

DEBORA: Developing an Interface to Support Collaboration in a Digital Library

David M. Nichols¹, Duncan Pemberton², Salah Dalhoumi³,
Omar Larouk³, Claire Belisle⁴, and Michael B. Twidale⁵

¹ Department of Computer Science, University of Waikato, Hamilton, New Zealand
daven@cs.waikato.ac.nz

² Computing Department, Lancaster University, LA1 4YR, UK
pemberto@comp.lancs.ac.uk

³ École Nationale Supérieure des Sciences de L'Information et des Bibliothèques,
69623 Villeurbanne Cedex, France
{dalhoumi, larouk}@enssib.fr

⁴ LIRE, Unité Mixte de Recherche 5611, CNRS-Université Lyon 2, France
claire.belisle@ish-lyon.cnrs.fr

⁵ Graduate School of Library and Information Science,
University of Illinois at Urbana-Champaign, IL 61820, USA
twidale@uiuc.edu

Abstract. Interfaces to library systems have largely failed to represent the inherently collaborative nature of information work. This paper describes how collaborative functionality is being implemented as part of the DEBORA project to provide access to digitised Renaissance documents. Work practices of users of Renaissance documents are described and the collaborative features of the client software are outlined. Functionalities discussed include annotation, the creation of virtual books and the inclusion of user-supplied metadata.

1 Introduction

This paper describes the development of collaborative functionality for users of digital libraries in the context of the EU *Telematics for Libraries* project DEBORA (Digital Access to Books of the Renaissance).

The aim of the DEBORA project is to make Renaissance books more generally available as digital resources and to examine the potential for novel collaborative functionality. The collection being created within DEBORA consists of digitised images of books from libraries in Lyon, Rome and Coimbra.

The first part of the paper outlines the nature of collaboration in digital libraries. Section 3 describes evidence gained from real life users of Renaissance materials. Section 4 describes the implementation of collaborative functionality in the DEBORA client followed by some initial user studies and a conclusion.

2 Collaboration in Digital Libraries

Digital Libraries offer new opportunities for collaboration and communication that were unfeasible in traditional libraries. Interfaces to information systems (including databases, library catalogues and information retrieval applications) have largely reflected single user stereotypes [27]. That is, the activities of other users have had almost no impact on the experience of any one user.

The technology associated with Digital Libraries allows us to consider new ways of working with library materials [13]. Specifically, for users to work in groups, rather than individually, and for users to add to collections, rather than simply reading.

2.1 Overview of Collaborative Work for Digital Libraries

Digital Libraries, in comparison with print-based libraries, more easily support the modification of their contents. Several researchers (e.g. [27, 8, 9, 10]) have recognised the potential for users, rather than librarians, to contribute to the development of a collection through user-supplied data (USD) [10]. Such USD can come in many different forms, although it can be split into two main groups: data automatically collected from users' activities and data explicitly generated by users. Implicit additions to a collection include: search term suggestion [10], ratings [19] and 'read-wear' [7]. There have been many proposals for explicit USD: annotation [30, 14, 2], keyword addition [4, 10], evaluative commentary [9, 10], hypertext links [3, 9, 8, 18, 15, 24, 23], ratings [23, 2], error correction [26] – for a review see [27].

The common thread amongst these ideas for collaboration is that one users' actions can be shared with other users *within the system* [27, 13]. In a paper-based library such sharing is more circuitous – via the publishing of cross-cited works that eventually physically arrive in libraries. The belief amongst researchers is that such collaboration will be more productive for the users [13]: e.g., by enhancing retrieval effectiveness through community rating of resources [2]. Although many forms of USD have been suggested the most common example is probably *annotation*.

