Martin Memmel

Socially Enhanced Access to Digital Resources

Aggregating Resources and Metadata in a Social Hub

Dissertation

genehmigt vom Fachbereich Informatik der Universität Kaiserslautern

zur Verleihung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)

Dekan:
Prof. Dr. Klaus Schneider, Universität Kaiserslautern

Vorsitzender der Promotionskommission:
Prof. Dr. Stefan Deßloch, Universität Kaiserslautern

Berichterstatter:
Prof. Dr. Andreas Dengel, Universität Kaiserslautern
Prof. Dr. Erik Duval, Katholieke Universiteit Leuven

Datum der Aussprache:
26. Juni 2015

D 386

Socially Enhanced Access

to Digital Resources

Martin Memmel

Acknowledgements

I thank all the many people who supported me during the development of this work. I thank them for their encouraging words, fruitful discussions, patience, and helpful critique.

This work was developed at the German Research Center for Artificial Intelligence GmbH in the Knowledge Management Department of Professor Andreas Dengel. I am grateful for his support, especially for giving me the chance to realise the CoMet project (where the basis for a lot of the results presented here was laid), and for pushing me to finish this work. I also enjoyed our inspiring conversations about Knowledge Management and the way we work with digital resources.

I also particularly like to thank Professor Erik Duval, the co-supervisor of this thesis. His work and the work of his group had an important impact on my professional career, and he was always ready to help me whenever I asked for support, even in tough times. I am very grateful that he joined me in the process of writing this work, and I am looking very much forward to further working with him.

The practical outcomes of this work resulted from a large number of projects, and I thank all the people working with me on those projects over the years. Regarding the ALOE project, my biggest thanks go to the "ALOE chief developers", namely Heinz Kirchmann and especially Rafael Schirru. I hope they forgive me my perpetual requests for improvements and new features.

I also want to thank my colleagues at the Knowledge Management Department that supported this work with their ideas and feedback, advices, and several inspiring discussions.

For proofreading and commenting on several versions of this thesis, I thank Prof. Dr. Thomas Roth-Berghofer, Prof. Dr. Martin Wolpers, Dr. Stephan Baumann, and Prof. Dr. Sandra Zilles.

Last but not least, this journey would not have been possible without the support of my family and friends – I consider myself to be a very lucky person to have such amazing people around me. I can not name all of them here, but I am very grateful for your love and your friendship.

Martin Memmel

Preface

In the digital era we live in, users can access an abundance of digital resources in their daily life. These digital resources can be located on the user's devices, in traditional repositories such as intranets or digital libraries, but also in open environments such as the World Wide Web.

To be able to efficiently work with this abundance of information, users need support to get access to the resources that are relevant to them. Access to digital resources can be supported in various ways. Whether we talk about technologies for browsing, searching, filtering, ranking, or recommending resources: what they all have in common is that they depend on the available information (i.e., resources and metadata). The accessibility of digital resources that meet a user's information need, and the existence and quality of metadata is crucial for the success of any information system.

This work focuses on how social media technologies can support the access to digital resources. In contrast to closed and controlled environments where only selected users have the rights to contribute digital resources and metadata, and where this contribution involves a social process of formal agreement of the relevant stakeholders, potentially any user can easily create and provide information in social media environments. This usually leads to a larger variety of resources and metadata, and allows for dynamics that would otherwise hardly be possible.

Most information systems still mainly rely on traditional top-down approaches where only selected stakeholders can contribute information. The main idea of this thesis is an approach that allows for introducing the characteristics of social media environments in such traditional contexts. The requirements for such an approach are being examined, as well as the benefits and potentials it can provide.

The ALOE infrastructure was developed according to the identified requirements and realises a *Social Resource and Metadata Hub*. Case studies and evaluation results are provided to show the impact of the approach on the user's behaviours and the creation of digital resources and metadata, and to justify the presented approach.

Contents

List of Figures

List of Tables

It is extremely difficult for computers to discern what's relevant, as you prove every time you use even the best Web search engines. Because we humans are embodied creatures who care about ourselves and our world, we are able to sniff out relevancy faster than bats find june bugs.

David Weinberger, Small Pieces, Losely Joined, p. 143

I
Background

CHAPTER 1
Introduction

The topic of this thesis is how the access to resources in digital environments can be supported in a generic way using social media technologies. Therefore, the potential benefits that can be provided by these technologies are being examined, as well as the challenges when aiming to realise them. The ALOE infrastructure[1] will be presented as an implementation of a generic and adaptable environment that allows for supporting access to digital resources in a wide range of scenarios. Evaluation results and several case studies will provide insights into its potential: to get one step closer to fully exploiting the potential when working with digital resources.

1.1 Thesis Overview

In the following, a description of the focus and objective of this work will be given, followed by a brief presentation of the applied methodology as well as main contributions and results. A description of the background of the thesis will close this section.

1.1.1 Focus

Today, people spend more time than ever in digital environments, where they can access an abundance of digital resources – whether this concerns working environments, organisational contexts or private use cases. These digital resources can be located on the users' devices, in traditional repositories such as intranets or digital libraries, but also in open environments such as the World Wide Web.

To be able to efficiently work with this large amount of information accessible in heterogeneous environments, users need to be able to quickly get access to resources that are relevant to them. This work focuses on how social media based

---

[1]See Section 5.1.1 for an explanation of the name "ALOE".

infrastructures and the digital resources and metadata created and provided in such environments can help to support users in this task.

In contrast to closed and controlled environments where only selected users are allowed to contribute digital resources and metadata, and where this contribution involves a social process of formal agreement of the relevant stakeholders, potentially any user can easily create and provide information in social media environments. This usually leads to a larger variety of resources and metadata, and allows for dynamics that would otherwise hardly be possible.

Metadata (i.e., data about data that can also comprise relations to other data) in general can either be generated by machines (e.g., by analysing the content of a resource) or by humans, and each of the approaches has certain benefits and limitations concerning production costs, quality, and possible application areas. Usually, machines can generate metadata very fast, but are very limited in what can be generated for which type of resource. Humans cannot compete with the speed of machines, but are not depending on the machine-readability of a resource or relevant background information. And what is even more important: humans can contribute with opinions, interpretations and subjective views, something machines will – most likely – never be able to. This is especially important because there is not only objective metadata (e.g., the title or technical format of a document), but also subjective metadata (e.g., the classification or quality of a resource). Deriving information or knowledge from the data itself (i.e., intrinsic metadata) as well as associating information that cannot be inferred from the data (i.e., extrinsic metadata) always requires the ability to understand and interpret the data. This process is influenced by various factors – among others, it often involves experience and individual, subjective opinions.

For a long time, human generated metadata was equated to expert generated metadata. Experts often care about complex tasks requiring knowledge about the domain and the process of describing a resource. As this approach to generate metadata is very time-consuming, it is simply impossible to generate a sufficient amount of up-to-date information about the abundance of digital content that is available and created each day. The situation changed with the increasing importance of the Internet and especially the development of advanced web based applications that can easily be used by almost anyone. The World Wide 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. With the developments usually subsumed under the term *Web 2.0*, new ways to interact with information have emerged in recent years. So called social media environments now also allow end users to easily create, publish and annotate digital resources without any expert knowledge. Such approaches are also becoming more and more popular in corporate environments, where they enhance

and often replace traditional knowledge management methods.

Resources and metadata in such contexts are created and provided by end users – in contrast to traditional, centralistic approaches. The user generated metadata (e.g., tags, comments, and ratings) is usually very informal, lightweight and simple, but has its strength when many users provide information about resources, allowing to treat aspects of subjectivity in a better way (e.g., by aggregating information from different users). The metadata generated in social media environments can also provide valuable information about the usage of a resource, e.g., by how many people it was used in which way, and who these people are. On the other hand, end users usually will not create complex metadata that is often very useful when working with resources.

Introducing social media technologies in existing environments allows for new ways of contributing and accessing digital resources, and of providing new services based on user generated and enriched metadata. Thus, new methods are possible to support the user in finding relevant content instead of simply facing a flood of information. This especially concerns browsing, searching, filtering, ranking, and recommendation techniques. The focus of this thesis will therefore be put on

- the analysis on how digital resources are accessed and which aspects are important in these processes,

- how social media technologies can be used to support the access to digital resources,

- how such technologies can be offered in a way such that they can potentially be applied for supporting the access to digital resources wherever this support is needed, and

- the potentials of introducing social media environments with digital resources and metadata contributed by end users in existing, especially traditional settings.

1.1.2 Objectives

Access to the ever growing amount of information can be supported in various ways. Whether this concerns technologies for browsing, searching, filtering, ranking, or recommending resources: what they all have in common is that they depend on the available information (i.e., resources and metadata).

The accessibility of digital resources that meet a user's information need, and the existence and quality of metadata is crucial for the success of any environment that aims to support the work with digital resources. Therefore, the characteristics of information systems as well as resource metadata will be examined, with the aim to realise and encourage the creation of better metadata for supporting the access to digital resources.

Main idea   Most information systems until now do rely on traditional top-down approaches where only selected stakeholders can contribute information. The main idea of this thesis is to enable the introduction of a generic, adaptable *social media environment* in such traditional approaches, so that information can easily be accessed and contributed by end users.

Main goal   The main goal of this work is the introduction and justification of a comprehensive, generic, and holistic approach that allows

- to support access to arbitrary kinds of digital resources by means of social media technologies, and

- to integrate social media technologies also in existing, especially traditional and non-social environments.

This approach is referred to in this work as a *Social Resource and Metadata Hub*.

Research questions   The main research questions addressed in this thesis are:

RQ1: What benefits can social media technologies provide for different aspects regarding the access to digital resources?

RQ2: What are the requirements and demands for a generic approach that allows

- to support access to digital resources by means of social media technologies, and

- to integrate social media technologies in existing environments?

RQ3: Is it possible to realise an environment that meets these requirements and demands, and that provides the expected benefits?

1.1.3 Methodology, Contributions and Results

First, a holistic architecture of information systems, as well as a general access process for digital resources in such systems is presented. This serves as the

basis for identifying the characteristics of a hypothetical, **ideal setting** for the access to digital resources in information systems. Based on these findings, a characterisation of what constitutes "good" resource metadata for such an ideal setting is provided.

Then it is discussed how the way users interact with information has changed, and an **overview of entities and relations in social media environments** is provided. This serves as the basis for the **identification of challenges and potentials of social media technologies** for supporting the access to digital resources.

After an identification of typical characteristics that hinder the access to digital resources, the results presented at this point are used to identify the **requirements for an infrastructure that allows the introduction of social media technologies in existing digital environments**. This leads to the **definition of a *Social Resource and Metadata Hub*** – an infrastructure that provides social media functionalities and allows to integrate existing resources and metadata for the specific needs of various application scenarios. A **system design checklist** with respective requirements is also provided.

According to this checklist, the **ALOE infrastructure** has been designed and developed, and is then introduced as a realisation of a Social Resource and Metadata Hub. Detailed information about the system architecture, features, and use cases are given, and main decisions as well as possible alternatives are discussed.

ALOE

To justify the presented approach, the following ALOE **case studies and results of according evaluations** are presented:

Case studies

MACE (`http://www.mace-project.eu`) is a European initiative aimed at improving architectural education, by integrating and connecting vast amounts of content from diverse repositories. MACE uses ALOE as a backbone for community related features.

Mindpool (`https://mindpool.dfki.de`)[2] is the ALOE-based social resource sharing platform available for all DFKI employees. It shows how ALOE can be adapted to support specific resource types such as press articles or visits, and how available metadata information (e.g., about projects or persons) can be integrated.

ALOE@KM (`http://projects.dfki.uni-kl.de/aloe`) is an ALOE instance specifically provided for the Knowledge Management group of DFKI in Kaiserslautern.

---

[2]Please note that this system is only accessible within DFKI.

Additionally, a variety of further use cases and applications is presented to illustrate how ALOE was adapted and extended in order to support the specific needs of different scenarios.

The results provide evidence that a Social Resource and Metadata Hub allowing the contribution of resources and metadata by end users enables significant progress when aiming to support access to digital resources.

1.1.4 Background

The work presented in this thesis has its origins in several projects conducted at the Knowledge Management Department of Prof. Andreas Dengel at the German Research Center for Artificial Intelligence DFKI GmbH and the Algorithmic Learning group of Prof. Rolf Wiehagen at the Department of Computer Science of the University of Kaiserslautern.

The thesis summarises the identified requirements and consequences, as well as the collected experiences both from the world of traditional approaches and social media environments. On the one hand, it is application and demand driven, but on the other hand, it also provides general insights about how to leverage the experience for users when working with digital resources.

Projects The starting point of this thesis was the research and development project DaMiT sponsored by the German Federal Ministry for Education and Research (BMBF). DaMiT is an abbreviation of "Data Mining Tutor" and is an Internet-based Intelligent Tutoring System for the domain of knowledge discovery and data mining (cf. [GLM03, JGLM04]). The DaMiT system is a traditional digital repository, where different approaches have been realised to adapt the content to different presentation styles and difficulty levels.

Building on the experience gained in the DaMiT project and motivated by the emerging field of social media, the development of ALOE was initiated in 2006 as an attempt to combine approaches based on traditional metadata with user generated content. Most of the work reported here was created in the context of designing and developing ALOE. The first project that laid the basis for the ALOE infrastructure was CoMet (Collaborative Sharing of Resources and Metadata), sponsored by the Stiftung Rheinland-Pfalz für Innovation from 2007 until 2008. The aim of the project was to develop an open platform to share metadata about arbitrary types of digital resources.

In the following years, several projects were conducted that made use of ALOE, and that allowed to further enhance the ALOE infrastructure:

- The European project MACE (Metadata for Architectural Contents in Europe) initiated in 2006 uses the ALOE infrastructure as a social backbone

to support a variety of community features.

- In the context of the Cluster of Excellency "Center for Mathematical and Computational Modelling" (CMCM) that was initiated in 2008 in Kaiserslautern, ALOE is used as a basis to build up a "Web of Mathematical Models".

- Mindpool is DFKI's internal project aiming to offer a social media suite for all DFKI employees (in Berlin, Bremen, Kaiserslautern and Saarbrücken).

- The aim of the project RADAR (Resource Annotation and Delivery for Mobile Augmented Reality Services) was the development of an infrastructure to contribute, organise and annotate multimedia resources that can be used within mobile augmented reality services. RADAR was sponsored by the Stiftung Rheinland-Pfalz für Innovation from 2010 until 2011.

- In the project NEXUS funded by the German Federal Ministry of Education and Research (BMBF) from 2012 until 2014, ALOE was extended with several features that allow for combining technologies from the Social and Semantic Web, especially regarding the annotation of digital resources.

Workshops, Tutorials

To better understand how digital resources are used in daily tasks and especially in the context of learning, the LOKMOL workshop series were organised. The first LOKMOL (Learner Oriented Knowledge Management and KM-Oriented E-Learning) workshop was part of the Conference on Professional Knowledge Management (WM 2005) in Kaiserslautern. One of its main results was that – when aiming to exploit the full potential of working and learning with digital resources – it is a central issue to encourage members of organisations to make their knowledge explicit by creating digital resources and sharing them with others [RMW05]. The findings and identified barriers for the integration of knowledge management and eLearning led to subsequent LOKMOL workshops at the European Conference on Technology Enhanced Learning EC-TEL in 2006 and 2007. There, the focus was also laid on the importance of context and Web 2.0 technologies (see [MRW$^+$06] and [MRWA07]). The Tutorial "Setting Up Your Own Learning Object Environment" affiliated to the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education E-Learn in 2005 [MRZ05] focused on different needs, requirements, and their impact when managing digital resources.

Related Work   Especially in recent years, a large amount of research has been conducted in the emerging field of social media and how it can be used to sup-

port access to digital resources. Yet, what is missing is a broad and comprehensive identification of the potentials that social media technologies offer for different steps in the process of accessing digital resources in information systems. Instead, isolated scenarios with solutions that focus on specific scenarios and domains are examined.

From a technology point of view, several solutions exist that offer social media functionalities, and that also allow to introduce them in existing environments. However, most of these tools are neither instantiable, nor can they be adapted to the specific needs of a scenario. Furthermore, incorporating existing information is usually very complex or not even possible at all.

What is missing is a comprehensive, holistic approach and a respective framework that allows to exploit the potentials of social media also in existing environments, with support for potentially arbitrary kinds of contributions.

Due to the fact that this thesis – as just described – has a comprehensive, holistic, and thus multidimensional character, related work is introduced and discussed throughout the whole thesis. Yet, a more detailed discussion of related technical approaches can be found in Section 5.7.

1.2 Prerequisites

In order to clarify the terminology being used and to prepare the ground for the subsequent chapters, a detailed introduction to digital resources, metadata and social media is provided in Chapter 2 and 3. Nevertheless, for obtaining a more profound understanding of these topics in general, it is recommended that the reader checks out introductory literature such as *Modern Information Retrieval* by Ricardo Baeeza-Yates and Berthier Reibeiro-Neto [BYRN99] for an introduction to information access, and Toby Segaran's *Programming Collective Intelligence* [Seg07], *The Wisdom of Crowds* by James Surowiecki [Sur04], and David Weinberger's *Small Pieces Loosely Joined: A Unified Theory of the Web* [Wei02] for an introduction into the field of social media. Another work of Weinberger, namely *Everything is Miscellaneous* [Wei07] is recommended to gain some deeper insights into how information can be organised by making use of social media concepts and technologies.

1.3 Guide to this Work

This thesis consists of three parts, each with a different focus. The first part provides background information. It gives an introduction into the problems one

faces when accessing digital resources, and into the employment of metadata for supporting users in such scenarios. The second part describes the paradigm shift from traditional to social media as well as the resulting new potentials and challenges for information systems aiming at supporting users. The requirements for an infrastructure that allows for the introduction of social media technologies into traditional environments will then be identified, leading to the definition of a *Social Resource and Metadata Hub*. In part three, the ALOE system is introduced as an implementation of a Social Resource and Metadata Hub, followed by a presentation of case studies and evaluations.

▷ In Chapter 2, the basic concepts required to understand how digital resources can be accessed, and how this can be supported using metadata are presented. The terminology that will be used throughout this thesis is introduced.

▷ Chapter 3 deals with the paradigm shift from traditional media to social media, focusing on the influence of this development on digital resource creation, publication, and access. The chapter concludes with the identification of new challenges and potentials this paradigm shift brings with it.

▷ Based on the findings so far, and on identified characteristics of scenarios in which difficulties to access digital resources occur, the concept of a Social Resource and Metadata Hub is introduced in Chapter 4. Consequently, the potential benefits that such a hub can provide plus the requirements for a realisation will be identified.

▷ Chapter 5 introduces the ALOE system that was developed as a realisation of a Social Resource and Metadata Hub. The main concepts, main features, and the system design and components will be described.

▷ In Chapter 6, selected use cases and evaluations will be presented to provide evidence for the benefits of the approach.

▷ The thesis is concluded with a summary and an outlook on future work.

The appendix offers further details, mainly about the ALOE system as well as the conducted evaluation:

▷ Appendix A offers an overview of several metadata schemas relevant in this work.

▷ DB schemas with information about existing entities and their metadata as well as the relations to other entities in ALOE are provided in Appendix B.

▷ The ALOE Terms of Service as well as the ALOE Privacy Policy is presented in Appendix C.

▷ In Appendix D, statistical information about created artefacts, tracked activities, and Web analytics data is presented for a variety of ALOE instances that have been deployed since 2008.

▷ The online questionnaire used for evaluating Mindpool and ALOE@KM is provided in Appendix E.

▷ Selected functionalities offered by the *ALOE Web Service* API are presented in Appendix F.

CHAPTER 2
Accessing Digital Resources in Information Systems

In private, educational as well as economical contexts, individuals can access an ever growing amount and diversity of information in their digital environments, using different kinds of information systems [ASPJ+03, LV09, Goo08]. Working on daily tasks that range from information seeking, information discovery and conceptual mapping to decision making and work process control, people spend large amounts of time with searching, finding, and employing appropriate content. A lot of efforts are undertaken to increase effectiveness and efficiency in information processing, and to make these tasks less difficult, time-consuming, and expensive. Therefore, not only the existence of relevant information has to be considered, but also the way it is described and can be accessed. Descriptions can contain information about the content (e.g., the title and the creator of a document) and relations to other resources, but also information about the way resources are used and by whom. Such information can allow users as well as machines to work with digital resources in a more efficient way.

People have much more information at hand than they can possibly use

In the following, a detailed introduction is provided into how *digital resources* and *information systems* are understood in this thesis, how digital resources can be accessed in such systems, and how users can be supported in these processes. After discussing the concepts *information need* and *relevance*, the chapter is closed by introducing the concept of metadata as a key enabler to support the access to digital resources.

2.1 (Digital) Resources

The objective of the work presented in this thesis is to support users to cope with the abundance of information that they can access in their digital environments. Therefore, a precise understanding of what kind of information is meant

What are the resources relevant in this work?

is required.

Information can be represented in documents such as text, audio and video files that are located on the user's desktop, in closed environments such as an Intranet, or in open environments like the World Wide Web. Furthermore, also "real" resources such as books, human beings, locations or events that might be relevant for a user's specific needs can be represented digitally. Therefore, none of these resource types is excluded in the context of this work; the only prerequisite that has to be met is that a resource can be digitally referenced. Therefore, appropriate resource identifiers are needed. Ideally, such identifiers are unique, persistent and do not only allow to identify the resource, but also to access it. In this work, the definition of *digital resources* refers to the well-established standard of *Uniform Resource Identifiers (URI)*.[1] A URI "is a compact sequence of characters that identifies an abstract or physical resource" [BLFIM98] and provides a simple and extensible means to identify resources.[2]

URIs are characterised by the following definitions:

Uniform "Uniformity provides several benefits: it allows different types of resource identifiers to be used in the same context, even when the mechanisms used to access those resources may differ; it allows uniform semantic interpretation of common syntactic conventions across different types of resource identifiers; it allows introduction of new types of resource identifiers without interfering with the way that existing identifiers are used; and, it allows the identifiers to be reused in many different contexts, thus permitting new applications or protocols to leverage a pre-existing, large, and widely-used set of resource identifiers." (cf. [BLFIM98, p.1])

Resource "A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., 'today's weather report for Los Angeles'), and a collection of other resources. Not all resources are network 'retrievable'; e.g., human beings, corporations, and bound books in a library can also be considered resources. The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content – the entities to which it currently corresponds – changes over time, provided that the conceptual mapping is not changed in the process." (cf. [BLFIM98, p.1–2])

---

[1]There are several other approaches to provide identifiers in different domains, but these can be neglected in the context of this thesis without loss of generality.

[2]See [BL94] for a detailed description of the syntax for URI as used in the World Wide Web.

Identifier "An identifier is an object that can act as a reference to something that has identity. In the case of URI, the object is a sequence of characters with a restricted syntax." (cf. [BLFIM98, p.2])

Consequently, the W3C in [MM04] provides the following definition of a resource: *A resource is anything that is identifiable by a URI reference.*[3] However, this might be confusing, as the term *URI* itself is already characterised by a definition of the term *resource*. The definition of a resource given there also does not distinguish between the resource itself (e.g., a person) and what can be retrieved in an information system (e.g., contact data connected to the URI of the person).

It is not important in the context of this work whether a resource "just" consists of a URI, or if it also refers to a digital artefact. Only the according digital information that can be accessed and retrieved is of interest. To avoid confusion, a difference is thus made between a *resource* and a *digital resource*, and they are defined as follows:

Term 2.1 *A* resource *is anything that has identity.*   **Resource**

This definition does not only include real objects and digital artefacts, but also abstract concepts such as an event, and even imaginary concepts (e.g., a figure in a fairy tale). This of course raises the question whether there is anything at all that is *not* a resource, i.e., has no identity. A person's thought, or a single entity in a set of indistinguishable entities (e.g., a water molecule in a glass of water) might serve as such examples.

Term 2.2 *A* digital resource *is anything that allows to be identifiable by a URI refer-*   **Digital**
ence and retrievable in an information system.[4]   **Resource**

It is important to note that this definition still allows a digital resource to refer to any kind of resource as mentioned above (e.g., a digital resource might represent a non-digital entity such as a real person or an event). Furthermore, it does also include resources that are not yet identifiable by a URI reference and retrievable in an information system, as long as they can potentially be integrated.

---

[3]The difference between *Universal Resource Identifiers* and *Uniform Resource Identifiers* and between *URI* and *URI reference* can be neglected in the context of this thesis.
[4]See 2.3 for the definition of the term *information system*.

2.2 Accessing Digital Resources

Users access digital resources in various environments, and they do it in very different ways. To understand how the access to digital resources can be supported, some *sample scenarios* will first be introduced. Then, a general architecture of *information systems* is provided, and typical *access types* are identified.

2.2.1 Sample Scenarios

The following scenarios will serve as exemplary cases for our further investigations.

Scenario 1 – Searching in a closed, organisational environment: A user in an enterprise only has access to the organisation's Intranet. He/she has to solve a given task, but does not have all required information at hand. Thus, he/she is looking for resources that contain the information he/she needs. First, he/she chooses the business area in the Intranet that the task is concerned with. Then, he/she uses the Intranet search engine for this area and enters some search terms. A list with several hundred documents is returned, ranked according to the estimated relevance of the keywords for each document. The user scans the first two result pages and considers the focus of the returned documents as far too broad. Thus, he/she refines the query by adding some more specific keywords. After checking some resources in the result list, the user finds some resources that he/she wants to check more thoroughly.

Scenario 2 – Exploratory browsing in an open environment: A user is using his/her PC at home, opening the feed reader to check whether some interesting article was published recently in one of the blogs he/she has subscribed. The user finds an interesting article about social resource sharing and reads it. As he/she is interested in more information about the aspects mentioned in the article, he/she clicks on a link in the article. The user bookmarks this article with his/her social bookmarking system. Doing that, he/she finds out that there are several other people that also bookmarked the article. The user checks what else was bookmarked by these users and finds some more interesting information about social resource sharing.

Scenario 3 – Searching in heterogeneous environments: A student in a university has to solve an exercise about decision trees for a course on Artificial Intelligence. He/she did not really understand how decision

trees can be generated, so he/she is looking for some additional material. He/she finds several courses on Artificial Intelligence that seem to contain information about decision trees. The courses are stored in a learning management system provided by the university as well as in different repositories in the World Wide Web, and they offer different navigation facilities and use different classification vocabularies. After checking several of the sources the user finds a platform where his/her search for "decision tree" returns a list ranked according to the ratings of other students using this platform. He/she decides to stop his/her inquiry and chooses the recommended source.

These examples show that access to digital resources can happen by various means using different systems. Sometimes a pre-defined categorisation is used, sometimes users actively formulate a query using a search engine, sometimes a system recommends digital resources, and sometimes users browse in an exploratory manner. Before identifying different *types of information access* and a *general access process*, an underlying *general architecture of information systems* for all these scenarios will first be introduced.

2.2.2 A General Architecture of Information Systems

Following the definition provided by the ISM3 consortium[5] the systems that are investigate and in which users can access digital resources will be denoted as *information systems*:

Term 2.3 *An* information system *is a human and technical infrastructure for the storage, organisation, processing, transmission, input and output of information.*

Information system

This broad definition includes all the scenarios described above, and in contrast to other definitions of the term it does not focus on organisational contexts only.

For the further investigations on how access to resources in information systems can be supported, the existence of a general architecture as depicted in Figure 2.1 will always be assumed.

General architecture of information systems

---

[5]ISM3 is an acronym for *Information Security Management Maturity Model*, see `http://www.ism3.com` and `http://www.ism3.com/index.php?option=com_content&task=view&id=10&Itemid=13`
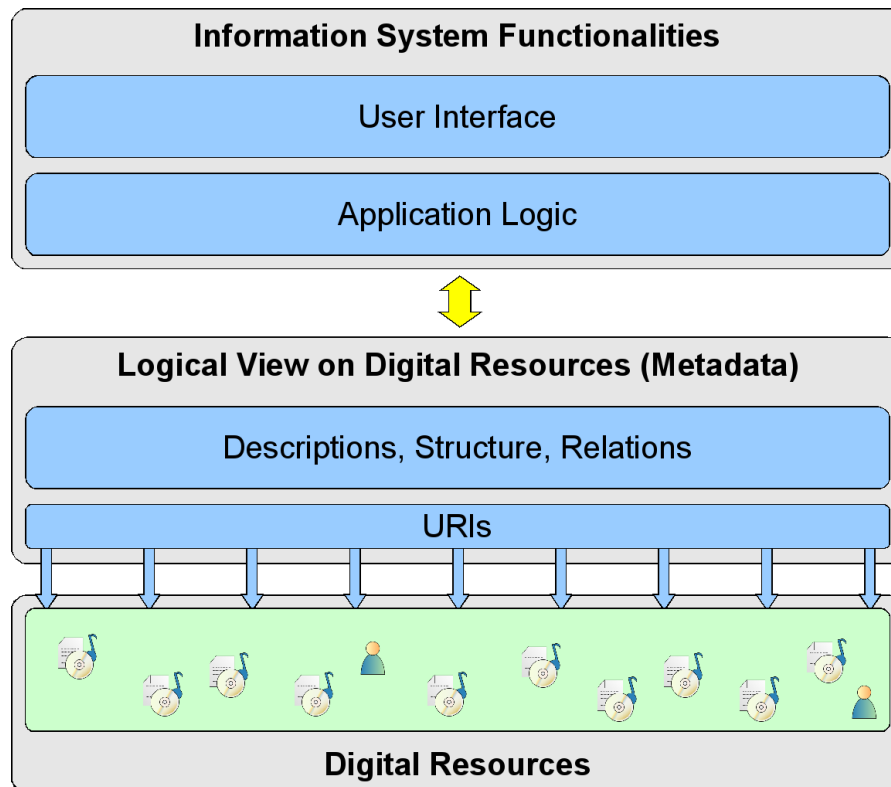
Figure 2.1: General architecture of an information system

In this architecture, the following components are present:

Digital Resources: The entities that can be accessed in the system as a possible result of an inquiry. In closed scenarios, e.g., an Intranet, these resources might be restricted to what is offered by different services in an organisation, and they might be represented as well as described in a homogeneous way following certain guidelines and conventions. In open scenarios such as the World Wide Web, the resources can be stored in widespread locations and come in very heterogeneous formats.

Logical View on Resources: Following the definition from Section 2.1, it is assumed that each resource in our information system has a URI to refer to it. Furthermore, resources may be arranged in folder structures or other types of hierarchies, relations between them may exist, and metadata can be associated with them (e.g., providing information about the content of the resource and technical specifications). It is also allowed that an information system offers more than only one logical view on resources.

Application Logic: In this component, the system's functionalities are defined. This might include a query engine, ranking algorithms, recommenders, etc. Here, a system might also draw on additional information (e.g., formal sources such as thesauri and ontologies, or semi-structured information extracted from large sources such as Wikipedia[6]).

User Interface: The user interface offers means to interact with the system's functionalities. Here, users can specify their information need (e.g., by entering a query), look at the results of an inquiry, browse, etc.

2.2.3 Access Types and the Access Process

Access is a term that is used in many different ways. For example, the UNESCO Open Educational Resources Community[7] lists the following access types and categories [Com09]:

Awareness, policy, attitude, culture Access in terms of awareness, local policy / attitude, languages

Legal Access in terms of licensing

Technical: Provision of digital resources Access in terms of file formats, disability

Technical: Receiving digital resources Access in terms of infrastructure, Internet connectivity, discovery, ability and skills

In the context of this work, access will only be referred to regarding the *provision of digital resources* in information systems.

In general, two fundamental types of information access can be distinguished: *pulling* and *pushing*. These terms are used as follows:

Term 2.4 Pulling *refers to the activity of interactively requesting information [BYRN99, p.5] in an information system.* **Pulling**

Term 2.5 Pushing *refers to the activity of automatically pushing information to the user without any need for inquiry for the user.* **Pushing**

---

[6]Wikipedia is a large online encyclopaedia; see Chapter 3 for more information
[7]See `http://oerwiki.iiep-unesco.org`

	Technology
Pulling	*Query-based searching:* Users provide an oral or written representation of their information need, the most common way is to type key phrases. This is usually enabled by indexing resources. Query-based searching requires users to be able to formulate queries in a syntax required by the respective information system – this can be non-trivial especially for advanced queries, e.g., when Boolean logic shall be applied. Furthermore – like any other pulling technology – users might not be aware that (potentially new) relevant resources exist and thus might not even try to actively search for information.
	Browsing: Users browse on a structured corpus of resources. This is usually enabled by linking documents, or by providing directories, taxonomies, etc. When exploring a corpus by filtering available information, this is also referred to as *faceted browsing* or *faceted search*.[8]
Pushing	*Static approaches:* Such approaches do not take into account a user's current context or maybe changing interests and needs. Examples are simple techniques such as feeds or search alerts.
	Proactive Information Delivery: This group of technologies aims to provide relevant information without explicit request [Hol02]. A very prominent example is *information filtering*. information filtering is typically applied to large amounts of data (often streams) and typically focuses on removing data that does not meet certain criteria [BC92]. Such criteria can be, e.g., aboutness, coverage, novelty, reliability, or timeliness [BBC$^+$07]. Usually, information filtering involves comparing a user's profile to the characteristics of digital resources (content-based filtering) or to the characteristics of other users (collaborative filtering). Hanani et al. distinguish *active* information filtering systems that actively seek relevant information for a user and *passive* information filtering systems that filter out irrelevant information from incoming data [HSS01].

Table 2.1: An overview of pushing and pulling technologies

---

[8]Query-based searching can also be combined with filtering technologies. This is sometimes called *filtered search*.

Please note that this characterisation is different to characterisations that make a distinction between pushing and pulling based on whether an information flow is excepted by a user or not (e.g., [CS01]). Usually, only a selected amount of information that is available in an information system is delivered to the user, so pushing techniques are sometimes also referred to as "Selective Dissemination of Information (SDI)" [Luh61].

Table 2.1 gives an overview of different pulling and pushing techniques.

Whether pushing or pulling technologies are used, one can usually divide the process of accessing resources in the following steps depicted in Figure 2.2 on Page 22:

Access process steps

Step 1 – Resource subset selection: It is usually not possible to manually check whether every resource that can be accessed in an information system is relevant for the current information need. Thus, a subset of resources has to be selected for further inquiries. Such a subset can be created using different pull- and push-technologies as presented in Table 2.1:

- A query-based search will usually return a ranked list of resources.

- Certain filter criteria can be selected when faceted browsing is applied (e.g., "only show me documents of user X generated in year Y").

- A user's feed reader presents new entries from feeds that a user has subscribed to.

- A recommendation engine automatically suggests a set of resources it considers as relevant for a user.

Step 2 – Resource subset delivery: A view on the resource subset selected in step 1 is created and presented to the user.

Step 3 – Resource and metadata retrieval: In this step, a user selects a resource with the purpose of examining it in more detail. The resource and respective metadata is then delivered to the user by the information system.

Step 4 – Resource examination: Finally, the user examines the resource and the respective in more detail. This might include a final decision whether it is considered as relevant or not.
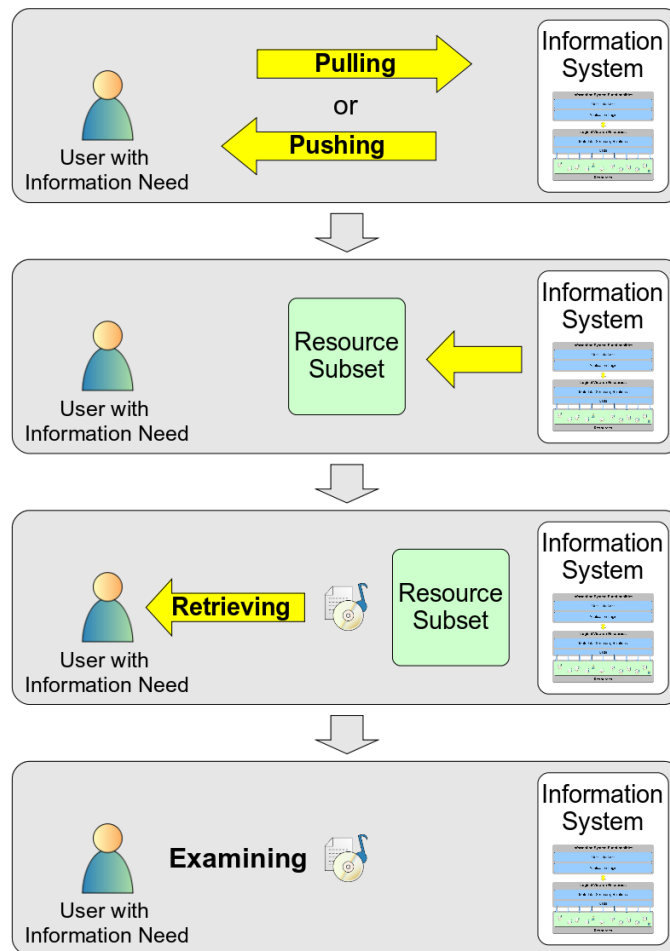
Figure 2.2: The resource access process in information systems

The disciplines *information retrieval* and *information seeking* investigate how this process can be supported, and how according information systems can be built.

Information seeking **Term 2.6** information seeking *refers to the human-oriented activity of attempting to find information in an information system.*

information seeking includes, e.g., activities such as filtering and browsing. While information seeking is usually more human-oriented, information retrieval is more technology-oriented and has the function to guide the users to the resources that will best enable them to satisfy their information need [Rob81]. In this work, the definition of information retrieval provided by Salton in [Sal68] is used:

Term 2.7 information retrieval *is a field concerned with the structure, analysis, organisation, storage, searching, and retrieval of information.*

<div align="right">**Information retrieval**</div>

While information retrieval was mainly concerned with the retrieval of text documents in the 1970s and 1989s, the huge amount of digital resource accessible led to a widened scope of the field also including topics such as multimedia retrieval, summarisation, or topic detection [AAB⁺03].

2.3 Supporting Users in Accessing Digital Resources

As a basis for our investigations on how users can be supported in accessing digital resources in information systems, it is first specified what would characterise an ideal setting, and which factors have to be considered. Then, a detailed discussion of two of the most important concepts when aiming to realise such a setting follows: *information need* and *relevance*. Thereafter, different ways to measure the quality of an information system's output will be presented.

2.3.1 An Ideal Setting

In this work, the focus is on supporting the access to digital resources for both information retrieval and information seeking scenarios. What all these types of information access – whether they are based on pulling or pushing – have in common is that they always are triggered by a certain (assumed) information need of a user. So what is the aim? How would an ideal setting look like?[9] For answering this question, three aspects have to be considered:

<div align="right">**An ideal setting**</div>

Existence of resources   meeting the user's information need

Accessibility of resources   that are potentially relevant for a user's current information need

Adequate information about resources   allowing the user as well as machines to efficiently find and access the relevant resources. Referring to the resource access process introduced above, this especially concerns

- the *resource subset selection*, and
- the *resource examination*.

Main problem Given that at least some resources that can satisfy the user's information need exist and are accessible, the main problem to solve is to provide – for users and machines respectively – adequate information about the resources that allow to access them and decide about their *relevance* for a certain *information need*.

To know what "adequate" means, a closer look on what characterises *information need* and *relevance* will now be taken in the following section.

2.3.2 Information Need and Relevance

No matter whether a resource is accessed as a result of an information retrieval or information seeking process – in the end it will always be the user who judges whether a resource is relevant for him/her and satisfies his/her information need. So when aiming to provide an information system that allows users with a certain information need to access digital resources relevant for them, it is crucial to understand the meaning of both the concepts *information need* and *relevance*. Thus, the following questions will be discussed now:

- What constitutes a user's information need? How can an information system capture according factors?

- What does relevance mean? Which factors influence the decision whether a resource is considered as relevant?

Information Need

Several efforts have been put into the definitional problem of *information need* and the difficulty to separate this concept from related ones such as "wants", "expressed demand" or "satisfied demand" [Wil81]. As shown in the sample

Information need is determined by a variety of factors scenarios, the information need of a user is determined by a variety of factors such as the user's knowledge, experience, and current context [MTW08]. The importance of these factors is also stressed by Wilson who claims that information need has been "the subject of much debate and no little confusion", and that this is partly due to the failure of not identifying the context of an information need [Wil81]. Barry states that "information need encompasses all factors a user brings to the situation: previous knowledge, awareness of information that is available, affective or emotional factors, the expected use of the information, any time constraints under which the user is working, and so on" and that "information need situations, as based on these types of factors, are dynamic and constantly changing" (cf. [Bar94, p.149]).

---

[9]Please note that personalisation and adaptation of resources is not in the focus of this work.

As information need is affected by such a variety of factors, and because a user might even not be aware of some of these factors when interacting with an information system, one has to consider that a transformation process will always take place when a user has to provide an explicit representation of his need to an information system. According to Taylor, one can distinguish the following four levels of information need (cf. [Tay62, p.392]):

Visceral need: The actual, but unexpressed, need for information.

Conscious need: The conscious within-brain description of the need.

Formalised need: The formal statement of the question.

Compromised need: The question as presented to the information system.

Although information systems must "deal with people in all their complexity: their tasks, knowledge, personality, motivation, etc." [Rut05, p.3], it is obvious that it is impossible for an information system – even in an ideal setting – to be aware of every aspect that influences the information need. A system can just make use of the explicit question of the user represented in a certain syntax, and the additional information that it has at hand about this user. In case such information exists, it is usually stored and represented in a *user model*.

> **Information systems can never be aware of all relevant aspects**

Term 2.8 *A* user model *is an explicit representation of some characteristics of a certain user.*

> **User model**

When designing an information system that aims to store information about a user in a user model, according characteristics have to be chosen in advance. Table 2.2 on Page 26 shows an example of possible user data categories for such a user model – the depicted categories are used in the IMS Learner Information Package (LIP)[10] designed to describe characteristics of learners in learning management systems, human resource systems, knowledge management systems, etc. A further example for a user model can be found in Section A.2, where the metadata captured about users in the DaMiT system is presented.

Information that is relevant for a user model can be gathered using different approaches. Therefore, *implicit* and *explicit* user model techniques can be distinguished:

Explicit user model: Information is acquired directly from the user, e.g., using editors or questionnaires. Downsides of such approaches are that they are usually time-consuming, distract the users from their real tasks, and may lead to a feeling of being diagnosed by a system [Ric83].

---

[10]See `http://www.imsglobal.org/profiles/lipinfo01.html`

Category	Content
Identification	Biographic and demographic data relevant to learning.
Goal	Learning, career and other objectives and aspirations.
Qualifications, Certifications, and Licenses	Qualifications, certifications, and licenses granted by recognised authorities.
Activity	Any learning-related activity in any state of completion. Could be self-reported. Includes formal and informal education, training, work experience, and military or civic service.
Transcript	A record that is used to provide an institutionally-based summary of academic achievement. The structure of this record can take many forms.
Interest	Information describing hobbies and recreational activities.
Competency	Skills, knowledge, and abilities acquired in the cognitive, affective, and/or psychomotor domains.
Affiliation	Membership of professional organisations, etc. Membership of groups is covered by the IMS Enterprise specification.
Accessibility	General accessibility to the learner information as defined through language capabilities, disabilities, eligibilities and learning preferences including cognitive preferences (e.g., issues of learning style), physical preferences (e.g., a preference for large print), and technological preferences (e.g., a preference for a particular computer platform).
Security Key	The set of passwords and security keys assigned to the learner for transactions with learner information systems and services.
Relationship	The set of relationships between the core components. The core structures do not have within them identifiers that link to the core structures. Instead all of these relationships are captured in a single core structure thereby making the links simpler to identify and manage.

