Sharing Digital Resources and Metadata for Open and Flexible Knowledge Management Systems

Martin Memmel and Rafael Schirru
(Knowledge Management Department,
German Research Center for Artificial Intelligence DFKI GmbH
e-mail: {memmel,schirru}@dfki.de)

Abstract: This paper discusses the requirements of a framework for sharing digital resources and metadata to meet the needs of open, flexible Knowledge Management solutions. The changing nature of the Web and its users as observed in recent years clearly establishes the need for new approaches and technologies to fully exploit the potential for working with existing digital resources. Formal metadata about the resources can be combined with information created in lightweight and user-centric approaches in order to significantly enhance resource descriptions and enable more efficient access to existing knowledge. The ALOE system, currently in development at DFKI, is one such solution and it is used here as the basis for a sample realization of an appropriate framework.

Key Words: digital resources, knowledge management, knowledge sharing, lifelong learning, metadata, resource profiles, social software, web 2.0
Category: H.3.0, H3.2, H3.3, H3.4, H3.5, H.5.0, H5.1, H5.2

1 Introduction

Today, the number of digital resources is growing at a speed never seen before [Lyman, Varian (03)], and people are confronted with the need to learn throughout their lifespan. In this situation, the role of Knowledge Management and lifelong learning becomes more and more important. It should support people in collaborating with others, and in getting just the right content in just the right time without having to leave the current task and context. Therefore, being able to share digital resources and information about them is a key factor for the success of any system trying to support users in their daily tasks.

Not only the amount of available resources has to be considered – it also has to be taken into account that the way content is created and accessed by the users has fundamentally changed in the last years. The Web is continuously evolving from a place where information was usually only consumed by most of the users to a more social and participatory system. More and more users contribute information in the Web (so called 'user generated content'), using tools like blogs, wikis, social bookmarking services or file sharing platforms such as Flickr and Youtube. Usually, this group of technologies, platforms and tools facilitating 'a more socially connected Web where everyone is able to add and edit the information space' [Anderson (07)], and where sharing of resources through

an individual's social network is eased, is referred to as 'Web 2.0'[1] [O'Reilly (05)]. Instead of connecting resources Web 2.0 is focussing on connecting users with what is commonly denoted as social software.

This paper discusses the requirements for a framework to share digital resources and metadata for various contexts and applications in section 2, the ALOE system, currently developed at DFKI, is presented in section 3 and will serve as an example for the realization of such an approach.

2 Sharing Digital Resources and Metadata

In the following we will discuss the requirements for a system to share digital resources and information about these resources. We will consider the following key questions: *(1) Which types of resources will be handled by the system, and how can the resources be integrated? (2) How can resources be described properly to enable access and further functionalities?* and *(3) How can users and other systems and applications interact with the system?*

2.1 Resources

As already stated, the amount of resources that can be used to support users in their daily tasks is continuously growing. There are numerous different types of multimedia resources (e.g., HTML, PDF, MPEG), and they are stored in distributed locations and repositories. For all these resources, and also for resources created by users themselves, integration possibilities should be provided. Integration should be possible by inserting the resources directly into the system or by referencing them via a URI. In the latter case, the problem of URI persistence has to be considered.

2.2 Metadata - Describing Resources

There are different ways and standards to describe digital resources, e.g., Dublin Core[2] or LOM[3]. However, these approaches usually suffer from several problems (see [Doctorow (01)]) that can only be partly solved with technology. The main problem is that there is the implicit assumption in the structure of most metadata formats which suggests that there is a one-to-one relationship between a resource and the metadata that describes it [Downes (04)]. But there is no 'single and correct' way to describe a resource. A lot of the information depends on the context in which a resource was created, by whom it will be

[1] For a concrete definition see
http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html
[2] http://dublincore.org/documents/dces/
[3] http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

used, and as a means to what end. Wiley et al. therefore distinguish between objective (e.g., the size of a file) and subjective (e.g., the degree of interactivity of a resource) metadata [Wiley, Recker, Gibbons (05)]. A one-to-one relationship also neglects that metadata may change during the lifecycle of a resource [Cardinaels, Duval, Olivié (06)].

