference between these two modes is slowly disappearing, yet, in case of editors, synchronous or asynchronous joint editing provides different user experience, and requires quite different implementation. In Section 4, we provide analysis about more common characteristics of collaborative functions.

Library 2.0

The term Library 2.0 appeared in 2005, a year after Web 2.0 popped up as a new buzzword. Library 2.0 (or shortly L2) is not only about applying Web 2.0 in libraries, but it has also a strong social aspect: how libraries can reach more people.

As Michael Casey (possibly the first one who used the term Library 2.0) wrote in his blog: 'L2 is, to me, a service philosophy built upon three things; a willingness to change and try new things; a willingness to constantly re-evaluate our service offerings; and finally, a willingness to look outside our own world for solutions, be they technology-driven or not (this is where Web 2.0 fits in).'[4]

The OPAC with L2 is still in a central position, yet it is no longer an isolated information source, but an entry to the web of world wide knowledge, with many connections in and out. Furthermore, Library 2.0 is not only about the OPAC, but about new services, which may let the information flow smoothly in all possible directions between librarians and users. Such services need careful design, and the idea is to make the users participate in the design and let them add their own services to libraries.

For the illustration of these nice concepts and theories, some real examples are listed to show what librarians think as 'L2 in practice':

[4] http://tametheweb.com/2006/01/defining_library_20_ii_is_it_m.html

- Raising current issues to the library users through his/her blog,
- Allowing users to comment on the catalogue contents,
- Creating social networks about visitors' special interests,
- Using Twitter to inform users about ongoing activities,
- Introducing online chat with the reference desk service,
- Using Podcasts, YouTube or Flickr to distribute information on community-interest topics.

Library 2.0 does not differentiate between books and digital content; it handles traditional and digital libraries uniformly, and as we can read in librarian blogs, the focus is on the social and collaborative services and their further invasion into libraries. Maness summarizes the key features of Library 2.0 as being user-centered, provides multi-media experience, socially rich and communally innovative (Maness 2006). He goes even further on this last issue, by saying that 'they must allow users to change the library'. In the next section we investigate what these necessary changes mean in terms of modelling of digital libraries.

Modelling collaboration

As we saw, in the new library following Library 2.0 patterns, almost anything can be a result or target of collaboration. The library has to be prepared to support emerging new collaborative or social functions, which appear quite fast. This requires a very dynamic software infrastructure for the library, with high level of customization support. We hope that our investigation can help to find the right basic paradigms or principles for such library software. For that, our goal is to summarize the Library 2.0 trends in a generalized model. First, we need to find a base DL model, upon which we can make our extensions.

There are few candidates for such model. The DELOS Classification and Evaluation Scheme (Fuhr et al. 2001) took a broader scope than the usual evaluation context, and therefore can be seen as a seed of a reference model. The model emerges from three non-orthogonal components of digital libraries: the users, the data/collection and the technology used. The interaction of these three defines the fourth component: usage. Each component connects the DL domain to different fields of research, with different interests and evaluation 'culture'. However, this model is not detailed enough to serve as the basis for our modeling effort.

The 5S model (Goncalves et al. 2004) is based on mathematical formalism, and has been used in various case studies, including the generation of a taxonomy of DL terms. According to this model a digital library consists of a repository, metadata catalogues, services and a society of users. The 5S refers to *streams* and *structures* for the construction of digital objects, *spaces* for the description of digital object collections and their interrelations, *scenarios* for the definition of how services and activities change the state of the system, and finally *societies* for the interconnection of roles and activities within the user community. However, we think that the scenarios and societies in this model are too abstract and do not provide a natural and easy way for the description of collaborative functions.

The DELOS Digital Library Reference Model (Candela et al. 2007) seems to be the most elaborated and most suitable for our purposes. It is segmented into six domains: content domain, user domain, functionality domain, policy domain, quality domain and architecture domain. Within the domains concepts and relations

are defined, which can be visualized well as concept maps. In the following, we will rely most of the time to the functionality domain of the DELOS DL reference model.