Table 2.2: Categories used in the IMS Learner Information Package

Implicit user model: Information is captured by observing the user, or by inferring new information based on existing data in the user model:

User observation: An information system can capture data about potentially all actions it can track. This can include, e.g., information about resources a user has paid attention to, about searches that were carried out, and about data that was contributed [Sch10].

Inferring new information: Here, one can distinguish *shallow models* and *deep models* taking into consideration short-term or long-term interaction of a user [Ric83]. Furthermore, online and offline modelling techniques are distinguished.

While these approaches have the advantage not to distract the user, they bring with them the challenge to interpret the gathered information for generating assumptions about the user [ZD01].

Due to the complexity of information need and the limited possibilities to gather information about users, one can state that today's information systems are still "not yet able to resolve more than a portion of users' real information needs" [SEN90, p.758].

Relevance

The concept of relevance is crucial for the functioning and evaluation of information systems [Bor03]. Most information systems that try to deliver relevant digital resources for a specified information need primarily rely on subject matching, i.e., they try to retrieve topically appropriate documents. The underlying assumption in this approach is that terms can represent meaning, so matching the subject terms in the specification of the information need as well as the representation of the resource (this can be an index, but also associated metadata) is sufficient. Yet, although topicality is of course a very important aspect, users judge relevance using criteria beyond just that. For example, user-oriented aspects such as satisfaction or usefulness cannot be covered just relying on topicality.

Users judge relevance using criteria beyond just topicality

As for information need, the way individuals conceive the information contained in a resource also depends on a variety of factors such as their knowledge, experience and the current task. As Maron states in [Mar65, p.9]:

> *"information is not a stuff contained in books as marbles might be contained in a bag – even though we sometimes speak of it in that way. It is, rather a relationship. The impact of a given message on an individual is*

> *relative to what he already knows, and of course, the same message could convey different amounts of information to different receivers, depending on each one's internal model or map."*

These internal models are also known as "mental models" [Den06]. Together with the user's information need they are decisive for whether a resource is considered relevant or not. The meaning of relevance "is largely dependent on users' perceptions of information and their own information need situations' [SEN90, p.774]. Also, "a user's perception of the relevance of an object can change over time" [Rut05, p.19]. Thus, relevance is obviously not simply the property of a resource. Rees emphasised this in [Ree66, p.318]:

> *"We reject the notion that six ounces of relevance can be wrung like juice out of a document . . . There is no such thing as* the *relevance of a document to an information requirement, but rather the relevance judgement of an individual in a specific judging situation recording his judgement . . . at a certain point in time."*

Lewin's equation　The importance of a user's specific situation was also expressed by the German behavioural psychologist Kurt Zadek Lewin who stated that behaviour B is a function of the person P and his or her environment E [SMP03, p.119]. This is known as *Lewin's equation*:

$$B = f(P, E)$$

This environment is crucially influenced by other people, i.e., social aspects play a very important role. Knowledge is socially constructed, and the social context of a user is very important for knowledge-intensive processes such as learning [LW91].

There is no consensus on the meaning of relevance　As with information need, defining relevance is very difficult. Schamber et al. state that "although relevance judgements are fundamental to the design and evaluation of all information retrieval systems, information scientists have not reached a consensus in defining the central concept of relevance" [SEN90, p.755], and that in spite of several efforts that have been made to understand the topic, "serious questions about the nature of relevance remain" [SEN90, p.756].

Relevance is subjective, multidimensional, dynamic and situational　Yet, one can state that it is "a subjective, multidimensional, dynamic and situational phenomenon" [Rut05, p.4]. Furthermore, relevance is no binary concept, i.e., there is no dichotomy "relevant or not". While some aspects of the information need might be covered by a digital resource, some others may be not. And although some resources might be considered as relevant, others might be

considered as even more relevant for a certain information need. Barry summarised the results of several studies about relevance starting in the 1960s as follows [Bar94, p.150]:

- *"The aboutness or topical appropriateness of documents is not a sufficient condition for judgements of relevance by the person who initially requested the information;*

- *The evaluation of relevance is closely tied to the requester's experience, cognitive state and perceptions, and relevance can only be judged by the requester;*

- *The requester's information need situation is typically a dynamic and fluid state which will be updated and revised as new information is received; and*

- *Evaluations of relevance will involve interactions among various factors including but perhaps not restricted to the requester's situation and goals; the requester's knowledge level and beliefs; the information being evaluated; the way in which information is represented; the availability of other information within the environment; and the time, effort and cost involved in obtaining information."*

Based on [Sar96], Borlund defines the following independent types of relevance in [Bor03]:

Systemic or algorithmic relevance representing the relation between the query and the resources retrieved by the system;

Topical relevance representing the relation between a resource and the topic represented by the query;

Pertinence, cognitive or personal relevance representing the relation between the resource and the information need as perceived by the user;

Situational relevance representing the relation between the resource and the user's current task.

A very important aspect that has an impact on the decision whether some resource is considered as relevant or not is the *credibility* [RD07] of the resource, i.e., the information provided about it and the respective creator(s). This is especially emphasised by different studies highlighting the importance of people when trying to find information. For example, Tom Allen's findings show that employees in R&D organisations were approximately five times more likely to turn to a person for information than to an impersonal source [All84]. Cross

Credibility plays an important role

et al. conducted a study with 40 managers, asking them to indicate where they obtained information critical to the success of a recent important project. The results are depicted in Figure 2.3 and show that other people play are far more important role than impersonal sources [CPPB01].

Sources of Important Information
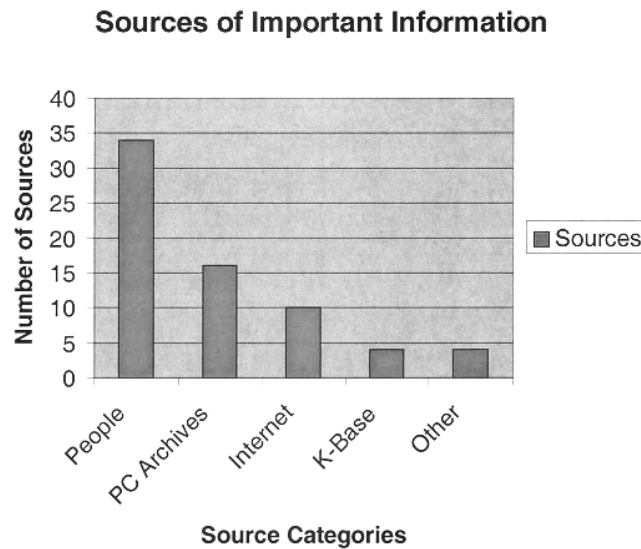
Figure 2.3: Sources of important information according to Cross et al. [CPPB01, p.7]

Although no clear definition of credibility exists, "the overarching view is that credibility is the believability of a source or message, which is made up of two primary dimensions: trustworthiness and expertise" [FM08, p.8]. In contrast to the large number of credibility studies that focused on domain-specific scenarios, Hilligoss and Rieh identified common aspects of credibility assessment that are valid for a broad range of scenarios [HR08]. They identified three levels of credibility judgements that are shown in Table 2.3. The *construct level* refers to how a user conceptualises credibility, the *heuristics level* is concerned with general rules of thumb for credibility judgements, and the *interaction level* involves specific source or content cues used for credibility judgements.

Level	Definition	Types	Influence
Construct	Conceptualisations of credibility	• Truthfulness • Believability • Trustworthiness • Objectivity • Reliability	Provides a particular point of view for judging credibility
Heuristics	General rules of thumb that are broadly applicable to a variety of situations	• Media-related • Source-related • Endorsement-based • Aesthetics-based	Provides useful ways of finding information conveniently and making credibility judgement quickly
Interaction	Specific attributes associated with particular information objects and sources for credibility judgements	• Content cues • Peripheral source cues • Peripheral information object cues	Provides specific information source or object characteristics on which to base a judgement

Table 2.3: Credibility assessment levels identified by Hilligoss and Rieh [HR08, p.1473]

As for relevance in general, one can state that credibility is also a very individual, context-dependent, and multidimensional concept.

2.4 Metadata – Organising and Describing Digital Resources

As shown in the last section, it is decisive for the quality of an information system to offer means that allow to find resources relevant for a certain information need. Referring to the process as introduced in Section 2.2.3, this especially concerns the steps *resource subset selection* and *resource examination*.
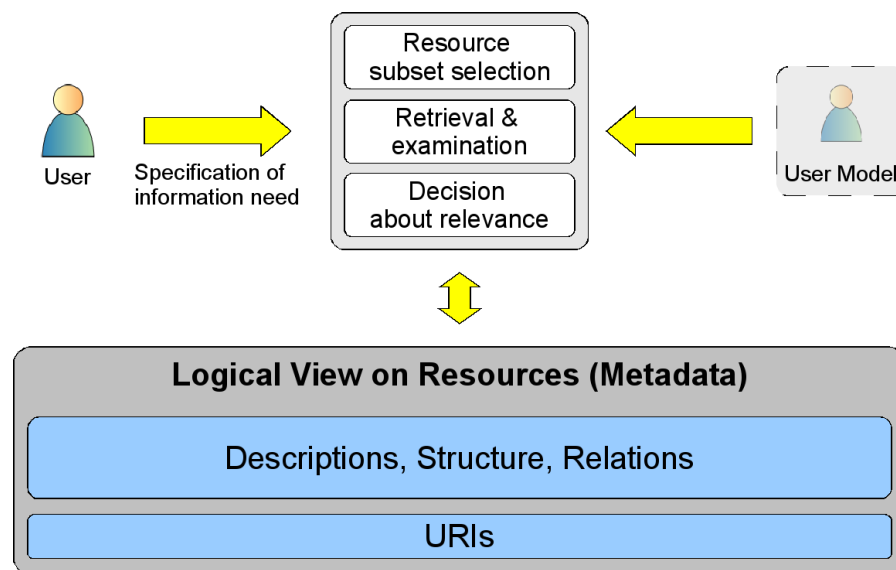
Figure 2.4: Available input to access resources and to judge their relevance in an information system

Figure 2.4 shows that the information about digital resources is key for machines as well as users when aiming to access digital resources and to identify whether a digital resource is relevant for a given information need. Such information can concern the content of a digital resource, relations to other digital resources, prerequisites for its use, the context of its creation, etc. Usually, it is referred to as *metadata*:

Metadata   **Term 2.9**   Metadata *is data about data.*

Please consider that this definition neither demands that metadata is machine-readable nor that it is represented in a certain format (although this is the most common type of metadata). In fact, it includes any conceivable information about digital resources at any level of aggregation.

Metadata can provide information about various aspects of a digital resource, it can be generated and used by different stakeholders and for different purposes, and numerous approaches exist to store and represent it. Thus, an overview of different metadata types, involved stakeholders and usage scenarios, standards, and ways to represent metadata will be given now.

2.4.1 Metadata Types

Metadata can provide information about a whole range of factors associated with a resource. In literature, one can find several attempts to classify metadata types, many of them based on how the respective metadata is used. For example, one can find distinctions such as *descriptive*, *structural*, and *administrative* metadata (see [Org04]). Yet, classifying metadata in such a way may cause problems, because metadata is often used for more than just one purpose.

The Functional Requirements for Bibliographic Records (FRBR) model developed by the International Federation of Library Associations and Institutions[11] (IFLA) is a approach that offers "structure and relationships of bibliographic and authority records, and also a more precise vocabulary to help future cataloguing rule makers and system designers in meeting user needs" [Til04, p.2]. This holistic approach is especially useful for supporting retrieval and access in bibliographic databases, but does not provide a fine-grained distinction regarding what kind of information is described by specific metadata terms.

Another approach is the distinction between *intrinsic* and *extrinsic* metadata. By intrinsic metadata all kind of information is meant that can be inferred from an analysis of the respective content itself. Extrinsic metadata, on the other hand, denotes the metadata that is inherited due to associations with something else. Yet, it is not always possible to clearly distinguish whether certain metadata are intrinsic or extrinsic, as the process of inferring metadata from the content itself will almost always require certain background knowledge – there is no such thing as a self-containing document! The only information at hand is the representation of the content, e.g., a bitmap or characters that are encoded in a specific way. Even when talking about a simple example such as a text document where the author of the document is mentioned on the title page, this information can only be inferred with knowledge about the respective language and the knowledge about how information about authorship is commonly represented in a document.

Intrinsic vs extrinsic metadata

Therefore, a categorisation that is focusing on what kind of information about the resource is described will now be introduced. It is based on the categorisa-

---

[11]See `http://www.ifla.org`

tion used in the Learning Object Metadata (LOM) standard ([oEL02]) and consists of the following six main categories: *Creation*, *Technical Aspects*, *Content*, *Rights*, *Relations* and *Usage*.

Creation: Information about how, when, by whom, for what reason, and in which context the resource was created and made available. This can also include lifecycle information about the resource history or different versions.

Technical Aspects: Here one finds information about the (digital) representation of the resource. Among others, this can contain information such as

- a resource identifier (e.g., a URI)

- the resource type (e.g., doc, pdf, mp3)

- the file size

- how the resource can be accessed

- hard-/software prerequisites

Content: This covers all aspects related to the content of the resource that can be inferred from the content itself, i.e., it is *intrinsic metadata*. Examples are

- title

- language

- topic of the content

Rights: Here, one finds information about intellectual property rights and conditions of use for the resource.

Relations: This *extrinsic metadata* covers relations to other resources and structures. This especially includes information about how the resource can be classified according to a given schema (this can be an entry of a taxonomy, but also a numeric value representing the degree of difficulty). Relations can be defined explicitly (using relations such as is-a or part-of) and implicitly (e.g., referring to the structure of a repository). Here one can also find structural metadata for compound objects.

Usage: The usage metadata concerns information about when, by whom and in which contexts the resource has been used. Such information can be derived from log files or user observation components [MD07].

Please note that this categorisation does not consider by whom metadata is created. For example, both the formal subject classification generated by a resource author in a digital library and a rating generated by an end-user in a social resource sharing platform can be considered as classifications of a resource.

2.4.2 How is Metadata Used?

Metadata has been proven to be important for a variety of aspects associated with (digital) resources. Based on [BS04] and [otFRfBR98], the most important scenarios for the use of metadata will now be presented.

Resource discovery: Metadata about various facets of resources (e.g., the authors, the title or relevant topics and keywords) can facilitate the discovery of relevant resources that correspond to a user's expressed information need. Such metadata can also be very helpful to filter out inappropriate resources and to provide the most appropriate versions of a digital resource.

Resource understanding: Metadata can help users and machines respectively to get a better understanding of a resource, and thus – as already mentioned in Section 2.3.1 – is key to judge the relevance of resources for a given information need.

Resource credibility: As already mentioned in Section 2.3.2, credibility plays an important role in the decision whether some resource is considered as relevant or not. Metadata about the provenance of a resource (e.g., information about who created and published it in which context?) thus can help to judge the trustworthiness and quality of a resource.

Resource identification: The task to identify a resource or to distinguish between different resources can be supported through the provision of respective metadata.

Resource location: Metadata can help to get information about where a resource is located. Thus, it can provide the necessary information for an information system to deliver the according digital resource, and it can facilitate the access for human users.

Resource organisation: Digital resources can be grouped according to a variety of metadata.

Resource interoperability: The task to exchange and (re)use or repurpose digital resources from different contexts and applications can be eased a lot

through the use of metadata that follows certain specifications and provides the necessary information. This aspect will be discussed in more detail in Section 4.3.4.

Resources archiving, preservation, and maintenance: This traditional and administrative task from the library domain is also important in the era of digital resources. As digital information is often not static and persistent, but dynamically changing or can even disappear, there is a strong need for metadata that provides necessary information, e.g., about the lifecycle of a resource.

2.4.3 Metadata Schemas and Standards

Numerous schemas and standards exist for different means

There are numerous metadata schemas and standards to describe digital resources. Such schemas are created for different target audiences, domains, and purposes. MARC[12] (MAchine-Readable Cataloging), DC[13] (Dublin Core) by the DC Metadata Initiative (DCMI), LOM[14] (Learning Object Metadata) by the Institute for Electrical and Electronics Engineers (IEEE) LOM Working Group, and the multimedia content description standard MPEG-7[15] are just a few prominent examples. DC and LOM are now presented as examples illustrating how metadata can be represented.

Dublin Core

Dublin Core (DC) is a set of terms to describe any kind of resources, primarily on the World Wide Web. It was designed by the Dublin Core Metadata Initiative[16] (DCMI) in 1995 and consists of a basic set of elements to facilitate resource discovery and retrieval. The focus was to provide a simple means for any kind of individual to describe resources in arbitrary domains, and the simplicity of the approach should also facilitate the exchange and retrieval of resources in a networked environment. The Dublin Core Metadata Element Set consists of the 15 elements presented in Table 2.4 [DC003]:

---

[12]See `http://www.loc.gov/marc`
[13]See `http://dublincore.org/documents/dces`
[14]See `http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf`
[15]See `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm`
[16]See `http://dublincore.org`

Element name	Explanation
Title	A name given to the resource
Creator	An entity primarily responsible for making the content of the resource
Subject	The topic of the content of the resource
Description	An account of the content of the resource
Publisher	An entity responsible for making the resource available
Contributor	An entity responsible for making contributions to the content of the resource
Date	A date associated with an event in the life cycle of the resource
Type	The nature or genre of the content of the resource
Format	The physical or digital manifestation of the resource
Identifier	An unambiguous reference to the resource within a given context
Source	A reference to a resource from which the present resource is derived
Language	A language of the intellectual content of the resource
Relation	A reference to a related resource
Coverage	The extent or scope of the content of the resource
Rights	Information about rights held in and over the resource

Table 2.4: The 15 elements of the Dublin Core Metadata Element Set

Learning Object Metadata (LOM)

LOM is intended to allow the description of digital resources that can be used to support learning. It defines a hierarchy of elements with the following nine top-level categories:

Nr.	Category name	Category explanation
1	**General**	Groups the general information that describes this learning object as a whole
2	**Life Cycle**	Describes the history and current state of this learning object and those entities that have affected this learning object during its evolution

Nr.	Category name	Category explanation
3	**Meta-Metadata**	Describes this metadata record itself (rather than the learning object that this record describes)
4	**Technical**	Describes the technical requirements and characteristics of this learning object
5	**Educational**	Describes the key educational or pedagogic characteristics of this learning object
6	**Rights**	Describes the intellectual property rights and conditions of use for this learning object
7	**Relation**	Defines the relationship between this learning object and other learning objects, if any
8	**Annotation**	Provides comments on the educational use of this learning object, and information on when and by whom the comments were created
9	**Classification**	Describes where this learning object falls within a particular classification system

Table 2.5: The nine top level categories of the LOM v1.0 base schema [oEL02]

A complete overview of the LOM schema is given in Appendix A.1.

2.4.4 Representing Metadata

Metadata can be represented in a variety of ways. Several non-digital examples are known from everyday lives such as information written on the packing of goods or washing instructions written on signs inside of clothes. For digital resources, several possibilities exist to provide metadata that is associated with a resource. Thus, different association models and representation formats will now be introduced.

Association Models

Embedded vs associated metadata

Metadata can be stored separate from the resource it describes, or in can be embedded within it. One therefore distinguishes *embedded metadata* and *associated metadata* [DHSW02].[17] These different types are already known from real objects

---

[17]Duval et al. [DHSW02] also distinguish *third-party metadata*, which is a special kind of associated metadata usually stored in separate repositories.

such as books: information about them can often be found inside the book (e.g., a CIP[18] entry), but also in a library on an index card. In the world of digital resources, well known examples of embedded metadata are the metadata in pdf documents, in html pages, or in JPEG files (EXIF information).

Embedded metadata has the advantage that the metadata is very closely tied to the resource and will not be lost. The disadvantage is that working with large amounts of embedded metadata (e.g., to manage or access resources) will be more difficult and will raise performance issues.

Representation Formats

Metadata can be represented using any kind of syntax. Popular examples are SGML, XML or RDF. Some schemas require certain syntax, while others are syntax-independent.

Standards like the LOM declaration often define a concept, but not how the according data should be represented. Yet, to allow machine readability, data exchange and the development of tools to support a certain metadata format, according specifications or implementation guidelines are required. Therefore, bindings, i.e., mappings from the conceptual data model to a machine readable representation, are usually provided. For example, the IEEE LTSC is working on has three different bindings for LOM:

- LOM ISO/IEC 11404 Binding (see `http://grouper.ieee.org/groups/ltsc/wg12/par1484-12-2.html`)

- LOM XML Binding (see `http://grouper.ieee.org/groups/ltsc/wg12/par1484-12-3.html`)

- LOM RDF Binding (see `http://grouper.ieee.org/groups/ltsc/wg12/par1484-12-4.html`)

It is important to note that metadata does not necessarily mean machine-processable data about data. Although this is often very useful, very important information is also found expressed as natural language. To transform this into a machine-processable representation that conveys the same amount of (often subtle) information is a very complex and sometimes impossible task. Furthermore, a lot of information represented in a machine-readable way might not be interpretable for humans in a convenient way (e.g., think of the description of the content of a resource by means of an RDF graph). Thus, the metadata that has to be considered in this thesis is not necessarily machine-processable.

---

[18]CIP = Cataloguing in Publication

2.5 Metadata Quality - What is "Good" Resource Metadata?

From the information provided so far, it is clean how digital resources can be accessed, and respective process steps as well as the characteristics of an ideal scenario have been identified. For such a scenario, the decisive role of adequate metadata was pointed out – for machines and humans respectively – to allow to find digital resources and to decide whether a resource can be considered as relevant or not for a certain information need. Different metadata types were introduced, and information was provided on how metadata can be used and represented. Yet, the key question is: *How would ideal metadata about a resource look like?* And what can be concluded for the generation of "adequate metadata" for our ideal scenario?

   To identify which information about resources should be provided, it is important to remember what this information is required for: It should provide for users as well as machines information that allow to *find* and *access* the resource, and to *decide whether it is relevant for a user's information need*. More concrete, this concerns the *resource subset selection* and the *resource examination* as presented in Section 2.2.3. As already discussed in Section 2.3.2, information need as well as relevance are highly individual, multidimensional, and dynamic concepts. Moreover, it is simply impossible to anticipate all possible needs and scenarios. Furthermore, the decision about the relevance of a resource is very complex and can require information that allows to understand the meaning of a variety of facets of a resource. But what is the "meaning" of a resource? Is there something like *perfect metadata* about a resource? To discuss this, some philosophical aspects related to this question will first be considered.

2.5.1 Some Philosophical Considerations about the Meaning of a Resource

Question about "truth" is one of the oldest in philosophy

The question about "truth" and the "true meaning" of something is one of the oldest questions in philosophy. In the context of this thesis, especially *relativism* and *perspectivism* provide interesting insights. Relativism denotes a group of philosophical doctrines for which reasons for statements and justifications for acts respectively are only possible by assuming principles that do not have superior or universal validity [Car95], and where knowledge is considered as relative and only valid for a certain viewpoint. Relativism is a superordinate concept of positions such as perspectivism, historism and scepticism [Kön92]. Perspectivism was developed – among others – by Leibnitz and Nietzsche, who stated

(according to [Bla05, p.13]):

> *"there are no facts, just interpretations".*

Transferred to the problem of providing adequate information about a resource, this means that there is not one true and perfect description or classification, but potentially one for each imaginable interpretation. Certainly this does no imply that all these creations are equally useful or good (measured according to some predefined goals).

The position that there is not simply one and definite meaning of a resource is also supported by the Austrian Philosopher Ludwig Wittgenstein who emphasises that meaning is constructed, saying ([Wit01, p.43]):

> *"The meaning of a word is its use in the language."*

One can translate this into the world of digital resource by stating:

> *"The meaning of a resource is its use in the community."*

This statement also underlines that the user's current context and (social) environment are of decisive importance, as already stated in Section 2.3.2.

2.5.2 Metadata Problems and Misconceptions

Meaning is a dynamic concept, and individual viewpoints and interpretations play a crucial role. This is frequently neglected in the context of metadata – misconceptions, wrong assumptions and expectations can often be observed. The way metadata has been used, and the expectations often connected with its use lead to several criticisms. One of the earliest and most popular criticisms is the article "Metacrap - putting the straw to the seven strawmen of Meta-Utopia" from Cory Doctorow [Doc01]. The aspects he mentions in this article can be considered as representative and provide a comprehensive overview of misconceptions and problems:

Use of metadata and connected expectations often lead to several criticisms

People lie: Depending on how metadata is used, a lot of people will always try to annotate metadata the way it best suits their interests, not the way it describes the content accurately. The HTML "Meta" element provides a good example of how metadata is often abused to gain more attention.

People are lazy: When there is no reason for users to annotate metadata (e.g., tagging a resource can help users to find it later), there is almost no motivation for people to provide information.

People are stupid: Apart from literacy aspects, providing complex information about resources often requires a deep understanding of a subject. This cannot be expected from the vast majority of people.

Mission: Impossible – know thyself: As Doctorow states, "People are lousy observers of their own behaviours", e.g., when programmers are asked to estimate the time to develop a certain module.

Schemas aren't neutral: Each way to categorise concepts implies the importance of some concepts over others. Yet, these hierarchies can be completely different depending on the context in which a schema was developed, and in which it is used.

Metrics influence results: Whatever metric is chosen, there is always the danger that some items will be privileged, regardless of their overall quality.

There's more than one way to describe something: This refers to the problems just depicted in the philosophical considerations about the meaning of a resource. There is the implicit assumption in the structure of most metadata schemas which suggests that there is a one-to-one relationship between a resource and the metadata that describes it [Dow04]. But as already argued, 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, and by whom it will be used for what reasons. Wiley et al. therefore distinguish between objective (e.g., the size of a file) and subjective (e.g., the degree of interactivity of a learning object) metadata [WRG00]. A one-to-one relationship also neglects that metadata may change during the lifecycle of a learning resource [CDO06]. Sen also argues that "the most neglected area in the metadata management is the notion of managing changes in metadata" [Sen04, p.152].

Still, Doctorow states that this does of course not mean that metadata is necessarily bad, and that it can at least be considered as a good means to make assumptions about several aspects of a resource. Although it is very difficult to precisely distinguish between *subjective* and *objective* metadata, one can surely consider some types of metadata (e.g., technical metadata) to be more context-independent and less influenced by an individual's interpretation than others (e.g., content metadata).

2.5.3 Quality Criteria for Metadata

Although one can state that there is no perfect metadata for a resource, there are nevertheless quality criteria that can be applied in relation to what the metadata shall be used for. For example, inconsistency and a lack of precision in description and data entry can have a negative impact on precision and recall. Concerning *subjective* metadata, one mainly has to focus on aspects that allow the interpretation of metadata, and a good match between information need and relevance. Consequently, the OECD definition of metadata quality is followed, and it is defined here as *"fitness for use in terms of user needs"* [OEC03].

Several efforts haven been made to define quality criteria for metadata (e.g., [BH04, OEC03]). Some of the most important ones are introduced now:

Accessibility: This can refer to potentially all categories as introduced in Section 2.2.3.

Accuracy: This measuring of correctness and precision can concern typographical aspects, but also a classification (especially numerical values).

Completeness: Given a certain schema, completeness can partly be measured as the degree to which the respective entries exist. Yet, completeness regarding subjective metadata is very difficult, if not impossible to measure. A resource in general can hardly be semantically complete in terms of being fully self-contained – additional information such as the knowledge of symbols is always required to interpret the information that is provided.

Conformance to Expectations: This of course depends on the information need of the users, and thus also involves whether the metadata is considered as relevant or not. Metadata should contain all information that is expected, in a way that is eventually expected by a certain user group or domain.

Cost-efficiency: The creation of metadata can be a time-consuming and expensive task. This has to be considered in relation to the usefulness of the created metadata.

Credibility: This was already introduced as one of the main factors when estimating the relevance of a digital resource. It involves expertise and motivation of the creator or the organisation that is responsible.

Interpretability: Metadata of course has to be presented in way that allows the respective users to understand its meaning.

Logical consistency and coherence: The information given should not be contradicting.

Timeliness: On the one hand, this refers to the time lag between the provision of a resource and metadata about it. On the other hand, it also involves the fact that resources may change over time, so up to date information is an important aspect.

As metadata shall here be used in a variety of different scenarios and domains, criteria as defined by the ADL[19] in [Sco04] also have to be taken into account. These so-called *ADL ilities* define demands for learning objects in the Sharable Content Object Reference Model.

Accessibility: *"The ability to locate and access instructional components from one remote location and deliver them to many other locations"*

Adaptability: *"The ability to tailor instruction to individual and organisational needs"*

Affordability: *"the ability to increase efficiency and productivity by reducing the time and costs involved in delivering instruction"*

Durability: *"The ability to withstand technology evolution and changes without costly redesign, reconfiguration or recoding"*

Interoperability: *"The ability to take instructional components developed in one location with one set of tools or platform and use them in another location with a different set of tools or platform"*

Reusability: *"The flexibility to incorporate instructional components in multiple applications and contexts"*

Yet, these criteria sometime pose conflicting demands as illustrated in Figure 2.5 [MRZ05]. It shows some interdependencies of the *ADL ilities* – while some of the demands do not affect each other, others can have a positive, but also a negative influence.[20]

---

[19]The Advanced Distributed Learning Initiative, see `http://www.adlnet.gov`

[20]Please note that this table is not complete – it only provides some insights into the interdependencies and illustrates some potential conflicts.

	Accessibility	Adaptability	Affordability	Durability	Interoperability	Reusability
Accessibility		OK (impact of adaptability on accessibility)	OK	OK	OK	OK
Adaptability	OK (impact of accessibility on adaptability)		easy authoring decreases possibilities to enable adaptability (-)	conforming to standards is not sufficient to enable adaptability (-)	conforming to standards is not sufficient to enable adaptability (-)	conforming to standards is not sufficient to enable adaptability (-)
Affordability	OK	enabling adaptivity means considerably more authoring effort (-)		OK	effort to enable interoperability vs ability to use other components	lower authoring costs by using the same content in different scenarios (+)
Durability	OK	adaptability is only possible by not conforming to standards (-)	easy authoring possibly means not conforming to standards (-)		OK	OK
Interoperability	OK	adaptability is only possible by not conforming to standards (-)	easy authoring possibly means not conforming to standards (-)	OK		OK
Reusability	OK	adaptability is only possible by not conforming to standards (-)	easy authoring possibly means not conforming to standards (-)	OK	interoperability supports reusability on a technical level (+)	

Figure 2.5: Interdependencies between different demands on metadata, here related to on *ADL's ilities*. Positive influence is marked with a '+', negative influence with a '-'.

2.6 Approaches to Generate Metadata

Human generated vs automatically generated metadata

Metadata can be created in various ways and for different purposes. In the context of this thesis, the focus is on the metadata generated in social media environments. Nevertheless, to understand how this metadata can support the access to digital resources, other approaches to generate metadata also have to be taken into account. In general, two categories of metadata creators can be distinguished: *human metadata generation* and *automatic metadata generation* [Gre03]. Then, one can also distinguish which stakeholders such as content providers or end users contribute metadata. Each of these approaches has its benefits and limitations concerning various dimensions such as quality, scalability, areas where it can be applied, etc. Usually, machines can generate metadata very fast, but are very limited in what can be generated for which type of resource. Humans cannot compete with the speed of machines, but are not depending on the machine-readability of a resource or relevant background information. Different approaches and their characteristics are now briefly discussed.

2.6.1 Automatic Metadata Generation

Statistical analysis and information extraction

Several tools and technologies exist to extract information from digital resources, mainly from the fields of statistical analysis and information extraction. Statistical analysis particularly applies tf-idf[21] based methods that rely on the vector space representation of documents (see [BYRN99] for a detailed description of these approaches). Furthermore, technologies such as Latent Semantic Indexing, Principal Component Analysis or Independent Component Analysis can be used to extract topics [KZ01].

Information Extraction aims at obtaining information by conducting a context-sensitive text analysis and by using dictionaries and rule-based approaches. Heterogeneous content of a digital resource is transferred into a structured format, only presenting information that was requested by the user beforehand. Most information extraction technologies have been developed within the context of the Message Understanding Conferences (MUCs) [GS96] and the Automatic Content Extraction (ACE) program[22].

Application-supplied and creator-supplied metadata

Metadata that was already provided by users or applications can also be automatically extracted by machines (Duval et al. therefore distinguish between application-supplied and creator-supplied metadata [DHSW02]). This espe-

---

[21]The *term frequency-inverse document frequency* (tf-idf) is a commonly used weight in information retrieval. It is used to evaluate the importance of a term for a document in a collection.

[22]See http://www.nist.gov/speech/tests/ace/

cially concerns embedded metadata that is provided for a variety of resource formats. Such metadata can contain bibliographic and content-based information, but also almost anything about the multimedia data (i.e., technical information) [FM06]. Simple examples are the document type and the size of a resource. Image metadata may identify the device used to capture an image or the dimensions of the image or refer to low-level content analysis in pixel-based values, shapes, colours and texture. Moving image metadata can be bibliographic, describing the file name, date and creator, and content-based, such as scene-shot-frame identifications, shot start and end time-stamps, duration, camera motion etc. Closed caption decoders may also produce textual metadata, moving into high-level semantic content descriptions. Metadata provided in audio files can also be either bibliographic, e.g., MP3 files include metadata ID3 tags, such as the duration, title, creator, or content-based, referring to speech or music. While automatically generated metadata attached to audio-visual resources has been used so far for indexing, management and retrieval, it is still hard to extract high-level features (see [UKBB09] or [DUBqW09]).

During the last decade, various schemas have been designed to generate metadata for multimedia documents. Dublin Core was initially designed to generate metadata for textual documents and then extended for image, video and sound. The MPEG standard is globally used for video metadata, while web multimedia presentations can be encoded in XML-based MPEG [Hun98]. Flash video can also generate XML metadata descriptions.

A number of frameworks exist to automatically generate metadata, and four of the most prominent ones are now briefly introduced :[23]

Frameworks for automatic metadata generation

- **Tika**[24] is an open source Java framework for extracting and querying full-text content and metadata from various information systems (e.g. file systems, web sites, mail boxes) and the file formats (e.g., documents, images) occurring in these systems.

- **Lucene**[25] is an established open source framework that allows to create and search an index of documents. Contents are evaluated using statistical methods, and Lucene also offers to extract relevant terms from documents.

- **GATE**[26] (a General Architecture for Text Engineering) is being developed

---

[23]There are of course many more useful tools and frameworks to automatically generate metadata (e.g., ALOA, LingPipe, SAmgI, UIMA or SProUT), but in the context of this thesis, the presented tools are sufficient.

[24]See `http://tika.apache.org`

[25]See `http://lucene.apache.org/java/docs`

[26]See `http://gate.ac.uk`

by the University of Sheffield since 1995. This Java-based open source framework offers a variety of technologies from the fields of Natural Language Processing and Information Extraction.

- **TagTheNet**[27] is a GATE-based Web Service that offers Named Entity Recognition.

- **OpenCalais**[28] is an initiative by Thomson Reuters[29]. The OpenCalais Web Service can identify entities, facts and events within text-based resources. It uses Natural Language Processing, machine learning and other methods to analyse document and to find entities.

Characteristics   To sum up, automatic metadata generation can especially be applied as a fast and cheap means to extract existing metadata, and to extract content-based information from text-based resources. It often also provides a solid basis for recommendations that can be used, e.g., when aiming to help users to provide metadata. Yet, the creation of high-level features is still difficult, if not impossible. Furthermore, machines usually are not able to take into account the specific demands of users and their specific information needs.

2.6.2 Human Metadata Generation

Although Meire et al. state that "Manual creation of metadata is tedious, error-prone and doesn't scale" [MOD07, p.1], and despite the existence of several technologies that allow for the automatic generation of metadata, it was already determined that meaningful metadata – especially subjective metadata – can often only by created manually by humans.

Different tools can be used for the manual annotation of metadata. They are mostly template-based and produce an according mark-up in a predefined syntax of the metadata schema. Popular examples are the *Properties*-dialogues in document creation software such as OpenOffice or Microsoft Word, or web pages that offer the opportunity to tag or comment on a resource.

Different metadata creators   The manual generation of metadata can be executed by different kinds of creators. The following distinction is based on [Org04], and also takes end users into account:

---

[27]See `http://tagthe.net`
[28]See `http://www.opencalais.com`
[29]See `http://thomsonreuters.com`

Experts – Resource creators: As with authors of scientific publications, creators are usually most familiar with the their resources, so they can provide good information about the content of a resource.

Experts – Technical staff: Structural and administrative metadata that cannot be created automatically is often provided by technical staff, e.g., employees responsible for scanning documents.

Experts – Professional indexers: Metadata experts such as librarians or archivists often have the task to create metadata for resources. They are familiar with metadata schemas, creation tools, conventions, the need for consistency, classifications, etc. that is required for the creation of metadata in the respective environment.

End users: Especially in social media systems, end users can provide a variety of metadata about resources, e.g., tags and ratings.

Both expert- and end user generated metadata are now briefly discussed.

Metadata Generated by Experts

For many years, metadata about resources was mainly produced by experts, especially in the field of libraries. An expert is considered to be familiar with the resource, the respective domain, and the metadata creation process. The quality of such manually generated metadata depends on a number of factors such as the expert, the motivation and purpose of the metadata, and the respective review process [Doc01]. It is commonly assumed that metadata generated by experts is of higher quality than those generated by regular users [Gre03].

Yet, the manual production of expert metadata is usually a highly time-consuming and expensive process. This is mostly due to the characteristics of the environments in which such expert metadata is used. Very often, complex metadata schemas are used, and contribution processes involve a social process of formal agreement of the relevant stakeholders. As a consequence, many resources in such scenarios are fitted with only a minimum of possibly relevant metadata, and this metadata remains static. Corrections and editions require extensive work.

Metadata Generated by End Users

Especially in social media platforms like Flickr and YouTube, the process of generating metadata such as tags, comments, and ratings is very simple and thus

not very time-consuming (this social metadata will be presented in more detail in the next chapter). As a consequence, the number of potential users that can contribute metadata is much higher than in traditional environments with complex processes. Thus, not only more resources can potentially be described, but information about new resources can also become available much faster than in traditional environments [YJNT07].

In general, three different kinds of human generated metadata can be distinguished:

Metadata created explicitly: Here, users are explicitly providing metadata about resources, and there is a 1:1 mapping between the contribution and the respective metadata element. Example for such metadata are tags, comments, and ratings.

Metadata created implicitly: On the one hand, this includes usage data (e.g., "User 'Martin' viewed resource XYZ on 2010-05-01-12-34"). On the other hand, it also includes metadata that is a by-product of user activities such as creating a collection and herewith aggregating resources.

Although end users can usually not cope with complex scenarios as described above, they can still provide lots of meaningful metadata, especially subjective metadata. For example, Zhang and Jastram have shown that "most of the subject-oriented metadata on the Internet is not fraudulent or incorrect" [ZJ06], focusing on titles, descriptions and keywords.

Tagging   In many social media platforms, users can assign tags (i.e., freely chosen keywords) to the resources in the system. Tagging can be considered as one the most important types of end-user generated metadata about resources, and it is used in a wide range of scenarios for a variety of different purposes.

In contrast to systems where information about resources is only provided by a small set of experts, collaborative tagging systems take into account that the way individuals conceive the information contained in a resource differs a lot. A folksonomy ([Wal07]), i.e., the result of the collaborative tagging effort, can reflect this diversity. As pointed out in [GH05], collaborative tagging is most useful when there is nobody in the "librarian" role or if there is just too much content for a single authority to classify. For the web, where collaborative tagging has grown popular in the recent years, both cases apply.

Tags can convey information about potentially any facet of a resource. This concerns information about the content and creation of a resource, about the way it should be or was used, etc. In [GH05], the following kinds of tags for resources are identified:

- *identifying what (or who) it is about,*

- *identifying what it is,*

- *identifying who owns it,*

- *identifying qualities or characteristics,*

- *self reference,* and

- *task organizing.*

People use tags for different purposes. But not only the purposes of tags, but also the motivation of users to tag resources has to be considered. Marlow et al. identified the following criteria ([MNBD06]):

- *future retrieval,*

- *contribution and sharing,*

- *attract attention,*

- *play and competition,*

- *self presentation,* and

- *opinion expression.*

Characteristics   When creating metadata, humans can usually not compete with the speed of machines, but they are not depending on the machine-readability of a resource or relevant background information. And what is even more important: humans can contribute with opinions, interpretations and subjective views, something machines will – most likely – never be able to. This is especially important for subjective metadata like the quality or classification of a resource. Deriving information or knowledge from the data itself (i.e., intrinsic metadata) as well as associating information that can not be inferred from the data (i.e., extrinsic metadata) always requires the ability to understand and interpret the data. This process is influenced by various factors – among others, it often involves experience and individual, subjective opinions. Of course, humans need a motivation to contribute metadata, and depending on the number of resources, a critical mass of users has to be attracted.

2.7 Summary and Conclusion

This chapter presented the basic concepts required for an understanding of how access to digital resources in information systems can be supported. Therefore, a definition of resources, digital resources, and information systems was provided, and a respective, holistic architecture has been developed. Different ways to access digital resources in information systems were presented, and a general access process was developed and introduced.

As a basis to identify how users can be supported in accessing digital resources, the key characteristics of an ideal setting were then identified: *existence of resources, accessibility of resources*, and *adequate information about resources*. The concepts *information need* and *relevance* were then discussed in detail to understand how one can solve the task to provide adequate information about resources. Both concepts are highly individual, multidimensional, context-dependent, and dynamic. When aiming to find relevant information and deciding on the relevance of digital resources, credibility and other individuals and perspectives play a very important role.

Finally, the concept of metadata as a key to support the access to digital resources – for humans and machines respectively – was introduced. Therefore, a new categorisation of metadata types was introduced, and ways to create metadata, usage scenarios, ways to represent metadata as well as interoperability aspects were presented, compared, and discussed.

Several important quality criteria for metadata given a certain usage scenario and metadata schema are known, e.g., completeness, correctness, and coherence. Yet, a perfect description of a resource does not exist – it is simply impossible to anticipate all possible usage scenarios, and to judge the quality of subjective metadata in a satisfactory and unbiased way. Like information need and relevance, subjective metadata such as the quality or the degree of difficulty of a resource are determined by a variety of factors that are influenced by the context in which they are annotated, and this context can dynamically be changing. Individual views can only be provided by humans – they can contribute with real and authentic interpretations and opinions. While human generated metadata has its strengths especially in the provision of different views and metadata that requires a thorough and deep understanding of a resource, it also has its limits. Humans are slow, and they need to be motivated to contribute information. Automatic metadata generation, on the other hand, usually produces information very fast and efficiently. The extraction of existing metadata using tools such as Tika or the extraction of (named) entities with tools such as OpenCalais are just two technologies that can contribute a lot of very useful information about a resource. Still, automatic approaches are by their very nature restricted to ob-

jective metadata and cannot contribute with different views or opinions.