It is obvious that centralised approaches, where only a single authority is responsible for the definition of metadata, are a bad idea. Instead, any attempt to describe resources should embrace diversity. The use of application profiles that 'consist of data elements drawn from one or more namespace schemas combined together by implementors and optimised for a particular local application' [Heery, Patel (00)] is one possible solution. Using less restrictive approaches (e.g., application profiles do not allow to introduce new data elements) such as 'resource profiles' instead of single metadata sets [Downes (04)] can also be considered. A resource profile is defined as a 'a multi-faceted, wide ranging description of a resource'. It does not conform to a particular XML schema, instead, it is a patchwork of metadata formats (potentially created by different authors) which are assembled as needed in order to form a description that is most appropriate for the given resource.

In any case, the possibility to annotate various descriptions for each resource should be offered. Nevertheless, there should be some mandatory metadata. Firstly, to *enable basic functionalities* such as search and access (containing, e.g., the name and location of a resource). Secondly, about the *technical format of a resource and the technical requirements to use it*, and thirdly for *intellectual property rights* with information about the way in which a resource may be used.

2.3 Access by users

As already discussed in section 2.2, centralistic approaches have a lot of weaknesses. Thus, the aim must be to attract enough stakeholders that can provide valuable information about resources, at best working as a self-sustained community. But it is not enough just to attract a lot of users, there are certain conditions enabling the 'Harnessing of collective intelligence' as one of the Web 2.0 key principles [O'Reilly (05)]. A 'wise crowd' can be characterized by four conditions [Surowiecki (04)]: *Diversity of opinion* (each person should have some private information), *Independence* (people's opinions are not determined by the opinions of those around them), *Decentralization* (people are able to specialize and draw on local knowledge), and *Aggregation* (some mechanism exists for turning private judgements into a collective decision).

If we want to ensure that these conditions are fulfilled, first of all we have to attract enough users that will contribute to the system. This of course requires dissemination efforts, but it is also very important to provide a user interface following the principles of simplicity [Nielsen (00)] and joy-of-use [Reeps (04)],

and to encourage users to participate, e.g., by using reward mechanisms. To ease the flow of information, possibilities to import (e.g., tags from a social bookmarking system) and export (i.e., users 'own' their information like their profile or their tags) of existing information should exist. Microformats can be used for the export of, e.g., contact details of users or reviews of resources. The aggregation of information can be realized in several ways: E.g., allowing access to all comments concerning a resource in an easy way, showing existing tags in form of lists or clouds, showing average ratings, etc. Last but not least, users should be offered the possibility to use functionalities in their usual contexts and applications, so that they can contribute with different views on resources. Using widgets is one way to realize such an integration.

2.4 Access by systems

Not only users, but also other systems and components can provide and use resources and information about them. So we need more than 'just' good user interfaces or widgets. An infrastructure allowing an easy creation of mash-ups and complex functionalities using the data provided from the system is required.

Access to the data as well as the functionalities of the system may be provided by making use of Web services. As an alternative to robust Web services which make use of 'heavyweight' techniques like SOAP[4] and WS-*, often more lightweight or simplified programming models are demanded [Anderson (07)]. REST[5] is a technology which is likely to meet these requirements.

Furthermore, the system should offer means of notification, e.g., about new resources concerning a certain topic, as well as the transmission of the new content itself. This can be realized by providing feeds using formats such as RSS or Atom.

3 The ALOE system

The ALOE system[6] is currently being developed in the project CoMet[7] at the Knowledge Management Department of DFKI. The aim of the project is the development of a system which offers possibilities to share digital resources and metadata about them. ALOE provides a rich user interface (see figure 1) to motivate users to participate, and allows data to be exchanged via a Web service API. Gathering information in this way is intended to achieve a richer description of resources, thus enabling the development of advanced retrieval and personalization techniques in the future.

[4] http://www.w3.org/TR/soap/
[5] http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm
[6] http://aloe-project.de/
[7] http://www.dfki.uni-kl.de/comet/