Annotations. Annotation has been frequently proposed as a technique for users to add content (and so share ideas) within information systems – for a review see [22]. However, as Wilensky recently notes: “despite its evident usefulness, digital annotation capabilities are not very widespread” [30]. There are several annotation systems in different contexts [22] but the ‘broad territory’ of annotation [15], from free text to metadata, has not been conducive to the development of an accepted standard for annotations [22]. Marshall [15] characterises annotations along seven dimensions: formal/informal, explicit/tacit, permanent/transient, public/private etc. Wilensky [30] suggests four system requirements; annotations should be:

- Able to be placed *in situ*
- Expressive, extensible and composable
- Format and platform independent

- Free of permission and registration requirements

That annotations should be placed (and viewed) *in situ* accords with evidence from real world studies [14, 15]. The second and third requirements follow from the variability of users, usage contexts and documents; although they also reflect the generality of the *multi-valent document* approach [21] and the desire to support spontaneous ubiquitous collaboration [30]. The requirement that annotations should not be dependent on prior registration, illustrates a particular perspective on the usage of annotations. If we wish to collaboratively construct cataloguing information [4] (the formal end of Marshall's dimension), then user registration may be necessary to maintain the metadata authority.

The different interpretations of 'annotation' cover many of the forms of explicit USD. Consequently they share attributes with other novel collaborative functionalities: presenting significant added complexity to designers, a lack of accepted standards and their adoption would entail significant changes to users' work practices.

2.2 Designing for New Ways of Working

By studying existing practice and comparing it with evolving practice in other disciplines (especially in the sciences) that have enjoyed better computational resources over a longer period of time, we can explore the design space and create systems that 'add value' to existing work activities. Our design challenge is to support new ways of working by adding collaborative features, but they must be incorporated in a thoughtful manner. A digital library that allowed users to contribute and share information around collection items would be a radical change for many users – especially in a domain such as Renaissance texts. Humanities scholars mostly do not perceive themselves as working collaboratively [20]. Thus just asking them about their collaborative work is insufficient; that work must also be observed in order to see the substantial 'invisible collaboration'. The system must allow different kinds of use, from the currently conventional solitary forms of work, to supporting more effectively existing kinds of collaborative work, to supporting kinds of collaboration wholly new to this group. To be acceptable, the system needs to support graceful (and if necessary, slow) transitions in use along that scale.

3 Work Practice and Design Implications

Renaissance books represent a turning point not only because the advent of printing brought about a gradual re-organisation and standardisation of the textual material and its presentation [16], but also because the political, social and religious upheavals which characterise the century profoundly influenced the use of books.

Due to their rarity and fragility, the availability of Renaissance texts has generally been limited to acknowledged scholars. But even these researchers can have access

difficulties. The research work done on such book corpuses is basically comparing different versions, identifying and tracing originality and influences [25, 29].

At the same time, there is an increasing number of requests for access to 16th century material coming from a variety of users including: educators and their students, linguists, book historians, social and cultural historians ('histoire des mentalities'), specialists in literary studies, illustrators, wardrobe designers etc. [5].

This section summarises a usage study undertaken through observations, interviews and questionnaires with users familiar with Renaissance or other old books. The main findings are taken from the results of a survey through a written questionnaire (83 questions) answered by 62 scholars working on old or 16th century books.

3.1 Observations of the Use of Renaissance Documents

Working with Renaissance books implies, for specialists, accessing a specific copy, for there are almost no two identical copies that have survived (even copies of the same text from the same publisher). Each copy will reveal unique information on where, when and how it was printed, through explicit information in the book or through the book's material composition.

Two groups of users of Renaissance books can be identified: book specialists (who may eventually require access to the physical copy) and those interested in the content. This second (larger) group, including users of the existing digitised networked material [29], has working habits that can be summarised in four main areas:

- They need to be able to identify the specific copy of the book they are accessing and very carefully read the text or study the illustrations [17].
- While doing this, they take notes, either handwritten or on their computers, if these are allowed in the library (and if using local electrical outlets is permitted).
- Scholars find themselves alone, or almost, in their speciality: for example, the study of the 16th century dialect of Lyon in the writings of Paradon, a regional bishop. Consequently, they will not tend to exchange information - often because there is no-one to share it with. This may also be related to a university tradition of individual evaluation based on personal, and not collaborative, publications.
- Each scholar usually maintains a system of indexed cards where she stores all the information patiently gathered from many trips to libraries and archives; together with any personal notes. Some scholars now use computers for this task.