When examining human generated metadata in the context of this work, the metadata created by end users within social media environments is of special importance. Typically, such metadata is "flat", i.e., the environments only allow for the provision of simple metadata terms (e.g., tags, comments, and ratings) without using complex structures. As a consequence, and comparing this approach with more complex schemas like LOM, such social metadata is of course restricted regarding its potential expressive power – this especially concerns aspects like granularity and precision. Where one might have very specific attributes to describe certain aspects of a resource in a complex schema, one usually only has very coarse means (e.g., a tag) for the same purpose in social metadata. On the other hand, the simplicity of the approach and the fact that potentially any user may contribute information can lead to a much larger amount of metadata in comparison to a traditional environment. This allows for a variety of new approaches that are illustrated in the next chapter.

II

Challenges and Potentials of a Social Hub

The Paradigm Shift from Traditional to Social Media – Challenges and Potentials

The way that digital resources can be accessed, created, used, and modified has changed a lot in the last decade. Users turned more and more from consumers to producers. Numerous labels

Understanding the developments in the recent years as well as the characteristics of what is now referred to as *Web 2.0* is only possible within a much broader context. Humans very early tried to represent and hand down information in a more durable way than just by means of communication, i.e., not only using spoken language or gestures. The availability of such information and the way it can be accessed, created, used, modified, and made available to others has undergone deep changes from the days of cave-painting until *Web 2.0*. Therefore, a brief overview of the history of *writing and printing* will first be given, before the developments that led to what is called the *World Wide Web* will be presented, and the following paradigm shift leading to the introduction of the term *Web 2.0*. After a characterisation of this phenomenon, *potentials* and *challenges* for supporting the access to digital resources will be identified.

3.1 Writing and Printing

Already in prehistoric times (the oldest European cave-paintings were created approximately 32,000 years BC), mankind tried to create artefacts containing or representing information in some way. Yet, it was the introduction of *writing systems* that first allowed to represent complex information persistently, and to hand it down over generations in a reliable way. With the invention of *movable type printing* as a second key step in the development of information society, it was also possible to provide written information to a very large audience. In the following, information about both *writing systems* and *movable type printing* will be given.

Date	Event
≈ 32000 BC	Earliest known cave paintings in Europe
≈ 4000 BC	Development of a first writing system in Sumer
≈ 3200 BC	Earliest known hieroglyphic inscriptions
593	First printing press invented in China
700	First newspaper available in Beijing
868	Earliest known complete woodblock printed book (the "Diamond Sutra") is printed in China
≈ 1040	Invention of movable type printing in China by Bi Sheng
1450	Introduction of movable type printing in Europe (invented independently by Johannes Gutenberg)
1455	Release of the Gutenberg Bible

Table 3.1: History of writing and printing

3.1.1 Writing Systems

Writing is a "form of human communication by means of a set of visible marks that are related, by convention, to some particular structural level of language" [Enc08]. It enables "the representation of language in a textual medium through the use of a set of signs or symbols" [Wik08n]. Especially the use of alphabets, i.e., standardised sets of letters, allowed to efficiently represent such information. Important steps in the development of writing systems are shown in Table 3.1 [Wik08n, Wik08h, Wik08i, Wik08d].

Although writing systems provided an efficient way to represent information, the creation and thus also the copying of such information was still very time-consuming. Thus, resources that contained information were rare, and one can speak of an era of scarcity concerning the availability and accessibility of information. In Europe, this scarcity was often even reinforced by the political systems in which the writings existed – in these systems, access to information was only allowed for privileged members of society.

3.1.2 Movable Type Printing

Techniques to reproduce texts (and also images) were first invented in Asia, where woodblock printing was used to print information on textiles and paper. In the following centuries, these techniques were refined, wherein the invention of movable type printing was the major step towards the ability to reproduce

information on a large scale in a quick and durable way. Technically it was then possible to make information available to very large audiences (of course, this usually only included alphabetised circles) in a far more convenient way than before. Still, information was tied to a physical medium such as a newspaper or a book.

Table 3.1 gives an overview of the most important events towards the development of modern printing technology.

3.1.3 Electricity and Computers

Although industrialisation led to several technological advances in the 18th and 19th century, communication above the level of smoke signals, semaphore systems or similar techniques was still tied to physical transportation, i.e., the information had to be physically carried from a sender to a receiver. With the invention of electrical telecommunication in the 1830s systems such as electrical telegraphs appeared [Wik08l, Wik08f]. Information could be transformed into electric signals, and these could be transformed into information again. So it was now no longer tied to a physical medium. With the existence of respective infrastructures, information could then be provided to a very large audience over very large distances almost immediately. Simple telegraphy was followed by the invention of the telephone, and more complex technologies such as radio and TV followed. Table 3.2 gives an overview of the most important steps in this development.

Still, the transformation and distribution process usually was only possible with a loss of quality (at least when it concerned more complex information than just a telegram). And whereas almost everyone could be a consumer, the publishing of information was only possible through a complex and controlled process, where few people in the broadcasting media acted as gatekeepers. For example, producing a TV program and transmitting it to a large audience required the existence of a very expensive technical infrastructure, and it was only possible for organisations that had access to the available distribution infrastructure required for broadcasting the signals to the end users.

Computers

Calculating machines such as the abacus existed a long time before the first devices existed that are today called "computers". They allowed to manipulate information represented in form of discrete values. Such a discrete representation is called *digital* (from the Latin word *digit* for finger or toe), although one

Date	Event
1206	Invention of the water clock, a water-powered astronomical clock, by Al Jazari. It is considered to be the earliest programmable analogue computer.
1623	Invention of the first digital mechanical calculator by Wilhelm Schickard
1792	First fixed visual telegraphy system between Lille and Paris by Claude Chappe
1820	Creation of the first commercially successful mechanical calculator by Charles Xavier Thomas
1839	First commercial electronic telegraph by Charles Wheatstone and and William Fothergill Cooke
1832	Classroom demonstration of wireless telegraphy by James Lindsay
1837	Electrical telegraph by Samuel Morse
1866	First transatlantic telecommunication using a telegraph cable
1876	Conventional telephones invented independently by Alexander Bell and Elisha Gray
1884	First electromechanical television system patented by Paul Nipkow
1900	Wireless transmission of a human voice by Reginald Fessenden
1901	Establishment of wireless communication between Britain and the United States by Guglielmo Marconi
1920	Broadcast of the first radio news program in Detroit, Michigan
1925	First contemporaneous transmission of moving, monochromatic images with continuous tonal variation by John Logie Baird
1928	Broadcast of the first transatlantic television signal between London and New York
1929	First publicly announced television broadcasts in Germany and the United States
1936	Invention of the first freely programmable computer (Z1) by Konrad Zuse
1946	Development of ENIAC, the first electronic general-purpose computer
1971	Invention of the microprocessor by Ted Hoff, Federico Faggin, and Stanley Mazor at Intel

Table 3.2: History of telecommunication and computers [Wik08l, Wik08f, Wik08j, Wik08e, Wik08m, Wik08g, Wik08b, Wik08c]

today mostly refers to a binary code when using this term. More complex machines later also allowed to automatically manipulate information. But the last technological quantum leap in the way one can work with information was the ability to represent, transport and manipulate complex information based on a list of instructions using electricity. This kind of machines is usually referred to as *computers* [Wik08b]. Table 3.2 gives an overview of the according developments.

Where the distribution and manipulation of information was almost always time-consuming and involved a loss of quality (especially concerning complex information), information represented digitally can be manipulated and distributed very fast and almost without any loss of quality using today's computers.

3.2 The World Wide Web

In the same way that an infrastructure exists for transmitting radio or TV signals, one today uses the *Internet* and especially the *World Wide Web* as means of accessing and providing arbitrary digital information. Visions of a networked knowledge-base existed long before the World Wide Web: Amongst others, intellectual fathers of the World Wide Web are Paul Otlet as the inventor of the "Mundaneum" [Ray94], Vannevar Bush with his famous article "As we may think" [Bus45] and Douglas Engelbart [Eng62], who later invented "hypertext" together with Ted Nelson.

The starting point of the World Wide Web that was introduced by Tim Berners-Lee in 1990 was laid in the 1960s, when first computer networks such as the ARPANET were developed for research and military. Table 3.3 gives an overview of some of the most important steps towards what is today called the *Internet* and the *World Wide Web*.

Date	Event
1958	Researchers at the Bell Labs invent the modem as a means of enabling the communication between computers [MPS99].
1965	The terms *hypertext* and *hyperlink* are first coined by Ted Nelson.
1969	Creation of the ARPANET, initiated by the **A**dvanced **R**esearch **P**rojects **A**gency that sponsored research institutes to save or extend the US' technologically dominating position. The intention is to securing fail-safe communication.

Date	Event
1973	First international connections to the ARPANET
1979	Start of the bulletin-board system USENET
1982	Invention of the TCP/IP (Transmission Control Protocol / Internet Protocol) transfer protocol (sometimes referred to as the Internet's hour of birth in a technical sense)
1983	All ARPANET hosts switch from NCP (Network Control Protocol) to TCP/IP.
1984	Creation of the Global Domain Name System (DNS) to connect Internet addresses via nodes in a network
1989-1992	Creation and release of the *World Wide Web (WWW)* by Tim Berners-Lee at the European Organisation for Nuclear Research (CERN) (March 13, 1989 is considered as the birthday of the World Wide Web.)
1993	Release of the Mosaic Browser
1995	15 million people with Internet access.
March 2008	1.4 billion (21 per cent of the world population) with Internet access[1]

Table 3.3: The development towards the World Wide Web

With the World Wide Web, an infrastructure existed that allowed potentially anyone to connect to it, and to freely access and publish information for a worldwide audience – with the benefits of digital representations as presented in the previous section. This switch from broadcast media to networked media fundamentally changed the structure in which information flows.

But although the infrastructure and technology existed, it took several years until the World Wide Web was not only used as a means to mainly consume information passively, and where end users created content (e.g., an html page) only in exceptional cases. The term Web 2.0 originated in the attempt to describe the change from a mostly passive medium with a focus on textual resources to a multimedia platform with lots of participation possibilities.

---

[1]See `http://www.internetworldstats.com`

3.3 Web 2.0

Along with the prevalence of high-speed Internet connections, the dropping of costs for being online, and the development of advanced web based applications that can easily be used by almost anyone, the World Wide Web has been – and still is – changing from a place where information was usually only consumed by most of the users to a more social and participatory system. As many studies show (see Section 3.4.4), more and more users contribute information in the World Wide Web (so called *user generated content*), using tools like blogs, wikis, social bookmarking services or file sharing platforms such as Flickr[2] and YouTube[3]. In addition to these rather object-centric networks, people-centric networks (also often referred to as social networks) such as Twitter[4] and Facebook[5] became more and more popular in the last years. 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" [And07], and where sharing of resources through an individual's social network is eased, is denoted as *Web 2.0*[6]. In [O'R05a], Tim O'Reilly gives the following definition of the term Web 2.0:

The World Wide Web became a more social and participatory system

Object-centric networks and user-centric networks

> *"Web 2.0 is the network as platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an 'architecture of participation,' and going beyond the page metaphor of Web 1.0 to deliver rich user experiences."*

According to O'Reilly, there are seven principles which characterise Web 2.0 applications [O'R05b].

Seven principles of Web 2.0 applications

The Web As Platform: Web 2.0 applications are hosted on the World Wide Web and accessed entirely via a browser. Thus, not only platform independence can be ensured, but the applications can also be permanently improved

---

[2]Flickr is resource sharing platform focusing on photos, see `http://www.flickr.com`.

[3]YouTube is a video sharing platform, see `http://www.youtube.com`.

[4]See `http://www.twitter.com`

[5]See `http://www.facebook.com`

[6]The term Web 2.0 was officially coined by Dale Dougherty (a vice president of O'Reilly Media Inc.) in 2004 during a conference brainstorming session between O'Reilly and MediaLive International [O'R05b].

("perpetual beta") without requiring the user to install updates or new versions of a software.

Harnessing Collective Intelligence: The power of the World Wide Web and the existence of various participation possibilities such as rating, tagging, writing reviews etc. allow to harness Collective Intelligence (this concept will be introduced in detail in Section 3.6.3). According to O'Reilly, "Network effects from user contributions are the key to market dominance in the Web 2.0 era" [O'R05b].

Data is the Next Intel Inside: Specialised databases are a key aspect in most of the significant Web 2.0 tools. For example, the success of applications such as Amazon[7] or Google Maps[8] can to a large extent be attributed to the large amounts of data that were created and made available.

End of the Software Release Cycle: Software in a Web 2.0 context is delivered as a service and not as a product. Thus, operations must be a core competency and the software has to be maintained on a daily basis (otherwise it will cease to perform), and it is important to treat users as co-developers.

Lightweight Programming Models: Lightweight Programming Models, especially in the area of Web Services, enable to easily syndicate and integrate data and functionalities from different applications (often referred to as "mashups"), while still allowing to only loosely couple the involved components. A well-known example is the Representational State Transfer (REST) approach. REST is often contrasted to the "heavyweight" and rather formal techniques of Web Services.

Software Above the Level of Single Devices: By delivering software as a service, Web 2.0 applications are no longer limited to the Desktop PC as a platform. They can be accessed with numerous clients platforms, e.g., portable devices.

Rich User Experiences: By using techniques such as AJAX[9], full scale applications can be delivered through the World Wide Web. Due to quick response times, users can be offered an experience that is similar to what

---

[7]See `http://www.amazon.com`

[8]See `http://maps.google.com`

[9]Ajax stands for Asynchronous JavaScript and XML and is a collection of several technologies. For further information see:
`http://www.adaptivepath.com/publications/essays/archives/000385.php`.

they are used to when interacting with Desktop applications. The rise of Desktop applications ports such as Google Docs[10] or Adobe Photoshop Express[11] provides good examples for this principle.

The changes leading to the creation of the term Web 2.0 can also be illustrated with the comparisons itemised in Table 3.4 [WD06, p.4].

	Web 1.0	**Web 2.0**
Paradigms	- Web is unidirectional - Home pages - Web form interaction - Services sold over Web - API and IPR ownership	- Web as platform - Collective Intelligence: "Wisdom of crowds" - Web Services - Real Web applications - Power of Data - Data management and enrichment - Conversation
Key Success Factors	- Own high-quality content - Strong brand - Large user base	- Community leveraging - Robust platforms - Open systems

Table 3.4: A comparison of characteristics in Web1.0 and Web 2.0

3.4 The Paradigm Shift from Traditional to Social Media

The developments subsumed under the term Web 2.0 have caused a paradigm shift mainly in two dimensions:

---

[10]See `http://docs.google.com`
[11]See `https://www.photoshop.com/express`

1. In spite of the existence of the Usenet and tools such as bulletin boards, the World Wide Web was for a long time mainly focused on connecting resources. In contrast to that, Web 2.0 is focusing on connecting users by offering communication and collaboration opportunities. Thus, this kind of software is often referred to as *social software*.

2. Instead of complex publishing processes with gatekeepers only allowing a few people to publish information for a larger audience, it is now possible for almost everyone to publish information easily and to reach potentially everyone with access to the World Wide Web (it has to be noted that this was – at least potentially – also possible in the traditional World Wide Web, but it was not until Web 2.0 that this potential was exploited on a large scale by ordinary end users without a specific technical background).

This is why one talks about a *paradigm shift from traditional to social media*. The following description of both social software and social media will explain this paradigm shift in more detail. Some numbers describing the size and impact of this phenomenon will then be provided in Section 3.4.4.

3.4.1 Social Software

In a broader sense the term social software subsumes all applications that support human communication, interaction, and collaboration. This also includes long existing tools like e-mail, instant messaging, and groupware applications. In the last years a narrower understanding of the term evolved: Today, social software stands for web applications such as blogs, wikis, and social networks that are based on the many-to-many communication of the Internet [Six05].

A definition of the term social software also including applications that are not web-based is provided by Tom Coates [Coa03]:

Social software

> *"Social software is a particular sub-class of software-prosthesis that concerns itself with the augmentation of human social and / or collaborative abilities through structured mediation (this mediation may be distributed or centralised, top-down or bottom-up/emergent)."*

Coates has identified three enablers for the abilities which are described in his definition:

1. social software removes limitations which are present in the real world (e.g., language, geography, background or financial status) on social as well as collaborative behaviour.

2. Users of software systems are often confronted with an information overload. Social Software can compensate for human inadequacies in processing, maintaining or developing social as well as collaborative mechanisms by providing appropriate filter mechanisms. Further structures which prevent, e.g., blame-culture and recrimination may be developed.

3. social software enables the creation of environments or distributed toolsets which pull useful end results out of human social and collaborative behaviour. For example, software might be generated that supports human creative processes in groups in order to obtain a distinct and productive end result.

In his second point Coates refers to the development of structures which prevent, e.g., blame-culture and recrimination. This aspect of social software is also mentioned in [Six05]. By providing feedback and rating mechanisms social software supports the development of a digital reputation. The idea that users can rate contributions of other users is not new. It follows conventions of web forums like Slashdot[12]. What is new is the interaction of these mechanisms with a variety of possibilities to connect and comment contributions. This way not only a personal atmosphere may be created for resource sharing platforms and similar applications, it also provides a basis for mechanisms that want to harness Collective Intelligence.

3.4.2 Social Media

Traditional media such as print, television or radio are usually controlled by large institutions and organisations. The process of publishing information in such a context is very complex, and it only offers very few possibilities for users to contribute information that will then be provided to the consumers. In contrast to that, Web 2.0 online tools and platforms such as blogs, wikis, forums and resource sharing sites allow a many-to-many communication where anybody with an Internet connection can share any kind of digital resources, their opinions, experiences, perspectives, etc.

Term 3.1 *We refer to participatory online media where information can be made public by end users through submission in a respective system as* social media.   **Social media**

---

[12]Slashdot is a mixture of a web based news service and a blog with a focus on technical news, see `http://slashdot.org`.
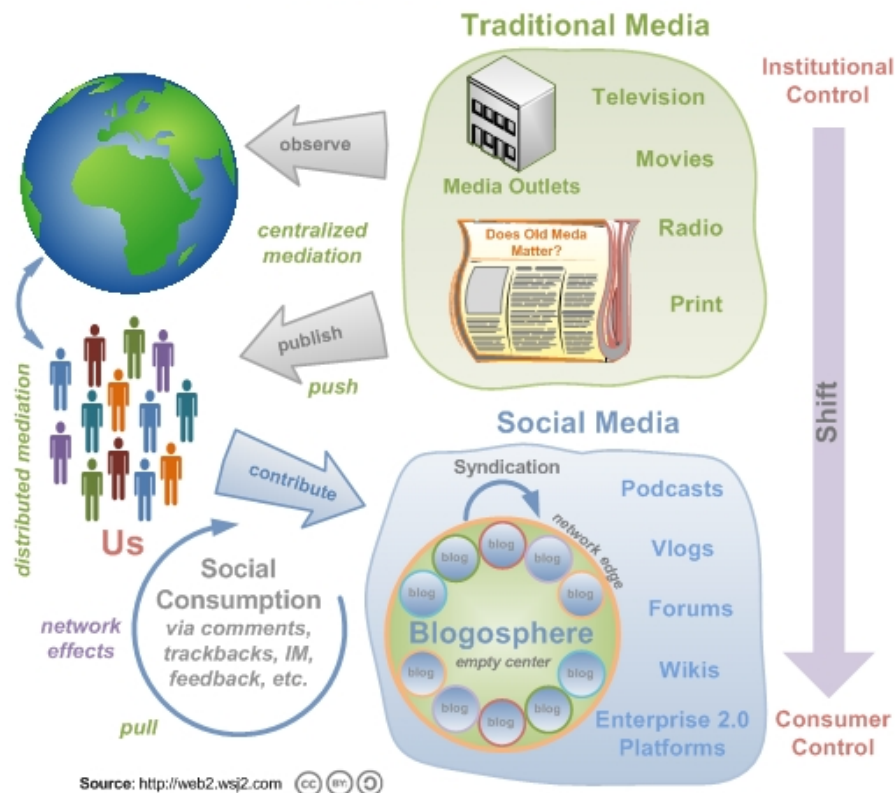
Figure 3.1: The paradigm shift from traditional media to social media

A shift from broadcasting to a many-to-many model

The paradigm shift from traditional media to social media (see Figure 3.1[13]) has the potential to lead to "democratisation of information, transforming people from content readers into content publishers. It is the shift from a broadcast mechanism to a many-to-many model, rooted in conversations between authors, people, and peers" [Wik08k]. The role of social media in the political movements subsumed under the term "Jasmine revolution" is a recent example for this potential.

An issue that is very closely interrelated with the new means to create and publish contents is the question of who is allowed to access or modify these contents in which contexts. As this question has several implications on how to support access to digital resources by means of social media technologies, it will briefly be discuss it in the following section.

---

[13]Source: `http://web2.wsj2.com`

© *Martin Memmel, 2015*

3.4.3 Intellectual Property Rights

A growing amount of content is freely available on the Internet, e.g., Open Access (OA) scientific publications or Open Educational Resources (OER). But "freely available" does not necessarily mean that it is allowed to modify or remix the content. So questions about the way resources may be used by whom and in which way play a very important role, especially because there are many applications that allow the creation of mash-ups in a rather simple way.

Intellectual property enables people to own their creativity and innovation in the same manner that physical property can be owned. The intention is to further support innovation and creativity by giving owners the possibility to control their intellectual property and to be rewarded for its use. Owners of intellectual property that want other people to use their content may license its use. Nowadays, Creative Commons can be considered as the most important license models for digital resources. It has proven its maturity and has been adopted by many types of media. Thus, this licensing model is now briefly introduced.

Creative Commons   Creative Commons[14] licenses are intended to help creators to retain the copyright over the work they provide. As delineated in [Com] licensees have to keep any copyright notice on all copies of the licensor's work and they also have to link to the licensor's license from their copies of the work. Further they are not allowed to alter the terms of the license or to use technology to restrict the lawful use of the work by other licensees. If licensees want to do anything that is restricted by the license (e.g., use the license owner's work commercially) then they have to ask for the licensor's permission. Every Creative Commons license allows licensees to copy, distribute, and display or perform a work of a licensor publicly. Further they can make digital public performances of it or make a verbatim copy of the work in a format which is different from the original format. Creative Commons licenses apply worldwide, are valid for the duration of the work's copyright, and are further not revocable.

A user can choose between six main license types which are composed of four key terms:

- **Attribution**: The licensor has to be credited for the original creation. This key term applies to each of the six main licenses.

- **Non-commercial**: Forbids the commercial use of the licensor's work

---

[14]CEO and Chairman of the Board of Directors of Creative Commons is Lawrence Lessig, see `http://www.lessig.org`

- **No Derivatives**: The licensor's work has to be passed along unchanged and in whole.

- **Share Alike**: All new content based on the licensor's work has to be published under the same license.

For an overview over the six main Creative Commons licenses as well as a detailed description see `http://creativecommons.org`.

3.4.4 Hype or Reality?

Although an abundance of new web-based applications emerged in the recent years, and social media has got a lot of press coverage (e.g., the Time Magazine voted *You* as the person of the year in 2006 [Gro06]), the question remains to which extent these tools are really used and by whom, and how big the impact of the above mentioned paradigm shift really is. Therefore, several numbers will now be presented for illustrating the role that social media currently plays, and is expected to play in the future. The numbers are derived from several studies and reports (sometimes international, but also nationwide studies) as well as from information given by social media providers.

Social Media in Open Environments

With open environments, we here refer to use cases outside an organisational context and where users are free to choose the information they want to access. The following facts illustrate the impact of social media in such environments:

- On the photo sharing platform Flickr, more than 92 million users from 63 countries shared around 1 million photos every day in February 2014.[15]

- According to the market-researcher comScore[16],
  - 770 million people visited a social networking site in July 2009, which is an increase of 18% compared to 2008 [com10]. The major social networks MySpace, Facebook, Hi5, Friendster, Orkut, Bebo, Tagged already had substantial growth from 2006 to 2007 (regarding the global reach). Whereas the sum of total unique visitors for these networks

---

[15]According to `http://techcrunch.com/2014/02/10/flickr-at-10-1m-photos-shared-per-day-170-increase-since-making-1tb-free/`, accessed March 12, 2014

[16]See `http://www.comscore.com`

was more than 135 million in June 2006, it increased more than 100% to more than 274 million users in June 2007 [com07].

– in 2009 social networking sites reached nearly 70% of Internet users worldwide [com10].

- In 2014, Facebook had more than 1.32 billion active users[17] that were uploading an average of 300 million photos per day[18], and was the 2nd most-trafficked website in the world[19].

- According to a report from the Pew Internet & American Life Project[20], released on January 14th, 2009, the share of adult Internet users who have a profile on an online social networking site has more than quadrupled from 2005 to 2008 [Len09]. The report further states that less than a third of social network users are using these networks for professional purposes. Instead, socialising with friends and people that the users already know plays the most important role for almost 90% of both teenagers and adults.

- Lenhart et al. report that in 2007, 64% of online users in the age from 12 to 17 have participated in one or more content-creating activities on the Internet, compared to 57% at the end of 2004. [LMMS07].

- A December 2006 survey by the Pew Internet & American Life Project found that 28 % of American Internet users have tagged content on the Internet, and that 7% of them tag or categorise online content on a typical day online [Rai07].

- A study funded by the John D. and Catherine T. MacArthur Foundation examining the participation of young people in the United States in the new media ecology states that social media tools "are now fixtures of youth culture"[IHB+08, p.1], and "have altered how youth socialise and learn, and raise a new set of issues that educators, parents, and policy-makers should consider"."Our values and norms in education, literacy, and public participation are being challenged by a shifting landscape of media and communications in which youth are central actors."[IHB+08, p.4].

---

[17]According to `http://newsroom.fb.com/company-info`, accessed September 11, 2014

[18]According to `https://zephoria.com/social-media/top-15-valuable-facebook-statistics/`, accessed October 13, 2014

[19]According to `http://www.alexa.com/topsites`, accessed October 14, 2014

[20]See `http://www.pewinternet.org`

- A study about German Internet users[21] showed that while on average, each grown-up in Germany spends 58 minutes in the Internet in 2008, users in the age from 14-19 spend 120 minutes each day. The same study states that the demand for Web 2.0 applications is rising, and that 23% of users already visited a photo sharing community, and 32% of them also uploaded pictures [FG08].

Social Media in Enterprise Contexts

The application of social media technologies such as blogs and wikis in enterprise contexts has increased a lot in the recent years. Sometimes this is referred to as "Enterprise 2.0" or "Knowledge Management 2.0". The rise of these technologies can be underpinned, among others, with the following numbers:

- The information technology research and advisory company Gartner[22] identified social software as one of the "top 10 technologies and trends that will be strategic for most organisations", stating that "social software technologies will increasingly be brought into the enterprise to augment traditional collaboration" [Gar08].

- An independent, market report released by the telecommunication company AT&T states that "social networking in the workplace increases efficiency" [T08]. The pan-European survey conducted by Dynamic Markets[23] shows that 65% of the employees that use social networking tools in the workplace "say that it has made them, and/or their colleagues more efficient". Furthermore, "46% say that it has sparked ideas and creativity for them personally". The study further shows that 65% of the participants say that social networking was adopted by their company as part of their working culture. AT&T executive director Martin Silman states that "the research shows that there is a clear trend across Europe for business users to embrace the benefits of 'web 2.0' technology to underpin collaboration, improve productivity and embrace business efficiency. It is clear that CIO's and their colleagues need to think about the implications this has for their own internal networking strategy and ensure that they are equipped to make the most of the opportunities created by social networking."

---

[21]The ARD/ZDF-Onlinestudie 2008
[22]See http://www.gartner.com
[23]See http://www.dynamicmarkets.co.uk

- For the report "Millennials at the Gates" [Acc09], the consulting company Accenture[24] examined how the Millennials (defined as those born between 1977 and 1997) views and uses social media. A survey with more than 400 US students and employees was conducted, and the results showed an increasing demand for high-tech devices to connect with colleagues, peers, friends etc. both inside and outside an educational or business setting. One-third of the Millennials between 18 and 22 responded that they expect to be able to access their preferred technology applications also in the workplace. Furthermore, the findings of the survey "point to a disconnect between the enterprise technology that organisations provide their employees and how young workers actually want to use technology and collaborate in the workplace" [Acc09, p.2].

Critical Voices

It may not be neglected that although it seems obvious that social media is used more frequently by a constantly growing number of users, there are also several critical voices. The main points of criticism are the underestimation of the role that traditional media like TV still plays, and the doubts concerning several of the characteristics often attributed to the so called *net generation*.

- Participation inequality – the phenomenon that people usually rather consume and only contribute to a comparably small extent – does also exist in Web 2.0: For example, a study about German Internet users[25] shows that only 15% of the users (yet, 57 % of the teenagers) are interested in active participation [FG08].

- A representative study about the adoption of the Internet in Germany conducted by the BAT Stiftung für Zukunftsfragen[26] in 2008 revealed that the proportion of private Internet users in Germany has more than doubled since 1999 (from 16& to 42%), and more than two third (71%) of users in the age between 14 and 29 are using the Internet at least once a week. Still, the majority of Germans (58%) are never using the Internet, in the generation older than 55 years even 83% never use it[27].

---

[24]See `http://www.accenture.com`

[25]The ARD/ZDF-Onlinestudie 2008

[26]See `http://www.stiftungfuerzukunftsfragen.de`

[27]See `http://www.stiftungfuerzukunftsfragen.de/de/forschung/archiv/2008/forschung-aktuell-210-29-jg-20112008.html`

- Schulmeister in [Sch08] questions the existence of the so called "net generation", "millenials" or "digital natives" with properties such as being very tolerant towards ethnical minorities, being particularly curious, having more self confidence and being more independent and smarter than earlier generations, etc. He states that there is often a lack of a sound scientific basis for these claims, and that it is mostly based on single experiences, anecdotes and episodical knowledge. Several contra-positions are listed by Schulmeister, and he concludes with a quotation from the Canadian Media Awareness Network: "The Internet just is" [Net04, p.8]. Amongst other, he summarises his examinations with the following theses ([Sch08, p.92–94]:

  - The primary aim of youths still is to win and meet friends.

  - Traditional media such as TV and video are still dominating. When examining the use of new media, music and videos are once again playing a very important role, besides social software to communicate and exchange information with others.

  - Several subgroups exist that highly differ concerning their motivations, frequency of use and competences. Furthermore, ethnical and social background play a very important role in this differentiation, so the so-called digital divide does not seem to be reduced.

The results of a Pew Internet Personal Networks and Community survey from 2009 [HSHR09] show that concerning American adults, the use of technology does not lead to social isolation. Instead, online participation leads to people having more and larger discussion networks, and social media activities are associated with several beneficial social activities. Furthermore, "the extent of social isolation has hardly changed since 1985" (it is about only 6% of the adult population).

Conclusion

Surely not all the myths or claims about the *digital natives* and other users of social media are true, and the importance and impact of Web 2.0 tools might sometimes be exaggerated. But one can definitely state that the rate of adoption is continuously rising, and that it will be totally self-evident for the next user generations to use such tools – also in enterprise contexts. Still, the phenomenon of participation inequality can still be observed, and outside of social networking tools people still rarely contribute lots of information. But even the small percentage provides enormous potential (as Wikipedia shows). And the

adoption rate in closed scenarios with much smaller user groups also provides enough evidence that the use of social media in such contexts will be more and more self-evident.

3.5 Entities and Interaction in Web 2.0 Applications

Figure 3.2: Web 2.0 themes represented in a figure by Markus Angermeier[28]

As depicted in Figure 3.2, there are numerous terms and concepts used in conjunction with the term Web 2.0: *user generated content*, *social media* and *social software* are just some of them. Also including the above mentioned key aspects that characterise Web 2.0 applications, these concepts are highly connected and interrelated. Moreover, it is very difficult to clearly separate them. What is missing in social media research is a clear terminology and structuring of the entities and interactions in Web 2.0 applications on a general level. Yet, the provision of such a framework is essential for a systematic identification of potentials and

A general and clear terminology and structuring is missing

---

[28]Source: `http://kosmar.de/archives/2005/11/11/the-huge-cloud-lens-bubble-map-web20`

challenges for supporting the access to digital resources. A structured overview of key components in Web 2.0 applications will thus be introduced in Figure 3.3.
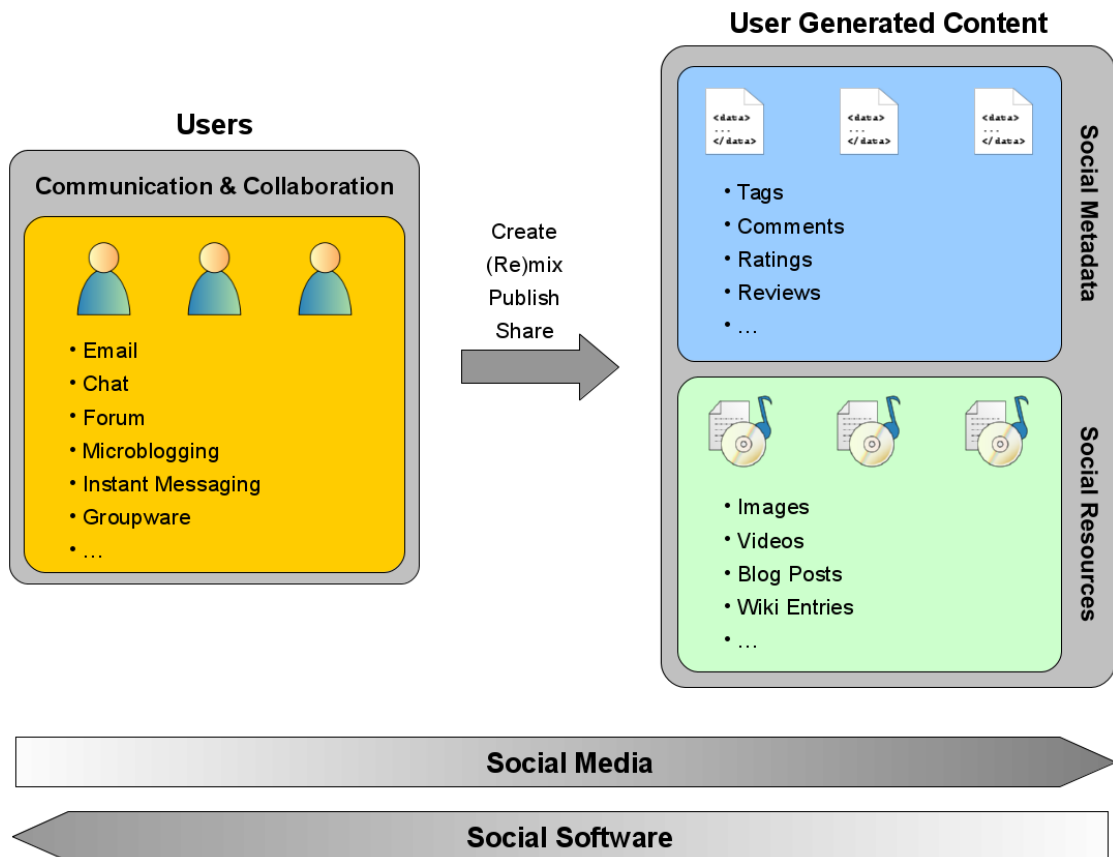
Figure 3.3: A structured overview of key components in Web 2.0 applications

On the left side, the users interacting with social software resp. social media can be seen. At this, two main types of interaction can be distinguished:

1. Communication and collaboration with other users by means of different mechanisms such as instant messaging or groupware. This is the focus of the concept *social software*.

2. Creation, publishing, (re)mixing and sharing of digital resources and metadata. The concept *social media* is mainly focusing on these aspects.

In the context of this thesis, it is the latter type of interaction that is most important. The result of these interactions is depicted on the right side of Figure 3.3: the *user generated content*. Although this term is used in different ways

(see, e.g., [Kos03] or [VWV07]), the common characteristics are reflected by the following definition:

Term 3.2 *We will refer to the sum of artefacts generated from users in social media environments by creating, publishing, (re)mixing and sharing of digital resources and metadata as* user generated content.

User generated content

Although this definition also includes digital artefacts originating from communications (e.g., emails, chat messages, or status updates) as user generated content, in this thesis such artefacts created mainly within people centric networks will be neglected, instead the focus is put on artefacts created within object centric networks. On the one hand, this concerns *digital resources* such as wiki pages, blog entries, images, or videos, partly causing or at least enhancing the *Long Tail* phenomenon that will be described in Section 3.6.2. On the other hand, this concerns *metadata* such as tags, comments, ratings or reviews, but also about the usage of resources. We will refer to these kinds of user generated content as *social resources* and *social metadata*.

Term 3.3 *A* social resource *is any resource provided by users in social media environments.*

Social resource

Term 3.4 social metadata *is any metadata created through user activities in social media environments.*

Social metadata

Note that the use of the term *social* here is generalising the characteristics of social media environments – it does neither claim that each social resource or social metadata is the result of a collaborative process, nor that any social activities exist around them.

3.6 New Phenomena, Potentials, and Challenges

For our aim to support access to digital resources, the developments described so far in this chapter involve several new phenomena and potentials, as well as ambitious challenges. These aspects will now be described, referring to the *resource access process* and *ideal setting* as described in Chapter 2. They are the central motivation for the topic of this work – the use of social media technologies to support the access to digital resources in various digital environments.

3.6.1 Interaction Possibilities – Means to Create and Provide Digital Resources and Metadata

Whether we talk about Web 2.0, the social web or social media– a common characteristic of the developments associated with these concepts is that they offer rich interaction possibilities to users. As depicted in Figure 3.3 above, we here can distinguish interaction with other users and interaction concerning user generated content. Typically, users can communicate and collaborate with others, they can create, re(mix), publish and share digital resources, and they can provide information about these resources (e.g., tags, comments and ratings). These interaction possibilities can turn users from simple consumers to active participants who also produce information (hence the term *prosumer* turned up).

Potentials and Challenges

Potentials As shown in Section 3.4.4, the interaction possibilities provided as a de-facto-standard in Web 2.0 applications as well as the resulting data are a main reason for their popularity. More and more users expect these functionalities as self-evident [Acc09]. Thus, providing such interaction possibilities in existing, traditional environments also has the potential to attract more users, and to encourage their participation.

In the context of this thesis, we focus on interaction possibilities related to the creation and provision of digital resources and metadata. First of all, means that allow users to create and provide digital resources in an easy way lead to a larger variety of digital resources that can be accessed by others.

Second, information about the usage of digital resources and other social metadata that is created explicitly (e.g., tags, comments, and ratings) or implicitly (e.g., through the creation of personal collections) allows for new or enhanced ways of

- generating views on digital resources such as *ordered by number of page views*, *ordered by rating* or *ordered by number of times bookmarked*,

- supporting users and machines respectively in judging the relevance of a digital resource,

- maintaining underlying structures of an information system (e.g., by taking into account end user contributions such as tags to add, remove, or modify elements from a taxonomy), and of

- navigating the content – this so-called *social browsing* (see scenario 2 in Section 2.2.1) allows users to *tap into the Long Tail* (this phenomenon will be discussed in the next section) and to find niches that are relevant for them.

The term *social browsing* was used by Lerman and Jones in [LJ07] to describe the browsing behaviour of users on the photo sharing platform Flickr. There, users mainly find resources by browsing through the photo streams of their contacts. To have a more precise understanding, we will use the following, broader definition:

Term 3.5 *We will refer to browsing processes that make use of the existence of social metadata as* social browsing. **Social browsing**

Applying and adapting existing technologies from the field of information retrieval that make use of traditional domain metadata as well as social metadata has the potential to allow for new ways of interacting with digital resources, and of providing new services based on these metadata. Especially tags as introduced in Section 2.6.2 play an important role, as they have proven to be an effective means of appropriately describing the content of a resource and can thus significantly improve resource discovery [LGZ08]. This is of special importance for videos and images, where several automatic approaches to analyse contents are investigated (see [UKBB09] or [DUBqW09]), but where it is still difficult to automatically extract appropriate textual representations. Furthermore, automatic trend detection [HJSS06] can be applied to find out about emerging and relevant topics.

The challenges when aiming to support access to digital resources by means **Challenges** of the discussed interaction possibilities will be:

- to provide them in an adequate way so that users are encouraged to make use of them,

- to offer them in existing, traditional environments in scenarios with different characteristics than the World Wide Web (e.g., intranets), and

- to develop new services that meet the users needs and exploit the full potential of all information at hand.

3.6.2 The Long Tail Phenomenon

The term *Long Tail* was first coined by Chris Anderson in an article in the Wired magazine[29] [And04]. It describes a phenomenon that can be observed in many markets (especially in eCommerce) and online platforms: a declining importance of very popular "mainstream resources" or "hits" in favour of niches or so-called micromarkets that are only attractive for a few users. Figure 3.4[30] depicts this Long Tail, a power law function: Every point on the x-axis stands for a single entity, whereas the y-axis represents the popularity of an entity (i.e., the number of times it was bought or accessed).
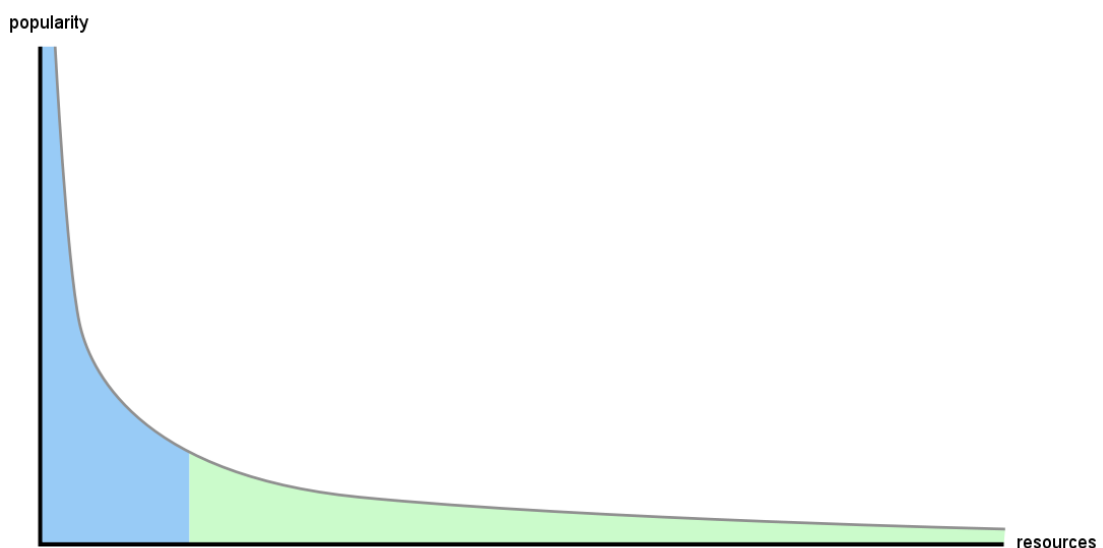
Figure 3.4: The Long Tail – a power law function describing the distribution of resources and their popularity

The online store Amazon[31] is a very good example for such a Long Tail distribution. In contrast to a traditional store where every good that can be sold also requires some physical space, and where only a few goods can be presented prominently to a very restricted number of customers, Amazon can offer basically any kind of product to any user with access to the World Wide Web. Although the majority of books is sold only a few times, about a quarter of Amazon's sales of books comes from outside the top 100,000 bestselling titles [And06].

