

# Evaluation of a Digital Library System

Maristella Agosti, Giorgio Maria Di Nunzio, and Nicola Ferro

Department of Information Engineering  
University of Padua  
{agosti,dinunzio,nf76}@dei.unipd.it

**Abstract.** Starting from the general picture of the evaluation of a Digital Library, we place ourselves in the System/Technology view of a Digital Library. In this context we point out some issues regarding the design of an evaluation infrastructure that need to be addressed, and we consider whether the 5S model is a possible framework for analyzing those issues.

## 1 Introduction

Digital library systems generally include collections of multi-media digitalized data and services that help storage, access, retrieval and analysis of the collections of data. The final aim of the DELOS network is of “enabling any citizen to access human knowledge any time and anywhere, in a friendly multi-modal, efficient and effective way, by overcoming barriers of distance, language and culture and by using multiple Internet-connected devices” [1].

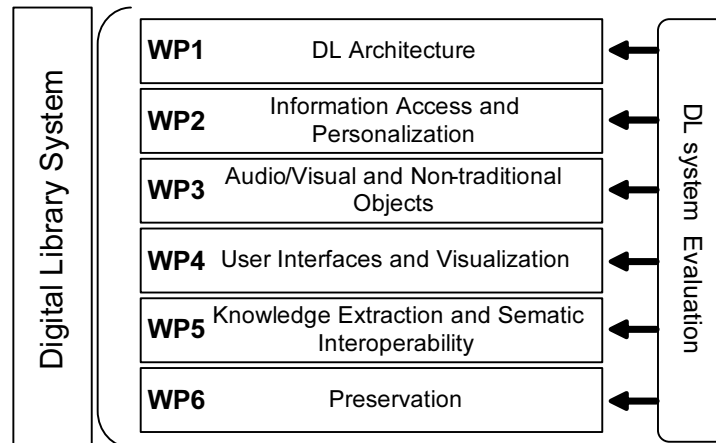


Fig. 1. Main aspects of a DL System

During the DELOS Evaluation Work Package (WP7) meetings, the fact that WP7 plays an important role emerged more than once. In particular, if we con-

sider the topics of the first six workpackages of the DELOS NoE as the main subjects that cover all the aspects of a Digital Library system then the evaluation of it is orthogonal to all the six points, shown in Fig.1.

Thus, the orthogonality of the WP7 with respect to the other WPs implies that every single aspect of a DL system needs to be carefully evaluated with appropriate methods and metrics.

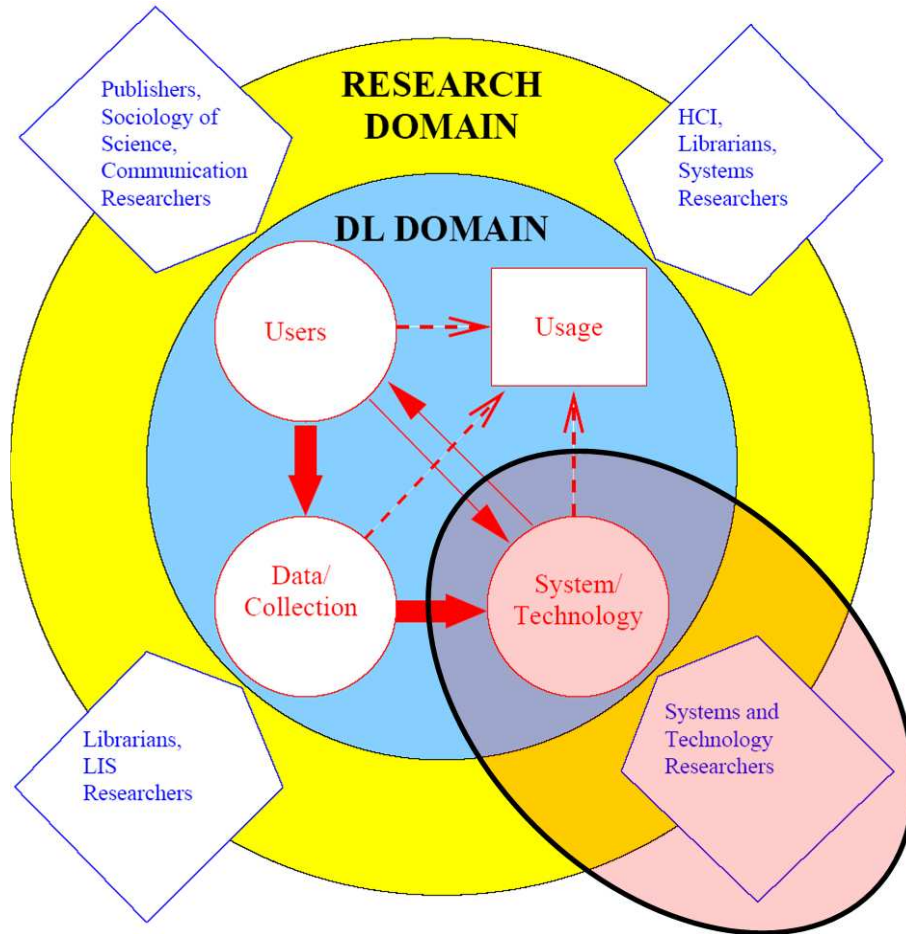


Fig. 2. General Digital Library Scheme (adapted from [2])

In particular, with respect to the the classification scheme of a DL given in [2], we will focus our attention on the System/Technology view of a DL, as highlighted in Fig.2, and especially we are interested in the evaluation aspects of information access and extraction evaluation tools [3, 4].