3.2. Current Collaborative Practices

There is an increasing recognition of the collaborative aspects of many forms of work [12, 27, 28]. We expect that the benefits potentially available from networked interaction will increase collaboration and modify the manner in which users work. We found that almost 80% of the scholars use some form of Web access to find and

exchange information. Although they rarely have the occasion to collaborate specifically in their research, email is used widely. These collaborations often involve passing on, to a colleague, information found by chance (serendipitous finds).

Collaboration in real time (synchronous conferencing) is not seen as an essential function. Answers focus on a more restricted sense of collaboration; finding out how the work of colleagues is going and asking information on certain aspects of their own work. A majority of scholars are willing to share their notes with colleagues and, of those who intend to collaborate, most express a desire for these tools to be integrated into the interface of the access software.

3.3 Metadata

Image collections, such as DEBORA, often lack detailed metadata. The librarians in the project are supplying typical book-level MARC metadata. Beneath this is the internal structure of a book, specific to 16th century texts: the location of indices, prefaces etc. However for effective retrieval using conventional searching the detail of individual pages (illustrations, decorative elements etc.) is needed to allocate indexing terms. Although basic structuring, such as differentiating illustrations from normal text, can be achieved with image analysis tools, most detail must be contributed by content specialists. As for most image collections, this amount of effort is infeasible. It is at this level, of detailed page-specific metadata, that collaborative contributions by the users of image collections could prove most valuable.

3.4 Design

Based on our findings, we focus on annotation mechanisms to explore collaboration. Annotation is already a part of existing solitary scholarly practice, and so potential users are most likely to be willing to learn to use the system in order to obtain the benefits of familiar work practices. This is our 'Trojan Horse' for studying new collaborative features based on those same annotations. Firstly, we explore the kinds of annotations that scholars find useful, through iterative prototyping.

4 The DEBORA Client

DEBORA is based around a client-server architecture with two distinct types of server. A Z39.50 based server is used for the storage of 'official' metadata – including the location of the catalogued images. In common with many other annotation systems a separate server is used to store and retrieve user annotations [22]. The client has two main functions: to provide access to the images of the collection and to support the collaborative functionality of the system.

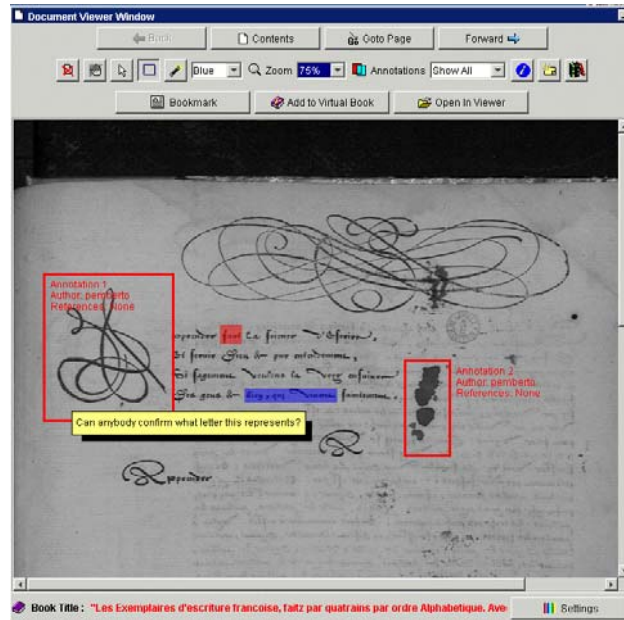


Fig. 1. The DEBORA client interface showing annotation and highlighting

4.1 Client Interface

The main window of the DEBORA client contains several image viewing tools (magnification, contrast etc). Fig. 1 shows an annotation attached to a rectangular area of a document. Annotations are currently free-text, as opposed to structured thesaurus terms [2]. The personal-public dimension of an annotation (author specified) is shown by colour and can be used to view a subset of all of the annotations.

The client also provides facilities for highlighting areas of an image in a variety of colours – typical of paper annotations [14, 15]. User-definable workspaces are provided to allow users with similar interests to structure their collaboration activities. Alternatively a user can restrict their additions to be personal and private.

Any set of annotations can be chained together to provide trails, or paths, [3, 24, 23]; following a path may involve moving to any part of any book in the collection. These hyperlinked annotations and associated images can be gathered together by a user to create a *virtual book*. We expect that this aggregation will help to reduce the adverse navigational effects of traversing a trail that spans many collection items.