---

[29]See www.wired.com
[30]Source: http://en.wikipedia.org/wiki/The_Long_Tail
[31]See http://www.amazon.com

The reasons for the emersion of Long Tail distributions are manifold. In [And06], Anderson identified the following forces:

- Democratisation of production (lengthens the tail)

- Democratisation of distribution (fattens the tail)

- Connections between supply and demand (drives business from hits to niches)

Although Long Tail phenomena could be observed long before the World Wide Web (e.g., in the emergence of big supermarkets offering a lot more products than small stores), the tools and infrastructure of the Web 2.0 significantly accelerated this development. First of all, we have a democratisation of production realised with tools that allow to create digital resources like images or videos by almost everyone. Second, using the World Wide Web and according platforms, resources or information about them can easily be published and shared (i.e., distributed). This also holds true for physical goods such as books. Last but not least, each resource can be offered to potentially any user or customer through the World Wide Web, thus further connecting supply and demand.

The Long Tail phenomenon can also be observed in the *production* of resources. In [OD08], Ochoa and Duval provide a quantitative analysis of several Web sites that are based on different types of user generated content. They showed that the production of user generated content follows a Long Tail distribution – this is often referred to as *participation inequality*.

Potentials and Challenges

In the context of this thesis, the Long Tail phenomenon first of all provides an opportunity. In contrast to environments where only mainstream resources could be accessed, the growing number of niches with very specialised digital resources gives users more choice, and also raises the chance that resources exist and are accessible that exactly meet the information needs of a user (of course, niche resources are not necessarily of better quality). **Potentials**

Yet, when it's about finding and accessing content, traditional mechanisms based on broadcasting (e.g., advertisements in the mass media) or on the popularity of a resource (e.g., Google's Pagerank) will not work or are at least not sufficient in such a situation. **Challenges**

The challenge is – besides the above described means to create and provide digital resources for the democratisation of production and distribution – to lead

and guide users to resources that are interesting for them. This especially concerns means to provide resource subset selections, where simple mechanisms just based on a user's query are not sufficient. Alternative methods to navigate and filter the content are needed – e.g., social browsing can be applied to offer intuitive access to resources in the Long Tail. Automatic approaches to estimate the relevance of a resource have to take into account a user's personal interests (e.g., by applying information filtering techniques [HSS01]).

3.6.3 Collective Intelligence

The harnessing of Collective Intelligence was already mentioned in Section 3.3 as one of the key principles of Web 2.0. Collective Intelligence is not at a new phenomenon from the field of computer science but has its roots in different disciplines such as biology (there sometimes referred to as *swarm intelligence*), sociology, and economy. Generally, it refers to a phenomenon where the collective behaviour of many entities (such entities can be humans, but also animals, bacteria or even particles) leads to a result considered as intelligent. Such phenomena can be found, e.g., in the World Wide Web, in science, in politics or in business [TW06]. A helpful characterisation emphasising the idea that Collective Intelligence is "more than the sum of its parts" emerged during the "Collective Intelligence FOO Camp" organised by O'Reilly at Google in 2008:

> *"The network knows what the nodes do not."* [Tor08]

To illustrate the concept of Collective Intelligence, we will now provide some examples where Collective Intelligence occurs or where it is harvested to provide additional value:

Ant societies: Sometimes referred to as "superorganisms", ant societies are highly organised and consist of millions of ants that communicate with each other and perform very different tasks (division of labour). Ant societies are able to solve complex problems such as constructing nests and cultivating food.

Google PageRank: The PageRank algorithm developed by Page et al. in 1998 [PBMW98] still is the basis for how Google[32] is estimating the importance of search results, thus ranking the results accordingly. The PageRank of a web site depends on its location in the Web's graph structure – in simple terms, the underlying assumption is that the more people link to a web page, the more important it is.

---

[32]See `http://www.google.com`

Wikipedia: This online encyclopaedia is built up only from contributions of volunteers. Basically any user with access to the World Wide Web can create and modify entries in the Wikipedia. The English Wikipedia site contains the remarkable amount of more than 3.2 million articles.[33] In 2005, the Nature magazine carried out a study comparing Wikipedia and the Encyclopaedia Britannica concerning their coverage of science. In the study, entries about numerous scientific fields from both sources were sent to experts for a peer review. The result showed that the difference in accuracy was surprisingly small: science entries from Wikipedia contained 3.86 inaccuracies on average, entries from the Encyclopaedia Britannica 2.92 [Gil05].

Digg: Digg[34] is a news service that allows users to provide feedback (so-called "diggs") about news they consider important. Based on this feedback, news are ranked accordingly, thus allowing users to discover relevant news.

Delicious The social bookmarking tool Delicious[35] offers the possibility to store and tag bookmarks online, as well as to share this information with others. Users' bookmark lists and the collective annotation of resources with tags and descriptions allow for social browsing and efficient retrieval of relevant content. As of November 2008, Delicious has more that 5.3 million users and over 180 million unique URLs saved.[36]

Akismet Akismet[37] initially was a plugin for the blog tool and publishing software WordPress[38]. It allows users to classify comments in their blogs as spam and to classify new comments accordingly based on the feedback of all users. Today, Akismet can also be used to classify arbitrary strings. According to Akismet, more than 15 milliard spam comments have already been identified by the system as of April 2010.[39]

Collective Intelligence can be considered as an emergent property that sometimes can simply be observed (e.g., in an ant society), and sometimes requires more sophisticated methods (e.g., Google's PageRank algorithm) to collect and aggregate the results of the single entities behaviours.

---

[33]According to `http://wikipedia.org`, accessed April 29, 2010
[34]See `http://digg.com`
[35]See `http://delicious.com`
[36]Source: `http://blog.delicious.com/blog/2008/11/delicious-is-5.html`
[37]See `http://akismet.com`
[38]See `http://wordpress.org`
[39]Source: `http://akismet.com/stats`

Several definitions and characterisations of Collective Intelligence can be found in literature, and there is no consensus about the exact definition of the term itself. For example, the MIT Center for Collective Intelligence[40] provides this definition:

> *"Collective Intelligence is groups of individuals doing things collectively that seem intelligent."*

Obviously, this definition is very fuzzy and excludes cases in which the entities are not individuals. It is also not necessarily true that Collective Intelligence can only be harvested when the things that groups of individuals do "seem intelligent" – actions might also seem non-intelligent until a proper and maybe complex aggregation method is applied.

Wikipedia defines Collective Intelligence as follows:

> *"Collective intelligence is a shared or group intelligence that emerges from the collaboration and competition of many individuals."* [Wik08a]

This definition is restricting Collective Intelligence to cases where collaboration and competition is required. Yet, such a definition would exclude scenarios where individuals act independently with different motivations than competition. For example, Google PageRank would be excluded in this case.

As we want to include all scenarios in which Collective Intelligence shows up, we will thus follow the definition provided by Tom Atlee, a pioneer in the field of Collective Intelligence [Atl04]:

Collective Intelligence   **Term 3.6** *Collective Intelligence is the intelligence of a collective, which arises from one or more sources.*

Here, "collective" refers to any entity constituted by other entities. Of course this definition also raises the question "what is intelligence?" There are numerous definitions of this term which will not be discussed here. In our context the most important facet of intelligence is the ability to solve problems.

When Does Collective Intelligence Occur?

In the context of this thesis, we are interested in Collective Intelligence as a means of supporting the access to digital resources. Thus, we need to identify

---

[40]See `http://cci.mit.edu`

which prerequisites have to be met when aiming to harness Collective Intelligence.

Following our notion of Collective Intelligence, we first of all need "a collective, which arises from many sources." Such sources in our context are pieces of information, derived from actions and contributions by individuals (e.g., tags annotated by users in a social bookmarking system). Collecting and processing such information was usually a complex and time-consuming task. Nowadays, by using online tools and systems, it is no longer a technical problem to gather and process data from and about millions of users [Seg07].

Yet, just having the data at one's disposal is not enough when aiming to harness Collective Intelligence. Of course not every contribution that is produced from any collective provides information from which intelligence arises. In [Sur04], Surowiecki identifies four conditions that characterise what he calls a "wise crowd", i.e., a number of individuals that contributes information allowing the harnessing of Collective Intelligence:

1. *"**Diversity of opinion** – each person should have some private information, even if it is just an eccentric interpretation of the known facts."* When such a diversity does not exist, contributions of individuals will not be different enough to allow that intelligence arises out of them (see [LDP08] for a detailed discussion on diversity in open social networks).

2. *"**Independence** – people's opinions are not determined by the opinions of those around them."* This condition is important, because otherwise people might be influenced too much by other's actions (*social proof* is a well known psychological phenomenon that occurs when the behaviour of others is considered as a model for the own behaviour). A lack of independence can lead to the unwanted occurrence of what is often referred to as *cascade effects*. Especially in the context of web applications, the use of *aggregate displays* (i.e., displaying information in a way that shows what other users did) influences people's opinions [Por07].

3. *"**Decentralisation** – people are able to specialise and draw on local knowledge."* To ensure that the diversity of people's contributions also reflects all information that is required for a result that is considered as intelligent, decentralisation is an important aspect. The infrastructure of the World Wide Web provides a good basis to ensure this, as it is accessible from basically everywhere.

4. *"**Aggregation** – some mechanism exists for turning private judgements into a collective decision."* Depending on the concrete scenario, the contributions

by the crowd will have to be aggregated in an according way. Well-known examples in the area of Web 2.0 are to rank resources by ratings or number of views, and to aggregate the tags people used in a tag cloud.

Potentials and Challenges

Challenges Fulfilling the above mentioned criteria to allow the harvesting of Collective Intelligence is a non-trivial issue. It requires mechanisms that:

- attract and motivate users, and that allow decentralised contributions in a way that fosters independence and diversity of these contributions,

- aggregate and process the information contributed by the users in way that produces an added value.

Potentials Once this is realised, Collective Intelligence can be a very important source for supporting the access to digital resources. This concerns two aspects:

1. The provision of social resources (e.g., Wikipedia articles) and social metadata (e.g., tags and ratings) that might be retrieved as relevant output of an information retrieval or information seeking process.

2. The creation of social metadata that allows to harvest Collective Intelligence for
    - filtering, searching and ranking, and for
    - recommendations (e.g., using collaborative and social filtering).

3.6.4 Crowdsourcing

As already stated, the World Wide Web offers an infrastructure and technology allowing to reach very large audiences and to aggregate contributions of users on a large scale. In the last years, these capabilities were used by more and more organisations and platforms to outsource tasks formerly conducted by themselves to potentially any user – so called *Crowdsourcing*. The term *Crowdsourcing* was first coined by Jeff Howe in the June 2006 issue of Wired magazine [How06b]. He provides the following definition:

> *"Crowdsourcing is the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."* [How06a]

Howe further details that Crowdsourcing "can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals' and that "the crucial prerequisite is the use of the open call format and the large network of potential laborers." This also makes clear that Crowdsourcing may not be confused with Collective Intelligence – the contributions of people in a Crowdsourcing process might results in the occurrence of Collective Intelligence, but this is not required.

To illustrate the Crowdsourcing concept, we will now provide some examples:

Goldcorp Challenge: In 2000, the Canadian gold mining company Goldcorp initiated the "Goldcorp Challenge". Participants were encouraged to examine parts of Goldcorp's geologic data, and to submit proposals identifying potential targets where the next 6 million ounces of gold will be found. As prize money, Goldcorp offered more than 500 thousand US Dollars to the 25 finalists that identified the most gold deposits. Goldcorp managed to attract "more than 400 online prospectors from 51 countries registered as Challenge participants" [Gol01, p.2]. The submitted solutions identified a total of 110 deposits, confirming many of Goldcorp's suspected deposits, but also identifying several new ones [Bra08].

SETI@home: SETI@home[41] is hosted by the Space Sciences Laboratory at the University of California, Berkeley. It uses Internet-connected computers to search for extraterrestrial intelligence. People can participate by downloading a program that automatically downloads and analyses radio telescope data.

Mechanical Turk: Amazon's Mechanical Turk[42] offers to businesses and developers the opportunity to ask workers to complete so-called "HITS" (Human Intelligence Tasks). Workers on the other hand can choose according HITS they want to work on and get paid for. As of November 2008, Mechanical Turk offered more than 110 thousand of such HITS.

Adobe knowhow: Adobe knowhow[43] is a technology preview that uses Delicious as a platform to collect user generated content that is then provided in the Adobe Creative Suite 3. Users can suggest new content like tool descriptions or tutorials.

---

[41]See `http://setiathome.berkeley.edu`
[42]See `https://www.mturk.com/mturk/welcome`
[43]See `http://labs.adobe.com/technologies/knowhow`

Potentials and Challenges

Potentials   Crowdsourcing can be applied for any kind of task that could otherwise only be realised with a lot of work by specialised experts or employees. In our context, we have to consider the social resources and social metadata created explicitly with the interaction possibilities as introduced in Section 3.6.1. As a result, we can derive the following tasks to apply Crowdsourcing:

Provision of digital resources: The creation and the provision of digital resources that meet some specified needs (e.g., courses in an eLearning context, manuals that provide information about how to use a product, or contents related to a certain topic) is often a time-consuming process that requires a lot of expertise. As mentioned in the Adobe knowhow example, such tasks can sometimes also be carried out by end users.

Creation of metadata: When the amount of resources that have to be annotated with metadata is very large (and eventually grows each day), traditional approaches to describe the content usually cannot be applied. This especially holds for the World Wide Web, and with the rise of Web 2.0, even more resources are being created. As Guenther and McCallum state: "The rapid proliferation of digital resources demands both rapidly produced descriptive data and the encoding of more types of metadata" [GM03, p.12]. In traditional environments such as libraries, the metadata was usually created by experts such as librarians or archivists. Such metadata provides information such as the title and author of a resource, or a classification. But not only the Crowdsourcing of this kind of metadata can be used to allow for an efficient retrieval of resources. As already stated, also social metadata such as ratings or tags can be a very useful source.

Challenges   In order to realise a successful approach based on Crowdsourcing, one faces two main challenges:

- A sufficient amount of users with appropriate resources to contribute to the solution of the given task or problem has to be addressed. Although we can reach potentially any user by using the infrastructure of the World Wide Web, this obviously is not enough – the attention of according target groups has to be attracted, and the users have to be motivated to contribute. This might be realised by offering rewards such as money, but also with game-based approaches as introduced by Luis von Ahn (for more information see [vAD04, vAGK+06]).

- We must of course offer the opportunities that allow to contribute information in the desired way.

3.7 Summary and Conclusion

The opportunities for accessing, creating and sharing information have changed a lot with the rise of the World Wide Web, and especially with social media tools available in what we call the Web 2.0. Users turned more and more from consumers to producers. Without requiring any special expertise or permissions, people can actively contribute and share information as well as communicate with others on a world-wide scale. This paradigm shift from traditional to social media entails several phenomena that provide potentials as well as challenges when aiming to support the access to digital resources.

In this chapter, a holistic framework with a clear terminology and structuring of entities in Web 2.0 applications was provided, in order to then identify these potentials and challenges.

Table 3.5 provides a summary of the identified potentials and challenges, with a special focus on the creation and provision of social metadata and social resources. If direct links exists to the characteristics of an ideal setting as presented in Section 2.3.1, the respective characteristics are mentioned accordingly.

Phenomenon	Potentials	Challenges
Interaction possibilities	• Attracting users and encourage participation • **Existence of resources:** creation of social resources • **Accessibility of resources:** provision of social resources and of social metadata that eases the discovery of digital resources • **Adequate information about resources:** creation of social metadata to support *resource examination* and *resource subset selection* – Generation of views on resources – Maintaining underlying structures – Navigation facilities (especially social browsing)	• Provision of interaction possibilities in an adequate way so that users are encouraged to make use of them • Offering interaction possibilities in existing, traditional environments
The *Long Tail* phenomenon	• **Existence and accessibility of resources:** a growing number of niches with very specialised digital resources	• Techniques to navigate and filter resources in the Long Tail (e.g., social browsing) • Techniques to provide personalised filtering and recommendations (e.g., using collaborative and social filtering)

Phenomenon	Potentials	Challenges

Collective Intelligence	• **Existence of resources:** creation of social resources (e.g., Wikipedia articles)
	
	• **Accessibility of resources:** provision of social resources and provision of social metadata that eases the discovery of digital resources
	
	• **Adequate information about resources:** creation of social metadata (e.g., tags and ratings) that allow for the harvesting of Collective Intelligence as a means to support *resource examination* and *resource subset selection*
	
	  – Filtering, searching and ranking (e.g., based on aggregated user ratings)
	
	  – Recommendations (e.g., using collaborative and social filtering)
		• Providing mechanisms that attract and motivate users, and that allow for decentralised contributions in a way that fosters independence and diversity of these contributions
		
		• Aggregating and processing the information contributed by the users in way that produces an added value

Crowdsourcing	• **Existence of resources:** creation of social resources • **Accessibility of resources:** provision of digital resources (e.g., collections of resources about specific topics) • **Adequate information about resources:** creation of social metadata about resources	• Attracting the attention of a sufficient amount of users with appropriate resources to contribute to the solution of the given task of problem • Motivating users to contribute

Table 3.5: Web 2.0 related potentials and challenges to support access to digital resources

Enabling Socially Enhanced Access to Digital Resources – Requirements and Demands

The topic of this thesis is the use of social media technologies for supporting the access to digital resources. Therefore, discussed the following questions have been discussed in the preceding chapters:

- How can digital resources in information systems be accessed, and what characterises an ideal setting (see Chapter 2)?

- What characterises the paradigm shift from traditional to social media, what is the impact of these changes on how digital resources can be accessed, and what are the according potentials and challenges for supporting the access to digital resources (see Chapter 3)?

Based on these findings, the requirements and demands for the main goal of this work will now be identified: a generic approach that allows

- to support access to digital resources by means of social media technologies, and

- to integrate social media technologies also in existing, especially traditional environments with no or only few interaction possibilities.

Following the general architecture of information systems presented in Section 2.2.2, the following key questions will therefore be discussed in this chapter:

1. Which application scenarios will be considered for such a generic approach, how is the targeted approach defined, and what distinguishes it from related movements and approaches (see Section 4.1)?

2. Which types of digital resources shall the environment support (see Section 4.2)?

3. Which information about digital resources shall be provided (i.e., how should the metadata for digital resources look like) (see Section 4.3)?

4. Which information about users shall be provided (see Section 4.4)?

5. Which functionalities shall be offered (see Section 4.5)?

6. Which interfaces shall be provided – to humans and systems – for the offered functionalities (see Section 4.6)?

Finally, the main research hypotheses of this work will be presented. The validity of these hypotheses will be investigated in Chapter 6.

4.1 Definition of the Approach and Targeted Application Scenarios

When aiming to support access to digital resources, it is a key question to decide for which scenarios this support should be provided. This decision has a huge impact on the key questions discussed in the following sections in this chapter.

The aim is a generic approach   In this work, the aim is a generic approach, and therefore the relevant aspects to consider are examined without focusing on a selected scenario. This is an essential difference to the majority of information systems– whether they are commercial or non-commercial – in the field of social media. To put the work presented here into context, such specialised will now first be discusses, followed by a specification of the targeted approach and facilitating conditions for its introduction. The section will close with a comparison to related generic approaches.

4.1.1 Specialised vs Generic Approaches

Most of these systems are explicitly targeting specific resource types or user groups, e.g.:

- Google's YouTube[1] only supports video resources,

---

[1]see `http://www.youtube.com`

- Yahoo's Flickr[2] focuses and photos (while also offering video uploads),

- Delicious[3] is only managing bookmarks,

- SlideShare[4] is only intended for uploading and managing presentations,

- BibSonomy[5] developed at the University of Kassel is only supporting the management of publication data,

- etc.

Concentrating on a very specific scenario (e.g., "knowledge workers in a research department") in the system's design can certainly provide the benefit of a customised solution that takes into account the very specific characteristics of this scenario and the needs of the involved users. If designed and implemented in a proper way, such a specialised system is also very likely to outperform a generic solution that has to be adapted. Yet, a very targeted approach almost inevitably has several downsides:

- A huge modelling effort is required, e.g., for specifying and generating complex and tailored structures such as ontologies.

- A created model can always only be a snapshot – yet, people and organisations evolve. Thus, maintenance is required, which is usually a very complex and time-consuming task. Furthermore, no model is able to anticipate all possible needs and scenarios (see Section 2.5).

- The restriction to a very specific scenario and model hinders interoperability with other components (e.g, tools, technologies, and other data sources) that might be used in such a scenario. Although the adaptation of such components is sometimes possible, this is once again a usually complex and time-consuming task.

- The resulting information system can only be repurposed in a very limited way, thus restricting means to reuse the system in different application scenarios.

---

[2]see `http://www.flickr.com`
[3]see `http://delicious.com`
[4]see `http://www.slideshare.net`
[5]see `http://www.bibsonomy.org`

Instead of focusing on a specific scenario in this thesis, the aim is to provide a generic approach that can potentially be applied to support access to digital resources wherever this support is needed. So instead of defining prerequisites that have to be met for infrastructures, domains or user types, characteristics of scenarios with a special need to support access are now identified, and where the benefits of social media technologies have the biggest potential. In Chapter 6, respective scenarios will be presented where implementations of the presented approach were introduced.

4.1.2 Generic Approach: A Social Resource and Metadata Hub

Typical characteristics that hinder access to digital resources

It would be a futile task to provide universally valid characteristics of scenarios where access to digital resources needs to be supported. The existence of a "perfect system" is hardly conceivable. Yet, there are some typical characteristics that can make access to digital resources difficult, and that frequently occur in digital environments:

Abundance of available digital resources: This holds true for most information systems, independent from a specific domain or provided access types. The World Wide Web is of course the most prominent example. The more digital resources exist in an environment, the more difficult it usually is:

- to be aware of the existence of relevant information,
- to find all digital resources that are relevant for a certain information need, and
- to filter out the unwanted ones.

Heterogeneous access to digital resources: There are many situations without a single point of access for all information that is potentially relevant. Digital resources can be stored in a variety of different, and often isolated systems and repositories. Hence, users have to know about these different sources and their specifics, have to access them separately and, in general, are not able to easily find and retrieve appropriate digital resources. Furthermore, the single sources often use different metadata for resources, both concerning how digital resources are structured and described. Each community, and sometime each single source uses its own standard to describe digital resources. In addition to that, even when suitable digital resources are found, they often do not provide a link to further information,

so that the search for additional information potentially has to be carried out for each source of information. Thus, accessing information is difficult and time-consuming for users.

Missing metadata for digital resources: Often no metadata exists about a digital resource except for what could be extracted from the digital resource itself (e.g., a full text). In these cases, an information system only has very few possibilities (often only focusing on topicality) at hand to allow to generate meaningful resource subset selections and to decide on the relevance of a digital resource.

These are the key issues that shall be tackled by means of the aimed generic approach.

A Social Resource and Metadata Hub

Ideally, this approach would

- offer the social media technologies for all existing components in an environment,

- incorporate all information that is relevant in an environment, and

- enable a *single point of access* that allows for accessing all digital resources that exist in an environment.

Thus, this approach will from now on be referred to as a **Social Resource and Metadata Hub**. Figure 4.1 depicts a typical application scenario for a respective hub.

A Social Resource and Metadata Hub offers the possibility to be used as

Usable as stand-alone system or social backbone

- a *stand-alone system* with appropriate user interfaces that can be adapted to the needs of the specific scenario, and also as

- a component only running in the background of an environment, thus realising what is referred to as a *social backbone* from now on.

4.1.3 Facilitating Conditions for Introducing a Social Resource and Metadata Hub

Introducing social media does not necessarily provide benefits for all scenarios. For example, sometimes it is desired that only information provided by selected experts is used (e.g., for liability reasons), and sometimes not enough users can be involved that could potentially contribute. Yet, based on the challenges and
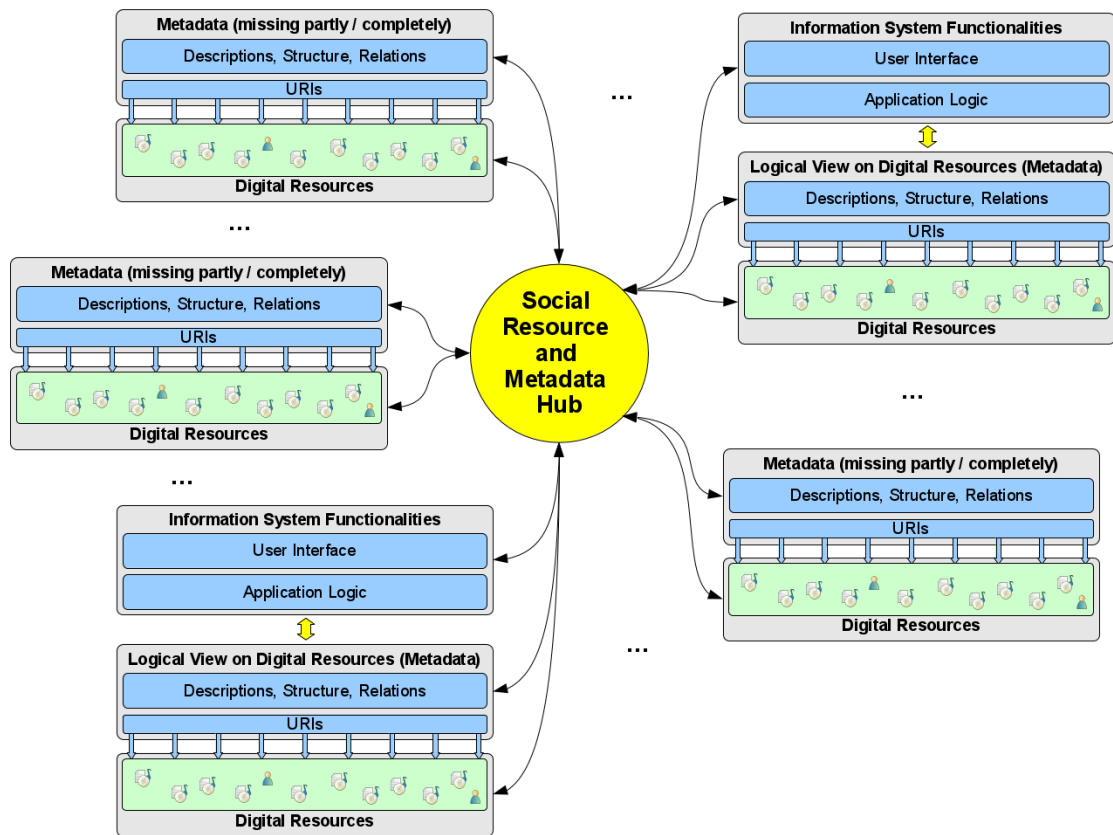
Figure 4.1: A typical application scenario for a Social Resource and Metadata Hub

Criteria fostering success
opportunities identified in Chapter 3 and the characteristics of metadata identified in Section 2.6, the following criteria that can foster the success of a social media based approach can be deduced:

- There are enough users that can (potentially) contribute. The size of such a *critical mass* very much depends on the concrete application.

- The environment does not yet provide social media functionalities. This is the state in most organisations with intranets, traditional repositories or digital libraries (e.g., ARIADNE[6] or OAIster[7]). They often follow a top-down approach with controlled publishing processes, and they only offer very few interaction or contribution possibilities to end users.

---

[6]See `http://www.ariadne-eu.org`
[7]See `http://www.oclc.org/oaister`

- Diversity is of importance. This is typically the case when

  - scenarios are dynamically changing and new patterns of use emerge regularly,

  - available information is used in different use cases, and

  - different and individual views on digital resources are of importance.

4.1.4 Relation to the Semantic Web

While the Semantic Web also follows the vision of creating an ecosystem where all resources can be accessed in a uniform way through URIs and RDF metadata, there are some fundamental differences to the presented approach:

- The Semantic Web in general mainly promotes common data formats and does not focus on the application layer.

- In the Semantic Web approaches existing so far, the idea is still dominant that everything is uniquely describable. Dynamics, different views, and different contexts of meaning are usually disregarded.

- The tool support for Semantic Web technologies is still not in a state where a significant ecosystem of tools and applications has been built up. Even basic tools such as Virtuoso still have to fight with severe problems when more complex use cases are tackled.

Still, a Social Resource and Metadata Hub should provide means to be used within Semantic Web architectures (e.g., by providing respective interfaces), and it should also be open to make use of existing Semantic Web data whenever possible.

4.2 Resource Types

To identify which types of digital resources a Social Resource and Metadata Hub should support, the implications of each aspect that has to be considered for the ideal setting presented in Section 2.3.1 will now be discussed:

Existence of resources   Creation means can be helpful. Yet, at least usable in an environment where creation tools are available (Desktop, Web).

Accessibility of resources   To make resources accessible, it should be possible to incorporate any type of resource as defined in Section 2.1. This includes arbitrary types of multimedia resources (e.g., HTML, PDF, MPEG), but also services or even physical resources just represented by a URI in an information system. "Incorporate" here means:

- When a digital resource is newly created or not yet accessible in the respective environment, it should be possible to contribute this digital resource, and to make it accessible. A system that offers this realises a *repository*.

- For digital resources that are already accessible in the respective environment, it should be possible to integrate them into the system without having to physically copy them. Otherwise, the following problems are very likely to arise:

    Maintenance issues: When digital resources are copied from a source where new contents are added, or existing contents are deleted and modified, the system will have to react to this changes. This is usually an expensive and time-consuming task.

    Memory requirements: Every digital resource that is copied will require some memory capacity. For large collections or certain resource types such as videos, this can result in very high memory requirements.

    Legal concerns: Sometimes it is simply forbidden to physically copy existing digital resources and to provide them in a different system.

    A system that offers this realises a *referatory*.

- If the content of a resource is accessible in a machine-readable way, it can be separately stored within the hub to allow for efficient data processing, e.g., to support better search and retrieval processes.

Adequate information about resources   Here, means for *resource examination* need to be offered. This on the one hand concerns the resource metadata that is discussed in the next section, but it also involves means that allow humans to examine a digital resource. In some cases (e.g., for text documents that can only be opened with specific tools or for audio and video files), a system should provide support for easily accessing the contents of a digital resource. E.g., this can be realised by providing preview images for the resources, or by integrating players for multimedia contents such as audio or video files.

Summary of Requirements and Demands

- Incorporation of any type of resource as defined in Section 2.1, with means for resource examination

- Realisation of a *repository*

- Realisation of a *referatory*

4.3 Resource Metadata

Several quality criteria and the question of what constitutes "good" metadata were already discussed in Section 2.5. To identify the requirements concerning the metadata to be used for digital resources in a Social Resource and Metadata Hub, the focus will now be put on the aspects most relevant for the presented approach approach:

- subjectivity and diversity,

- interpretability,

- core metadata and extensibility, and

- interoperability.

4.3.1 Subjectivity and Diversity

We have already shown that metadata about a resource can not capture the "true meaning of a document", and that there are objective and subjective metadata. We always have to consider that metadata is created for certain purposes in certain contexts, and that it is impossible to anticipate for whom and for what reasons a resource might be considered as relevant in the future. We have to accept and to embrace the fact that there is no "single and correct" way to describe a resource. As a consequence we should allow *subjectivity*, and also *diversity* in the metadata about resources, instead of a *metadata monoculture*[8]. The need to support diversity is also motivated by the fact that we aim to support the access to digital resources in a variety of application scenarios, especially with heterogeneous components and most likely also heterogeneous metadata formats used

---

[8]The term *metadata monoculture* was coined by Randy Goebel in 2008.

for resources. Furthermore, we identified independence and diversity of contributions as two of the challenges when aiming to support the access to digital resources by means of social media technologies in Chapter 3.

These requirements are also supported by Nilsson et al. in [NPN02], where the authors identified the needs for a Semantic Web architecture, concluding that it should be:

- Evolving, supporting a dynamic metadata eco-system

- Extensible, allowing introduction of new vocabulary with new semantics

- Distributed, supporting descriptions by anyone about anything, anywhere

- Flexible, supporting unforeseen uses of resources

- Conceptual, supporting the evolution of human knowledge

It is clear that a one-size-fits-all solution for metadata about resources will not fit these needs. Instead, an ideal infrastructure would be generic in a way that allows for the generation of adequate resource descriptions for different users in different scenarios. Therefore, potentially any existing metadata might be incorporated. As we have shown in Section 2.6, the different approaches to generate metadata can only be applied successfully in certain scenarios and for certain types of digital resources and metadata, and each of them has its benefits and limitations. Ideally, a digital environment should allow in each scenario to combine the benefits of each of the metadata generation approaches and to avoid the limitations. To allow for subjectivity and diversity, human generated metadata is most important, as only humans can contribute with different views and opinions. The need for diversity demands a non-authoritarian approach, supporting different views of the same resource. Thus, social metadata as defined in Section 3.5 is most likely to meet these requirements, because it allows any user to contribute metadata about a resource.

What are the consequences for a Social Resource and Metadata Hub? Of course social metadata has to be supported. But beyond that, we should be able to make use of potentially any metadata existing in the environment where the hub is introduced. Moreover, it should be possible to contribute a variety of different metadata for digital resources. Such metadata can immediately be helpful for end users (e.g., bibliographic information about a resource), and it can also be an important source for several functionalities (e.g, search or recommendations).

4.3.2 Interpretability

Besides subjectivity and diversity, we also have to consider that relevance can only be estimated to a small extent by a machine. A lot depends on the interpretation of the end user. Thus, provenance metadata should exist that can be interpreted by humans to judge the relevance of a resource. When we want to be able to interpret the metadata about a resource, this means it must be human-understandable, and it is also very important to know who created and contributed the resource as well as the metadata (credibility was already introduced in Section 2.3.2 as an important concept when judging the relevance of a resource). So any information that allows the user to get an understanding of the respective creator or contributor is important.

4.3.3 Core Metadata and Extensibility

We identified that we should offer the possibility to annotate various metadata from different sources for each digital resource. Nevertheless, a minimal set of core metadata elements (not necessarily mandatory) has to be defined. This is required for realizing a standalone component with means for resource contribution, resource display, and functionalities to (efficiently) compute and process information in a meaningful way, and furthermore to support interoperability with other components.

Such a core set can, e.g., be based on Dublin Core (see Section 2.4.3), and it should cover all metadata types as introduced in Section 2.4.1. Such basic metadata is required, among others,

- to enable basic functionalities such as search, retrieval and display (including, e.g., the identifier, title, and location of a resource),

- to allow the generation of different views on resources (e.g., *ordered by rating* or *ordered by number of views*),

- to allow users to get an understanding about who provided a resource and metadata,

- to information users about the technical format of a digital resources and the technical requirements to use it, and

- to provide information about intellectual property rights with information about the way in which a digital resources may be used.

In order to improve the probability of cross-disciplinary interoperability and to allow for an easy generation of metadata, such a basic schema should be kept as simple as possible, although this may result in more effort in the part of searchers to identify relevant resources [DHSW02]. While complex schemas might result in better services, they require a higher investment in the creation of metadata, and make it more difficult to promote consistency. The metadata schema as provided for resources in the DaMiT system provides a good example for a partly too complex and too fine-grained specification (cf. [GLM03, JGLM04, JDG⁺04]). For example, DaMiT aimed to offer four different presentation styles for learning objects, ranging from "formal" to "informal". During the project, it turned out that it was an impossible task for the authors of the different courses and objects to realise a consistent use of these classifications. A quality control process that was introduced turned out to be an over complex and very time-consuming task. As a consequence, the DaMiT consortium agreed to reduce the available presentation styles to "formal" and "informal" only.

Extensibility　In order to enable the support of specific scenarios and resource types in a Social Resource and Metadata Hub, it should be possible to extend the defined core metadata set accordingly. Typical examples for such extensions can concern all metadata types as identified in Section 2.4.1:

- creation and rights metadata (e.g., involving scenario-specific user groups and roles),

- technical information (e.g., for multimedia contents of a specific type),

- content information (e.g., for domain-specific attributes),

- relation metadata (e.g., for classifying resources with respect to a given vocabulary or taxonomy), and

- usage metadata (e.g., taking into account actions involving scenario-specific extensions).

4.3.4 Interoperability

As already stated, a Social Resource and Metadata Hub should allow to integrate social media technologies in existing environments. Thus, interoperability

is a key issue, because it must be able to provide the metadata about digital resources in the hub to other components. The IEEE[9] defines interoperability as follows [Ger91]:

Term 4.1 Interoperability *is the ability of two or more systems or components to exchange information and to use the information that has been exchanged.*

Consequently, drawing upon standards concerning metadata elements and representation formats is required. Furthermore, modularity "allows designers of metadata schemas to create new assemblies based on established metadata schemas and benefit from observed best practice, rather than reinventing elements anew" [DHSW02, p.2].

Duval et al. provided the following fundamental principles for interoperability [DHSW02] that where enhanced by Nilsson et al. [NND+08] who added the principle "Machine-processability".

Extensibility: The ability to create structural additions to a metadata standard for specific needs of a domain, community or application

Modularity: The ability to combine different, heterogeneous metadata fragments

Refinements: The ability to create more fine-grained descriptions compatible with more coarse-grained metadata, and to translate a fine-grained into a more coarse-grained description

Multilingualism: The ability to express, process, and display metadata in a number of different linguistic and cultural circumstances

Machine-processability: The ability to automate processing of different aspects of the metadata specifications (e.g., to handle extensions, or understand refinements)

The more the digital environment in which a Social Resource and Metadata Hub is integrated follows these principles, the higher the chances are that the respective metadata from the hub can be integrated accordingly.

---

[9]See `http://www.ieee.org`

Application Profiles

When aiming to use metadata in a concrete scenario, developers often have to modify or even combine existing schemas to adapt them to their specific needs. Heery and Patel introduced the term *application profile* in 2002 [HP00], aiming to provide some sound basis to combine different metadata schemas and to adapt them to the needs of a specific scenario. They define an application profile as follows:

Application profile

Term 4.2 Application profiles *are schemas which consist of data elements drawn from one ore more namespaces, combined together by implementers, and optimised for a particular local application.*

They further demand that schema application profiles have to conform to the following characteristics:

- Elements from one or more different namespaces may be used.

- No new data elements not defined in existing namespaces may be introduced.

- Permitted schemas and values may be specified.

- Standard definitions can be refined.

Some of the oldest examples for application profiles are Z39.50 application profiles[10]. For DC and LOM, the DC-Education Proposal[11], UK LOM Core[12] and CanCore[13] might serve as examples.

In the context of this work, it is important to allow to use metadata from a hub in application profiles, too.

Summary of Requirements and Demands

- Support for social metadata

- Ability to make use of potentially any metadata existing in the environment where the hub is introduced, and consequently offering the possibility to contribute a variety of different metadata formats

---

[10]See http://www.loc.gov/z3950/agency/profiles/profiles.html
[11]See http://projects.ischool.washington.edu/sasutton/dc-ed/Dc-ac/DC-Education.html
[12]See http://zope.cetis.ac.uk/profiles/uklomcore
[13]See http://www.cancore.ca

- Provision of information that allows users to get an understanding of the respective creator or contributor

- Definition of a core set of metadata elements to enable interoperability and to efficiently compute and process information

- Support for scenario specific metadata extensions

- Following principles supporting interoperability

- Drawing upon standards concerning metadata elements and representation formats to ease information exchange

4.3.5 Parenthesis: Resource Profiles

Related to the requirements as introduced above, Stephan Downes introduced the concept of *resource profiles* in 2004 [Dow04]. A resource profile is described as "a multi-faceted, wide ranging description of a resource" which is characterised by the following features:

- A resource profile does not conform to a particular XML schema. On the contrary it is a patchwork of metadata formats which are assembled as needed in order to form a description that is most appropriate for the given resource.

- A resource profile is not authored by a particular author. It consists of a large set of information that is authored by many people. As these people do not have a particular stake and they further correlate different claims with each other and with the original resume the trustworthiness of a resource profile is much higher than that of a single metadata set.

- Traditional resource descriptions are intended to be instantiated as a single digital file and located in a particular place whereas resource profiles may be distributed, in pieces, across a multitude of locations.

- There is no single canonical or authoritative resource profile associated with a given resource.

One of the roles that a Social Resource and Metadata Hub can play is to be one source of information for the creation of such resource profiles.

4.4 User Metadata

The concept of a user model was already introduced in Section 2.3.2, where it was defined it as an explicit representation of some characteristics of a certain user. It can comprise a variety of information that can help to estimate a user's information need. Furthermore, it was just shown that it is important to provide information that allows users to get an understanding of creators or contributors of resources and metadata. Thus, the metadata about users should – beyond typical information required to login to a system – comprise at least elements that allow to store

- basic profile information such as a user's name and address,

- information about a user's interests,

- preference information of a user (including system specific settings),

- information about a user's activities (captured by observing the user),

- information about a user's contributions, and

- information that allow to connect to other components in the scenario where the Social Resource and Metadata Hub is deployed.

For all user metadata, privacy is a key for the acceptance of users (cf. [Int97], [PD03], or [LHDL04]). This has to be reflected in the user metadata, too (e.g., means that allow to specify who can access certain data).

4.5 Functionalities

In Chapter 3, several challenges were determined, including the provision of typical social media interaction possibilities for users. Based on these challenges and the findings presented so far, the basic functionalities that at least have to be offered will now be identified. This involves the deployment of a Social Resource and Metadata Hub as a stand-alone system as well as its use as a social backbone.

User management, networking facilities, and resource organisation:

> **Registration, login, and logout:** The system needs to provide means to register, to login, and also to logout.

Profile management: A facility is needed that allows users to provide information about themselves. Furthermore, the possibility to retrieve information about other users' profiles has to be offered, of course respecting privacy concerns.

Contacts, messaging: Typical social software functionalities such as the maintenance of contact lists and the possibility to communicate with other users should be offered.

Privacy and control of data: Users have to be able to control by whom any of their contributions and information about them can be accessed. To ensure transparency, information has to be provided about the data the system stores about the user, and it must also be possible to delete this data.

Furthermore, it should be self-evident that users are able to organise and manage their contributions. This includes, e.g., the possibility to get an overview of contributed digital resources or metadata, but also further facilities to structure contributions.

Contributing digital resources and metadata: A Social Resource and Metadata Hub certainly has to offer means to contribute digital resources and metadata. As a repository as well as a referatory has to be realised, means have to be offered to

1. upload digital resources, and
2. to refer to resources (so called *bookmarking*).

Following the requirements and demands identified in 4.3, first of all ways to contribute core metadata (including the social metadata) have to be provided, but also a variety of other metadata formats.

