

Metadata in a European Perspective

Makx Dekkers, PricewaterhouseCoopers Technology Consultants, Luxembourg,
mdekkers@ip.lu

Abstract

This article looks at international issues related to metadata and digital libraries from a European perspective. Europe can be seen as a testbed for these international issues in a multilingual and multicultural environment, in a continent with a common cultural and scientific heritage. Important issues that are identified through experience of European co-operation are quality, sustainability and handling of rights. Solutions to problems in these areas can only be found through global co-operation and standardisation to establish a level playing field for information providers, technology vendors and users. Finding these solutions will establish a truly global information infrastructure to the benefit of everyone.

Keywords

Digital Libraries; Metadata; Multilingual; Multicultural; Global Information Infrastructure

The European landscape

In the European environment, many issues related to digital libraries in general and metadata in particular, come up in an international context. The European continent is essentially multilingual and multicultural, but throughout history has developed a common cultural and scientific heritage. The various treaties underlying the European Union [1] have now also established a common market in which this heritage is being made accessible across national boundaries.

Various activities in the European Union and associated states, from the Framework Programmes for Research and Technological Development to industry-related cross-border co-operations, have acknowledged this fact by addressing these aspects in a co-ordinated way, involving many commercial and non-profit organisations. In fact, the co-ordination between national initiatives and supranational activities allows the investigation of these aspects as part of a framework of international co-operation. Because of this, Europe can be seen as an interesting testbed for issues that are relevant in the global arena.

Serving a diverse audience

The metadata landscape in Europe is an example of this framework. All players are aware of the need to build bridges across national boundaries and across professional domains. Among the players involved we find all the organisations that are stakeholders in the management and distribution of

information and knowledge, such as the publishing industry, government agencies, professional societies and the so-called memory institutions (museums, archives and libraries).

The objective of distributing the information might vary across these players: generating a profit for the author or the publisher, informing the citizen by the government, advancing knowledge through research libraries, sharing resources for educational institutions. All of these applications will have to take into account the realities of serving an audience with a multitude of linguistic and cultural backgrounds.

When we talk about metadata, as resource description and discovery mechanism for information on the Internet, it is clear that in this distributed environment solutions have to be found for problems that emerge from the heterogeneity in this marketplace.

Research and practical implementations

In Europe, as in other regions, there are two starting points that we can see in current activities. On the one hand there is research into technical issues in the context of digital libraries, on the other hand there are practical implementations of services.

In the research field, important activities are going on the level of the European Union in the Frameworks for Research and Technological Development. In the last four years, research issues have been addressed under the Fourth Framework Programme [2] in the Telematics for Libraries [3] and Information Engineering [4] Programmes. In the next four years, the research will take place in the Key Action for Multimedia Content and Tools under the Information Society Technologies Programme [5] in the Fifth Framework Programme [6]. Other examples are eLib, the UK Electronic Libraries Programme in Higher Education [7], and the Nordic Metadata Project in Scandinavia [8]. The amount of funds being made available for these programmes underlines the importance of research in this area.

From the perspective of service development, many services have been or are being established around Europe. Worth mentioning in this context are the Scottish Cultural Resources Access Network (SCRAN) [9], the Arts and Humanities Data Service (AHDS) [10] in the UK, MeDoc [11] and the Göttingen Digitisation Centre (GDZ) [12] in Germany, Netpublikationer [13] in Denmark, and the Swedish EnviroNet [14].

The current activities in the European programmes in many cases address the integration issues of the national initiatives to build European services.

Quality – sustainability – handling of rights

In many of the activities listed above, three major areas of concern have appeared: quality, sustainability and handling of rights. These three areas are crucial to the building of an information infrastructure, as they are at the very heart of the usability of the information by a wide audience now and in the future.

Quality and dependability of resource description is a crucial factor enabling the users to be confident about the information they find. In the non-digital environment, information providers have taken responsibility for the quality of the information, either through product catalogues, advertising in newspapers and journals or through traditional library catalogues. In that environment, it has been relatively easy for users to derive confidence from factors such as brand names, quality of the newspaper and physical entry into the shop or library. In the networked environment where resources are distributed around the world, it has become more difficult to determine the quality and dependability. We all know the cases of Web sites attaching metadata to pages in a way that ensure a high ranking in a search engine's hit list, where the metadata does not necessarily give a fair description of the contents. A mechanism is needed to establish the quality of the information so users can be confident about the validity of the information. As in the past, libraries could play an important role as a Trusted Third Party. They will need tools and mechanisms for metadata creation, validation and maintenance, to be able to play that role in the electronic environment with its ever-faster rate of growth.

Sustainability of metadata and resources in this environment is also a growing concern. Many of the resources and their descriptions that are created on the Internet at the present time are being produced with a short horizon, based on current technology and current usage patterns. However, technology develops at an increasing rate and usage patterns develop as we all learn new methods of work and adapt to new possibilities. Already, organisations are struggling to deal with the effects of the changes around us. Common approaches are desperately needed. In the European environment, NEDLIB, a project involving several national libraries [15], is looking at the usefulness of the Open Archival Information System (OAIS) reference model [16], providing a common framework of terms and concepts, to contribute to the solving of digital

preservation problems. Another initiative that looks at long-term availability of resources is the Digital Object Identifier [17], which aims at providing a solution for persistent identification of resources that is independent of the technology. A major factor in the area of sustainability is the underlying business model: which resources need to be preserved and who will maintain the metadata. In the end it comes down to the question: who will pay the bill?