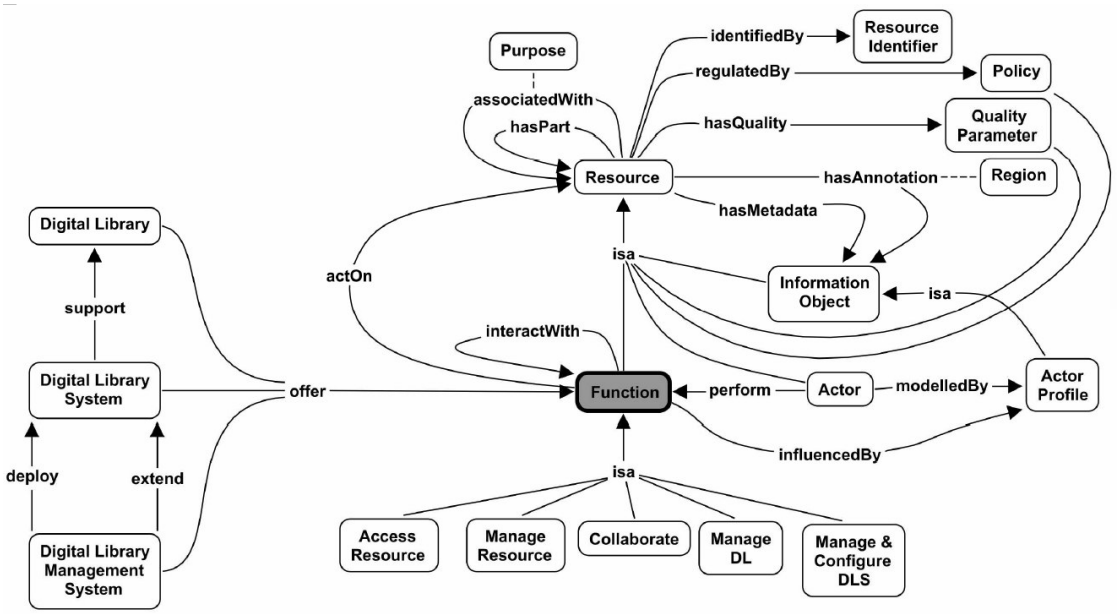

Figure 1 Functionality Domain Concept Map in the DELOS DL Reference Model (Candela et al. 2007:180)

In Figure 2, the concept of functions and functionality are presented as a concept map. The main concepts are Function, Actor and Resource. We can find Collaborate as a sub-category of Function, which can be provided by Digital Library instances.

For our further investigations we need a working definition of collaboration in a DL system: 2 or more people act together in the system towards a common goal. This definition reflects the fact that these persons are consciously collaborating, not by chance doing things which later prove to be useful, therefore, the common goal can be captured. Secondly, the collaboration is mostly supported by the system, and not by external tools. If people discuss on mobile phone how to fill in metadata fields, that is unavoidable, but need not be included in the model.

Figure 3 zooms in on the collaborative functions of the reference model. Four basic collaborative functions are identified:

- Find Collaborator: supports finding other Actors to collaborate with,
- Converse: supports conversation with others through the DL system,
- Exchange Information: supports the sharing and exchange of Information Objects in the DL system,
- Author Collaboratively: allows several Actors to edit Information Objects jointly and to create new Versions of these objects.

The DELOS DL model says: a Function provided by a DL System, performed by an Actor, acts on a Resource (which is an Information Object), or interacts with other Functions. On a too high level, the collaborative nature of Functions remains hidden as some sub-functionality. For example, 'Manage DL' is a function, which may contain internal workflows, yet the collaborative aspect is not revealed here. On a too low level, the composition of Functions creates the collaborative functionality (for example writing a message and reading a message enables chatting). If we look at the functions supporting blogging or social tagging functionality, they are just simple variations of very basic functions: create new information object, aggregate new information objects, read/list/sort information objects.