Editing digital resources and metadata: This especially involves means to edit provided metadata (e.g., in order to correct mistakes), and to delete contributions. If possible, tools that allow for editing the digital resources might also be provided.[14]

Retrieval of digital resources and metadata: Users as well as machines have to be able to retrieve a digital resource, and accordingly metadata that is available for a resource. When different instances of metadata exist, it

---

[14]This is usually only possible with specific tools that support the respective resource format. Yet, for some formats such as ODF (OpenDocument Format), tools exist for embedding editing functionalities.

should be possible to retrieve each of them separately. Following the principles of modularity, it should also be possible to retrieve single metadata elements.

Means to provide resource subsets: The *resource subset selection* and *resource subset delivery* are two of the access process steps as introduced in Section 2.2.3. Such a resource subset can be created using different pull- and push-technologies. At least, the hub should offer standard functionalities from information systems, such as

- search facilities (e.g., a simple keyword search),

- means to filter contents according to different criteria, and

- different sorting and ranking facilities.

Navigation facilities: Means have to be offered that allow users to navigate within the information created in the Social Resource and Metadata Hub, using potentially any available relation. This especially includes social browsing as introduced in Chapter 3.

Information exchange and notifications: Ideally, the hub would be able to exchange information with potentially any other component in a digital environment. For example, export facilities are needed that offer information from the hub in different formats. Furthermore, the hub should offer means of notification, e.g., about activities that might be of interest for a user.

4.6 Interfaces

After the functionalities a Social Resource and Metadata Hub should provide have been identified, the way these functionalities should be offered to users and machines respectively will now be discussed. In order to allow the usage of the hub in as many scenarios as possible, and to foster the adoption of as many users as possible, the following aims should be followed:

Low technical barriers for system usage: Users should be able to use functionalities with minimal efforts. This means as few restrictions as possible concerning the technical environments in which the hub can be used, as well as minimal installation efforts.

Low conceptional barriers for system usage: Conceptional prerequisites for system usage such as the use of certain metadata formats should be kept to a minimum, while still allowing to provide added value for as many scenarios as possible.

4.6.1 Access by Systems

The targeted Social Resource and Metadata Hub should allow to be integrated in existing environments with different systems and components. Thus, more than "just" an adequate user interface is needed. Interfaces allowing for an easy creation of mash-ups and complex functionalities using information from the hub are required. Thus, interoperability is a very important aspect, and the hub should offer access to potentially any of its data and functionalities, regarding privacy aspects at the same time.

Pulling Technologies

Here, a service-oriented architecture (SOA) is most likely to be applied. It provides a set of services that are loosely integrated and can be used in a variety of contexts. A SOA also provides means to retrieve information about which services are available. An endpoint is the entry point for a SOA implementation, and interfaces are described in terms of protocols and functionality.[15]

A well established way to offer a SOA is to make use of Web Services. As an alternative to robust Web Services which make use of "heavyweight" techniques like SOAP[16], often more lightweight or simplified programming models are demanded [And07]. REST[17] is such a lightweight technology that meets these requirements.

Pushing Technologies

In order to enable notifications, feeds can be provided using *RSS* or *Atom*. RSS is a family of formats that allows users to find out about updates to the content of, e.g., RSS-enabled Web sites without having to visit the site [And07]. The information is collected within a feed and "piped" to the user. RSS is an XML-based data format for Web sites for the exchange of files that contain publishing information and summaries of the content of the site. As there are a number of different RSS formats with issues of incompatibility between them, a

---

[15]For a more detailed introduction into SOA, please refer to [Bel08].
[16]See `http://www.w3.org/TR/soap`
[17]See `http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm`

new syndication system was proposed and developed under the name Atom in 2003. Atom consists of two standards: First there is the Atom Syndication Format which is an XML language used for Web feeds. Second there is the Atom Publishing Protocol (APP) which is a HTTP-based protocol for creating and updating Web resources. Atom has particularly been developed in order to meet the requirements of Web sites containing news and weblogs.

Data Formats

For any data exchange with the hub, the data has to be represented in a machine-readable format that allows to transport also complex structures, and that can be processed by any connected component. This especially regards the exchange of resources and metadata. Popular formats that meet these demands are *XML* or *JSON* (JavaScript Object Notation) that are widely used and offer tool support for most common programming languages.

4.6.2 User Interfaces

Different kinds of users can be involved in accessing information from a Social Resource and Metadata Hub. The focus will mainly be put on *end users*, but there are also two other types of users:

Application designers have to be able to determine which functionalities are available. Furthermore, they need to know how these functionalities can be accessed, i.e., they have to know how the interface has to be accessed, which protocol has to be used, which parameters can or have to be provided, which data types and values are returned, etc.

Creators and managers of metadata who want to use metadata available in the hub (e.g., to assemble resource profiles) need to know which metadata is available for each resource and how it was created. Information about schemas and vocabularies also should be provided.

Offering Access to End Users

The task of designing proper user interfaces is very complex and depends on a variety of factors such as the main focus of an application scenario, the users that are involved, the characteristics of involved resources, and the context of use. A Social Resource and Metadata Hub is a generic approach intended to be deployable in a variety of scenarios, so only very common and scenario-independent aspects will be discussed here.

In Chapter 3, the need to provide interaction possibilities in an adequate way so that users are encouraged to make use of them was determined. Besides the necessity of dissemination efforts, it is thus important to provide an intuitive user interface that can be easily be used without expert knowledge, following principles such as simplicity [Nie00] and joy-of-use [Ree04]. Furthermore, mechanisms that attract and motivate users (e.g., by using reward mechanisms or game-based approaches) can be offered.

In order to allow decentralised contributions of digital resources and metadata in a way that fosters independence and diversity of these contributions (a challenge also identified in Chapter 3), users should be offered the possibility to use functionalities of the hub in their usual contexts and applications, so that they can contribute with different views. This can of course be realised if the persons in charge integrate functionalities offered by the hub into the respective applications. A more lightweight approach that allows to integrate information or functionalities including user interfaces is to use *widgets*[18]. A widget is an element of a graphical user interface providing information and/or interaction possibilities [SA88], and that can be embedded into existing environments (e.g., a lot of widgets exist that can be embedded in HTML pages).

Furthermore, as the aim is a generic approach that can be used in a variety of scenarios, the user interfaces should be adaptable in a way that allows to address specific needs of a scenario (e.g., concerning a corporate identity or a certain terminology).

4.7 Research Hypotheses

Referring to the summary of identified phenomenons of social media in Chapter 3 and considering the typical characteristics when facing problems in accessing digital resources presented in Section 4.1.2, the following hypotheses will now be introduced:

H1: A Social Resource and Metadata Hub can successfully tackle typical problems when accessing digital resources

- Hypothesis H1.1: A Social Resource and Metadata Hub can successfully tackle the problem of *abundance of available digital resources*

- Hypothesis H1.2: A Social Resource and Metadata Hub can successfully tackle the problem of *heterogeneous access to digital resources*

---

[18]The term *widget* is an abbreviation of *window gadget*

- Hypothesis H1.3: A Social Resource and Metadata Hub can successfully tackle the problem of *missing metadata for digital resources*

H2: A Social Resource and Metadata Hub allows for exploiting social web phenomenons

- Hypothesis H2.1: A Social Resource and Metadata Hub allows for exploiting *interaction possibilities*

- Hypothesis H2.2: A Social Resource and Metadata Hub allows for exploiting the *Long Tail* phenomenon

- Hypothesis H2.3: A Social Resource and Metadata Hub allows for exploiting *Collective Intelligence*

- Hypothesis H2.4: A Social Resource and Metadata Hub allows for exploiting *Crowdsourcing*

4.8 Summary and Conclusions

In this chapter, the requirements and demands for an approach have been identified that allows to support access to digital resources by means of social media technologies, and to integrate social media technologies in existing environments.

Instead of focusing on a specific scenario, it is a generic approach that can potentially be applied in any digital environment. Ideally, it would

- offer the social media technologies for all existing components in the environment,

- incorporate all information that is relevant in the environment, and

- enable a *single point of access* that allows for accessing all digital resources that exist in the environment.

This approach is referred to as a *Social Resource and Metadata Hub* that can be used as a *stand-alone system* or a *social backbone*.

Then, the requirements and demands for such a hub haven been identified, regarding types of digital resources to support, metadata about resources and users, as well as functionalities and interfaces to be offered.

Finally, the following hypotheses were introduced:

H1: A Social Resource and Metadata Hub can successfully tackle typical problems when accessing digital resources

H2: A Social Resource and Metadata Hub allows for exploiting social web phenomenons

System Design Checklist

The requirements and demands for a Social Resource and Metadata Hub that have been identified result in a system design checklist that will now be presented. It is an important contribution of this work to the field of information systems design and can serve as a basis for the realisation of any related system in the field that focuses on the application scenarios presented in this chapter.

Resource Types

- Incorporation of any type of resource as defined in Section 2.1, with means for resource examination

- Realisation of a *repository*

- Realisation of a *referatory*

Resource Metadata

- Support for social metadata

- Ability to make use of potentially any metadata existing in the environment where the hub is introduced, and consequently offering the possibility to contribute a variety of different metadata formats

- Provision of information that allows users to get an understanding of the respective creator or contributor

- Definition of a core set of metadata elements to enable interoperability and to efficiently compute and process information

- Support for scenario specific metadata extensions

- Following principles supporting interoperability

- Drawing upon standards concerning metadata elements and representation formats to ease information exchange

User Metadata

- Basic profile information such as a user's name and address

- Information about a user's interests

- Preference information of a user (including system specific settings)

- Information about a user's activities (captured by observing the user)

- Information about a user's contributions

- Information that allow to connect to other components in the scenario where the Social Resource and Metadata Hub is deployed

- Metadata allowing to specify who can access data about the user

Functionalities

1. *User management, networking facilities, and resource organisation:*

   Registration, login, and logout: The system needs to provide means to register, to login, and also to logout.

   Profile management: A facility is needed that allows users to provide information about themselves. Furthermore, the possibility to retrieve information about other users' profiles has to be offered, of course respecting privacy concerns.

   Contacts, messaging: Typical social software functionalities such as the maintenance of contact lists and the possibility to communicate with other users should be offered.

   Privacy and control of data: Users have to be able to control by whom any of their contributions and information about them can be accessed. To ensure transparency, information has to be provided about the data the system stores about the user, and it must also be possible to delete this data.

   Resource organisation: Features are required that allow users to organise and manage their contributions.

2. *Contributing digital resources and metadata:*

   - Upload of digital resources
   - Referring to existing resources (*bookmarking*)

- Means to contribute different kinds of resource metadata, with a focus on the core metadata (including social metadata)

3. *Editing digital resources and metadata:*
   - Editing metadata
   - Deleting digital resources and metadata

4. *Retrieval of digital resources and metadata:*
   - Retrieval of digital resources
   - Retrieval of complete metadata instances as well as single elements for each resource

5. *Means to provide resource subsets:*
   - Search facilities
   - Means to filter contents according to different criteria
   - Different sorting and ranking facilities
   - Means to provide subsets based on the resource organisation facilities offered

6. *Navigation facilities:*
   - Means that allow users to navigate within the available information using potentially any available relation, especially focusing on social browsing

7. *Information exchange and notifications:*
   - Exchange facilities (import and export) for potentially any kind of available information
   - Notification means for events in the system

Interfaces

- Low technical barriers for system usage

- Low conceptual barriers for system usage

1. Access by Systems
   - Focusing on interoperability, especially regarding protocols and data formats

- Offering access to potentially any data and functionality (of course regarding privacy aspects)
- Notification means (e.g., feeds)

2. Access by Users

- Provision of information about available functionalities, and how they can be accessed (with a focus on application designers)
- For each resource, information about which metadata is available and how it was created
- Information about schemas and vocabularies used
- Intuitive user interface that encourage participation and can be easily be used without expert knowledge, following principles such as simplicity and joy-of-use
- Offering the possibility to use functionalities of the hub in other contexts and applications
- Adaptability of user interfaces to allow to address specific needs of a scenario

III

Realisation of a Social Resource and Metadata Hub

CHAPTER 5
The ALOE System

In this chapter, the ALOE system that realises a Social Resource and Metadata Hub is presented. First, a short overview of the system, selected use cases, and the main layers and architecture is given, followed by a description of the main concepts used. After an introduction of the main features as provided through the ALOE Web interface, further details about the system design are presented, also describing the way ALOE can be adapted and integrated as a hub into existing environments. After a third hypothesis is introduced, it is finally proven step-by-step that ALOE meets the (technical) requirements presented in the system design checklist in Section 4.8.

5.1 What is ALOE?

ALOE is a social resource sharing infrastructure that offers numerous user management facilities and allows users to contribute, manage, organise, and share arbitrary digital resources. The system is realised as a Web application that is accessible via a Web interface, and a Web Service API offers the functionalities for other systems or components.

ALOE is a social resource sharing infrastructure

Users can upload resources, or they can refer to existing resources via bookmarks. Resources can be found by using various filter criteria and search modes, users can tag, rate and comment resources, initiate and join groups, organise contact lists, send messages to each other, and use various other social media functionalities.

ALOE is a generic and adaptable infrastructure suitable to realise systems such as social (Intranet) portals, sharing and communication platforms, and administration and organisation of arbitrary digital contents. It can be used as a stand-alone system, but also as a backbone to integrate social media paradigms in existing environments.

ALOE is generic and adaptable

5.1.1 Initial Idea and Evolution

The development of ALOE started in 2006 as a project conducted with interns at DFKI. The initial aim was to allow enhanced descriptions of learning resources[1] created in the DaMiT project (see [GLM03, JGLM04]), and to integrate them into Learning Management Systems such as Moodle[2] and Sakai[3]. In this first phase, a system was created that allowed to annotate different kinds of metadata not only for DaMiT learning resources, but also for arbitrary other types of digital resources. Content could also be searched and navigated.

In the next phase, ALOE was enhanced with various annotation functionalities. Amongst others, tagging, commenting, and rating was now possible. Furthermore, a first version of the ALOE Web interface was designed. In this phase, first parts of the API were also developed.

From April 2007 until April 2008, ALOE was enhanced mainly within the project CoMet[4] (**Co**llaborative Sharing of Resources and **Met**adata) funded by the Stiftung Rheinland-Pfalz für Innovation[5]. The aim of CoMet was to develop an open platform for sharing metadata about arbitrary types of digital resources. ALOE then no longer focused on learning contents but potentially any kind of digital resource.

With this shift in focus towards a generic system, the basis for further developments and usage scenarios was established. In the following years, ALOE was deployed, used, improved, and enhanced in a variety of projects and scenarios (see below). The underlying concept as well as the resulting platform architecture have proven to be a successful basis for a wide range of scenarios, from prototype and pilot development to productive software deployed in real life scenarios. Several of these scenarios will be presented in more detail in Chapter 6.

5.1.2 Selected Projects, Instances, and Usage Scenarios

ALOE-public (`http://aloe-project.de/AloeView`) is an ALOE instance that is publicly available since 2008. It is used in real application scenarios (e.g., by the CRP Henri Tudor[6]), but also as a simple playground. It is a

---

[1]This also explains the name *ALOE*, which was at this time chosen as an abbreviation of *Adaptable Learning Object Environment*. As of now, ALOE is only a name and does no longer refer to this.

[2]See `http://moodle.org`

[3]See `http://sakaiproject.org`

[4]See `http://www.dfki.uni-kl.de/comet`

[5]See `http://www.mwwfk.rlp.de/stiftung_innovation`

[6]The CRP Henri Tudor (`www.tudor.lu`) is a public research center in Luxembourg

plain ALOE instance without any adaptations.

C-LINK ("Conference Link") was an internal DFKI project funded from 03/2008 until 12/2008. The aim of C-LINK was the development of a Web based tool to support conference participants in their conference preparation, and to offer networking facilities for attendees. With C-LINK, users can share papers and presentations, generate individual conference schedules, get personalised recommendations to find interesting events and attendees, etc. C-LINK is based on ALOE and serves as a showcase for DFKI technologies (e.g., the information retrieval system DynaQ[7]). C-LINK was already used for two conferences: The 31st Annual German Conference on Artificial Intelligence in 2008 (KI 2008[8]), and the 10th International Conference on Document Analysis and Recognition (ICDAR 2009[9]).

OPENEER (`http://openeer.dfki.de`) is a research project initiated in 2009 by Stephan Baumann[10] within the DFKI Competence Center for Computational Culture (C4)[11]. It aims at gathering personal experiences with music on a worldwide scale. This information shall then be used as a basis to provide "emotional recommenders" for music. The OPENEER platform offers the possibility to volunteers to share their personal experiences in order to enable research based on this groundwork.

MACE (`http://www.mace-project.eu`) is a European initiative aiming at improving architectural education by integrating and connecting vast amounts of content from diverse repositories. MACE (Metadata for Architecture Contents in Europe) was funded from 2006 until 2009 and uses ALOE as a backbone for community related features.

ALOE@KM (`http://projects.dfki.uni-kl.de/aloe`) is an ALOE instance specifically provided for the Knowledge Management group of DFKI in Kaiserslautern since 2009.

ALOE-IAO (`http://aloe-iao.dfki.uni-kl.de`) is an ALOE instance that was set up for the Fraunhofer IAO (Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO) in 2009.

---

[7]See `http://dynaq.opendfki.de`
[8]See `http://ki2008.dfki.uni-kl.de`
[9]See `http://www.cvc.uab.es/icdar2009`
[10]See `http://www.dfki.de/~baumann`
[11]See `http://www.computationalculture.de`

DFKI Mindpool Treasures (`http://mindpool.dfki.de/AloeView`) is an adapted ALOE instance that is offered to all DFKI employees (in Berlin, Bremen, Kaiserslautern, and Saarbrücken). It is part of DFKI's internal social media suite *Mindpool* that was initiated in 2010, and that also serves as a platform for managing DFKI specific resources such as press articles and visits.

Web of Models (`http://webofmodels.org`) was a project within the Cluster of Excellency "Center for Mathematical and Computational Modelling"[12] (CMCM) funded by the Government of Rhineland-Palatinate. In this project that started in 2009, ALOE was used as a basis for the Web of Models platform. To support mathematical models, the ALOE instance used in this scenario offers specific metadata for mathematical models, as well as specific detail pages for model visualization. Furthermore, objects can formally be classified based on mathematical taxonomies.

RADAR (`http://radar-project.de`) was a project sponsored by the Stiftung Rheinland-Pfalz für Innovation from 03/2010 until 02/2011. The aim or RADAR (Resource Annotation and Delivery for Mobile Augmented Reality Services) was the development of an ALOE-based infrastructure to contribute, organise, and annotate multimedia resources that can be used within mobile augmented reality services. Therefore, ALOE was among others enhanced to also process spatial information such as coordinates, and to provide means to contribute and search geocontents.

NEXUS (`http://nexus.dfki.de`) was a project funded by the German Federal Ministry for Education and Research (BMBF) from 2012 until 2014. In NEXUS, ALOE was mainly enhanced with improved means to allow formal classifications of entities, and with OWL-based reasoning functionalities.

ALOE-UNIFARM (`http://unifarm.dfki.de`) is part of the EU-funded project UNIFARM (`http://www.project-unifarm.eu/`) initiated in 2012 and serves as platform that allows users to present and defend the needs of farmers in the development of Global Navigation Satellite System (GNSS) applications and services.

MOBIL-IN-KL (`http://mobil-in-kl.quertex.de`) is a RADAR-based platform that allows citizens to contribute or support suggestions for mobility services (e.g., car sharing stations or charging stations for e-cars) in

---

[12]See `http://cmcm.uni-kl.de`

Kaiserslautern. The platform was initiated in 2014 and is part of the project "Lautern macht mobil" realised with several partners in Kaiserslautern.

5.1.3 Main Layers and Architecture

ALOE is designed as a server-based application where information is exchanged via HTTP[13]. On the one hand, the system's functionalities are offered via a graphical user interface that can be accessed with any common Web browser (e.g., Mozilla Firefox, Google Chrome, Internet Explorer, Opera or Safari). On the other hand, a Web Service API is offered allowing to access the ALOE functionalities.

ALOE is a Web-based application

Figure 5.1 provides an overview of the ALOE system architecture that comprises the following layers:

ALOE Backend Services: In this layer, all services are located that can be accessed by other components in the environment where ALOE is deployed. The main component is the *ALOE Web Service* that is containing the business logic, and that is responsible for exchanging data with the *ALOE Data Storage* or *External Services*. It offers access to more than 200 methods (see Appendix F) and is realised as a REST Web Service. Further components in this layer are the Tika-based *Extraction Web Service* with means to extract contents and metadata from a variety of different resource types, and the *ALOE Thumbnail Service* that can generate preview images for most common resource types.

ALOE Data Storage: Resources and metadata that are held within the ALOE infrastructure are stored in a relational database (MySQL 5.5) that can optionally be synchronised with a Lucene index using an observer pattern approach.

ALOE Servlets: This layer contains three servlets responsible for realising the user interfaces within the ALOE infrastructure:

- The *ALOE View* servlet realises the Web interface as the main means for users to interact with the ALOE functionalities. It accesses the required methods through the *ALOE Web Service*. Furthermore, *ALOE View* also comprises means to send information mails to users about selected activities within the system.

---

[13]See `http://www.w3.org/Protocols` for more information

125
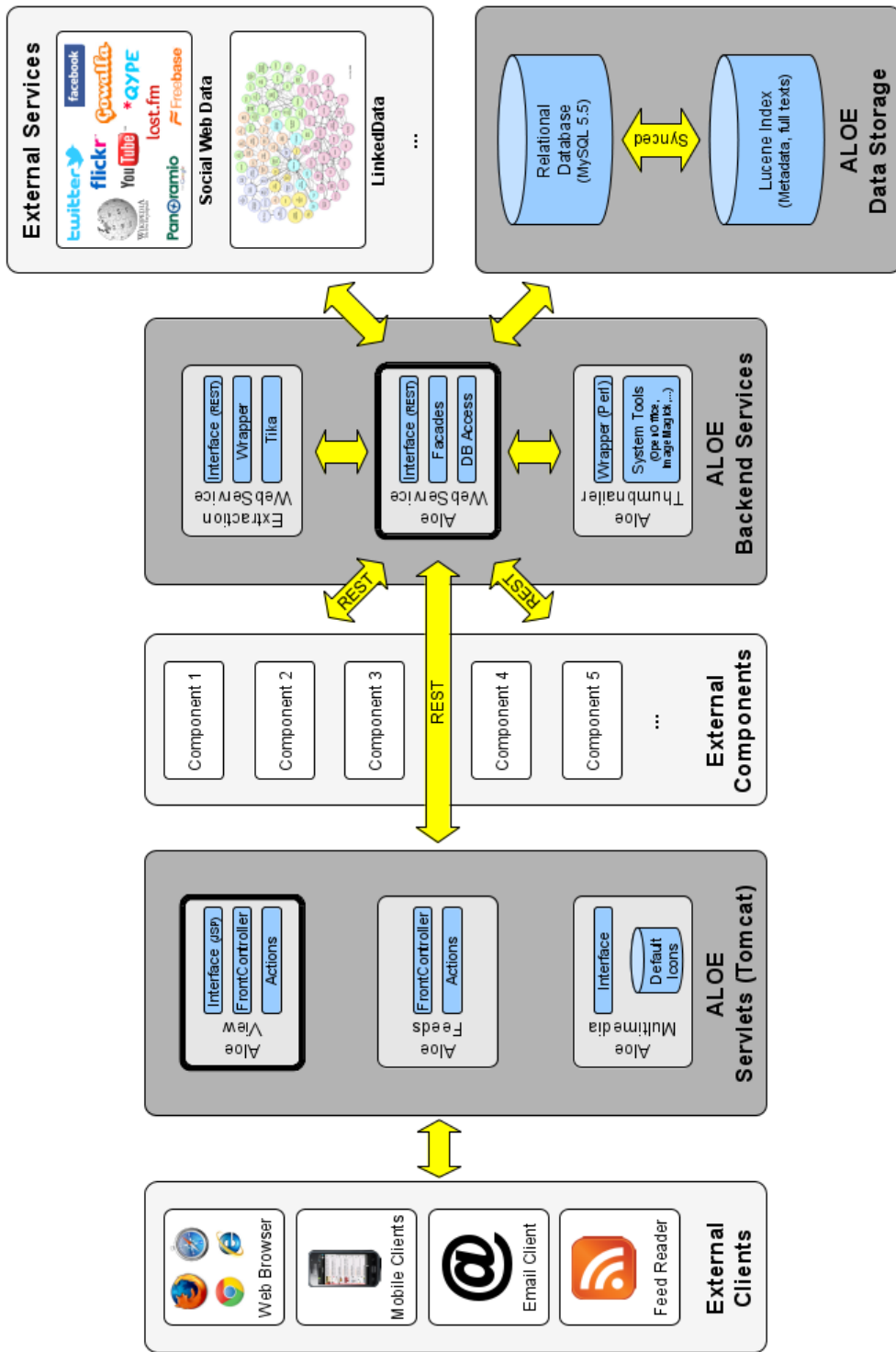
Figure 5.1: ALOE – system architecture and components

- The *ALOE Multimedia* servlet is responsible for the provision of all resources stored in the ALOE database (e.g., preview images or file resources that were uploaded).

- *ALOE Feeds* provides feeds based on the stored user interaction history.

Detailed information about the system design and the single components will be provided later in Section 5.4.

5.2 ALOE Main Concepts

Before an introduction to the Web Interface and the main features will be given, the following main concepts in ALOE will be presented: **resources**, **users**, **groups**, and **categories**. Further entities exist such as *collections* and *persons*, but they are not in the focus of this work. For an overview of different entity types and the respective database schemas, please refer to Appendix B.

5.2.1 Resources

Resources are the key entities within the infrastructure. According to the key questions discussed in Chapter 4, resource types and metadata which are supported in ALOE will now be introduced.

Resource Types

ALOE distinguishes two main types of resources that can be contributed, annotated, and organised:

Bookmark resources: A bookmark in ALOE is anything that can be referred to with a URI, e.g., a page from the World Wide Web, an Intranet page, or a resource in a repository.

File resources: File resources in ALOE are arbitrary files that users can upload to ALOE. For example, a file resource can be a text document, an image, an audio file or a movie, represented in any kind of format. Depending on system resources, ALOE just defines a limit for the maximum file size of such resources (typically 16MB).
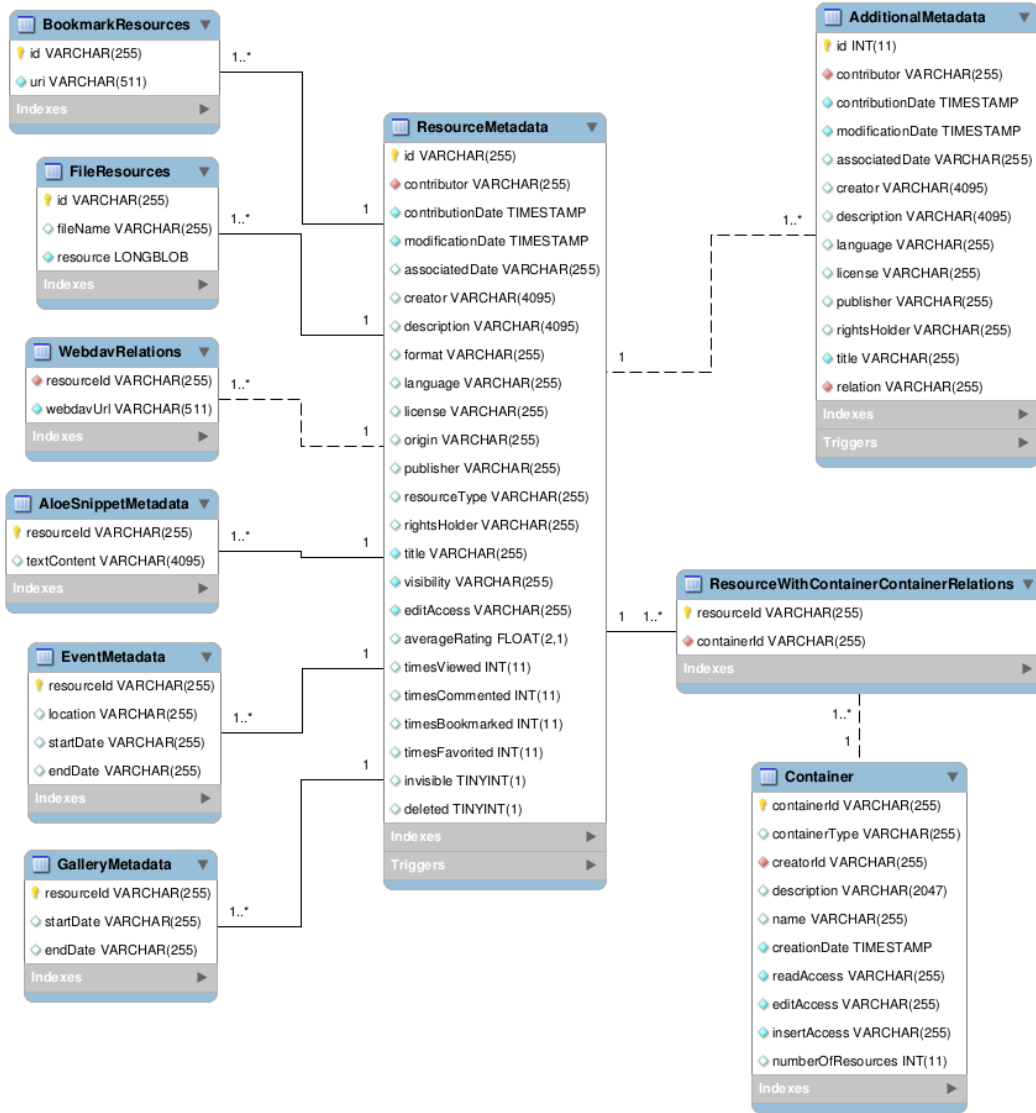
Figure 5.2: An excerpt of the ALOE database schema with tables for different resource types

ALOE is repository and referatory

As ALOE can handle both resource types, it realises a *repository* as well as as *referatory*.

Furthermore, specialised resource types such as events, galleries or snippets are supported (see Figure 5.2). Depending on the needs of the scenario where ALOE is deployed, further types can be introduced as shown later in Chapter 6. Specialised resource types do not only differ regarding their metadata, but also

regarding the following aspects:

- Some types are not necessarily related to any bookmark or file (e.g., a snippet where users can provide text contents on their own using a Wiki language).

- Some types can have associated resources stored in so-called containers (see Figure 5.2). E.g., users can provide additional information for events by associating resources such as an agenda or a picture.

Resource Metadata

In the design of the ALOE resource metadata schema, the aim was on the one hand to allow as much diversity as possible, while on the other hand providing a simple and easily understandable core set of elements still providing all necessary information as identified in this thesis so far. A lot of attributes from Dublin Core were chosen to ease the import and export of resources.

In order to allow the consideration of subjectivity in an adequate way, ALOE does not only provide means to store unique metadata for each resource. Instead, a mode can be activated where almost all metadata elements can exist in different versions. For example, different users may of course always assign different tags for a resource, but they may also provide a different title or description for it[14]. Sometimes such different versions of elements can optionally be aggregated by ALOE (i.e., to show all tags that exist for a resource), and sometimes only one selected version is returned (e.g., the resource title). This will be shown in more detail when introducing the respective functionalities later. The only metadata that is unique for each resource is the resource URI, the resource mime type that is automatically extracted, and the resource preview image that is automatically generated by ALOE.

Different versions of most metadata elements are allowed

Resource Administration and Rights Management ALOE distinguishes several access rights for resources that can be set only by the respective resource administrator. Initially, the contributor of a resource is the only resource administrator, but further ALOE users can be added as resource administrators.

To allow a maximum control of data, each resource can be contributed with a selected visibility. It decides whether the resource and the respective metadata

ALOE offers fine-grained rights management

---

[14]This mode is referred to as *bookmark mode*. The other mode that can be chosen is the *favourites mode* where it is usually not possible to contribute the same bookmark resources with different resource metadata.

can be accessed or returned as the result of a search operation. The following visibilities are possible in ALOE:

Private: The resource and metadata are only visible for the resource administrators, so private resources will also only appear in search result lists for the respective users, not for anyone else. This allows to use ALOE also for private resource organisation where certain information shall not be made available to other users.

Group: In this case, the identifier of a closed ALOE group (i.e., a group that requires the permission of a group administrator to be joined) also has to be provided. The resource and the respective metadata are then only visible for members of the closed group.

Public: This is the default visibility that allows any user to access the resource and the metadata for it.

If a bookmark is contributed to ALOE, and this bookmark does not yet exist with the chosen visibility (e.g., *public* or restricted to a closed group in ALOE), then a new identifier will be created within ALOE. As a result, the same bookmark can exist several times in ALOE, and metadata will always only be aggregated for bookmarks with the same visibility.

In addition to the visibility of a resource, the resource administrators can specify corresponding *edit access rights* (also using private, public, or an arbitrary number of closed groups). For resources with associated containers, administrators can also specify respective *insert access rights* in the same way. In contrast to the resource visibility that is fixed for each resource, *edit access rights* and *insert access rights* can be modified at any time.

Main Metadata Types Figure 5.3 provides an overview of different kinds of metadata that can be associated with a resource in ALOE. The main types are now briefly introduced:

Basic Metadata: Basic and automatically generated metadata about the resources

Individual Metadata: The basic metadata that can be specified by the resource administrators.[15]

---

[15]When ALOE is run in the *bookmark mode*, this metadata can be specified individually by each contributing user.

Tag: Tags that can be associated with the resource (ALOE allows duplicate tags from different users for the same resource[16])

Comment: Comments that users can add to a resource

Rating: Ratings that can be associated with a resource (for that, ALOE offers a discrete scale of five different rating values)

Group: The groups the resource has already been shared to

Collection: The collections the resource belongs to

Associated Metadata: Arbitrary metadata sets can be associated with a resource in ALOE (like a resource, such a metadata set can be a file that is uploaded to ALOE, or something that is referred to with a URI)

Category: A generic element that allows to specify relations to one or more entries (categories) of a selected vocabulary that was integrated in ALOE. A category is defined by a taxonomyId and a categoryId (referring to the specified taxonomy), and the relations are defined by a specific relation type. More details about categories as a means to allow the specification of formal relations between fixed vocabularies with entities in ALOE will be provided in Section 5.2.4.

The following table shows all ALOE resource metadata. The column *Creator* denotes whether the respective value or set of values is explicitly created by a user or extracted/generated automatically by ALOE. Please refer to Appendix A.3 for a detailed description of all ALOE resource metadata elements (including occurrences, value spaces, and data types).

Element name	Explanation	Creator
1 Basic Metadata		
1.1 dc:identifier	URI identifying the resource uniquely in ALOE[17]	ALOE
1.2 uri	URI identifying the resource uniquely	ALOE/ User[18]

---

[16]Marlow et al. therefore distinguish *bag models* and *set models* for the provision of tags [MNBD06]. In this sense, ALOE follows a bag model.

[17]This identifier will automatically be generated by ALOE. Every combination of *uri* and *visibility* is a unique resource within ALOE.

[18]In case of a file upload, the URI is automatically generated by ALOE.

Element name	Explanation	Creator
1.3 visibility	the resource visibility	User
1.4 dc:format	either MIME type or a proprietary format	ALOE
1.5 resourceThumbnail	an image file associated with the resource	ALOE
1.6 Individual Metadata		
1.6.1 contributor	ALOE identifier of the user who contributed the additional metadata (and added the resource to his/her ALOE portfolio)	ALOE
1.6.2 contributionDate	contribution date	ALOE
1.6.3 dc:publisher	entity responsible for making the resource available	User
1.6.4 dc:title	title of the resource	User
1.6.5 dc:description	a free-text description of the resource	User
1.6.6 dc:date	creation date	User
1.6.7 dc:creator	author(s) of the resource (not to be confused with the publisher)	User
1.6.8 dc:language	language of the resource content	User
1.6.9 dc:rights	creative commons license which is associated with the resource.	User
1.6.10 rightsholder	rightsholder of the resource	User
1.6.11 dc:source	origin of the resource	ALOE/ User
2 Tag		
2.1 dc:creator	ALOE identifier of the user who contributed the tag	ALOE
2.2 dc:date	contribution date	ALOE
2.3 dc:subject	the tag associated with the resource	User
3 Comment		
3.1 dc:creator	ALOE identifier of the user who contributed the comment	ALOE
3.2 dc:date	contribution date	ALOE
3.3 commentText	the comment text associated with the resource	User
4 Rating		
4.1 dc:creator	ALOE identifier of the user who contributed the rating	ALOE
4.2 dc:date	contribution date	ALOE

Element name	Explanation	Creator
4.3 ratingValue	rating value associated with the resource	User
5 Usage Metadata		
5.1 averageRating	the average rating of the resource	ALOE
5.2 timesBookmarked	the number of users that have the resource in their portfolios	ALOE
5.3 timesCommented	the number of comments left on the resource	ALOE
5.4 timesViewed	the number of times the resource was viewed (views by contributors are not counted in ALOE)	ALOE
6 Group		
6.1 groupId	the identifier of the group	User
6.2 sharingDate	date the resource was shared to the group	ALOE
7 Collection		
7.1 collectionId	collection the resource belongs to	User
7.2 addingDate	date the resource was added to the collection	ALOE
8 Associated Metadata[19]		
8.1 dc:identifier	URI identifying the metadata set uniquely	ALOE
8.2 dc:publisher	ALOE identifier of the user who added the metadata set	ALOE
8.3 dc:creator	author(s) of the metadata set (not to be confused with the publisher)	ALOE
8.4 dc:date	contribution date	ALOE
8.5 dc:description	a free-text description of the metadata set	User
8.6 dc:format	metadata format	User
8.7 relationType	the type of relation between the resource and the metadata set	ALOE/ User
8.8 provenance	provenance information	ALOE/ User
9 Category		
9.1 taxonomyId	the identifier of a taxonomy the category refers to	ALOE/ User

---

[19]Currently, no license information can be provided here. Yet, enhancements basing on approaches such as Open Data Commons (`http://www.opendatacommons.org`) are possible.

Element name	Explanation	Creator
9.2 categoryId	the identifier of a category in the specified taxonomy	ALOE/ User
9.3 name	the display name of a category	ALOE/ User
9.4 description	the description of a category	ALOE/ User
9.5 ResourceCategoryRelation		
9.5.1 contributor	ALOE identifier of the user who contributed the relation	ALOE
9.5.2 addingDate	relation contribution date	ALOE
9.5.3 relationType	the type of relation between the resource and the specified category	ALOE/ User
9.5.4 weight	a weight associated with the relation	ALOE/ User
9.5.5 provenance	provenance information	ALOE/ User

Table 5.1: ALOE resource metadata

5.2.2 Users

Users in ALOE can register and login to the system, they can maintain profile pages, and they can build up contact lists and send messages to each other. These functionalities will be presented in more detail in Section 5.3.2.

Access to information with public visibility as well as to lots of functionalities is also possible for anonymous users, i.e., users that did not register in ALOE or did not login.[20] Yet, contribution of information is only possible when logged in. Anonymous contributions are not desired mainly because of three reasons:

1. The possibility of anonymous contribution would allow to spam the system easily.

2. Information about a contributor is an essential factor when interpreting given information.

---

[20]This refers to the default mode. For specific scenarios, ALOE can also be run in a mode where no information at all can be accessed by anonymous users.
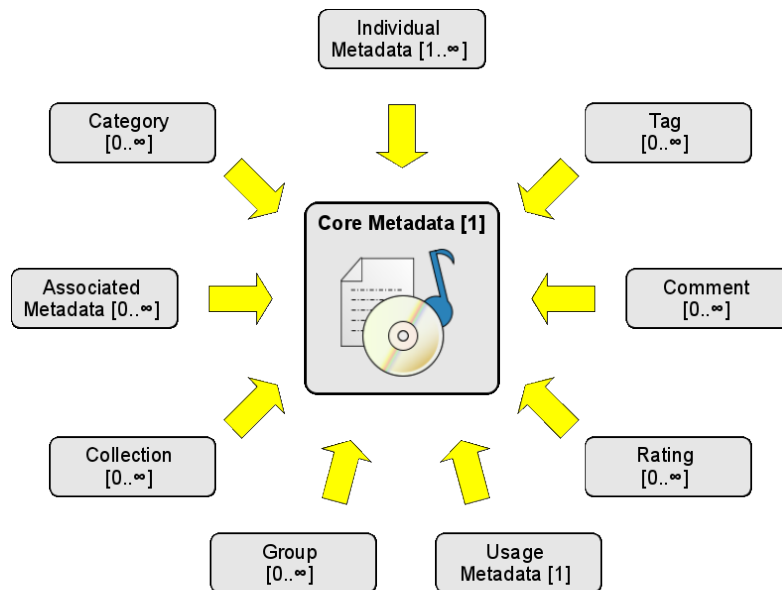
Figure 5.3: A schematic overview of the ALOE resource metadata

3. The unique identifier of a contributor allows to offer a variety of social browsing facilities (e.g., showing other contributions of the same user).

User Metadata

ALOE offers the user metadata as presented in Table 5.2. Information about which metadata is optional and which metadata is mandatory as well as information about possible values for the different elements will be given later when the respective functionalities will be presented.