Fig. 2 shows the virtual books in the lower left corner, beneath ‘real’ books from the collection. A user creates a virtual book by selecting elements (such as pages) from existing resources and arranging them in ‘virtual chapters’. This composite virtual document [18] can then become part of the collection for other users. A typical

educational scenario would be a professor tracing the historical development of an artistic style and collating examples into a virtual book for her students [24].

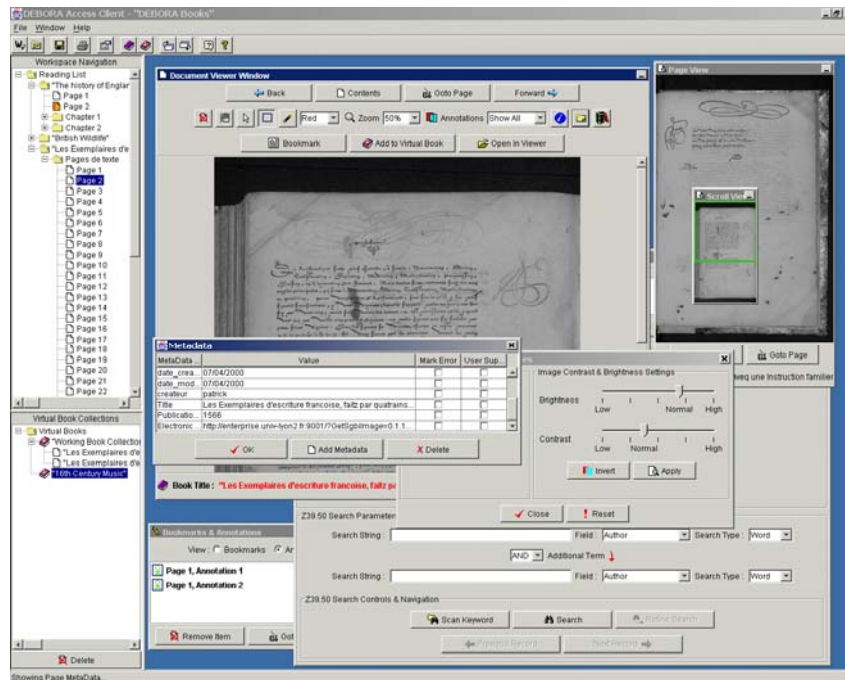


Fig. 2. The DEBORA interface showing the creation of a virtual book (bottom left)

The client currently displays virtual books separately from the main collection – however, if we are to take the promise of the digital library seriously then these books should be seen as *part* of the main collection. Extending virtual books to include items from other collections, and the issues of generating metadata for such documents, implies consideration of an extensible notation such as XML [22].

4.2 Collaboration and Metadata

In addition to shared annotations and virtual books there are at least three other methods for USD to enhance the metadata of the DEBORA collection.

Error Correction. Most database users have no way to record the presence of errors they detect in item descriptions. The client currently supports one-click boolean ‘error-present’ actions and allows users to suggest alternative descriptions. With a population of active users any data quality effort could prioritise those items with the most reported errors [26], or *error-wear* by analogy with *edit-wear* [7].

User-Supplied Metadata. In addition to correcting existing metadata, the client can accept new keywords from users. At present these terms are not easily integrated with metadata on the main DEBORA server but this is a small technical problem. A simple interface is required to allow human authorisation of additional metadata.

Re-purposing Annotations. The annotations are stored separately and so can be easily searched independently of the collection metadata. Annotation databases are potentially valuable sources of text [11, 22]; particularly for image databases. Golovchinsky et al. [6] use the text of freeform annotations as a source of query terms. Conversely, we can mine user annotations for indexing terms; as 'multimedia annotations...are simply meta-data associated with multimedia content' [1].

A major difference between conventional image annotation and this approach is the purpose of the annotation: collaborative annotations are not *intended* to describe the images. Although re-purposed annotations will generate index terms of lower quality than an expert cataloguer, it may well be better than their complete absence.

When user supplied data is used in conjunction with 'official' descriptions then search tools need to be aware of the differences in term authority [4]. One approach would be to attach less weight to USD terms in query matching. If the USD is accepted by collection maintainers then this increased trust could be reflected by increasing the weights of other USD terms from the same user [26].