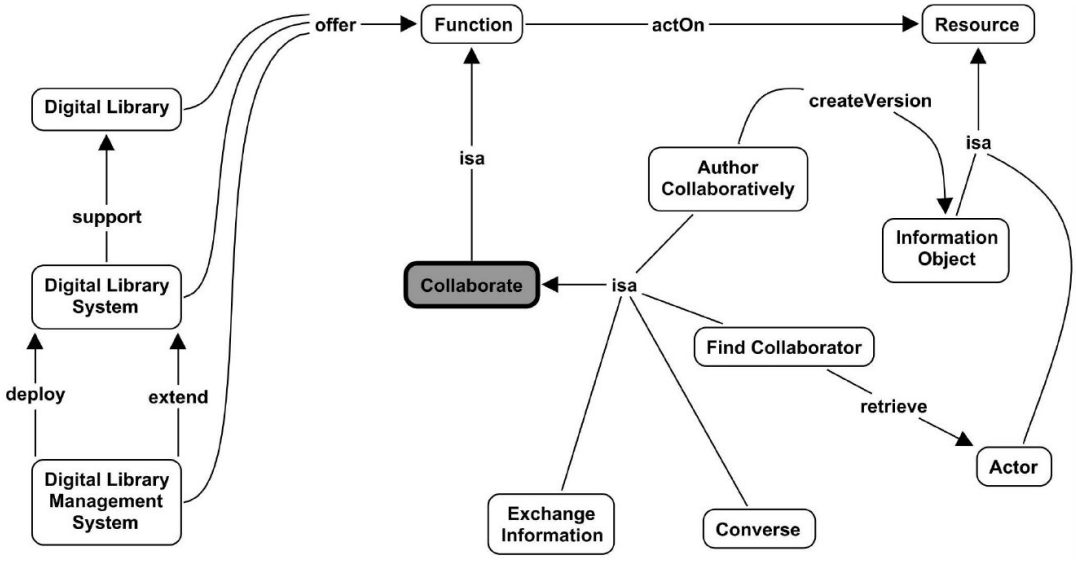

Figure 3. Sub-Functions of Collaborate in the DELOS DL Reference Model (Candela et al. 2007:186)

Therefore, to grasp the collaborative aspect, we need to focus on something else instead of functions. For example, the term collaborative space creates the necessary abstraction to characterize collaborations. We introduce the term CollabSpace (collaborative space) in the model as the grouping point of collaborative possibilities in certain scenarios (**Error! Reference source not found.**). A CollabSpace is supported by DL Functions, and itself supports some Goals of the collaborators. As an example, the virtual reference desk of a library is a CollabSpace, where users can find librarians, chat with them, ask questions, receive responses, etc.

All these are supporting functions, grouped together by their availability in various user interfaces, and also by sharing the same context.

The actors of CollabSpaces may form a Community, with relationships, reputation, history, and so on. A Community can be shared by more CollabSpaces. The Community is the result of the social functions, which determine the projection of social characteristics onto the community. For example, if users can rate each other, then this rating can be used to establish some kind of ranking for users.

A CollabSpace supports one or more Goals. These Goals justify the functionality of the CollabSpace. By analyzing the collaborative aspects of information consumption listed earlier, we can find several types of goals (or meta-goals):

- Information acquisition (e.g. question answering, virtual reference desk)
- Enrich content (e.g. joint authoring, annotated thematic collections, self-publishing)
- Improve quality (e.g. annotation or correction of bibliographic records, collaborative checking of digitized text, tagging of content to enhance quality of searching)
- Collecting or establishing common opinion (e.g. rating and tagging, discussions, chat rooms, feedback on service quality)
- Enhanced awareness (e.g. personal notifications on new contributions, available expert help)

These goals are achieved via functions acting on various information objects, some of those provide the data model for the collaborative functionality. We call them collaboration artifacts. In Section 2, the most typical examples of collaboration artifacts were mentioned. These artifacts can be very characteristic for the functionality (e.g. blog entry), and connects the DL world to already established paradigms of the social web. We can identify simple or compound artifacts, where the simple ones are usually parts of the compound ones.