Element name	Explanation	Creator
Basic Profile Information		
id	URI identifying the user uniquely in ALOE	ALOE
nickname	A nickname identifying the user uniquely in ALOE	User
password	The password required to log in to ALOE (stored as an MD5 hash)	User
email	The user's email (can never be accessed by other users)	User

Element name	Explanation	Creator
memberSince	The user's joining date	ALOE
gender	The user's gender	User
country	The country a user comes from	User
spokenLanguages	Languages spoken by the user	User
interests	The user's interests	User
buddyIcon	The user's buddy icon	User
resourcesCount	The number of resources the user added to his/her portfolio	ALOE
commentsCount	The number of comments the user left on ALOE resources	ALOE
ratingsCount	The number of ratings the user left on ALOE resources	ALOE
Additional Profile Information		
firstName	The user's first name	User
lastName	The user's last name	User
birthday	The user's date of birth	User
street	User address: street	User
zipCode	User address: zip code	User
city	User address: city	User
instantMessaging	Identifier of a messaging service	User
messengerId	The user's id in the respective messaging service	User
Diverse information		
visibility	Specifies whether additional profile information will be private, public, or only visible for users that are part of this user's contact list	User
confirmed	If registration confirmation is required, this specifies if the user confirmed his/her registration	ALOE
confirmationId	An ALOE internal identifier used to allow the confirmation of a registration	ALOE

Table 5.2: ALOE user metadata
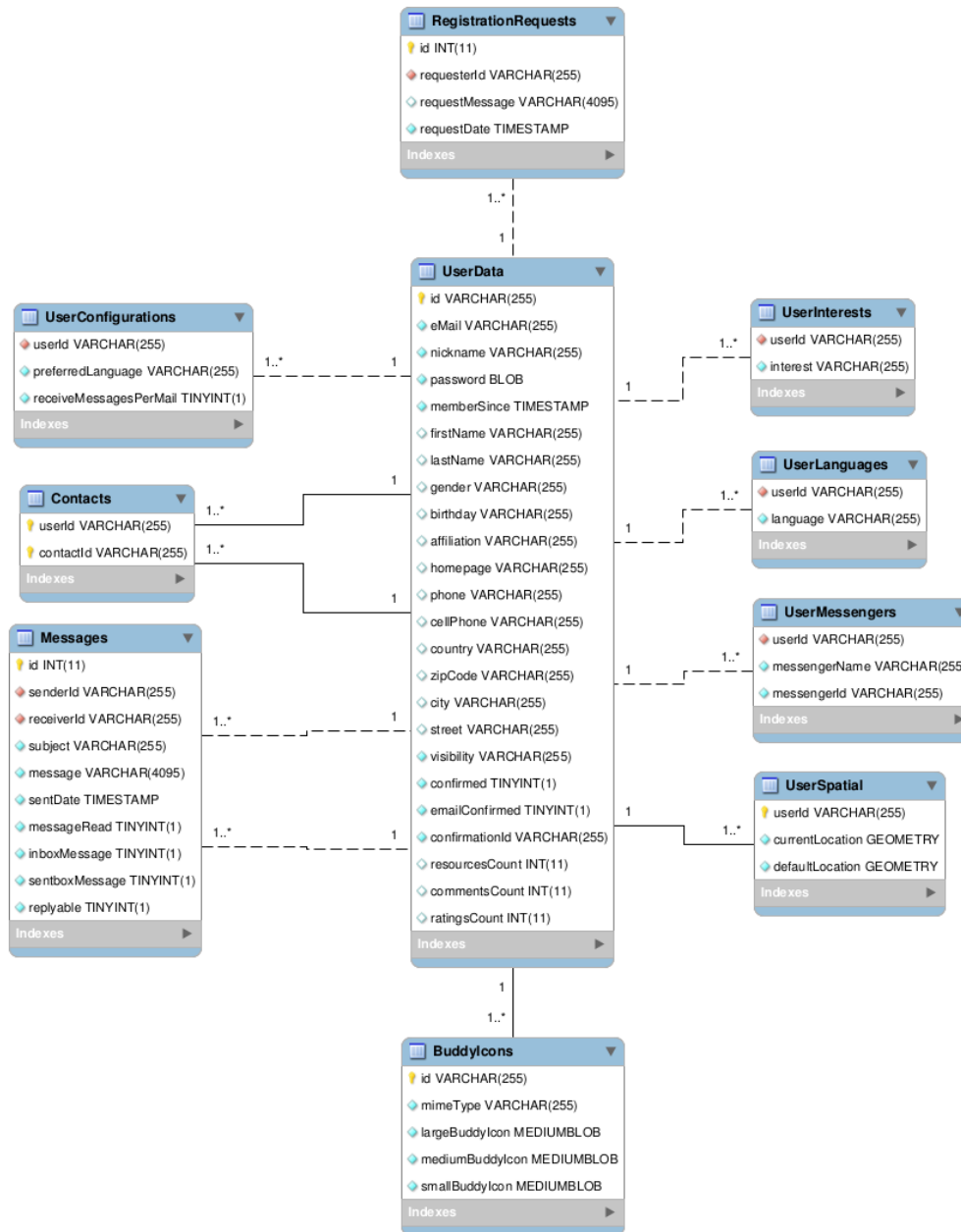
Figure 5.4: An excerpt of the ALOE database schema with key tables for users

5.2.3 Groups

Groups in ALOE are a means to organise groups of users, and to explicitly share resources with and send messages to the group members. ALOE distinguishes

ALOE distinguishes open and closed groups

open groups that can be joined by any user, and *closed groups* where joining requires permission of a group administrator. Furthermore, closed-private groups allow that resources can be contributed with restricted visibility, i.e., these resources as well as all annotations for them are only visible for the respective group members. Any information about such groups will only be visible for members of the group. Analogue to resource administrators, the creator of a group initially is the only group administrator, but further ALOE users can be added as group administrators later.

ALOE offers the following group metadata:

Element name	Explanation	Creator
id	URI identifying the group uniquely	ALOE
name	A name that identifies the group in ALOE	User
description	A free-text description of the group	User
originatorId	ALOE identifier of the user who initiated the group	ALOE
foundationDate	Date when the group was initiated	ALOE
status	Group status (open, closed, or closed-private)	User
numberOfMembers	Number of group members	ALOE
numberOfResources	Number of resources shared to the group	ALOE
groupThumbnail	An image file associated with the group	User

Table 5.3: ALOE group metadata

5.2.4 Categories

In many use cases, relations of resources (and other entities) to terms of a fixed vocabulary or to categories within specific taxonomies are desired and/or already existing. To allow the provision of such relations in a generic way, ALOE offers so called categories that are defined as shown in Table 5.4. Each category is uniquely defined by its categoryId and its taxonomyId.

Element name	Explanation	Creator
taxonomyId	Unique identifier of a taxonomy the category refers to	User

Element name	Explanation	Creator
categoryId	Unique identifier of the category in the specified taxonomy	User
name	A display name for the category	User
description	A free-text description of the category	User

Table 5.4: ALOE category metadata

In ALOE, relations to categories can be defined for resources, users, groups, and collections. The metadata that can be provided for such relations is shown in Table 5.5. It allows for the definition of arbitrary formal relations and can even be used as the basis for reasoning within an ontology as realised within the NEXUS project. This was possible as there are also means to define relations between categories.

Element name	Explanation	Creator
relationId	URI identifying the relation uniquely	ALOE
resourceId	ALOE identifier of the related resource	ALOE
taxonomyId	Unique identifier of the taxonomy the related category refers to	User
categoryId	Unique identifier of the related category in the specified taxonomy	User
contributorId	ALOE identifier of the user who contributed the relation	User
addingDate	Date the relation was contributed	User
relationType	Type of relation between the resource and the category	User
weight	A weight associated with the relation	User
provenanceInformation	Information about the provenance of the relation	User

Table 5.5: ALOE category relation metadata

5.3 ALOE Functionalities

In this section, the ALOE system and its main features as provided through the ALOE Web interface are introduced. For selected features, the names of the respective functionalities as provided through the *ALOE Web Service* are also be presented. Due to the focus of this work, the emphasis is put on the resource related functionalities. For a more detailed overview with selected functionalities, please refer to Appendix F.

First a general overview of the Web interface is given, then features structured into the following categories are presented:

User management: This includes functionalities like register and login, but also profile, contact, and message management.

Resource and metadata contribution: Different ways to contribute new resources and to add resources to one's own portfolio as well as means to provide information about them (e.g., tags, ratings, and comments)

Resource and metadata organisation: Features allowing to organise resources (e.g., groups and collections)

Search: Different search capabilities that ALOE offers for resources, users, and groups

Notifications, import and export of data: This entails all functionalities for importing and exporting data from and to ALOE.

5.3.1 A General Overview of the ALOE Web Interface

ALOE users are usually first directed to a start page as shown in Figure 5.5. It can be configured to provide information about a variety of different entities and activities in system (e.g., about newest members or recently added resources). The aim is usually to realise a dynamic and lively entry to the system, and also to offer the opportunity to immediately explore new information.

Like almost all ALOE pages, it starts with the page header as described in Figure 5.7, so access to all relevant functionalities is guaranteed. Some elements shown on the start page in Figure 5.5 will now be explained in more detail as they are also used on a variety of other pages:

- A *welcome message* that can be specified by any ALOE administrator. Different welcome messages for anonymous and logged in users can be defined. For example, a welcome message for a logged in user can also

Figure 5.5: Screenshot of an ALOE start page

contain information such as a personal greeting or information about new messages in the user's inbox.

● Information about *recently added resources* and *highest rated resources* in ALOE. In contrast to the resource detail page (see Page 148), this condensed view as shown in Figure 5.6 only provides a subset of the available information, namely:

– a small resource preview thumbnail with a link to the detail view

    page,

    – the resource title with a link to the detail view page,

    – the resource URL in case of a bookmark resource,

    – the number of times the resource has been viewed,

    – the average rating of the resource,

    – the resource description,

    – at most five tags that have been annotated for the resource (each of them linking to a respective tag search to get access to related resources),

    – the nickname of the first contributor of the resource (with a link to his/her profile page), and

    – the time when the resource was first contributed to the system.

    Further condensed views exist (e.g., a tile view that shows a larger preview image and fewer metadata, or a table view that simply shows a line of text without any preview image) and are also used to represent resources in result pages (see Section 5.3.1).

- A news text that can be specified by any ALOE administrator, e.g., containing information about latest updates or new features

- Nicknames as well as buddy icons of the three latest members that registered

- The 50 most popular tags displayed in a tag cloud (see Section 5.3.1). This tag cloud can be ordered alphabetically or by frequency.

1. FCK feiert Kantersieg gegen Rostock
http://www.fck.de/content/aktuell/news/15896_scrolltext.html
Views: 76   Average Rating: ★★★★★
Grossartig.
Tags: fck fussball kaiserslautern rostock
Added by Martin on 2008-11-18 08:07

Figure 5.6: Screenshot of an ALOE resource in a condensed overview mode

The Page Header

The page header (see Figure 5.7) is part of every page generated by ALOE. It offers immediate access to all important functionalities and consists of the menus as shown in Figure 5.8 that are activated by hovering the cursor above them:

MyALOE: This menu is only available for logged in users and contains links to functionalities that offer access to any information related to the user.

Explore: Here, links to different views on resources and members are provided, as well as a link to the advanced search (see Section 5.3.5).

Community: This menu provides information about users and groups in ALOE.

Search: Users can enter search terms here. Depending on which of the elements *Resources*, *Groups*, and *Members* is selected, a respective search will be carried out (see Section 5.3.5).
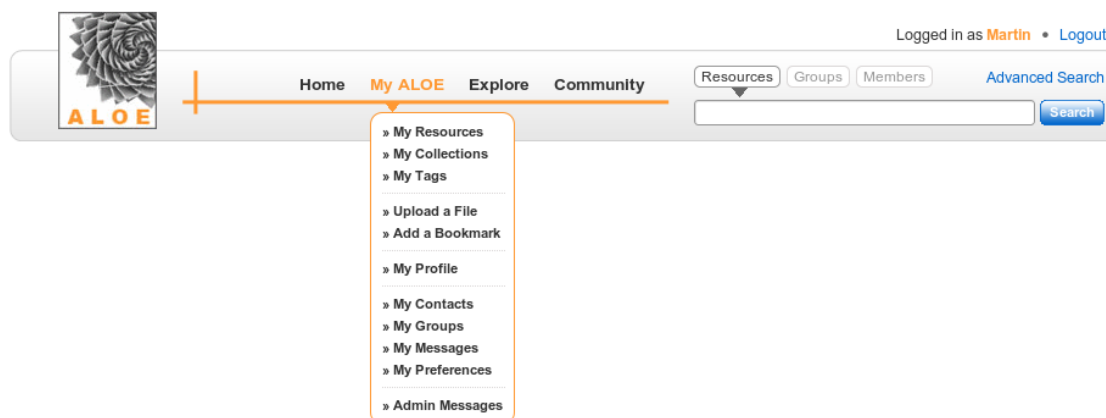
Figure 5.7: Screenshot of the ALOE page header with activated menu *My ALOE*

The Page Footer

Like the page header, the page footer is also part of each page generated by ALOE. It contains links to legal information (a disclaimer) and contact data. Depending on the ALOE instance, it can also contain a link to a blog with information about the respective ALOE instance, a link to a wiki page for feedback, or a link to a page with frequently asked questions, which can be maintained by any ALOE administrator.

» My Resources	» Recently Added	» Members
» My Collections	» Highest Rated	
» My Tags	» Most Viewed	» Show Groups
	» Most Commented	» My Groups
» Upload a File	» Most Bookmarked	» Initiate Group
» Add a Bookmark		
	» Advanced Search	
» My Profile		
	» Members	
» My Contacts		
» My Groups	» Categories	
» My Messages		
» My Preferences	» Spherical Tag Cloud	
» Admin Messages		
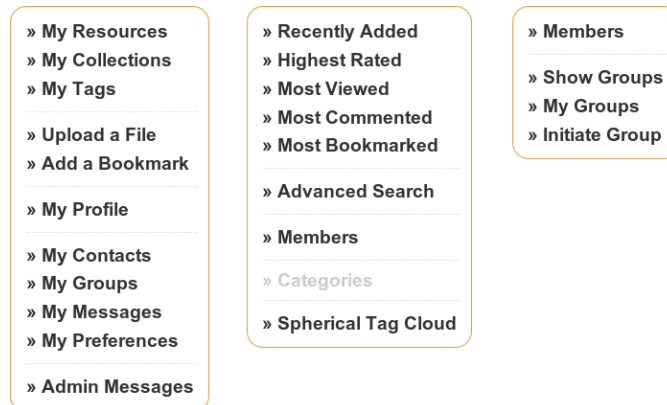
Figure 5.8: The page header menus *My ALOE*, *Explore*, and *Community*

Parenthesis: Tag Clouds

ALOE does not only show a tag cloud on the welcome page, but also in two other contexts:

- On each group page (see Section 5.3.4), a tag cloud with the most popular tags assigned to resources in this group is shown,

- On each user activities page, a tag cloud with the most popular tags added by this user is shown, and

- on the detail page of a resource, tags which have been assigned to this resource so far are shown (they are also displayed as a cloud because ALOE follows a bag model for tagging resources, i.e., the same tag can be assigned several times to a resource by different users).

In an ALOE tag cloud, a tag's frequency is mapped to a certain font size. A very straightforward approach that is often used works as follows:

Let n be the number of different font sizes $s_0, \ldots, s_{n-1}$ that shall be used, f_{min} the lowest and f_{max} the highest frequency of a tag in the cloud, and f_t the frequency of a given tag t.

Then one can determine a frequency range $r = \frac{f_{max} - f_{min}}{n}$ and assign the tags as follows:

- All tags t with $f_{min} \le f_t < f_{min} + r$ are assigned to font size s_0,

- all tags t with $f_{min} + r \le f_t < f_{min} + 2r$ are assigned to font size s_1.

- . . .

This approach is very simple, provides satisfying results, and was for some time also used in ALOE. Yet, in the MACE use case (see Section 6.4.1), a phenomenon occurred that is illustrated in Figure 5.9. One very large tag can be

Figure 5.9: ALOE-generated tag cloud in MACE (old approach)

seen – namely *iuav_test*, whereas all other tags are displayed with the same, smallest font size. The reason for this undesired behaviour was that the tag frequencies were not equally distributed within the tags to be displayed, and this is what is implicitly assumed when generating a tag cloud as explained. In the above example, the following frequencies existed:

- *iuav_test*: 176

- *glass*: 29

- *light*: 25

- *diagram*: 16

- *unreadable*: 14

- . . .

- *vanderrohe*: 4

In this scenario, five different font sizes were available ($s = 5$), so the frequency range r for each font size was $(176 - 4)/5 = 34.4$. As a result, each tag except for *iuav_test* was displayed with the smallest font size. Yet, outliers such as *iuav_test*, or big skips between frequencies of tags are not unusual, because often Long Tail distributions are found here. To solve this problem, the following key characteristics were identified for a new approach to generate tag clouds in ALOE:

- It should work like it did before for equally distributed tags, but also for other cases.

- Tags with a similar frequency should be displayed with a similar font size in the cloud.

- A big skip in the frequency distribution should result in a big skip in the font size.

- As many of the available font sizes as possible shall be used.

As a result, the following algorithm to determine ranges was implemented with $T :=$ the set with all tags to be displayed:

$i \leftarrow 1, \quad r \leftarrow \frac{f_{max} - f_{min}}{n}$
while $T \neq \emptyset$ **do**
$\quad T_i \leftarrow \{t \in T | f_{min} \leq f_t < f_{min} + r\}$
$\quad T \leftarrow T \setminus T_i$
$\quad i \leftarrow i + 1$
$\quad$ **if** $T_i = \emptyset$ **then**
$\quad \quad i \leftarrow i + 1$ // one font size will not be used
$\quad \quad f_{min} \leftarrow \min_{t \in T} f_t$
$\quad \quad r \leftarrow \frac{f_{max} - f_{min}}{n - i}$
$\quad$ **end if**
end while

This approach is still more or less simple, but fulfils all of the demanded criteria. As a default, it still uses the same frequency range for all font sizes. When a tag is found for each available range (i.e., font size), the algorithm does exactly the same as the original one. But as soon as no tag is found for a certain range, the following happens:

1. The range is skipped, so that there is a corresponding skip in the visualisation.

2. The remaining tags that still have to be distributed will now be examined, determining the lowest frequency of a tag. Then this new f_{min} is used to calculate a new default range for the remaining tags and font sizes.

3. Continue recursively.

Figure 5.10 shows the resulting tag clouds as visualised with the ALOE Web Interface for the old and the new approach.

　　　　　　　　　　　　　　　　　　　　　　　　　　© Martin Memmel, 2015

Figure 5.10: Tag clouds generated with the old (left) and new (right) approach. Tags that are displayed with a new size are marked accordingly.

Result Pages

Whenever a number of objects (resources, users, groups or collections) is returned as the result of an operation, a respective result page is returned by the system. In Figure 5.11, an example of such a result page is shown. It always contains the following components:

- A *pager* that allows to jump to different result pages. A link to the first and last page is always offered.

- A selector for the number of items to show on each result page (the minimum is 10, the maximum is 100).

- A selector for the different sort criteria offered for the respective object and search type. For example, when a search for resources in a group was conducted, the criteria *Most relevant* (in case Lucene is integrated in the respective ALOE instance), *Alphabetically*, *Contribution Date*, *Sharing Date*, *Average Rating*, *Most Bookmarked*, *Most Commented*, and *Most Viewed* are offered. When a user was searching for a group, the criteria will be *Date*, *Number of Members*, and *Number of Resources*.
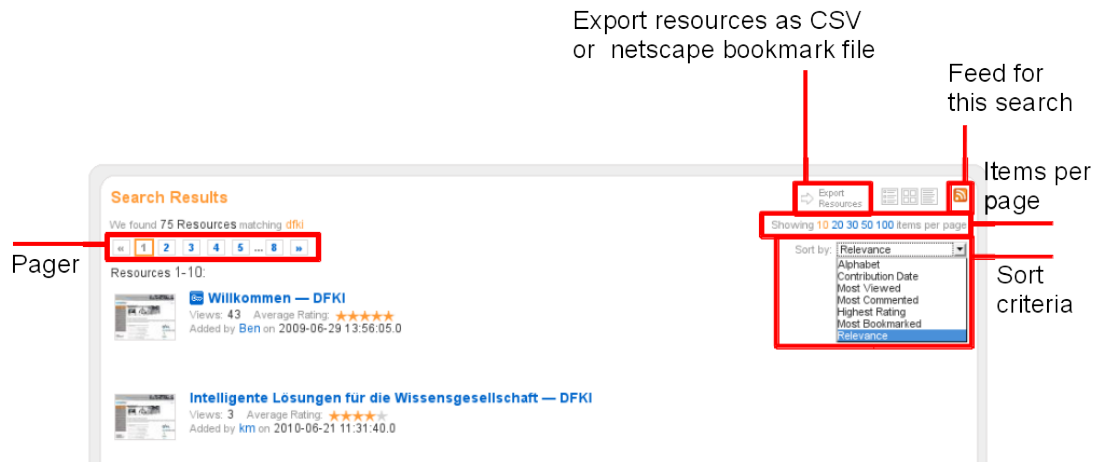
Figure 5.11: Screenshot of a result list in ALOE

When a search for resources was conducted, also the following elements can be found:

- A feed icon that links to the feed for the respective search. This allows to get notified when new resources matching the search criteria are contributed to ALOE.

- An export functionality that allows to store the search results as CSV file or as a Netscape bookmark file (this format can be imported in any common Web browser and other social bookmarking tools).

Which resources will be presented on resource result pages also depends of the access rights of the current user. For example, resources that were contributed with a different visibility than *public* will only appear if the user is allowed to access them.

Detail View of a Resource

In Figure 5.12, a detail view on a resource in ALOE is presented. It provides all resource-related information and functionalities.

In the upper left part, the title and a thumbnail of the resource as well as the following additional information are shown:

- the number of times the resources has already been viewed,

- the average rating of the resource (logged in users can also rate the resource here),

Figure 5.12: Screenshot of an ALOE detail page

- the resource visibility (public, private or group),

- copyright information,

- the resource creator, and

- an informal description of the resource.

Below users can comment on the resource and see other user's comments. In the upper right, tags that have been associated with the resource are shown. The

size of these tags corresponds to the number of times they had been associated with the resource. Users can add new tags here, and they can also remove own tags.

Below this tagging area, information about who else bookmarked the resource is presented. By clicking on the user's nicknames or profile pictures, their profile with information about their activity in the system is shown. By clicking on the icon below the profile picture, the information that this user annotated for the resource (i.e., title, creator, license, description) will be presented accordingly.

Next in this column, information about groups and collections the resource belongs to are provided. Finally, the icons in the lower right part show that two additional metadata sets had been associated with the resource.

Below the resource preview, some additional functionalities are offered. They are presented in a less prominent way in order to avoid confusing the user with too many interaction options. Figure 5.13 show these additional functionalities.

Figure 5.13: Additional functionalities available on an ALOE resource detail page

Report a Problem: In case of technical problems, wrong data, inappropriate contents etc., users have the opportunity to send a report. The respective email will always include a link to the resource detail page to allow for a convenient maintenance.

Send to a Friend: Logged in users here can specify an arbitrary email address to inform someone else about the respective detail page. A link to the page is automatically included, and additional text can be specified by the user.

Share to Group: A popup menu will appear allowing to share the resource to any of the groups the user is member of. Group administrators can also remove a resource from the respective groups here.

Add to Collection: Using this functionality, a resource can be added to or removed from any of the user's collections.

Edit Metadata: When a user bookmarked the respective resource, he/she can edit the metadata he/she associated with the resource (i.e., title, creator, description, and license).

Embedded Players   For selected resource types, ALOE provides an *embedded player* instead of a resource preview, i.e., the possibility to watch a resource or even interact with it within the ALOE page. Currently, this is offered for the following resource types:

- Bookmarks to videos provided by several online video platforms such as YouTube, Vimeo[21], or Dailymotion[22],

- bookmarks to maps provided by Google Maps[23],

- bookmarks to 3d models stored in the GoogleSketchup Warehouse[24]

- mp3 audio files uploaded to ALOE, and

- Flash video files that were uploaded.

5.3.2 User Management

Registration

The registration process in ALOE was designed to be as simple as possible, only asking the user for the information that is absolutely necessary to register:

- a nickname that is used to login and to represent the user in ALOE (here, the interface offers the opportunity to check whether the nickname that was entered is available),

- a password for the login that will also have to be confirmed to avoid misspellings,

- an email address used by ALOE to communicate with the user, and

- a confirmation of the ALOE terms of services as well as the ALOE privacy policy[25].

If the provided information is valid, an email will usually be sent to the user with a confirmation link that the user must click to confirm the registration and thus to finish the registration process (this feature can also be turned off). This confirmation process ensures that the provided email is valid and really belongs to the user. It further avoids registrations by spammers.

---

[21]See `http://vimeo.com`
[22]See `http://www.dailymotion.com`
[23]See `http://maps.google.com`
[24]See `http://sketchup.google.com/3dwarehouse`
[25]See Appendix C

Login

To login, a user simply has to provide his/her nickname and the respective password. The user further has the possibility to *stay logged in with cookie* (otherwise, logging in is required every time the browser is restarted). When a user has forgotten his/her password, *Forgot your password* offers the opportunity to define a new password. To do this, the user only has to provide a nickname; the system will then send an email to the corresponding email address with a link to an ALOE page where the user can set the new password.

LDAP support

ALOE also offers to connect to an existing LDAP server allowing the reuse of existing authentication mechanisms (see Section 6.4.2). In such a case, the registration process is slightly modified: The user has to provide the LDAP-credentials, and an additional nickname used to represent the user in ALOE. In the *ALOE LDAP Broker Web Service*, which realises this functionality, an ALOE password is automatically generated for the user, and the mapping between LDAP login and the ALOE credentials is stored. This way, ALOE avoids that the user's LDAP password needs to be stored in the ALOE database. When logging in to ALOE, the user simply has to provide the LDAP credentials.

User Profiles

As already described in Section 5.2.2, ALOE can store a variety of information about users. Such information is displayed and respectively entered on *user profile pages* (see Figure 5.14). In order to ease the input for the user, select fields[26] are provided wherever possible (e.g., for *gender*, *country*, and *spoken languages*). For the fields *interest* and *affiliation*, ALOE administrators can choose whether to allow free text, to define a fixed vocabulary (e.g., a given interest vocabulary or selected organisational units), or to use categories from a specific taxonomy.

To allow full control of their data, users can decide whether the *Additional Information* that is provided is always hidden for others, visible for every logged in user, or only visible for the users added to his/her contact list (see Section 5.3.2).

---

[26]A select field in an HTML form offers a select list (or drop-down list) to the user, so that no values have to be entered.

Figure 5.14: Screenshot of a user's own profile page in ALOE

User Activities

When checking the own profile or other user's profile pages, not only the data provided by them is shown, but also information about the user's activities in ALOE.

Figure 5.15 shows a screenshot of a page with an overview of latest user activities regarding different aspects such as resources, groups, and collections. For each of these aspects a detailed page with more information about the latest activities with respect to them exists.
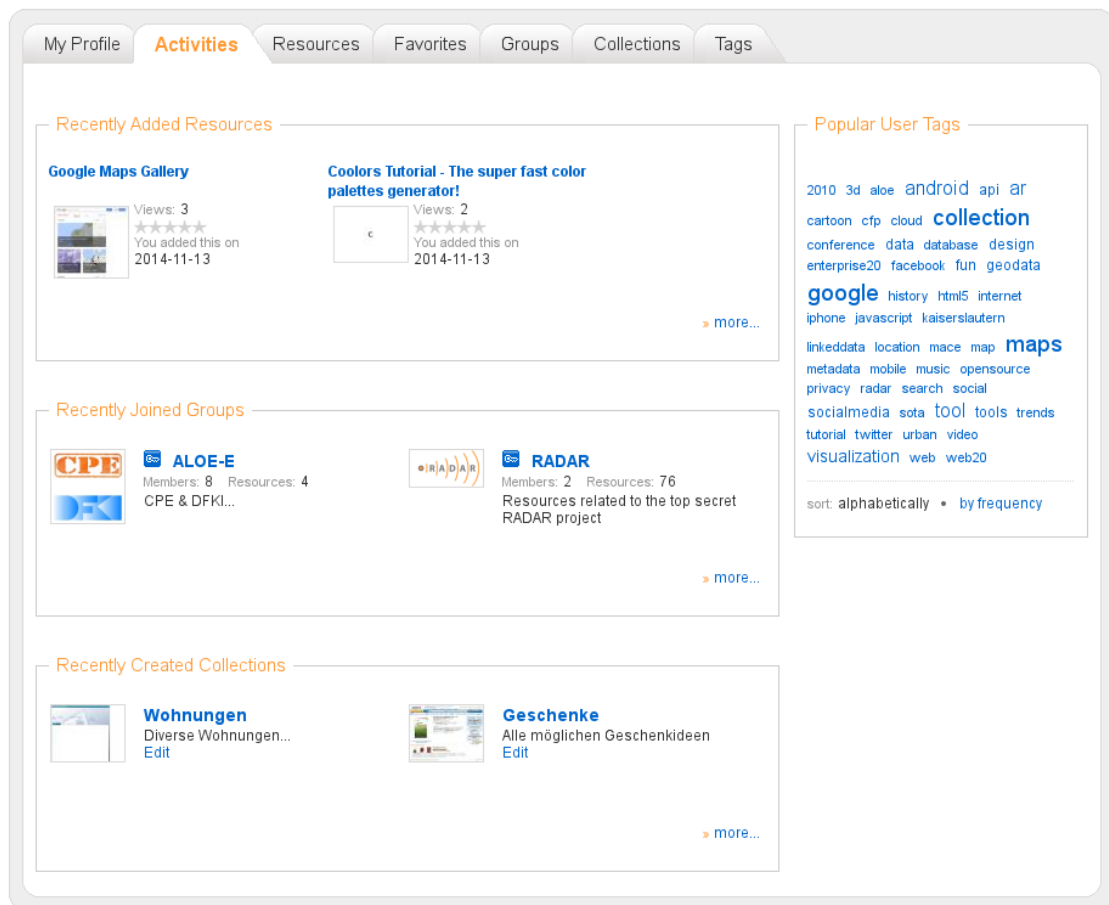
Figure 5.15: Screenshot of a user activities page in ALOE

Contact Lists

Users in ALOE can maintain contact lists by adding or removing other users. This option is always offered on the profile pages of other users. Contacts in ALOE are explicitly modelled in a non-symmetric way in ALOE because it is not wanted that people are forced to confirm contact requests. Nevertheless, when adding a user to one's contact list, this user gets notified with a respective email that offers a link to the profile page of the adding user, and with a note that it is also possible to add this user to the own contact list. This notification feature can optionally be turned off by an ALOE administrator, because in some scenarios, e.g., in an enterprise context, users could feel forced to put other users on their contact lists (see Section 6.4.2).

As of now, contact lists in ALOE are used for three purposes:

- As a shortcut to access information of users,

- to regulate who may access additional information in one's user profile, and

- to provide the basis of advanced functionalities that take this information into account.

Messages

Each user in ALOE can send and receive messages within the system. Sending a message is usually possible via using a message popup, where users can specify a subject and a message text.

An inbox and a sentbox are offered to manage ALOE messages. There, users can access and delete messages as well as reply to them. The following message types are distinguished:

Private messages are messages sent to one ALOE user by clicking the respective link on this user's profile page.

Group messages are messages that are sent to all members to a group (here, the additional option *reply all* is offered).

System messages are messages that were automatically sent by ALOE (e.g., information about being added to another user's contact list).

In order to avoid that users have to access the ALOE Web interface to check for new messages, ALOE offers to automatically forward ALOE messages to the user's email address as provided when registering.

Web Service API *acceptRegistrationRequest, addCategoryToUser, addContact, changePassword, confirmEmail, deleteFromContacts, deleteMessageFromInbox, deleteMessageFromSentbox, deleteUserAccount, denyRegistrationRequest, emailAvailable, getCategoryRelationsForUsers, getInboxMessages, getMessage, getRegistrationRequests, getSentboxMessages, getUserContacts, getUserData, getUserDataBeans, getUserRelatedCategories, getUserTagCloud, login, nicknameAvailable, registerUser, removeCategoryFromUser, requestResetPassword, resetPassword, searchUsers, sendMessage, setUserConfiguration, updateUserData*

5.3.3 Resource and Metadata Contribution

Contributing File Resources and Bookmark Resources

Via the *My ALOE* menu in the ALOE page header, users always can access the menu items *Add a bookmark* and *Upload a file*. Both types of contributions consist of two steps:

Specify your file / Specify your bookmark In this step (see Figure 5.16), the user has to provide a filename or a URL and the chosen visibility for the resource. The default visibility is *public*, but can also be set to *private* or *group*. When *group* is chosen, all closed groups the user is member of will be offered so that the user can choose one.
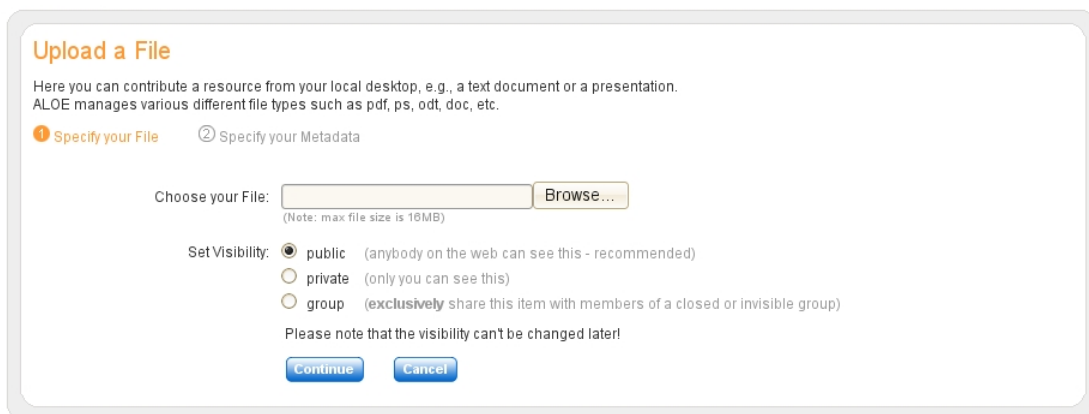
Figure 5.16: Step 1 of a resource contribution dialogue in ALOE

Specify your metadata In the second step (see Figure 5.17), the user can provide a title, tags, as well as a description, one or more creators and a license[27]. If the visibility was set to *public*, the user can also choose to share the resource to one or more of his/her groups. When adding tags, the system provides a *type ahead find* functionality (also called *autocompletion*), i.e., it offers a list with all tags the user already used that have the current input as prefix (e.g., when the user types *au*, the system might suggest tags like *audio* or *authority*, if these were already used by this user).

Thereupon, the *AloeThumbnailService* tries to generate a preview image for the resource. In case no preview image could be generated (e.g., for an audio file),

---

[27]Please note that is can be configured which of the attributes are mandatory or optional.

Figure 5.17: Step 2 of a resource contribution dialogue in ALOE

a default image is selected (depending on the resource type and resource mime type). As a final step, the resource as well as the specified metadata plus the preview image are stored in the ALOE database, and the user is forwarded to the respective resource detail page.

Contribution with the ALOE Bookmarklet In order to allow an easy contribution of Web pages without having to open a page in the ALOE system plus entering a URL, ALOE also offers a bookmarklet. A bookmarklet is small application that is stored as the URL of a bookmark in a Web browser and offers one-click functionality (the respective code could also be added as an element on a Web page). Figure 5.18 shows an example of an installed ALOE bookmarklet. When clicked, it automatically extracts the current site's URL and title and forwards the user to the first step of the ALOE bookmark contribution process, automatically using the extracted URL. In the second step, the extracted title is automatically put into the respective input field.
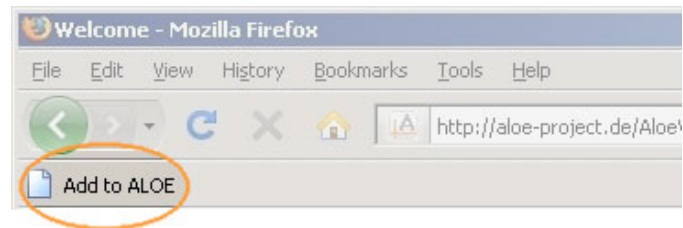
Figure 5.18: Example of an installed ALOE bookmarklet

Adding ALOE Resources to One's Portfolio   When a resource is not yet in a user's portfolio, ALOE always displays a button *Add this to my Bookmarks* on the respective resource detail page. When clicking it, a form as shown in Figure 5.17 will appear, pre-filled with the values that were currently shown on the detail page (except for the resource tags).

Automatic Metadata Generation in the Resource Contribution Process When the first step of the contribution process is finished, ALOE tries to automatically extract metadata from the given content. The user can then decide to adopt these metadata into the respective fields by clicking the blue arrows as depicted on Figure 5.17. ALOE currently uses a service based on the Aperture framework[28] to realise this functionality.

Contributing Metadata

ALOE provides several means to contribute metadata about a resource:

- The ALOE resource metadata as presented in Section 5.2.1 can be provided during resource contribution, and it can later be modified by clicking the link *Edit Metadata* in the menu shown in Figure 5.13.

- Users can add arbitrary metadata sets about a resource by clicking *Add new Metadata Set* on the respective resource detail page. Such additional metadata sets can be uploaded as file or be referenced with a URL. Further information about the contributed metadata set such as a schema that was used, information about the source, or a foreign can be provided. Certain formats (e.g., bibtex) will be displayed with a specific icon as shown in Figure 5.12. These additional metadata sets can be used for several different scenarios, especially to provide additional information helpful for other users (e.g., bibliographic information about a contributed paper) and to

---

[28]See `http://aperture.sourceforge.net`

offer machine-readable information that can be used by respective algorithms to provide enhanced search or recommendation facilities. For example, C-LINK used additional metadata sets to store information about specific events and papers that were then used to generate recommendations for resources and users.

- Categories (see Section 5.2.4) can be used to provide arbitrary formal classifications for a resource. E.g., in the MACE scenario (see Section 6.4.1), this was used to allow the storage of the association of MACE resources with concepts from the MACE taxonomy.

Web Service API

addComment, addResourceClassification, addTags, contributeBookmark, contributeBookmarkMetadataSet, contributeFile, contributeFileMetadataSet, deleteComment, deleteMetadataSet, deleteResource, deleteResourceClassification, deleteUserResourceTag, rate, updateResourceMetadata

5.3.4 Resource and Metadata Organisation

The delivery of resource subsets was already introduced as an important step of any resource access process in Section 2.2.3. Besides results of a search that was conducted (see Section 5.3.5), ALOE offers several means to provide resource subsets, based on how resources can be organised.

Groups

Group support is generally available via the *Community* menu in the page header. There, users can initiate new groups and request a list of all groups or only the groups they are member of.

Initiating Groups   To initiate a group, a user simply has to choose a group visibility and also has to provide a unique group name and an informal description of what the group is about. Furthermore, a thumbnail can be associated with the group. Group initiators in ALOE automatically become group administrators and can assign arbitrary other users as further group administrators.

Adding and Removing Resources   Users that are members of a group can share a resource to this group either while contributing the resource (see Section 5.3.3), or the corresponding resource detail page (see Section 5.3.1) using the

share to group functionality (see Figure 5.13). Removing resources from groups is currently only possible for group administrators. It is offered on resource detail pages (also in the right column where the groups the resource had been shared to are shown), but also as an additional link for each resource on *group resources* and *group resource search result* pages.

Group Lists   Group lists can be sorted according to the criteria *Alphabet*, *Date*, *Most Members*, and *Most Resources*. This allows to quickly find popular groups in the system. Single groups in this mode are presented as shown in Figure 5.19.

> **iGreen**
> Members: **11**   Resources: **92**
> Group for information relative to the BMBF iGreen project. ONLY FOR INTERNAL USE!
> Show Resources • Show Members • Send a Message to Members • Tell a Friend • Leave
>
> **Mobile Web**
> Members: **3**   Resources: **66**
> The mobile web will be big, and we'll have several use cases for iGreen as well as for the Stiftungsprojekt in 2010.
> Show Resources • Show Members • Send a Message to Members • Tell a Friend • Show Admin Page

Figure 5.19: Two groups as presented on a group list page in ALOE

It shows the group thumbnail, the group title (with an additional key icon in case the group is closed), the group description, as well as the number of members and resources. It can also offer the following interaction possibilities:

Apply for membership: When users are logged in, not member of the group, and when the group is closed, they can apply for group membership. A popup is then shown, allowing the users to provide a short explanatory statement for the group administrator why they should be accepted.

Join group: Joining an open group is possible for logged in users that are not yet member of the group.

Leave group: Leaving the group is possible for logged in users that are member of the group, as long as they are not the only group administrator left for the group.

Send a message to members: Users that are group members and logged in can send a message to all group members, using the message popup as already described.

Show admin page: Logged in group administrators can use this to access the group administration page as described below.

Show members: For open groups, this option is available for any ALOE user. For closed groups, it is only available for group members that are logged in. It shows a list of all group members.

Show resources: Exactly like for *show members*, this option is available for any ALOE user. For closed groups, it is only available for group members that are logged in. It forwards to a list of all resources that have been shared to the group.

Clicking on a group thumbnail or a group title leads to the respective group overview page that is described now.

Group Overview Pages The aim of a group overview page is to provide a variety of information about the respective group, and quick access to all group related functionalities. An example is shown in Figure 5.20.

In the top left, the group icon, the group title, the group description, the group initiator with a link to the user profile, the time the group was created, the group status (*open* or *closed*), and the number of members and resources is shown. Furthermore, the functionalities already presented for elements on group list pages (see Figure 5.19) can be accessed there.

Below, users can search for resources inside the group (in closed groups, this is only accessible for logged in group members). This functionality is described in Section 5.3.5. Furthermore, the two newest resources that were shared to the group are shown.

In the right column, links to group feeds are presented in case it is an open group. Two kinds of feeds are available: feeds for new resources, and feeds for activities in the group. As an alternative to feeds, logged in group members can modify their subscriptions to the available group email reports (also for new resources and activities). Group feeds and group email reports will be presented in detail in Section 5.3.6.

Finally, a tag cloud with the 50 most frequently used tags in the group is shown, with the tags linking to the corresponding group resource search. This tag cloud is shown for all open groups, and to logged in group members only for closed groups.

Group Administration The group administration page offers the following functionalities for group administrators:

Modify group description: The group description can be changed.

Figure 5.20: A sample group overview page in ALOE

Upload new group icon: An individual group icon can be uploaded.

Delete group: When an administrator wants to delete a group, the system will first show a message window that recommends to inform the group members first so that they have the chance to export group resources as described in Section 5.3.6. When a group is deleted, the following will happen:

- If the group status is *closed*, all resources and the corresponding metadata that were contributed with the respective visibility will automat-

ically be deleted.

- An email will be sent to all group members to inform them that the group was deleted by the group administrator. If the group status was *open*, the users are further advised to unsubscribe from the respective group feeds. If the group status was *closed*, the users are informed about the deletion of resources as described above.

Accept/deny membership requests: All pending membership requests for the group are shown here, and for each request the administrator can decide to accept or deny it.

Web Service API *acceptGroupMembershipRequest, addCategoryToGroup, addGroup, addGroupAdministrator, getCategoryRelationsForGroups, deleteGroup, denyGroupMembershipRequest, getGroup, getGroupAdministrators, getGroupMembers, getGroupMembershipRequests, getGroupResources, getGroups, getGroupsTagCloud, getGroupTagCloud, getUserGroups, joinGroup, joinGroupWithGlobalInvitationCode, removeCategoryFromGroup, requestGroupMembership, resignResourceFromGroup, resignUserFromGroup, searchGroups, searchInGroups, searchInGroupsSelectedFields, sendGroupMessage, shareResourceToGroups, updateGroupMetadata*

User Resources

As a central element to support social browsing, ALOE offers to browse each user's resource portfolio. Own portfolios can be browsed following the link *My Resources* in the menu *My ALOE*, and other user's resources can be browsed following the link *Show Resources of User* on the respective user's profile page.

Web Service API *getUserBookmarkResources, getUserFileResources, getUserResources*

User Favorites

As already mentioned, ALOE can be run in two different modes: The *bookmark mode* and the *favourites mode*. In the *bookmarks mode*, a bookmark for a URI only exists once in the system, but users can individually provide different metadata about the resource when adding it to their bookmarks. When the system is run in the *favourites mode*, users can either add a bookmark for the same URL as a completely unique resource where no metadata is aggregated, or simply add an existing resource in ALOE to their list of favourites (without contributing

any metadata at all). Consequently, own favourites can be browsed following the link *My Favorites* in the menu *My ALOE*, and other user's favourites can be browsed following the link *Favorites* on the respective user's profile page.

Web Service API *addResourceToFavorites*, *getFavoritesMetadata*, *getUserFavorites*

Tag Management

Tags are a central concept in ALOE and are one of the key elements to describe and organise resources. Therefore, specific support is provided for them under the link *My Tags* in the *My ALOE* menu. Figure 5.21 shows an example of an ALOE tag management page. Users can browse their tags here and see how often each tag was used by them and by all users. Furthermore, the following operations can be executed for each tag:

Rename: All occurrences of the tag will be renamed. This can also mean to unify different tags the user does not want to distinguish any more (e.g., *tool* and *tools*).