5 Initial User Studies with the Client Interface

The client has been scrutinised by several 16th century specialists and has yielded information on its functionality. Virtual documents, where each user creates her own document by selecting pieces of one or several books, are considered essential. A virtual book is perceived as answering important and specific user needs, and is seen as close to ways of working with physical documents. Annotation is appreciated as a fundamental functionality allowing collaboration. The three levels provided, private, group and public, all address specific user concerns. Users expressed concern about managing large numbers of annotations and recognising different annotation types.

The facility to have two pages side-by-side, or two illustrations taken from two different copies of the same book, eases comparison. This is particularly appreciated by specialists who need to authenticate a dubious copy or restore a copy.

The zoom function is seen as very useful. It allows a rapid assessment of the content of the different pages presented and is also an effective replacement for leafing through a book. The possibility of rapidly going back or forward facilitates memorising and global comprehension of the document. Users working on illustrations or typography suggested an option showing the actual size of a page element. These specialists also appreciated the highlighting facility that augments the contrast and sharpness necessary in their tasks. Some expressed fear that digitising books would result in a clean image that would not be sufficiently faithful to the original.

6 Conclusions

The paper considers how collaborative features can be added to a digital library. One pertinent problem is that the intended users do not perceive themselves as working collaboratively (even when they do), and so are unlikely to see the benefits of collaborative features. We are exploring the provision of annotation features as a mechanism to support a graceful transition from solitary use to collaborative work.

Future work will involve a continuation of the refinement of the basic annotation features for conventional use, along with an exploration of the additional collaborative use of these same annotations (e.g. as a source of index terms).

Conventionally, rare texts and their annotations were separate, in DEBORA we have brought them together. It will be important to integrate other aspects of work: the organisation of annotations, the creation of virtual books and texts written by users of the digital library. Such a richer environment will not only promote easier switching between its components, but it will itself afford new forms of collaboration, and in a networked digital library mean that the user's office then becomes as mobile as the Renaissance books themselves have become.

Acknowledgements

The DEBORA project is funded under the EU *Telematics for Libraries* programme.

References

1. Barger, D.M., Gupta, A., Grudin, J., Sanocki, E., Li, F.: Asynchronous Collaboration Around Multimedia and its Application to On-Demand Training. Microsoft Research Technical Report 99-66. (1999)
2. Bouthors, V., Dedieu, O.: Pharos, a Collaborative Infrastructure for Web Knowledge Sharing. In Proceedings of ECDL'99. Springer Verlag, Berlin (1999) 215-233
3. Bush, V.: As We May Think. *The Atlantic Monthly*. 6 (1945) 101-108
4. Chandrinos, K., Immerkær, J., Dörr, M., Trahanias, P.: A Visual Technique for Annotating Large-Volume Multi-Media Databases – A Tool for Adding Semantic Value to Improve Information Filtering. In Proceedings of the 5th DELOS Workshop: Filtering and Collaborative Filtering. ERCIM, Le Chesnay, France (1998) 125-129
5. Chartier, R.: *Les Usages de L'Imprimé*. Fayard, Paris (1986)
6. Golovchinsky, G., Price, M.N., Schilit, B.N.: From Reading to Retrieval: Freeform Ink Annotations as Queries. In Proceedings of SIGIR'99. ACM, New York (1999) 19-25
7. Hill, W.C., Hollan, J.D.: History-Enriched Digital Objects: Prototypes and Policy Issues. *The Information Society*. 10 (1994) 139-145
8. Kantor, P.B.: The Adaptive Library Network Interface: A Historical Overview and Interim Report. *Library Hi Tech*. 11 (1993) 81-92
9. King, G., Kung, H.T., Grosz, B., Verba, S., Flecker, D., Kahin, B.: The Harvard Self-Enriching Library Facilities (SELF) Project. In Proceedings of Digital Libraries '94 (DL'94). Texas A&M University (1994) 134-138