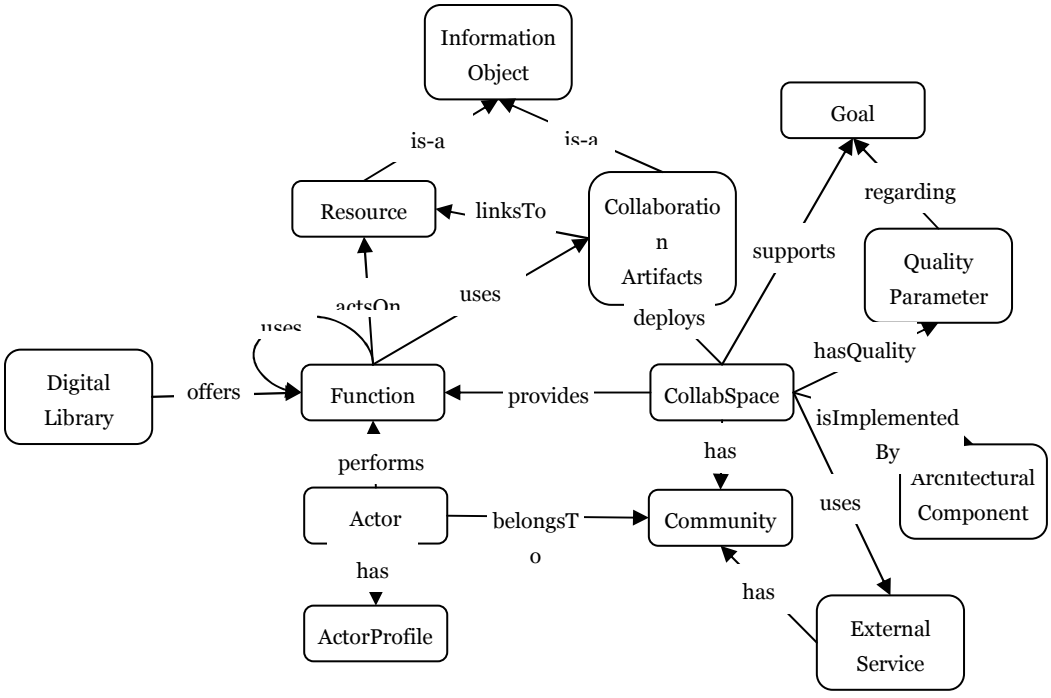

Figure 4. The concept of CollabSpace and its relations to other concepts

The simple ones are usually derivations of a message. Forums, chat rooms blogs use typical messages. An annotation is a message specially linked to other information objects. Tags, rates or votes are annotations with a strict semantics.

Compound artifacts represent aggregated views of simple objects. For example, a blog may contain blog entries, comments, tags, etc. As further examples, we can mention an annotated bibliographic record, a wiki page with several versions and corrections, or an annotated bibliography of a certain topic. It is very important to identify the links between these collaboration artifacts and the original DL resources.

Furthermore, we can observe some basic patterns in all collaborative functions. This can be tracked by the 'uses' relation in the concept map of **Error! Reference source not found.**. Most of the collaborative functionality can be composed of the following, so-called collaborative patterns:

Publish new information object: make the object accessible to a given group,

Aggregate new information: actualize, summarize and/or visualize certain characteristics of an object group (e.g. calculate the average rating or the tag cloud for a document),

Re-use existing information objects: make the connection explicit between a new information object and existing ones by defining their relations (e.g. reply in a discussion, create new version, cite, etc.).

We think that the abstraction of such common structures and patterns in the DL model helps us to better understand and to better support the large variety of collaborative functions in digital libraries.