Delete: All occurrences of the tag will be deleted.

Search: All resources are shown that were tagged respectively.

Web Service API *addTags*, *deleteTag*, *deleteUserResourceTag*, *deleteUserTag*, *getResourceTagCloud*, *getResourceTags*, *getTagCloud*, *getUserTagCloud*, *getUserTags*, *renameUserTag*, *searchUserTags*, *updateResourceMetadata*

5.3.5 Search

ALOE offers different keyword search facilities for resources, groups, and users that will now be introduced in detail.

Simple Resource Search As shown in Figure 5.11, users can enter search terms and additionally specify in which of the metadata fields *title*, *description*, *tags*, *creator* to search. When starting the resource search from the page header, all metadata fields will be selected as default. Standard search operators and wildcard symbols can be used for querying (in case the index search is enabled, Lucene queries can be provided).
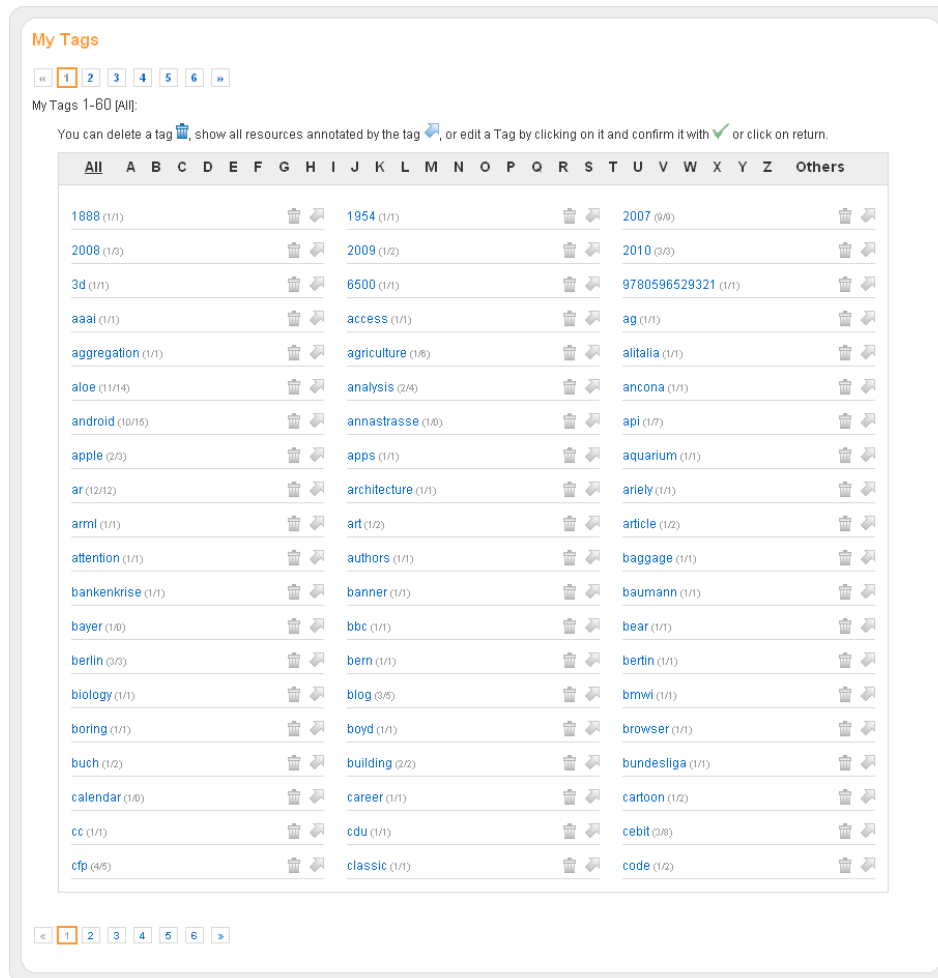
Figure 5.21: Tag management in ALOE

Advanced Resource Search The *advanced resource search* shown in Figure 5.22 allows a more fine-grained search in resources. It also realises an AND semantic, and offers prefix and suffix search (it is based on regular expressions, therefore performing slower than the simple resource search). Furthermore, the following search facilities that are not available in the *simple resource search* are offered:

Search fields: Different values for each of the attributes *title*, *tags*, *description* and *creator* can be provided.

Filter by scope: Logged in users can restrict the search scope to the resources in their own portfolio.

Filter by resource type: The search can be restricted to bookmark resources or file resources. If *only file resources* is chosen, users can additionally restrict the search to the mime type containers *audio*, *video* and *documents*.

Filter by date: Using the provided calendars or manually entering a date, the search can be restricted to resources that were first contributed in a certain period of time.

Filter by license: This allows to restrict the search to creative commons licensed resources. When this is chosen, it is also possible only to search for resources that allow a commercial use or that allow modifications.

Figure 5.22: The advanced search in ALOE

Group Resource Search As already shown in Section 5.3.4, it is also possible to search within resources of a specified group. This search works like the *simple resource search* explained above.

Group Search

Using the search in the page header, users can also search for groups. Any group where the specified search terms occur in the combination of the group title and the group description will be returned. This search works with regular expressions and thus also offer prefix and suffix search.

User Search

The user search is only available for logged in users, taking into account the metadata fields *nickname* and *language*. For users that have the user conducting the search in their contact list, also *first name*, *affiliation*, *city*, and *country* will be searched. As for the group search, the user search is accessible in the page header and works with regular expressions, offering prefix and suffix search.

Web Service API

search, *searchAdvanced*, *searchGroups*, *searchInGroup*, *searchInGroupSelectedFields*, *searchSelectedFields*, *searchUsers*

5.3.6 Notifications, Import and Export of Data

An important aim of this work is to allow that the information created within ALOE can be used in as many contexts as possible. Furthermore, users should not be forced to check the ALOE interfaces to find out if some new and potentially relevant information was contributed. Thus, several means to import and export data are offered, as well as a variety of notification mechanisms.

Importing Resources and Metadata

In order to import resources, ALOE of course offers the Web Service API that can be used to contribute bookmarks and to upload files. In addition to that, a specific XML schema (*ALOE-XML*) was developed allowing to specify any kind of ALOE resource metadata for an arbitrary number of resources in one file. Figure 5.23 provides an example for such a file, consisting of information about two resources. A specific Perl client was developed that can parse ALOE-XML, and that accesses the *ALOE Web Service* to contribute the respective resources to ALOE. The user just has to provide information about where to find the respective *ALOE Web Service* as well as a valid login and password.

Using this client as a basis, an import of resources was, amongst others, realised for the following scenarios:

```
<?xml version="1.0" encoding="UTF-8"?>

<resources>
   <resource>
      <visibility>public</visibility>
      <creator>Grieser, Gunter, Lange, Steffen, Memmel, Martin</creator>
      <title>DaMiT: Ein adaptives Tutorsystem für Data-Mining</title>
      <license>6</license>
      <description></description>
      <tags> martin publications </tags>
      <uri>/windows/d/work/diss/literatur/eigene/2003/GrieserLangeMemmel+03.pdf</uri>
      <sharetogroups>
         <group>KM Publications</group>
      </sharetogroups>
      <addtosets>
         <set>3</set>
      </addtosets>
      <additionalmetadatasets>
         <metadataset>
            <creator>Martin Memmel</creator>
            <format>bibtex</format>
            <description>bibtex</description>
            <uri>GrieserLangeMemmel+03.xml</uri>
         </metadataset>
      </additionalmetadatasets>
   </resource>
   <resource>
      <visibility>public</visibility>
      <creator>Mikael Nilsson, Matthias Palmér, Ambjörn Naeve</creator>
      <title>Semantic Web Metadata for e-Learning - Some Architectural Guidelines</title>
      <license>6</license>
      <description></description>
      <tags> martin publications </tags>
      <uri>/windows/d/work/diss/literatur/metadaten/NilssonPalmerNaeve+02.pdf</uri>
      <sharetogroups>
         <group>KM Publications</group>
      </sharetogroups>
      <addtosets>
         <set>3</set>
      </addtosets>
      <additionalmetadatasets>
         <metadataset>
            <creator>Martin Memmel</creator>
            <format>bibtex</format>
            <description>bibtex</description>
            <uri>NilssonPalmerNaeve+02.xml</uri>
         </metadataset>
      </additionalmetadatasets>
   </resource>
<resources>
```

Figure 5.23: An example for ALOE-XML

- To import information about more than 350 FlashMeeting video conference recordings in ALOE (see Section 6.4.4 for more information about this scenario), a converter was developed that can parse information from RSS feeds with information about the meetings to generate respective ALOE-XML.

- Information about bookmarks represented in the license free *Netscape bookmark file format*[29] can be imported using a converter that generates the re-

---

[29]See `http://msdn.microsoft.com/en-us/library/aa753582(VS.85).aspx` for a description of the Netscape bookmark file format.

spective ALOE-XML. As most common Web browsers and social book-marking tools offer the possibility to export information about bookmarks in this format, it is possible to easily import resources from these sources in ALOE.

- Bibliographic information specified in BIBXML (containing a link to a bookmark or to a file on the user's device) can be imported using a parser that converts the BIBXML to ALOE-XML. BIBXML was chosen because it can be generated from any common kind of format for bibliographic information (e.g., BibTex or EndNote) using tools such as jabref[30].

- Within the C-LINK system (see Section 6.4.4), ALOE-XML was used to allow the conference organisers to provide information about all conference events that were then imported in ALOE.

- In MACE (see Section 6.4.1), more than 150.000 resources were added to ALOE by using ALOE-XML and the respective client.

Exporting Resources and Metadata

For enabling to allow the import of information provided within ALOE in as many other contexts as possible, ALOE offers to export resources and the respective metadata in the above mentioned *Netscape bookmark file format*. This format can be imported in any common Web browser, and also by several social bookmarking tools such as Delicious or Diigo. An export is offered for the following cases:

User resources: Users can always export their complete portfolio. This is offered on the page *My Resources*.

Group resources: For each group (open or closed), all group resources can be exported by clicking the respective link on the *group resources* pages.

Search results: Results of a search can also be exported, following the link that is offered on the respective result pages.

Feeds and Email Reports

ALOE provides a variety of feeds (see Figure 5.26 on Page 174) and email reports (see Figure 5.24 on Page 172) to automatically inform users about selected

---

[30]See `http://jabref.sourceforge.net`

activities and resources in the system without having to check the ALOE Web interface. Email reports are sent to the respective user's email address as specified in his/her user profile. Feeds are provided in the Atom Syndication Format (see [NS05]) that allows respective feeds to be read by any common client.

Two different types of reports are provided in ALOE: *activity reports* provide information about a variety of activities that users can perform in ALOE, whereas *resource reports* only provide information about specific resources that have been contributed.

In order to respect the users privacy in ALOE, feeds can only contain information that a user without login could also see. Therefore, and also to address users that do prefer email instead of feeds, email reports were introduced. They allow to check if the user who subscribed to the report is allowed to see non-public contents, therefore such reports may also contain non-public information (e.g., concerning resources that were contributed with group visibility).

Activity Reports   As the basis for the provision of activity reports, ALOE uses a sophisticated mechanism to store information about a variety of activities in a history table in the ALOE database. The captured activities concern all entities within the system: resources, users, groups, and collections. Every activity involves several information required for the generation of reports. For every activity that is tracked, the following information is stored in ALOE:

userId: The ALOE identifier of the user who carried out the activity

actionId: Every activity that can be tracked in ALOE is represented by a certain identifier.

date: The point in time when the activity is performed

objectId: The ALOE identifier of the entity that is primarily affected by the user's activity. In some cases, only one entity is concerned (e.g., *view a resource*), but in others (e.g., *share a resource to a group*) more entities can be involved.

objectType: To allow efficient processing of the history entries and to make sure that ids of different entities cannot be confused, not only the concerned entity's identifier, but also the according entity type is stored.

additionalMetadata: Some activities that can be performed involve additional metadata (e.g., *comment on a resource* also involves the content of the comment, and *update resource metadata* involves the respective metadata) that

needs to be stored. In ALOE, XML is used here for the encoding of multiple entries, and the entries themselves are URL-encoded.

visibility: As not every activity a user performs in ALOE is intended to be visible for every user in the system, a privacy level is associated with every entry in the history table. This usually corresponds to the visibility of the affected entity (i.e., *public*, *group* or *private*). Information about activities shall only be made available when the user is aware of this, so some activities (e.g., *login* or *view a resource*) are always associated with the visibility *private*. Furthermore, activities that are not associated with the respective user in the Web interface (e.g., *rate a resource* on a non-private resource leads to a rating that can be seen by others, without showing the name of the user that rated) will be stored with the visibility *public-anonymous* or *group-anonymous*.

groupName: In case the visibility of an activity is *group*, the identifier of the according group is stored here.

guiClient: Entries in the history table are triggered by the *ALOE Web Service*. The services cannot only be accessed by human users via the ALOE Web interface, but also by other clients (e.g., when automatically contributing or annotating resources on a large scale). To allow a meaningful analysis of the performed activities, the parameter *guiClient* is used and provides information about the kind of client. Technically, this is realised as follows: To access ALOE, users first have to get a session identifier. When using the services that return such an identifier, a client has to provide information about whether it is a GUI client or not. The session will then keep this information, thus allowing to provide the according information for all activities carried out in this session.

Using this schema allows to introduce the tracking of further activities without having to adapt the respective algorithms. It is only necessary to specify a text for the description of the activity.

ALOE provides the following kinds of activity reports:

Group activities: All public activities related to the group (e.g., *join* or *unsubscribe*) and group resources (e.g., when someone commented on a group resource) are shown here. The feed URL is provided on the respective group page.

Activities on own resources: This feed provides information about all activities on resources in a user's portfolio. The feed URL can be found on the page *My Resources* or on the own profile page of a user (*My Profile*).
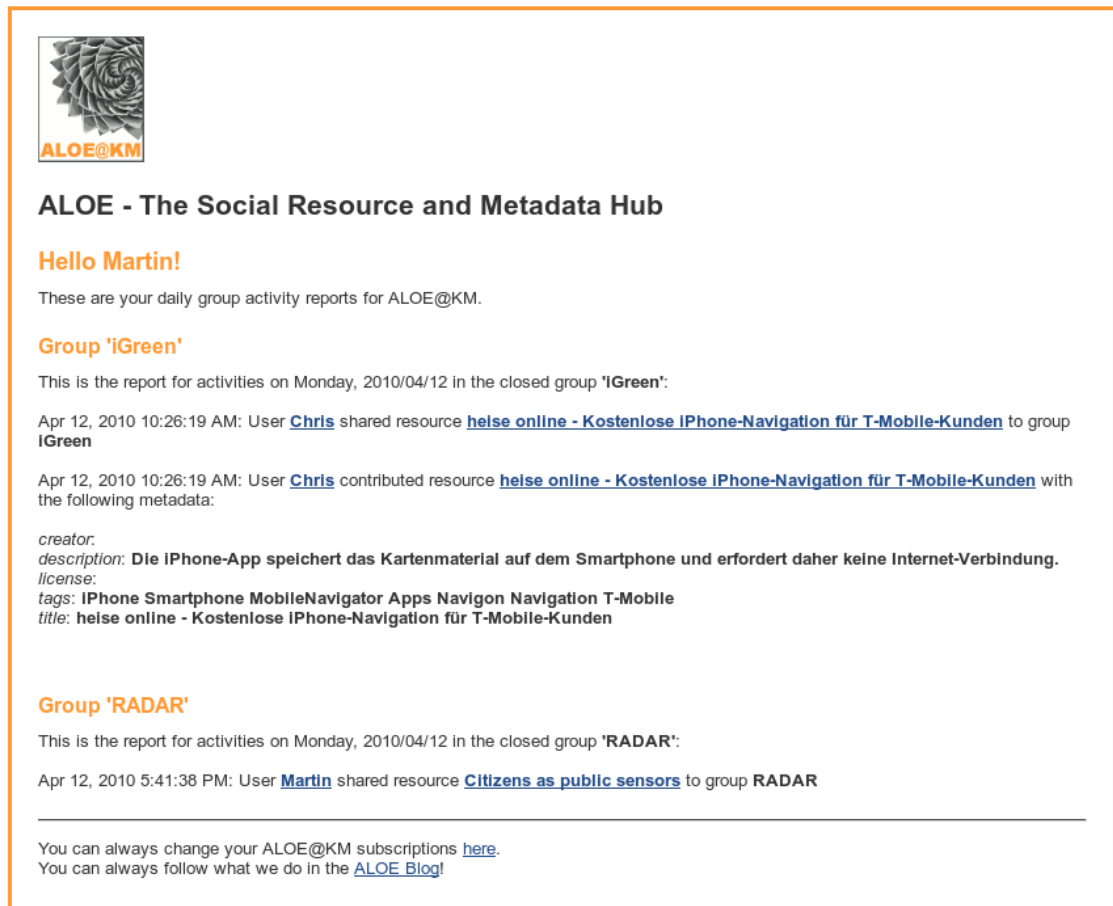
Figure 5.24: Excerpt of an email report for group activities

All activities: This feed delivers information about all public activities that have been tracked (Figure 5.26 shows an excerpt of such a feed).

Under the link *My Preferences*, each user can manage his/her subscriptions for all group related email reports (see Figure 5.25). Users here can choose whether to have none, daily or weekly reports.

Resource Reports   Resource reports provide information about certain resources and their associated metadata (see Figure 5.26):

Group resources: Resources shared to an open group. The feed URL is provided on the respective group page.

Figure 5.25: A sample overview of subscription settings in ALOE

User resources: All public resources that a user added to his/her portfolio. The feed link is provided on the respective user's profile page.

Search results: On the result pages of a resource search, a link to a feed is provided that provides information about recently contributed resources matching the current search criterion.

All resources: All resources that were contributed to ALOE

Parallel Contributions to External Services

Many users are already using other social sharing systems. ALOE does not aim to force users to switch from other systems to ALOE. Instead, diversity is

Figure 5.26: Excerpt on an ALOE feed for user resources

fostered, and users are supported in managing their information in the most convenient way. Thus, it is possible to contribute resources to ALOE, but simultaneously add the respective information to other systems with no extra effort. This is especially important if the ALOE system that is used is closed (i.e., not accessible for arbitrary users on the Web). Under the link *My Preferences*, users can manage their parallel contribution settings, i.e., they can select services and provide their credentials for these services (see Figure 5.27).

Once a user has chosen one ore more services, additional information as shown in Figure 5.17 is provided whenever a new resource is added. To have full control of what information is provided (e.g., to distinguish between internal and external resources), users can always unselect chosen services here. When a parallel contribution attempt fails, the user will be notified accordingly.

Figure 5.27: Sample parallel upload settings in ALOE

Amongst others, the following services are supported: Delicious[31], Diigo[32], and BibSonomy[33]. Furthermore, a connection to the microblogging service Twitter[34] can be established for publicly available ALOE instances. In this case, a status update with the title of the contributed resource as well as the corresponding link is performed.

Embedded Metadata in ALOE pages

ALOE offers embedded metadata, i.e., semantic markup with information about entities that is embedded in the XHTML pages generated by ALOE. Currently, two formats are supported:

Embedded metadata is offered as microformats and RDFa

RDFa (Resource Description Framework in attributes) is a W3C Recommendation that allows to embed RDF triples within XHTML documents [AB08]. It can reuse data models created for RDF and can be re(used) in a variety of Semantic Web based applications.

Microformats[35] is a format that was mainly developed in the context of the blog aggregation platform Technorati[36] [Dim05]. It reuses existing XHTML and

---

[31]See `http://delicious.com`
[32]See `http://www.diigo.com`
[33]See `http://www.bibsonomy.org`
[34]See `http://twitter.com`
[36]See `http://www.technorati.com`

HTML tags and allows to provide a variety of information such as contact data and calendar events. Information provided as microformats can be used by several tools. For example, the Firefox plugin *Operator*[37] allows to integrate such information in different Web and desktop applications.

Google's support of microformats as well as RDFa announced in 2009 also provides evidence about the growing importance of such embedded metadata.footnoteSee `http://radar.oreilly.com/2009/05/google-announces-support-for-m.html`

```
-<rdf:RDF>
  -<rdf:Description rdf:about="http://aloe-project.de/AloeView/action/resourceDetailed?resourceId=IzwX87Z">
      <xhv:icon rdf:resource="http://aloe-project.de/AloeView/images/favicon.ico"/>
  </rdf:Description>
  -<rdf:Description rdf:about="http://aloe-project.de/">
      <dc:subject xml:lang="en">martin_memmel</dc:subject>
      <dc:subject xml:lang="en">homepage</dc:subject>
      <dc:subject xml:lang="en">aloe</dc:subject>
      <dc:subject xml:lang="en">prototype</dc:subject>
      <dc:title xml:lang="en">ALOE Homepage</dc:title>
      <dc:creator xml:lang="en">Martin Memmel</dc:creator>
      <dc:description xml:lang="en">The Homepage of the ALOE project</dc:description>
  </rdf:Description>
</rdf:RDF>
```

Figure 5.28: Embedded RDFa as provided for a resource in ALOE

Embedded metadata in ALOE is offered for resource detail pages (see Figure 5.28), whenever a resource is shown in a condensed overview mode, and for user profile pages.

Widgets

ALOE offers Widgets enabling the provision of ALOE information and interaction possibilities in different contexts. For example, by using the code snippet shown in Figure 5.29, a widget as presented in Figure 5.30 can easily be embedded in existing HTML pages: It provides the possibility to add the respective page to one's portfolio, and it also shows the existing information in ALOE about this page.

OAI-PMH

To allow the harvesting of publicly available metadata about ALOE resources, ALOE offers an *OAI target* for selected instances (e.g., for the MACE scenario

---

[37]See `https://addons.mozilla.org/en-US/firefox/addon/4106`

```
<script type="text/javascript">
    // <![CDATA[
    var aloeUrl = 'http://aloe-project.de';
    var borderWidth = '2px';
    var borderColor = '#96DCFF';
    document.writeln( '<iframe width="190" height="120" frameborder="0" scrolling="no"
        marginheight="0" marginwidth="0" src="' + aloeUrl
        + '/AloeView/ajax/getResourceMetadataAsHtmlPage?bookmarkUri='
        + encodeURIComponent( location.href ) + '&amp;bookmarkTitle='
        + encodeURIComponent(document.title) + '&amp;lang=en'
        + '&amp;borderWidth=' + encodeURIComponent(borderWidth)
        + '&amp;borderColor=' + encodeURIComponent(borderColor) + '" ></iframe>' );
    // ]]>
</script>
```

Figure 5.29: A sample code snippet that can be embedded into an HTML page to show a widget as presented in Figure 5.30

Figure 5.30: An example of an ALOE Widget embedded into the sidebar of a Wiki

as described in Section 6.4.1). An OAI target provides metadata following OAI-PMH[38], a widely used protocol for metadata harvesting developed by the Open Archives initiative[39] [SNLW04]. ALOE supports the OAI-PMH verbs *Identify*, *GetRecord*, and *ListRecords*.

---

[38]See `http://www.openarchives.org/pmh`
[39]See `http://www.openarchives.org`

5.4 System Design

The main layers of the ALOE system architecture (*ALOE Backend Services*, *ALOE Data Storage*, and *ALOE Servlets*) have already been introduced in Section 5.1.3. They comprise the following components:

AloeFeeds: This component provides feeds in ALOE, based on the user history table in the ALOE database.

AloeInfoMail: Email reports as well as the optional ALOE newsletter are offered by this component. It is part of the *ALOE View* and invoked by respective cronjobs. Administrators can also send test mails via the ALOE frontend.

AloeMultimedia: The *ALOE Multimedia* servlet is responsible for the provision of all resources stored in the ALOE database (e.g., buddy icons or file resources that were uploaded).

AloeThumbnailService: This service is requested when preview images shall be generated. Depending on the resource type, it uses a variety of tools on operation system level (e.g., an OpenOffice Server, ImageMagick, and Firefox) to generate the respective images. For process control, access to tools, pre- and postprocessing steps and further management issues, a configurable service has been developed.

AloeView: The *ALOE View* realises the Web Interface as already presented.

AloeWebService: The *ALOE Web Service* is responsible for accessing and storing ALOE data and offers access to more than 200 methods (see Appendix F). It can be considered as the backbone of any ALOE system.

ExtractionWebService: The Tika-based metadata extraction is provided through this service. Optionally, also the full text of a resource is returned to allow further processing.

Except for the *ALOE Thumbnail Service*, all components are implemented in Java 1.6 and deployed as servlets within the Apache Tomcat Servlet Container[40]. Whenever possible, standard technologies like these where used for all components to allow the deployment of ALOE in any common infrastructure.

Figure 5.1 shows the different components and the way they interact with each other within ALOE. The main components – the *ALOE Web Service*, *ALOE View*, *ALOE Feeds*, and *ALOE Multimedia*– are explained in more detail now.

---

[40]See `http://tomcat.apache.org`

5.4.1 Structure of the ALOE Components

In the development of the ALOE components, several design decisions were made with respect to the given requirements.

Internationalisation and Configurable Messages

In particular in the *ALOE View* many labels, descriptions, and other messages are used in the user interface. In order to support the white label characteristics of ALOE, these messages should easily be adaptable, e.g., to allow for the usage of scenario-specific texts or even for internationalisation (i.e., for allowing to switch to a different language in the user interface). Therefore, *message resources* are used that are mainly intended to support internationalisation. Using this mechanism, all labels, descriptions, etc. can be specified in central files (the *message resource bundles*) that can be accessed in any JSP file or Java class. Among others, different default values offered in forms (e.g., *interests* and *affiliation* on the user profile page) can also be defined here. As of November 2014, the *ALOE View* used more than 1400 different labels.

Configuration Properties

To allow for an easy and comfortable adaptation of ALOE components, several properties of these components can easily be configured with respective service configuration files. This eases a convenient deployment of new ALOE instances with different specifics. A sample of such a file for the *ALOE View* is shown in Figure 5.31.

5.4.2 The ALOE Web Service

For the *ALOE Web Service*, REST is used as a standard and platform-independent protocol. It is realised using JSON-Delight[41], a middleware developed in the Knowledge Management Department of DFKI. Information is transferred using JSON which can be consumed by any common programming language. The *ALOE Web Service* is implemented in Java and deployed as a Tomcat servlet.

The *ALOE Web Service* comprises more than 200 methods (see Appendix F for an overview of selected methods). For several instances, a respective online Web Service documentation that can automatically be generated is offered for developers.

---

[41]See `/http://delight.opendfki.de` for a description of delight, which originally was developed as an XML-RPC addon, but now is also supports JSON-RPC

```
de.dfki.km.aloe.aloeview.webService=http://localhost:8080/axis2/services/AloeWebService
de.dfki.km.aloe.aloeview.secureWebService=https://localhost:8443/axis2/services/AloeWebService
de.dfki.km.aloe.aloeview.multimediaService=http://aloe-project.de/AloeMultimediaServlet
de.dfki.km.aloe.aloeview.feedService=http://aloe-project.de/AloeFeeds
de.dfki.km.aloe.aloeview.apertureWebService=http://localhost:8080/axis2/services/ApertureWebService/

# Defines which parallelUploadServices will be offered
# Important: The respective servicess have to exists in the DB table 'ParallelUploadServices'
de.dfki.km.aloe.aloeview.parallelUploadServicesOffered=Delicious,Diigo,Twitter

# Defines the visibility of the instance. Currently, everything else
# than 'public' will be considered as an instance that is not
# publicly visible. We may define some other distinctions whenever we need them.
de.dfki.km.aloe.aloeview.instanceVisibility=public

# which default language to use
de.dfki.km.aloe.aloeview.defaultLanguage=en

# use an LDAP server is the following property is not empty
de.dfki.km.aloe.aloeview.ldapServer=

# the email address where to send problem reports and the like
de.dfki.km.aloe.aloeview.webmasterEmailAddress=aloe@dfki.uni-kl.de

# the email address to use in the From: field of emails sent by the system
de.dfki.km.aloe.aloeview.eMailSender = aloe-noreply@dfki.uni-kl.de

# SMTP server to be used to send emails
de.dfki.km.aloe.aloeview.smtpServer=serv-4100.kl.dfki.de
```

Figure 5.31: A sample AloeView configuration file

As underlying database, the open source database MySQL 5.5[42] was chosen. The use of facades that separate the business logic from the database access allows for a convenient exchange of the used database. The relational database can optionally be synced with a Lucene index in which the full text of resources as well as the resource metadata is stored and synced using an observer pattern approach.

5.4.3 The ALOE View

The *ALOE View* is also implemented in Java and deployed as a Tomcat servlet. It is implemented according to the FrontController Pattern[43], i.e., a front controller manages all Web requests as a centralised entry point. For example, it ensures that a session identifier is valid or checks for information in ALOE cookies.

The single pages in the *ALOE View* are realised using JavaServer Pages (JSP)[44] that allow for the dynamic generation of HTML and XML documents of a Web

---

[42]See `http://www.mysql.com`
[43]See `http://java.sun.com/blueprints/corej2eepatterns/Patterns/FrontController.html`
[44]See `http://java.sun.com/products/jsp`

server. Using JSP, new pages in ALOE can quickly be created, and the usage of JSP actions (so called *tags*) allows to reuse a variety of existing interface elements such as the condensed view of a resource or a pager.

5.4.4 Aloe Feeds and ALOE InfoMail

In order to create reports as described in Section 5.3.6, *ALOE Feeds* and *ALOE Infomail* access report information via the *ALOE Web Service*. The code that is needed by both components (e.g., to generate text for a given action) is defined in an extra package *ALOE Utilities* that contains methods used in different ALOE components (including the *ALOE View* and the *ALOE Web Service*).

ALOE Feeds is realised as a Tomcat Servlet, also following the Front Controller pattern. Like in the *ALOE View*, internationalisation is possible, several configuration properties can be provided, and a DB facade is used.

ALOE Infomail is a Java application defined within *ALOE View* and is usually started as a cronjob on the respective ALOE system.

5.4.5 ALOE Multimedia

The *ALOE Multimedia* component accesses the *ALOE Web Service* to provide uploaded file resources, resource previews, and thumbnails for users, groups, and collections. In case no preview image could be generated or no thumbnail was uploaded, *ALOE Multimedia* automatically provides respective default thumbnails. For resources, the respective mime type is checked to provide different default thumbnails (e.g., an image with a loudspeaker if it was an audio file).

5.5 ALOE Adaptations and Enhancements

As already introduced, ALOE is a generic and adaptable infrastructure. To organise and manage the ALOE root system and the different variations of the system, Apache Subversion[45] is used as software versioning and revision control system. It allows for storing and distributing the code of the different ALOE components. When a new version of ALOE is needed, a new branch of the *ALOE View* is created, and the required adaptations and enhancements will be executed in this branch. Making use of a merge tool written in Perl, all changes affecting the root system can also be executed in the created branches. It is also

---

[45]See `http://subversion.apache.org`

possible to create branches of branches – for ALOE, this was used to create further versions of the RADAR system.

Enhancements and adaptations also might concern other components than *ALOE View*, but they are carried out without creating further branches:

AloeFeeds and AloeInfoMail: Reports that are only needed by specific ALOE branches will simply be added and are potentially available for all ALOE versions. Since the links for these reports will only be offered in the respective instances, there is no need to further separate them.

AloeMultimedia: When new default icons shall be used in a branch, a new subfolder containing these images is created, and the configuration properties can be set to this folder.

AloeWebService: If new methods are needed, a respective package will be created for them, and a new Web Service class will be deployed on top level that offers exactly these functionalities. The means to access this service are only provided in the respective *ALOE View* branch, and it is also possible to create a specific Web Service documentation.

Aloe Data Storage: In case new tables need to be created, specific SQL files are provided that have to be executed after the standard DB installation.

5.6 Research Hypothesis

ALOE was developed with the aim to implement a generic Social Resource and Metadata Hub. It is necessary to prove that this was successfully realised in order to use ALOE instances as a means to investigate the validity of the hypotheses presented so far. Thus, the following hypothesis is now introduced:

H3: ALOE technically realises a generic Social Resource and Metadata Hub

- Hypothesis H3.1: ALOE meets the technical requirements of a Social Resource and Metadata Hub as defined in Section 4.1.2

- Hypothesis H3.2: ALOE is a generic infrastructure that allows being adapted and extended for usage in arbitrary scenarios

5.7 Related Approaches

Several platforms exist that offer social media functionalities, and some of them also allow to introduce them in existing environments. Yet, most of them are limited in terms of the following characteristics:

Generic support: Mostly, only isolated use cases with solutions that focus on selected scenarios (e.g., support only for specific resource types such as images or videos) and domains are focused.

Creation of instances: Most Web-based platforms (e.g., Delicious) can not be instantiated at all, which not only means that it is not possible to adapt them in any way but also that an integration into existing environments is only possible to a small extent, if at all.

Adaptability: Most platforms cannot be adapted to the specific needs of a scenario, or they only offer very limited possibilities that usually only concern few aspects regarding the look and feel of the user interface.

Integrability: Frontend and backend technologies are mostly not separated in a way that allows to also use and integrate them as a social backbone in existing environments. If APIs are offered, they most of the time only include a small extent of the available functionalities. Furthermore, the incorporation of existing information is usually very complex or not even possible at all.

Access control: The ability to make use of different visibility levels and thus to allow a controlled sharing of information is often missing. If at all, most systems only offer a distinction between private and public contributions.

What is missing is a comprehensive approach and framework allowing to exploit the potentials of social media also in existing environments with support for potentially arbitrary kinds of contributions.

5.8 Summary

The ALOE system offers a wide range of social media functionalities via its Web Service interface or the user interface that can be accessed with any common Web browser. It can be used as a stand-alone system or just as a social backbone, and it offers a wide range of means to exchange information. The system was already deployed, used, improved, and enhanced in several projects and scenarios, where it has proven its maturity (see Chapter 6). Before these scenarios are presented, the following table sums up that ALOE meets the (technical) requirements identified in the last chapter, by iterating step-by-step through the system design checklist presented in Section 4.8.

Demand	ALOE
Resource Types	
Incorporation of any type of resource as defined in Section 2.1	ALOE only restricts the contribution of digital resources that exceed a certain file size.
Realisation of a *repository*	ALOE allows the upload of digital resources.
Realisation of a *referatory*	ALOE offers social bookmarking facilities.
Resource Metadata	
Support for social metadata	Users in ALOE can tag, rate, comment, create portfolios, collections, etc.
Ability to make use of potentially any metadata existing in the environment where the hub is introduced, and consequently offering the possibility to contribute a variety of different metadata formats	Metadata can easily be imported using the Web Service API or ALOE-XML. Arbitrary metadata sets can be associated with each resource, and ALOE allows to define arbitrary relations to freely chosen structures and respective classes.
Provision of information that allows users to get an understanding of the respective creator or contributor	ALOE provides information about respective user profiles and all other contributions of this user through the Web Service API and the user interface.

Demand	ALOE
Definition of a core set of metadata elements to enable interoperability and to efficiently compute and process information	A respective set is provided, and interoperability will be been proven in a variety of scenarios such as the usage of ALOE in MACE.
Support for scenario specific metadata extensions	Specific resource types with extended metadata can be introduced in ALOE.
Following principles supporting interoperability	The ALOE metadata is extensible and can also be refined, machine-processability and modularity is ensured by the accessibility of the metadata via the *ALOE Web Service*.
Drawing upon standards concerning metadata elements and representation formats to ease information exchange	Dublin Core was used as the basis of the resource metadata in ALOE, and the metadata is provided by the REST-based *ALOE Web Service* using JSON as a standard exchange format.
User Metadata	
Basic profile information such as a user's name and address	Respective metadata elements are offered
Information about a user's interests	Respective metadata elements are offered
Preference information of a user (including system specific settings)	Respective metadata elements are offered (e.g., for language settings and email notifications)
Information about a user's activities (captured by observing the user)	Respective metadata elements are offered
Information about a user's contributions	Respective metadata elements are offered for all possible types of contribution
Information that allow to connect to other components in the scenario where the Social Resource and Metadata Hub is deployed	Respective metadata elements are offered (e.g., for LDAP authentication)
Metadata allowing to specify who can access data about the user	Respective metadata elements are offered

Demand	ALOE
Functionalities	
User management, networking facilities, and resource organisation: • Registration and login • Profile management • Contacts, messaging • Privacy and control of data • Resource organisation	Facilities for registration, login, profile management, messaging and contact management are offered. To ensure privacy and control of data, different visibility modes for user data and contributions are offered, and information is only provided to others when the user is aware of it. Several means such as collections, groups, and tags allow to organise resources in ALOE.
Contributing digital resources and metadata: • Upload of digital resources • Referring to existing resources (*bookmarking*) • Means to contribute different kinds of resource metadata, with a focus on social metadata	Respective functionalities are offered
Editing digital resources and metadata: • Editing metadata • Deleting digital resources and metadata	Respective functionalities are provided

Demand	ALOE
Retrieval of digital resources and metadata: • Retrieval of digital resources • Retrieval of complete metadata instances as well as single elements for each resource	Respective functionalities are provided
Means to provide resource subsets: • Search facilities • Means to filter contents according to different criteria • Different sorting and ranking facilities • Means to provide subsets based on the resource organisation facilities offered	A variety of search functionalities and ranking mechanisms is offered for all entities in ALOE, and resources can also be filtered according to different criteria. Resources from specific groups, collections or users can be provided.
Navigation facilities: • Means that allow users to navigate within the available information using potentially any available relation, especially focusing on social browsing	ALOE offers, e.g., immediate access to a contributor's portfolio and to superordinate collections and groups.
Information exchange and notifications: • Exchange facilities (import and export) for potentially any kind of available information • Notification means for events in the system	Information can be imported and exported using several services, and ALOE offers means to export information via ATOM feeds and an OAI target, as well as the export of resource information as Netscape bookmark file. As notification means, ALOE offers info mails and ATOM feeds for a variety of information.

Demand	ALOE
Interfaces	
Low technical and conceptual barriers for system usage	Usage of standard technologies to provide access to data and functionalities; definition of a simple metadata element set.
Interfaces – Access by Systems	
• Focusing on interoperability • Offering access to potentially any data and functionality (of course regarding privacy aspects) • Notification means (e.g., feeds)	Access to functionalities is offered via a REST Web Services, information is pushed via email and ATOM feeds.
Interfaces – Access by Users	
Provision of information about available functionalities, and how they can be accessed	ALOE offers a WSDL description as well as an extensive online documentation of services (see Appendix F).
For each resource, information about which metadata is available and how it was created	Respective functionalities are provided through the user interface and via the Web Service API.
Information about schemas and vocabularies used	ALOE offers an extensive online documentation that also contains information about schemas and vocabularies.
Interfaces that encourage participation by end users	Proven through usage statistics and evaluation results (see Chapter 6)
Offering the possibility to use functionalities of the hub in other contexts and applications	ALOE offers widgets to show information about resources, and to contribute new information.
Adaptability of user interfaces to allow to address specific needs of a scenario	ALOE has been adapted for a variety of scenarios (e.g., Mindpool, Web of Models).

Table 5.6: Demands and corresponding information about ALOE

CHAPTER 6
Case Studies and Evaluation

6.1 Goals

The purpose of this chapter is to investigate the validity of the hypotheses presented in Chapter 4 and 5 by examining different ALOE instances that have been deployed since 2008. Where possible, the evaluation seeks to quantify any findings. For convenience, here the hypotheses are listed again.

H1: A Social Resource and Metadata Hub can successfully tackle typical problems when accessing digital resources

- Hypothesis H1.1: A Social Resource and Metadata Hub can successfully tackle the problem of *abundance of available digital resources*

- Hypothesis H1.2: A Social Resource and Metadata Hub can successfully tackle the problem of *heterogeneous access to digital resources*

- Hypothesis H1.3: A Social Resource and Metadata Hub can successfully tackle the problem of *missing metadata for digital resources*

H2: A Social Resource and Metadata Hub allows for exploiting social web phenomenons

- Hypothesis H2.1: A Social Resource and Metadata Hub allows for exploiting *interaction possibilities*

- Hypothesis H2.2: A Social Resource and Metadata Hub allows for exploiting the *Long Tail* phenomenon

- Hypothesis H2.3: A Social Resource and Metadata Hub allows for exploiting *Collective Intelligence*

- Hypothesis H2.4: A Social Resource and Metadata Hub allows for exploiting *Crowdsourcing*

H3: ALOE technically realises a generic Social Resource and Metadata Hub

- Hypothesis H3.1: ALOE meets the technical requirements of a Social Resource and Metadata Hub as defined in Section 4.1.2

- Hypothesis H3.2: ALOE is a generic infrastructure that allows being adapted and extended for usage in arbitrary scenarios

First, relevant challenges are discussed, and the evaluation setup and process are explained. Then, the ALOE instances and scenarios relevant for this chapter are introduced. Finally, a detailed investigation of each of the hypotheses is presented.

6.2 Challenges

The examination of the quality of the presented approach is a multidimensional and complex task. The following aspects have to be considered here and will be discussed briefly:

- Existence of functionalities

- Technical quality

- Possibility to exploit potentials

Existence of functionalities   The identified requirements and demands for the presented approach resulted in the system design checklist presented in Section 4.8. As shown in Chapter 5, these functionalities exist in ALOE. Consequently, it is guaranteed that the expected benefits can at least potentially be provided.

Technical quality   The mere existence of a functionality does not necessarily prove that it technically also works properly. This especially includes aspects such as scalability, stability, and performance. These quality dimensions can be investigated with different tools that are presented in Section 6.3. Furthermore, conclusions about the technical quality can also be drawn from data collected about user contributions as well as from users' satisfaction expressed in surveys.

Possibility to exploit potentials   To find out whether the targeted benefits can be realised with the presented approach is the most challenging question. Of course the existence of required functionalities with an adequate technical quality is the most important prerequisite considered in this work. However, the decision of whether an information system aiming to support users in accessing digital resources works properly can ultimately be answered best by the users themselves. Furthermore, several of the presented challenges require the involvement of users and their contributions. It is thus only possible to examine these challenges if enough users were involved. Yet, this means that certain preconditions need to be fulfilled that cannot be enforced by means of a technical setup. Whether users participate or not depends on a variety of factors that depend on the focus of a system, the context in which the application is deployed, the users that are involved, means to encourage participation, etc. Such questions are extremely important for the use of any information system, but they are beyond the scope of this thesis. For the scenarios in which ALOE was used, this could sometimes be influenced only to a small extent. The evaluation focus was thus put on the instances where enough users participated. If the effects can be observed there, it is obvious that they are conceptually possible in any ALOE instance.

6.3 Evaluation Setup and Process

To examine if ALOE provides the benefits expressed in the hypotheses, surveys as well as data collected about user activities and contributions were used. Different approaches were also used in order to provide evidence regarding the technical quality of the ALOE infrastructure. Load tests were conducted for the different components in the MACE use case (see 6.4.1), and cruise control components constantly check the availability of ALOE instances.

   The means that were used to collect quantitative as well as qualitative information are now presented.