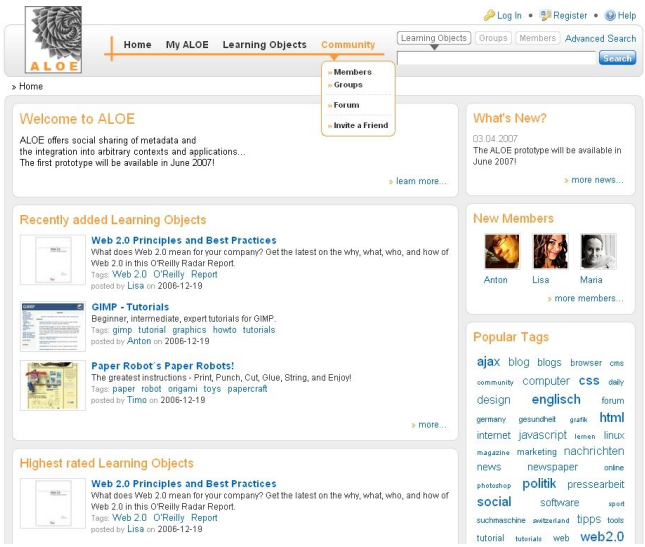

Figure 1: Screenshot of the ALOE user interface

In the following, the system architecture, the metadata which is used to describe resources, and ALOE's core functionalities are presented.

3.1 Architecture

ALOE consists of two components: The FLOR Web Service and the FLOR Connector. The **FLOR Web Service** is the Web interface to ALOE's **Fl**exible **O**bject **R**epository (FLOR). It has been implemented using the Apache Axis[8] technology thus providing access to the functionalities and the information stored in the system via services. The advantage of using this technology is that arbitrary clients can access the system by using SOAP. In ALOE, resources as well as user-defined metadata sets are stored in a Jakarta Slide content repository. Jakarta Slide offers full WebDAV support which enables direct access to the resources and user-defined metadata sets via their URI. Transactions and locking are provided so that data integrity can be ensured. The metadata which describes a resource and the metadata about a user-defined metadata set (see section 3.2) is stored in a MySQL database. That way metadata may be aggregated thus enabling different views on the resources registered in the system. The **FLOR Connector** realizes the user interface to ALOE's Flexible Object Repository. It has been implemented by using the JavaServer Faces[9] (JSF)

[8] `http://ws.apache.org/axis/`
[9] `http://java.sun.com/javaee/javaserverfaces/`

technology. Compared to technologies like JavaServer Pages[10] (JSP), JSF offers several advantages. For instance HTTP requests can automatically update the data of user interface components. Also fine grained event handling mechanisms are offered (e.g., for value change events in user interface components like list boxes) that could otherwise only be provided via client-side technologies. The use of JSF together with JSP causes issues concerning the rendering of user interface components ([Haiges, May (04)] pages 215-216). That's why the Facelets[11] technology was used in order to render the view. Facelets is an alternative view technology to JSP which has its strengths amongst others in templating and the composition of user interface components. As Facelets is based on XHTML and provides support for expression languages it can easily be integrated into our JSF application.

3.2 Metadata

To fulfill the requirements presented in section 2.2, ALOE stores a subset of the Dublin Core Metadata Element Set for every resource which is registered in the system (see table 1). It contains a minimal set of information which is necessary to work with resources of arbitrary formats (i.e., to make them retrievable, accessable and usable). In order to achieve a richer description, resources may additionally be described with metadata sets in arbitrary formats (e.g., LOM for learning objects or information about the context in which a resource was used). This allows for the creation of resource profiles which are likely to overcome the earlier stated problems of inappropriate metadata annotation.

3.3 Functionalities

ALOE provides the following basic functionalities:

- **Insert:** A resource or a user-defined metadata set can be inserted by uploading it as a file into ALOE's WebDAV repository, or by using a reference to the resource, i.e., its URI.
- **Annotate:** ALOE offers mechanisms to rate, tag and comment on resources. They can also manage own tags, favorite resources, etc.
- **Search & Navigate:** ALOE provides different search filters, e.g., users can search for resources which contain certain keywords in their title, description or tags. An advanced search is provided that allows to search for keywords in selected metadata terms. The information annotated by the users allows to browse content via tags (social browsing). Further, retrieved resources can

[10] `http://java.sun.com/products/jsp/`
[11] `https://facelets.dev.java.net/`

Mandatory Resource Metadata	
dc:contributor	Person who inserted the resource into ALOE.
dc:creator	Author of the resource.
dc:date	Date of insertion.
dc:description	Description of the resource.
dc:format	Either MIME type or a proprietary format.
dc:identifier	URI which identifies the resource uniquely.
dc:rights	CC license which is associated with the resource.
dc:title	Title of the resource.
Metadata of a User-Defined Metadata Set	
dc:contributor	Person who inserted the metadata set into ALOE.
dc:creator	Author of the metadata set.
dc:date	Date of insertion.
dc:description	Description of the metadata set.
dc:format	Metadata format (e.g., DC)
dc:identifier	Identifier of the metadata set.
dc:relation	URI of the described resource.