10. Koenig, M.E.D.: Linking Library Users: a Culture Change in Librarianship. *American Libraries*. 21 (1990) 844-849
11. Lawton, D.T., Smith, I.E.: The Knowledge Weasel Hypermedia Annotation System. In *Proceedings of the Fifth ACM Conference on Hypertext*. ACM, New York (1993) 106-117
12. Leplat, J.: L'Analyse Psychologique du Travail: Quelques Jalons Historiques. *Le Travail Humain*. 56 (1993) 115-131.
13. Levy, D., Marshall, C.C.: Going Digital: A Look at the Assumptions Underlying Digital Libraries. *Communications of the ACM*. 38 (1995) 77-84
14. Marshall, C.C.: Annotation: from Paper Books to the Digital Library. In *Proceedings of the 2nd International Conference on Digital Libraries*. ACM, New York (1997) 131-140
15. Marshall, C.C.: Toward an Ecology of Hypertext Annotation. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*. ACM, New York (1998) 40-49
16. Martin, H.: *Histoire et Pouvoirs de L'Écrit, Bibliothèque de L'Évolution de L'Humanité*. Albin Michel, Paris. (1996)
17. Martin, H.: *La Naissance du Livre Moderne, XIVE – XVIIe Siècles*. Electre, Paris (2000)
18. Myaeng, S.H., Lee, M., Kang, J.: Virtual Documents: a New Architecture for Knowledge Management in Digital Libraries. In *Proceedings of the Second Asian Digital Libraries Conference*. National Taiwan University, Taipei (1999) 171-181
19. Nichols, D.M.: Implicit Ratings and Filtering. In *Proceedings of the 5th DELOS Workshop: Filtering and Collaborative Filtering*. ERCIM, Le Chesnay, France (1998) 31-36
20. Palmer, C.L., Neumann, L.J.: *Exploration and Translation: the Research Work of Interdisciplinary Humanities Scholars*. *Library Quarterly* (to appear)
21. Phelps, T.A., Wilensky, R.: Multivalent Annotations. In *Proceedings of ECDL'97*. Springer Verlag, Berlin (1997) 287-303
22. Ovsianikov, I.A., Arbib, M.A., McNeil, T.H.: Annotation Technology. *International Journal of Human-Computer Studies*. 50 (1999) 329-362
23. Röscheisen, M., Morgensen, C., Winograd, T.: Beyond Browsing: Shared Comments, SOAPs, Trails, and On-line Communities. *Computer Networks and ISDN Systems*. 27 (1995) 739-749
24. Shipman, F.M., Furuta, R., Brenner, D., Chung, C., Hsieh, H.: Guided Paths Through Web-Based Collections: Design, Experiences, and Adaptations. *Journal of the American Society for Information Science*. 51 (2000) 260-272
25. Sordet, Y.: *Repérage et Navigation dans L'Espace du Livre Ancien*. Communication Présentée au 1er Forum de L'Édition et de la Documentation Spécialisé. (1997)
26. Twidale, M.B., Marty, P.F.: *An Investigation of Data Quality and Collaboration*. Technical Report UIUCLIS--1999/9+CSCW, University of Illinois at Urbana-Champaign. (1999)
27. Twidale, M.B., Nichols, D.M.: Computer Supported Cooperative Work in Information Search and Retrieval. In Williams, M.E. (ed.): *Annual Review of Information Science and Technology (ARIST)*, Vol. 33. Information Today, Inc., Medford, NJ (1998) 259-319
28. Twidale, M.B., Nichols, D.M., Paice, C.D.: Browsing is a Collaborative Process. *Information Processing & Management* 33 (1997) 761-783
29. de Ventabert, G.: *Représentation et Exploitation Électroniques de Documents Anciens (Textes et Images)*. *Document Numérique*. 3 (1999) 57-73
30. Wilensky, R.: Digital Library Resources as a Basis for Collaborative Work. *Journal of the American Society for Information Science*. 51 (2000) 228-245

Nichols, D.M., Pemberton, D., Dalhoumi, S., Larouk, O., Belisle, C. and Twidale, M.B. (2000) DEBORA: Developing an interface to support collaboration in a digital library, *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2000)*. LNCS 1923. Springer. 239-248.

http://dx.doi.org/10.1007/3-540-45268-0_22

This is an author-created version. The final publication is available at www.springerlink.com