There are also connections with the other domains of the DELOS reference model. The quality domain is connected via the Quality Parameter concept. CollabSpace, Community and individual Functions may have Quality Parameters, which reflect the subjective or objective measures of the implementation and the content. The quality is often assessed regarding the actual Goals supported by the CollabSpace, resulting in a triple: CollabSpace C has Quality Parameter Q for Goal G.

In the architecture domain of the reference model, an Architectural Component provides Functions for the Digital Library. Natually, a CollabSpace and its functionality are implemented by one or more Architectural Components, implementing the necessary functions and data model. The interesting question for future research is how these architectural components can share a flexible data model allowing them to compose and re-use collaborative functions in a service-oriented manner.

The integration and role of External Services is another related issue: we see that many social functions of DLs are outsourced to popular external services. Publishing photos on Flickr, sending notifications on Twitter, using a Facebook app are all handy solutions for DLs to extend their social functionality with little effort. However, doing this splits their collaborative spaces between their own system and external services, which may result in frittered user interfaces, or loss of social data (as the social interactions are stored in an external service). A common model for collaboration also opens the way to deal with such problems, for example by suggesting common interfaces and data models.

Conclusions

The current paper investigated the issue of modelling the collaborative aspects of digital libraries. We provided a structured overview of collaborative features used nowadays in digital libraries or in other web applications for information consumption. The great variety of emerging social use related to digital libraries means that almost every type of function can contain social features, and similarly, any type of information object can be the result or source of collaborative actions.

It is still a question how to model collaborations in the digital library. We propose an approach based on the DELOS DL reference model. The reference model is extended with 'collaboration spaces' representing the virtual spaces where collaboration may happen. These collaboration spaces aggregate the collaborative functionality of the system into a few spaces, making modelling of systems easier and more comprehensible. The structure of collaboration spaces was analysed, and some generalities were explored. We hope that such effort helps to better understand the collaborative aspect of DLs, and supports the improvement of DL architectures, which finally enables us to support collaboration more generally and also inject more collaborative functions into existing and future digital libraries.

References

Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., Pagano, P., Ross, S., Soergel, D., Agosti, M., Dobreva, M., Katifori, V., Schuldt, H. 2007
DELOS Digital Library Reference Model v0.98, 2007 December, http://www.delos.info/ReferenceModel

Farooq, U., Song, Y., Carroll, J.M., Giles, C.L. 2007
Social Bookmarking for Scholarly Digital Libraries
IEEE Internet Computing, Volume 11, Issue 6, Nov.-Dec. 2007 Page(s):29 – 35
http://dx.doi.org/10.1109/MIC.2007.135

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., Solvberg, I. 2001
Digital Libraries: A Generic Classification and Evaluation Scheme
ECDL 2001, September 4-9 2001, Darmstadt, Germany, Springer LNCS 2163, pp. 187-199
http://www.springerlink.com/content/w22qmkk09ncy59av

Goncalves, M. A., Fox, E. A., Watson, L. T., Kipp, N. A. 2004
Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries
ACM Transactions on Information Systems, 22(2), 2004, pp. 270-312
http://doi.acm.org/10.1145/984321.984325

Kwon, N. 2007
Public library patrons' use of collaborative chat reference service: The effectiveness of question answering by question type
Library and Information Science Research, 29 (1), pp. 70-91.
http://dx.doi.org/10.1016/j.lisr.2006.08.012

Maness, Jack M. 2006
Library 2.0 Theory: Web 2.0 and Its Implications for Libraries
Webology, Volume 3, Number 2, June, 2006,
http://webology.ir/2006/v3n2/a25.html

Morris, Meredith Ringel. 2008
A Survey of Collaborative Web Search Practices
In *CHI 2008*, April 5–10, 2008, Florence, Italy. ACM
http://research.microsoft.com/en-us/um/people/merrie/papers/collab_search_survey.pdf