Table 1: An excerpt of the metadata used in the ALOE system

be ranked according to different criteria, e.g., alphabetically, most viewed, best rated.

- **Preview:** Together with a preview image, various information about a selected resource (e.g., user-defined metadata files, comments, tags, ratings) may be displayed.
- **Group Management:** Users of ALOE can initiate and join groups of interest. Whenever they contribute a resource or a tag into the system, the users may decide whether the resource or tag is visible for all users of the system, for members of selected groups or only for themselves.

4 Summary and Future Work

Any approach that supports knowledge sharing and that meets the needs of today's users, should allow to incorporate any kind of digital resource available on the Web or a local repository, as well as content generated by the users themselves. To ease the use of the digital resources, information enabling basic functionalities, about technical requirements for the usage, and about the way in which a resource may be used (intellectual property rights) should always

be provided. Additionally, approaches such as application profiles or resource profiles are recommended to allow an adequate description of resources. A user interface that stimulates users to participate, options to import and export information in an easy way, methods to aggregate the collected information, and the possibility for users to use functionalities in their usual contexts and applications are very important to ensure that the harnessing of collective intelligence will be successful. Providing access to the system via services and making use of standardized formats for the representation of information ease the information exchange with other systems. The ALOE system realizes such an approach, and we plan to deploy and evaluate the first prototype at DFKI, where we focus on motivating researchers to share information about publications and conferences. ALOE will be enhanced with various functionalities in the future, e.g., feed support, semantic search and automatic generation of selected metadata.

References

[Anderson (07)] Anderson, P.: "What is Web 2.0? Ideas, technologies and implications for education"; JISC Technology and Standards Watch, 2007.

[Cardinaels, Duval, Olivié (06)] Cardinaels, K.; Duval, E.; Olivié, H.: "A formal model of learning object metadata"; in Wolfgang Nejdl and Klaus Tochtermann, editors, EC-TEL 2006 - Innovative Approaches for Learning and Knowledge Sharing, volume 4227 of Lecture Notes in Computer Science, pages 74-87. Springer, 2006.

[Doctorow (01)] Doctorow, C.: Metacrap: "Putting the torch to seven straw-men of the metautopia"; 2001, electronic document, date of publication: August 26, 2001, retrieved March 14, 2007, from http://www.well.com/~doctorow/metacrap.htm.

[Downes (04)] Downes, S.: "Resource profiles"; Journal of Interactive Media in Education, 5, 2004, ISSN:1365-893X.

[Haiges, May (04)] Haiges, S; May, M.: "JavaServer Faces. Web Development mit dem Standardframework"; entwickler.press, 2006.

[Heery, Patel (00)] Heery, R.; Patel, M.: "Application profiles: mixing and matching metadata schemas"; Ariadne, Issue 25, 2000, http://www.ariadne.ac.uk/issue25/app-profiles/intro. html.

[Lyman, Varian (03)] Lyman, P.; Varian, H. R.:"How Much Information"; 2003, retrieved July 12, 2007, from http://www.sims.berkeley.edu/how-much-info-2003.

[Nielsen (00)] Nielsen, J.: "Designing Web Usability: The Practice of Simplicity"; New Riders Publishing, 2000.

[O'Reilly (05)] O'Reilly, T.: "What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software"; 2005, electronic document, date of publication: September 30, 2005, retrieved February 12, 2007, from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

[Reeps (04)] Reeps, I. E.: "Joy-of-Use: eine neue Qualität für interaktive Produkte"; Master's thesis, University of Konstanz, 2004.

[Surowiecki (04)] Surowiecki, J.: "The Wisdom of Crowds". Doubleday, 2004.

[Wiley, Recker, Gibbons (05)] Wiley, D. A.; Recker, M.; Gibbons, A. S.: "Getting axiomatic about learning objects"; in David A. Wiley, editor, The Instructional Use of Learning Objects: Online Version. 2000, retrieved from the Internet on February 1, 2006 from http://reusability.org/axiomatic.pdf.