Handling of rights is an issue that has come to the forefront of people's attention. Previously libraries and research projects have addressed the issue in terms of licences for the use of information in negotiation with publishers and rights holders individually. Now there is a growing awareness that the handling of access and usage rights is a fundamental issue that has to be solved on a global scale. The <indec> project [18] under the European Commission's Info2000 programme [19], involving publishers and collecting societies is an important platform that addresses these issues. This project aims to develop a metadata model for intellectual property, a formal structure for describing and uniquely identifying intellectual property itself, the people and businesses involved in its trading, and the agreements which they make about it. Such a metadata model will help ensuring that rights are cleared for the use of resources and that rights are paid where applicable. Again, there is a need for a global solution.

Need for international co-operation

All of the issues related to the use of distributed resources on the Internet have an international component. On the Internet, there are no boundaries and material located anywhere is in principle accessible by users all over the world. This is an advantage but also requires common approaches. In this global marketplace there is a need to understand cross-border issues, such as diversity in languages, cultures, legal systems and financial environments. These issues can only be handled through international co-operation. It is a good sign that international groups such as the Dublin Core Metadata Initiative [20] have a truly global membership and that various research programmes around the world are starting to co-operate. Examples of such co-operation are the Delos working groups [21], the International Digital Libraries programme [22] of the National Science Foundation [23] in the US, the joint calls for proposals from NSF and eLib, emerging contacts between the European Commission [24] and NSF. It is important that these co-operation platforms are reinforced with involvement of a wider international community, including Asia and Oceania. The International Symposia on Digital Libraries will surely help to build bridges between the continents.

Standardisation

One aspect of international co-operation that deserves highlighting is international standardisation. Standardisation is important for metadata, again because of its global applicability. Standardisation provides the groundwork for co-operation by defining the rules in a stable and open environment. On the basis of standards, information providers can define their requirements, vendors can build products for the global marketplace and users can expect to get access to resources in a coherent way.

Because of the open and flexible environment of the Internet, standardisation processes have had to change. In the past, standardisation experts in standardisation bodies defined standards on more or less theoretical grounds. Now, needs and experience are leading the standardisation process thereby ensuring that the standards reflect the real requirements and provide adequate solutions. The standardisation process that the Dublin Core metadata specification is going through at the moment is an excellent example of this and I would encourage everyone with an interest in metadata to follow this initiative closely and contribute to its success.

Building the future

To summarise, I would like to reiterate three main points that I have made in this article.

First of all, further progress in the area of metadata (and digital libraries in general) require research activities as well as practical implementation of services. There are a number of questions still to be answered in technology that require specific research, but at the same time implementations will enable the identification of organisational aspects, such as business and legal issues, as well as verifying the validity of the solutions.

Secondly, specific issues that need to be addressed and solved on a global scale in the area of metadata are related to quality, sustainability and handling of rights. Solutions to these issues cannot be found today or tomorrow, but they are crucial to the usability of the information and knowledge base, also for future users.

Thirdly, developments in digital libraries need to be addressed in a multinational, global perspective, finding common approaches to build a global information infrastructure. Specifically, wide involvement in defining, testing and formalising standards is of the utmost importance to establish a level playing field to the advantage of information providers, vendors of technology and users.

It is my strong belief that these issues are worth solving and that we need to co-operate to find solutions that are applicable on a global scale. This way progress in building the global information resource landscape can be achieved to make the Internet a better place for all of us.

References

- [1] <http://europa.eu.int/abc/obj/treaties/>
- [2] <http://europa.eu.int/comm/dg12/fp4.html>
- [3] <http://www2.echo.lu/libraries/en/libraries.html>
- [4] <http://www2.echo.lu/ie/>
- [5] <http://www.cordis.lu/ist/>
- [6] <http://www.cordis.lu/fp5/>
- [7] <http://www.ukoln.ac.uk/services/elib/>
- [8] <http://linna.helsinki.fi/meta/>
- [9] <http://www.scran.ac.uk/>
- [10] <http://ahds.ac.uk/>
- [11] <http://medoc.informatik.tu-muenchen.de/>
- [12] <http://www.sub.uni-goettingen.de/gdz/>
- [13] <http://www.bs.dk/metadata/english.htm>
- [14] <http://smn.environ.se/miljonat/english/>
- [15] <http://www.konbib.nl/nedlib/>
- [16] http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html
- [17] <http://www.doi.org/>
- [18] <http://www.indecs.org/>
- [19] <http://www2.echo.lu/info2000/>
- [20] <http://purl.org/dc/>
- [21] <http://www.iei.pi.cnr.it/DELOS/>
- [22] <http://www.dli2.nsf.gov/intl.html>
- [23] <http://www.nsf.gov/>
- [24] <http://europa.eu.int/comm/>