An interesting perspective to information access and extraction is provided by the exploitation of annotations. Indeed, if we consider annotations in the context of digital libraries, this will call forth new possibilities which will enrich the semantics and expressive power of annotations. We can exploit the digital library by creating cross-linked annotations with a free flow of information between annotations and documents. A digital library can encourage cooperative work practices, enabling the sharing of documents and annotations, also with the aid of special devices

Annotations introduce a new content layer devoted to elucidate the meaning of an underlying document and they can make hidden facets of the annotated documents more explicit. Thus, we can exploit annotations for retrieval purposes and add the evidence provided by annotations themselves in order to better satisfy the users information needs [5]. Also this kind of new service requires the design and development of appropriate evaluation methodologies and metrics.

## 2 The CLEF Evaluation infrastructure

Today evaluation campaigns happen giving the participants one or more test-bed collections, a set of tasks to be performed, and a method by which evaluating the performances of their system with respect to the defined collections and tasks.

This has been the main approach to the evaluation of Cross-Language Information Retrieval (CLIR) systems until the very last CLEF Workshop [6]. However, something seems moving towards a more mature and aware way of evaluation of performances. In the break-out and panel sessions of CLEF-2004 two important arguments were raised by the research community:

- Every year data collections change constantly. In this way, comparing performance of two systems (or even the same system) of two different years is impossible. It has been proposed to overcome the lack of stability of collections by fixing some reference collections that do not vary in time.
- Closely related with the problem above, there is also the lack of an historical/temporal vision of the evaluation of CLIR systems. So far, we do not have a clear view of where we are in the process of performance improvement: have we already reached the asymptotic point of best performances or are we still improving?

We argue that the WP7 should meet these needs and should discuss how to design and provide an evaluation infrastructure able to satisfy such kind of request.

In particular, the Streams, Structures, Spaces, Scenarios, Societies (5S) model proposed by [7] could be useful in analyzing and formalizing such kind of evaluation infrastructure. Indeed, this infrastructure has to address the needs of the community of researchers that carry out evaluation campaigns. This community can be modelled using the concept of *society*; furthermore, the evaluation methodologies and the evaluation tasks can be formalized by using proper *scenarios* that describe the interactions among the researchers and the evaluated

systems. Finally, the metrics and the data collected during the experiments can be described and represented by means of appropriate *streams*, *structures* and *spaces*. The 5S model not only provides the definition of these concepts and the basic syntax for using them [7], but also defines the relationships among them [8].

Thus it could be worth discussing whether the 5S is an appropriate model for facing this kind of issues and whether it could further a better understanding in this research field.

### 3 Acknowledgements

This work was supported in part by the DELOS Network of Excellence on Digital Libraries funded by the European Commission.

### References

1. DELOS: a Network of Excellence on Digital Libraries (2004) Technical Annex 1, pag. 6.
2. Fuhr, N., Hansen, P., Micsik, A., Solvberg, I.: Digital Libraries: A Generic Classification Scheme. In: Proceedings European Conference on Digital Libraries (ECDL-01). LNCS 2163, Darmstadt, Germany, Springer, Heidelberg (2001) 187–199
3. Cleverdon, C.W.: The Cranfield Tests on Index Languages Devices. In Spack Jones, K., Willett, P., eds.: Readings in Information Retrieval, Morgan Kaufmann Publisher, Inc., San Francisco, California, USA (1997) 47–60
4. Cleverdon, C.W.: The Significance of the Cranfield Tests on Index Languages. In: Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Chicago, Illinois, USA, ACM Press (1991) 3–12
5. Agosti, M., Ferro, N., Frommholz, I., Thiel, U.: Annotations in Digital Libraries and Collaboratories – Facets, Models and Usage. In Heery, R., Lyon, L., eds.: Proc. 8th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2004), Lecture Notes in Computer Science (LNCS) 3232, Springer, Heidelberg, Germany (2004) 244–255
6. CLEF-2004: Cross Language Evaluation Forum 2004 (2004) University of Bath, Bath, UK. Web-page: [www.clef-campaign.org](http://www.clef-campaign.org).
7. Gonçalves, M.A., Fox, E.A., Watson, L.T., Kipp, N.A.: Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. ACM Transactions on Information Systems (TOIS) **22** (2004) 270–312
8. Gonçalves, M.A., Watson, L.T., Fox, E.A.: Towards a Digital Library Theory: A Formal Digital Library Ontology. In Dominich, S., van Rijsbergen, C.J., eds.: ACM SIGIR Mathematical/Formal Methods in Information Retrieval Workshop (MF/IR 2004), [http://www.dcs.vein.hu/CIR/cikkekek/MFIR\\_DLontology4.pdf](http://www.dcs.vein.hu/CIR/cikkekek/MFIR_DLontology4.pdf) (2004)