Collecting Quantitative Information

In the presented evaluation, three different sources for quantitative information were used: The *ALOE Statistics Dashboard*, the *ALOE User Histories*, and *Web Analytics Tools*. For the most important ALOE instances presented in this work, detailed information derived from these sources is assembled in Appendix D.

ALOE Statistics Dashboard Every ALOE instance offers a dashboard functionality (accessible for administrators only) that provides statistical information about a variety of aspects in ALOE. This includes information such as the number of users, resources, groups, tags, etc. Using this dashboard, it is also possible to specify the following parameters:

Time range: A time range can be specified to obtain information about activities that took place in a specific time period.

Excluded Users: The contributions of users specified here will be excluded from the provided statistics. This is of special importance as it allows to exclude contributions of the ALOE team when aiming to prove that ALOE successfully encouraged users to participate.

An example for a statistics dashboard is shown in Figure 6.1. The dashboard also shows specific information about what happened in the system in the last week and last month.

ALOE User Histories As already introduced in Section 5.3.6, ALOE uses a sophisticated mechanism to store information about a variety of activities in a user histories table in the ALOE database. This table is not only the basis for the different activity reports within ALOE, it also is a very important source for collecting quantitative information. For the scenarios presented, it was used to calculate the following information providing insights into whether a specific instance encouraged users to participate:

- the number of all tracked activities

- the number of all tracked activities carried out with a GUI client (to make sure that the actions were really conducted by humans and not by a script)

- the number of users with activities that were tracked

- the number of different objects that were affected by tracked activities

Furthermore, information about the ten most tracked activities is calculated to get an insight about the specific focus of an instance (e.g., if a lot of searches were conducted).
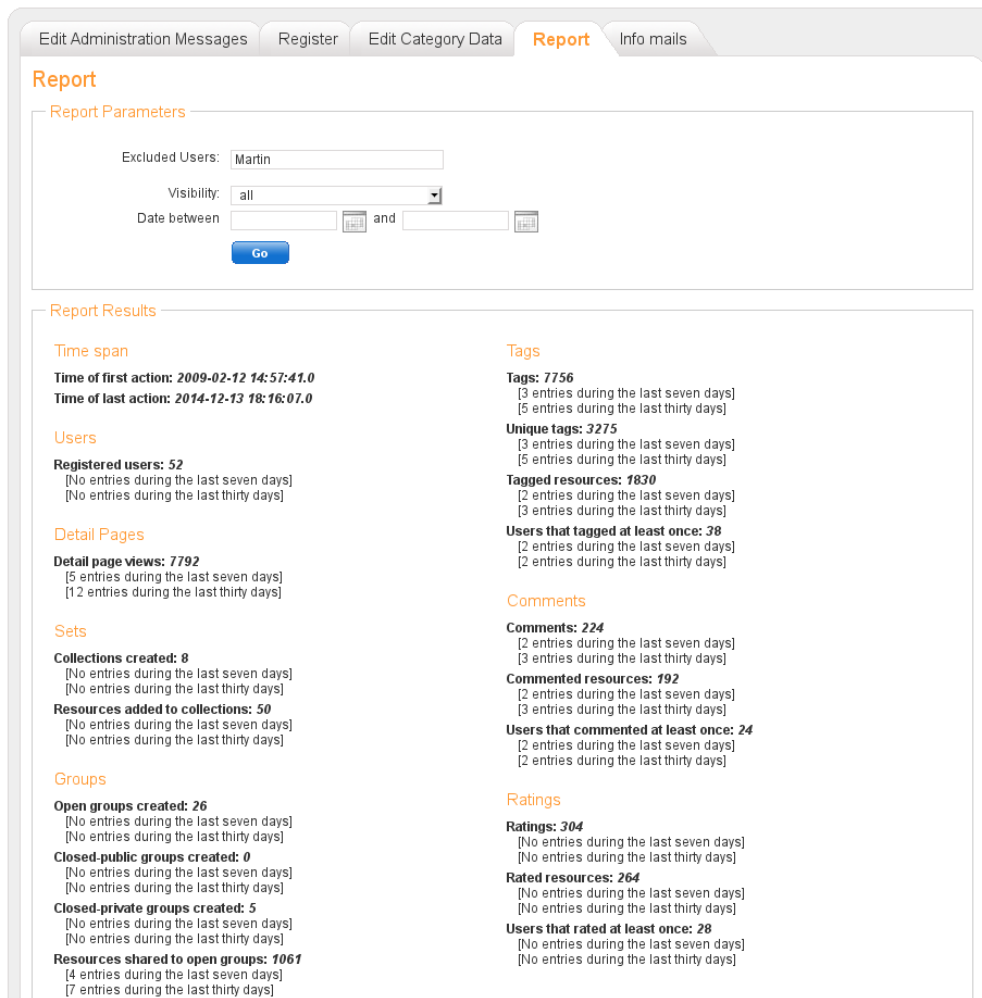
Figure 6.1: Excerpt of an ALOE Dashboard with statistical information

Web Analytics Tools Most ALOE instances were using Google Analytics[1] as an on-site Web analytics tool that measures visitor behaviour. The data collected using this technology is a further source for estimating whether ALOE instances attracted users. This especially concerns visits of anonymous users, as these are not tracked in the user histories table. Furthermore, some instances were going live before the tracking mechanisms in ALOE were established.

---

[1]See `http://www.google.com/analytics`

Gathering Qualitative Information

To gather qualitative information for ALOE@KM and Mindpool, two online questionnaires were prepared using the LimeSurvey application[2]. First, only users of the platform were informed in a respective ALOE newsletter for each system. Both newsletters were sent out on December 9th, 2014. For ALOE@KM, an email with information about the questionnaire was sent out to all members of the Knowledge Management group on December 12th, 2014, for Mindpool a further newsletter with a reminder was sent out on December 15th, 2014. The questionnaires that were used can be found in Appendix E.

Until December 17th, 2014, the questionnaire for Mindpool was filled out by 14 participants, whereas 13 users took part in the survey for ALOE@KM. To make sure that feedback regarding questions about system features and contents are only answered by people that know the respective system, answers of users that stated that they never use the system were not taken into account (this was the case for three users in ALOE@KM).

6.4 ALOE Instances and Scenarios

There are plenty of scenarios where ALOE instances were deployed and used as a stand-alone system or as a backbone infrastructure. The most important instances for investigating the presented hypotheses will now be introduced. First, the usage of ALOE as a social backbone within the European project MACE is described. This use case provides several results concerning the system infrastructure and the usage of social resources and social metadata. Second, the ALOE-based Mindpool infrastructure that realises an interactive knowledge sharing platform for all DFKI employees at all sites is presented. Mindpool will serve as an example of an adapted ALOE instance that can integrate heterogeneous information from various sources, and where a lot of user interactions led to a large number of artefacts created within an ALOE instance. Then, the ALOE instance used within the Knowledge Management Department of DFKI will be introduced as an example for an ALOE system that was deployed as-is in a real life scenario. Finally, several other use cases are briefly presented.

---

[2]LimeSurvey is a free and open source online survey application written in PHP. It allows users to create and publish online questionnaires, and to collect the responses. For further information see `http://www.limesurvey.org`.

6.4.1 MACE

MACE (Metadata for Architectural Contents in Europe) is a European Initiative aimed at improving architectural education, by integrating and connecting vast amounts of content from diverse repositories, including past European projects existing architectural design communities [SVC+07, WMS10]. It was co-funded by the EU eContentPlus program from September 2006 until October 2009. Work in MACE is continued within the MACE foundation.

Relevant learning material for the domain of architecture is available in a variety of repositories that are not related with each other. For example, educational material is scattered over many repositories like the Dynamo repository[3] providing information about architectural projects or ICONDA[4] providing access to legislative documents important to building construction and design.

Hence, students and teachers have to know about the various repositories and their specifics, have to access them separately and, in general, are not able to easily find and retrieve appropriate learning resources. Furthermore, the repositories use different terminologies and classifications to describe and classify their resources, and support for multilinguality is only partially provided. Thus, accessing information is difficult and time-consuming for users. In general, repositories have not agreed to use common standards to jointly describe resources. Each community, and sometimes each repository uses its own standard to describe resources. Furthermore, even when suitable learning resources are found, they often do not provide a link to further information so that the hunt for additional information begins all over again.

Within the MACE portal that is online and publicly accessible[5], searching through and finding appropriate learning resources from a variety of sources is enabled in a discovery oriented way.

By automatically and manually linking related architecture-related learning resources of various non-related repositories with each other, relations among them are established. These relations enable simple and unified access to the architectural learning resources scattered throughout repositories world-wide. Thus, users are able to discover new learning resources that serve as additional sources of inspiration and support reaching desired learning goals (see [WMG+12] for more details about the impact of the MACE benefits on learning).

For finding appropriate learning material, simple keyword search and result link presentation are not sufficient for architecturally motivated information

---

[3]See `http://dynamo.asro.kuleuven.be/dynamovi`
[4]See `http://www.iconda.org`
[5]See `http://www.mace-project.eu`

search. Instead, students need simple and personalised access to vast amounts
of architectural information using advanced, visually based discovery oriented
mechanisms for access to the learning material [Mar06]. Examples might be
image and location based search and classification browsing.

Figure 6.2: Filtered search interface of the MACE system – searching for learning
resources related to the architect "Antonio Gaudi"

MACE offers a variety of searching and browsing facilities that rely on the
metadata associated with the learning resources. The *filtered search* interface
of the MACE system is shown in Figure 6.2 as an example of how a user can
search for resources. The user is able to qualify the search with several addi-
tional facets that describe the context of the learning resource(s) in question: the
repositories in which to search, the language of the results, the resource media
type, the resource classification, and the associated competency. When choosing

a respective facet, the interface is dynamically updated by providing the numbers of results for each facet that match the selected criteria. The results of a search are shown below the context filters. A short summary of each result provides further selection criteria, e.g., the resource title, a short description, and the repository.

MACE not only focuses on architectural content, but also on the users that are interested in the provided contents. Therefore, one of the key aims was to attract users, and to encourage these users to contribute information. Furthermore, the building of communities around MACE and MACE contents is a key for the MACE sustainability efforts. Therefore, the provision of social media functionalities that allow end users to contribute had to be realised. The MACE project consortium decided to choose ALOE as a platform to introduce these functionalities, and DFKI became an official partner of the consortium in spring 2008.

Targeted problems   MACE mainly addresses these problems of abundance, heterogeneity, and missing interaction possibilities[6].

Abundance: More than 175,000 existing resources from other repositories have been integrated into MACE.

Heterogeneity: The integrated resources were located in different repositories using different metadata schemas and access mechanisms.

Missing interaction possibilities: Most of the integrated repositories only offered few or no interaction possibilities for end users.

MACE addresses the problems of abundance, heterogeneity, and missing metadata

Integration of ALOE into the MACE Infrastructure

The MACE system builds on a distributed service oriented architecture with a three-tier structure. The front-end with its graphical user interfaces and widgets forms the client tier. The business logic that is responsible for the provision of functionalities is organised in the application-server tier while the metadata stores form the data-server or back-end tier [WMK+09].[7]

Through the integration of ALOE into the MACE infrastructure, a variety of social media technologies are provided. Users can add new contents to MACE, they can maintain personal resource portfolios, they can contribute information

---

[6]Some of these phenomenons and respective strategies in MACE are also described in [WMN+11].

[7]See [SVC+07] and [PTdJ+07] for a complete overview of the overall system architecture.

about resources (e.g., tags, comments, and ratings), and they can search within this information. Furthermore, it is possible to maintain contact lists and to send messages to other users.
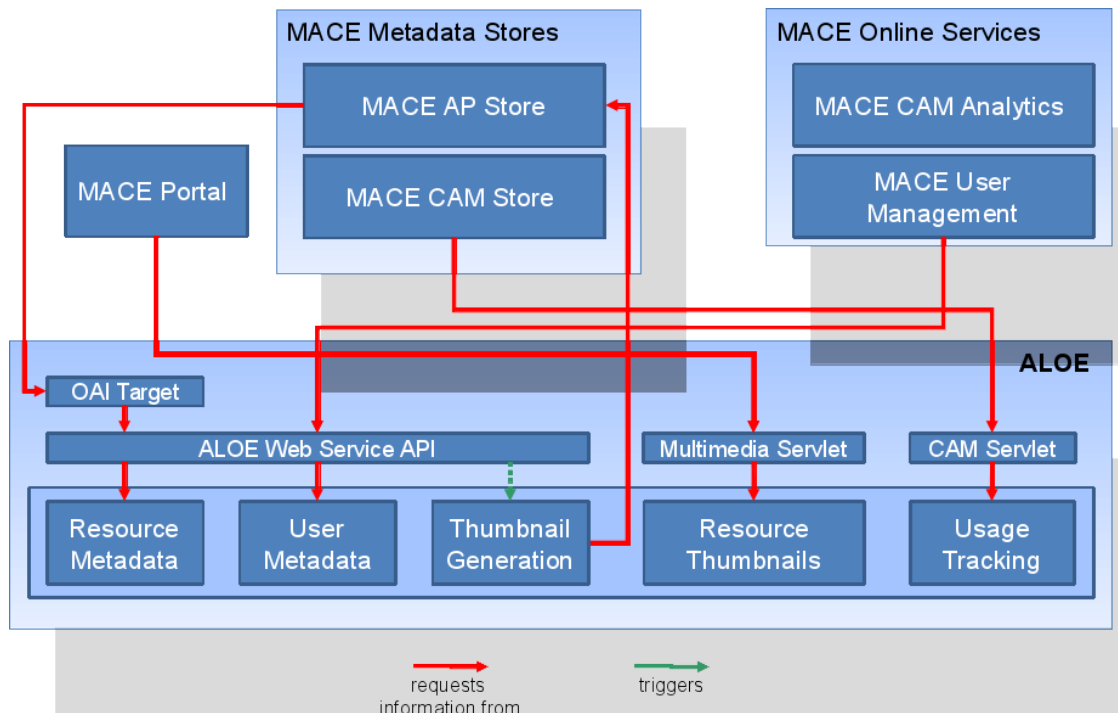
Figure 6.3: ALOE services and their integration into the MACE infrastructure

Figure 6.3 provides an overview of the ALOE services and their integration into the MACE infrastructure. The ALOE API is accessed by the MACE portal and the MACE user management for the following main functionalities:

User management and authentication: Basic user profile data such as login name, email address, and password is stored centrally in a database at Fraunhofer FIT and made accessible within the MACE infrastructure through a Web Service. In addition to this minimal profile information, more user data can be stored in the ALOE instance set up for MACE, e.g., a profile picture and an affiliation.

When a user registers, an internal MACE user id is generated, and the ALOE registration process is triggered. If the registration process in ALOE was successful, the ALOE user id of the newly generated user account will be returned and stored in the MACE database. It can then be used to retrieve information about the respective user from ALOE.

Contributing resources with the MACE bookmarklet: To allow an easy contribution of new learning resources, MACE offers a bookmarklet, i.e., a small application that is stored as the URL of a bookmark in a Web browser and offers one-click functionality (the respective code could also be added as an element on a Web page). When clicked, it automatically extracts the current site's URL and title and forwards the user to the contribution process, where the title can be changed, and where additional information such as a description and language can be provided. Finally, the ALOE contribution process is triggered, and the respective information is stored in ALOE. In addition to that, a preview image is generated for the resource.

Retrieving and storing social metadata: When a detail page about a resource is shown in MACE, existing social metadata is retrieved from ALOE, and logged in users can add new metadata that will then also be stored in ALOE.

Social search: ALOE offers to search for tags that have been associated with resources.

Furthermore, ALOE offers an OAI target that allows to harvest all social metadata and integrate it into the MACE metadata store. A CAM (Contextualized Attention Metadata, see [WNVD07, Wol08, WMS+09]) servlet provides usage data that is accessed by the MACE CAM store. This store contains a CAM instance for almost every event occurred at the portal, e.g., when a user conducted a search, viewed the metadata of an object or accessed the object itself. Each CAM instance includes at least the id of the respective user, the type of the event (e.g. filtered search or social search), the value of the event (e.g. the keyword used to search) and a timestamp. If an object is involved in the event (e.g. when the metadata of an object is viewed), the id of this object is stored as well.

ALOE Services Used in MACE   From the *ALOE Web Service* that at this time already comprised more than 150 methods, only a subset is used in MACE. Within this subset, the following main services are provided.

Method	Description
addComment	Adds a comment to a resource.
addTags	Adds tags to a resource.
contributeBookmark	Inserts a bookmark and its associated metadata into the system, without generating a preview image.

Method	Description
deleteUserResourceTag	Deletes a user's tag that is associated with a resource.
getResourceComments	Determines the comments that have been posted for a certain resource.
getResourceMetadata	Determines metadata about the specified resource.
getTagCloud	Determines the 50 most frequent tags in the system.
searchSelectedFields	Determines metadata about the resources that match the specified search string in selected metadata fields (in MACE, this is only applied with the field "tags").
searchUsers	Determines user data of the members in the system which match the specified search string in their eMail address, nickname, first name or last name.

Table 6.1: ALOE core services used in MACE

The delivery of preview images for MACE resources as well as buddy icons of MACE users is realised using the *ALOE Multimedia* servlet which is accessed by using simple HTTP requests. Only the following *ALOE Multimedia* servlet methods are used in MACE:

Method	Description
resourceThumbnailLarge	Returns a large thumbnail for the specified resource.
buddyIconLarge	Returns a large thumbnail for the specified user.
buddyIconSmall	Returns a small thumbnail for the specified user.

Table 6.2: *ALOE Multimedia* services used in MACE

6.4.2 Mindpool

The ALOE-based Mindpool infrastructure was initiated in 2010 as an internal DFKI project. It realises an interactive knowledge sharing platform for all DFKI employees at all sites (in Bremen, Kaiserslautern, Saarbrücken, and Berlin) and offers access to a variety of specifically tailored DFKI contents such as press articles, visits, and galleries that before were only accessible via static HTML pages. With Mindpool, these contents can now be accessed using different search and filter technologies, and users can provide own metadata such as tags, comments, ratings, and classifications for the contents. Initially, Mindpool consisted of the ALOE-based component *Mindpool Treasures* and the Microblogging component *Mindpool Hints*. As Mindpool Hints was stopped in 2013, the terms Mindpool and Mindpool Treasures now refer to the same component.

Mindpool is an example that illustrates how ALOE can be adapted to the needs of a specific scenario. This mainly concerns the introduction of new resource types with special metadata and corresponding functionalities, as well as the frontend design that was changed several times after Mindpool was initiated. These design changes are illustrated in Figure 6.4 and 6.5.
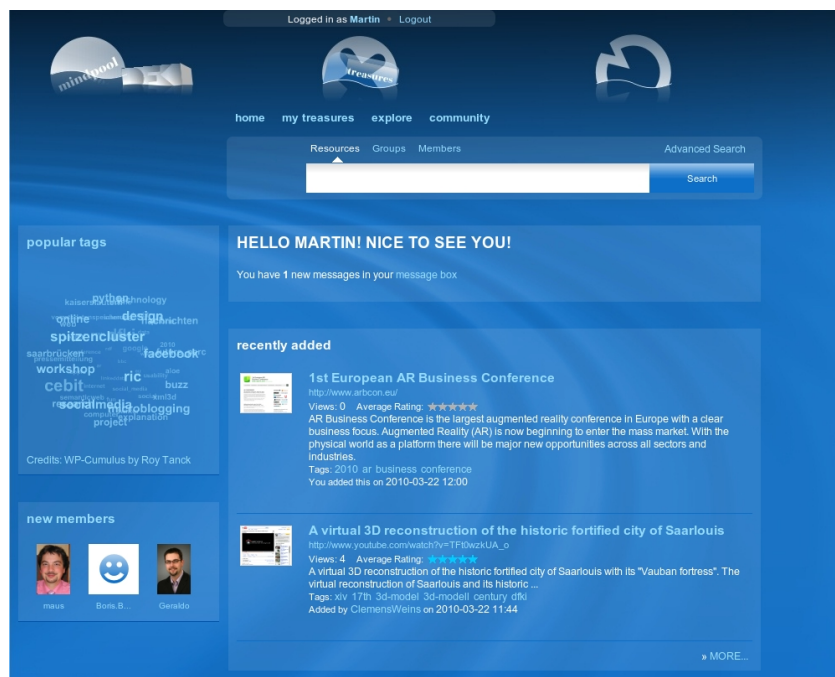
Figure 6.4: The Mindpool Treasures welcome page (state March, 2010)

As specific resource types, *press articles*, *visits*, and *galleries* were introduced.
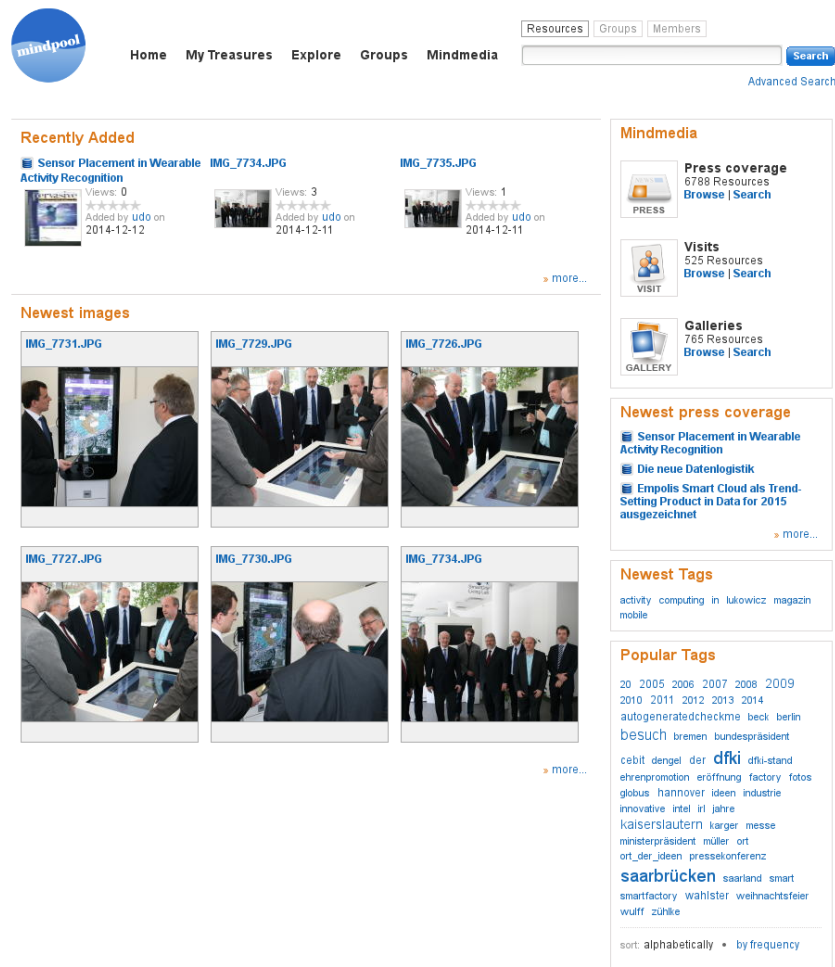
Figure 6.5: The Mindpool Treasures welcome page (state December, 2014)

They are supported with the following services and corresponding frontend elements only existing in Mindpool:

- Means are provided to contribute and edit the new resource types with the respective metadata.

- For each resource type, an advanced search functionality allows for searching within the offered metadata. An additional browsing mode basing on the search services can be used to browse the resources per year.

Furthermore, existing knowledge was integrated into Mindpool:

- Information about more than 35,000 existing resources (mainly press articles, visits, and photos) was integrated into Mindpool. The metadata about these objects was stored in tables on static HTML pages that were manually maintained by the enterprise communication unit. Existing files were also integrated either as uploads in ALOE (regarding scanned press articles and visit agendas) or as links pointing to the company WebDAV[8].

- More than 15,000 contacts from DFKI's internal contact database were imported into Mindpool (using the entity type *person*) to allow for an easy and convenient provision of persons involved into resources (especially regarding press articles, visits, galleries, and photos). Additionally, it is also possible to search for resources related to specified persons.

- Information about more than 500 projects from the DFKI Intranet were added as categories, so that they can be associated to resources and also be used as search fields.

For Mindpool, ALOE was enhanced with a component that allows for connecting to DFKI's LDAP server, i.e., users can simply login with their existing credentials for using internal DFKI resources.

Targeted problems   Several problems were tackled in Mindpool:

Abundance: More than 35,000 existing resources from other repositories have been integrated into Mindpool.

Missing metadata: The integrated contents only had very few or no metadata at all.

Missing interaction possibilities: Before being integrated into Mindpool, the contents were only accessible via static HTML pages.

6.4.3 ALOE at the Knowledge Management Department of DFKI

In February 2009, a standard ALOE instance without technical adaptations called ALOE@KM was deployed for the members of the Knowledge Management group in DFKI Kaiserslautern. By restricting the access on the DFKI Web

---

[8]WebDAV stands for "Web-based Distributed Authoring and Versioning" and allows users to collaboratively edit and manage files on remote Web servers, see `www.webdav.org`.

server, the instance was only accessible within DFKI. The aim was to provide a social sharing platform for the group that is also connected to other existing infrastructures (e.g., DFKI's Litfass system). ALOE@KM was introduced to the group in a "Brownbag session" with a presentation about what is ALOE, the functionalities offered, anticipated FAQs, and how it can be integrated in other contexts and applications. The group members could provide feedback personally, via email to the ALOE team, or using a Wiki page that was set up for ALOE.

6.4.4 Further Use Cases

ALOE-Public

ALOE-Public[9] is a standard ALOE instance without any adaptations and was the first instance that was made publicly available for any user in 2008. It was since then used in real application scenarios (e.g., by the Public Research Center Henri Tudor in Luxembourg[10]), but also as a simple playground.

OPENEER

OPENEER[11] is a platform that allowed users to share their emotions about music. The project was initiated by Dr. Stephan Baumann[12] in the context of research about music recommendations that also take into account people's emotions and individual experiences. Therefore, OPENEER served as a platform to collect respective data. Users can upload music or refer to sources like YouTube or last.fm[13] and are asked to leave information about their relations to the respective song ("What does this song do with you? (When, Where, Why)"). Although most of the ALOE interaction design could be used as-is, several adaptations were carried out. As depicted in Figure 6.6, the OPENEER system has a different "look and feel" than the ALOE screenshots presented so far. The most important differences are:

- A new logo as well as different backgrounds, colours, fonts, and icons (the respective design was realised by mess GbR — mobile einsatztruppe stadt und stil[14]).

---

[9]See `http://aloe-project.de/AloeView`
[10]See `http://www.tudor.lu`
[11]See `http://openeer.dfki.de`
[12]See `http://www.dfki.de/~baumann`
[13]See `http://www.last.fm`
[14]See `http://www.m-e-s-s.de`

- Adapted labels and texts (e.g., "songs" instead of "resources" or "Artist" instead of "Creator").

- New elements on the start page (e.g., the possibility to enter a URL, and a spherical tag cloud presented in the right column).
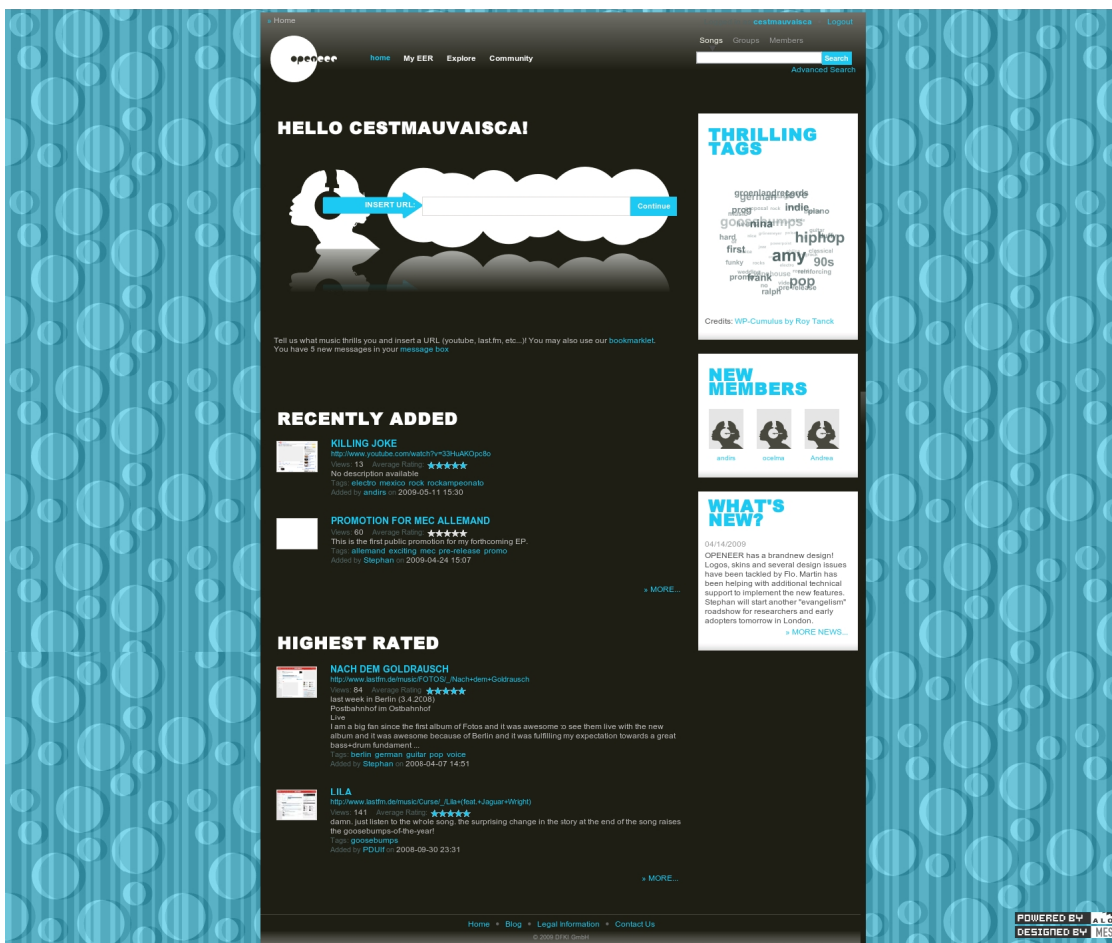
Figure 6.6: Screenshot of the welcome page in OPENEER

The adaptation process was very simple and straightforward. Most of the changes only concerned the ALOE CSS, the ALOE page header, and the images used in the platform. Further adaptations could be realised with simple changes in the respective Java Server Pages.

C-LINK

The aim of C-LINK (Conference Link) was the development of a Web based tool to support conference attendees. With C-LINK, users could share papers and presentations, generate individual conference schedules, get personalised recommendations to find interesting events and attendees, etc. The ALOE-CLINK branch was used during the KI 2008 (the Annual German Conference on Artificial Intelligence) in Kaiserslautern and the ICDAR 2009 (the International Conference on Document Analysis and Recognition) in Barcelona. C-LINK supports an event-specific provision of interests on user profile pages and can provide recommendations for conference events based on the user's profile and contributions.

The Web of Models

Within the Cluster of Excellency "Center for Mathematical and Computational Modelling" (CMCM), the ALOE branch ALOE-CMCM is used as a basis for the platform *Web of Models*. To support mathematical models, the new resource type *mathematical model* was introduced. It offers specific metadata, especially the definition of persons as own entities within ALOE that can have different relations with a model. Using this mechanism, different roles such as authors, contributors, and reviewers could be specified. Each mathematical model has four elements defining its content: *Definition and Model*, *Visualisation*, *Classification*, and *References*. The classifications could be provided by making use of the more than 6,500 elements of the *AMS 1991 Mathematical Subject Classification* and the *WZ08 Classification of Economic Activities* that were added as categories in ALOE-CMCM. The other contents are stored in a separate Wiki that offers versioning as well as the provision of mathematical contents, and that accesses the *ALOE Web Service* to check whether a user is authorised to modify a mathematical model.

Figure 6.7 shows several screenshots of the ALOE-CMCM instance with the interactive classification browser that was added for this instance, and several facets (description, visualization, and classification) of a mathematical model. For further details about the Web of Models system please refer to [GGM$^+$11] and [GGS$^+$11].

RADAR-Kaiserslautern

RADAR (Resource Annotation and Delivery for Mobile Augmented Reality Services) is an ALOE branch that was created in 2010 and 2011. The aim was to re-
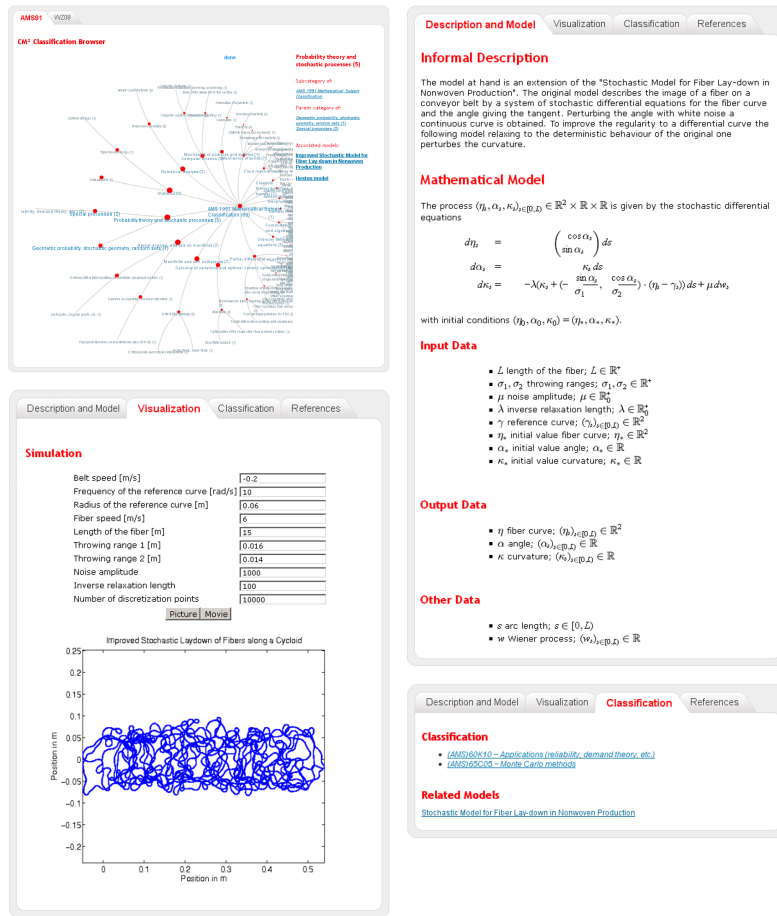
Figure 6.7: Screenshots from ALOE-CMCM

alise an open and flexible infrastructure that allows users to contribute, manage, and share arbitrary geocontents (i.e., contents that have a spatial dimension). Furthermore, it should be possible to access RADAR contents with Augmented Reality Browsers that are available for any common mobile device. To realise these functionalities, ALOE was enhanced to also allow the association of spatial information with any kind of digital resource, and to offer spatial searches also using map visualisations. Adapters were developed for providing contents to the popular Augmented Reality Browsers Junaio[15], Layar[16], and Wikitude[17]. For further information about the RADAR system, please refer to [MG11] or [AMZS11].

---

[15]See `http://www.junaio.com`
[16]See `http://www.layar.com`
[17]See `http://www.wikitude.com`

Several instances of RADAR were deployed since 2011, and have been used for a variety of application scenarios.

Event support With the RADAR instances RADAR-Berlin and especially RADAR-Kaiserslautern, several events were supported. This entails the annual ″Lange Nacht der Kultur″ in Kaiserslautern where RADAR was used since 2011. Together with the cultural department of the City of Kaiserslautern, all venues were added to the system, and a newly introduced resource type *event* was used to provide information about all events taking place. The single events could further be enriched – also by the involved artists – with arbitrary digital resources such as an artist's homepage or multimedia information. Users could then make use of RADAR in the following ways:

- To prepare for the event or to obtain additional material afterwards, users could search for all involved venues, events, and event material. An interactive map with all venues was offered, as well as an interactive timeline and an event overview with means to select event categories.

- During the event, users could use their mobile devices and start the specifically created channel for the event in their Augmented Reality Browser. They could then retrieve information about nearby venues and the events taking place there.

Figure 6.8 shows an example for event support in RADAR, referring to the ″Lange Nacht der Kultur″ in 2014.

Urban Sensing Using the specifically developed Android application ″RADAR-Sensing″, users can rate locations according to several *urban quality dimensions*. The results for each dimension can be examined using the RADAR Web frontend that offers visualisation means for locations such as interactive maps and heatmaps [ZME12].

The visualisation means for spatial information in RADAR were also used extensively within further urban sensing projects were participants were equipped with wearables that collect physical information (see [BEM⁺13b] and [BEM⁺13a] for further details).

GeoEvents GeoEvents is an interactive tool to analyse and visualise spatial information from the social web. The application provides two different data

Figure 6.8: Event support with RADAR

retrieval modes, the scenarios and the exploring mode. The scenarios mode accesses information that was already harvested into a local storage, while the exploring mode retrieves information on the fly using the respective APIs. In both modes, information about the returned resources is aggregated on a RADAR map and displayed using different visualisation means. In addition, the retrieved information is analysed using different analysis techniques. The system and a first evaluation were presented in [SMA13].

ALOE-Unifarm

Within the EU-funded project UNIFARM[18] that was initiated in 2011, an ALOE branch was created for realising a pilot system that allows users to present and defend the needs of farmers in the development of Global Navigation Satellite System (GNSS) applications and services. Apart from an adapted design with a different logo and color schema, a specific vocabulary to classify documents was used and represented as categories within ALOE-Unifarm. Furthermore, specific resource types were introduced that allow the contribution, management and display of *use cases*. A use case in UNIFARM is realised as a snippet that could be filled by the project stakeholders. Each snippet has a fixed structure and contains three further subelements in the respective resource container: *interface requirements*, *system requirements*, and *non-functional requirements* (each of them once again realised as a snippet). An example of such a use case is shown in Figure 6.9.

ALOE at Fraunhofer IAO

The implementation of ALOE in the Fraunhofer IAO (Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO) was initiated by Margit Beutler[19] and started as a grassroots project in May 2009. The exploratory initiative was triggered by the following observation: An ever growing number of tools are developed to support knowledge workers. Yet, very few of them are actually adopted – although their features seem to match needs expressed by potential users. By inviting employees to test ALOE and another social search tool (Qitera[20]), the aim was to achieve four major goals:

1. Introducing the institute's associates to the concept of "Social Search".

2. Realise the benefits of "Social search" tools (faster recall, increased collaboration, etc.).

3. Providing feedback to the software developers.

4. Initiating the discussion on technology acceptance within the organisation.

Members of six different research groups would receive a teaser, could book a group or individual presentation. Easily readable manuals were provided,

---

[18]See `http://project-unifarm.eu`
[19]See `http://www.mendeley.com/profiles/margit-beutler`
[20]See `www.qitera.com`

Figure 6.9: An example of a use case visualised within ALOE-Unifarm

assistance was offered (analogue and in form of a wiki), and curated content was uploaded.

The initial interest – brought forward by about 30 researchers – didn't translate into a mentionable engagement or user acceptance, of neither ALOE nor Qitera. Further marketing activities were cancelled. Theories like the Technology Acceptance Model ([Dav86]) and innovator roles ([Wit73, Cha74, GSH07]) would provide a comprehensive and somewhat less biased view on the stopblocks that were encountered.

Using ALOE for the Professional Translation Domain

Professional translators continuously have to increase the efficiency of their work processes to be able to compete with lay translators. This also entails research for information that consumes a considerable part of the time needed for producing a high-quality translation. In a joint work of professional translators and DFKI, Web 2.0 applications were evaluated towards their potential to support the professional translator in his daily work. Here the main focus of attention is on a possible increase in efficiency of search and knowledge management processes. Therefore, user tests on ALOE were conducted followed by open interviews aiming to identify the potential of respective applications in the professional translator's business. The evaluation showed that tagging and rating facilities finding relevant information, and that especially the possibility to find domain experts and get into contact with them was highly estimated. For a detailed overview of this scenario, please refer to [GAB09].

FlashMeetings in ALOE

Flashmeeting[21] is an academic research project by the Knowledge Media Institute (KMi) of the Open University[22]. It aims at understanding the nature of online events and helping users to meet and work more effectively. The Flash-Meeting technology allows to have online meetings using a standard browser. During a meeting, one participant is allowed to speak at a time, while others can simultaneously contribute information via a chat window. Meetings are recorded (including a chat protocol and metadata about participants and topics), and these meeting replays can optionally be made available to the public.

During a visit in 2007 at KMi, over 350 Flash video files from the FlashMeeting public event repository[23] as well as their associated metadata have been harvested as bookmarks in the ALOE platform. The underlying question was how browsing, retrieving, recommending and sharing of FM replays among peers can be enhanced.

On the the Flashmeeting project research pages[24], this is described as follows:

> *"By integrating the FM public event repository with the ALOE system, new services are enabled for the user. annotate their learning resources with tags and comments, collect references to them in their collections and share them*

---

[21]See `http://flashmeeting.e2bn.net`
[22]See `http://kmi.open.ac.uk`
[23]See `http://flashmeeting.open.ac.uk/public`
[24]See `http://flashmeeting.open.ac.uk/research/social_media.html`

among friends. This has not been possible before and adds to the experience the user has with the resources. The tags enable the users to easily find the learning resource within their collections. Annotations enable them to take notes of the concepts of the learning resource important to them. In addition, comments may reveal experts on certain areas, whilst users may benefit from their feedback or even form groups of relevant contacts around certain resources. The bookmark collection itself represent a reflection on the interests of the users, thus does not only enable easy access but also provides them with an insight on how their interests have changed over time. Furthermore, by sharing single bookmarks or collections of bookmarks with friends, the user collaborates with other users and thus supports the relevant learning process. Last but not least, using ALOE simplifies the task to find learning resources through sharing them with others, through tags and annotations for the user.“

For a detailed description of the integration process, please refer to [MTW08] and [MSWT08].

6.5 Technical Requirements and Technical Quality

As shown in Chapter 5, ALOE provides all functionalities identified in the system design checklist presented in Chapter 4. In this section, evidence is presented that ALOE meets the technical requirements of a Social Resource and Metadata Hub as defined in Section 4.1.2 (Hypothesis 3.1).

The evaluation of the technical quality of the ALOE infrastructure is based on software quality attributes as they are defined by Bass et al. in 2003 [BCK03]. Examples of such attributes are:

- availability

- performance

- interoperability

- modifiability

- reliability

- portability

- scalability

The performance, reliability, and scalability of ALOE is ensured using a variety of strategies and tests. For other software quality attributes, the evaluation is based on a more theoretical framework, e.g., quality tactics used to achieve the quality attributes.

To measure some of the software attributes, stress tests, time to respond tests, and availability tests of services were used. To calculate these measurements, the use of different tools was required:

SoapUI: A Web Service testing tool made for service-oriented architecture and used to evaluate ALOE within MACE. This can be used for simulation, load and functional testing, etc.

Apache JMeter: A load testing tool developed by the Apache Jakarta Project. This tool provides an effective way to analyse and measure the performance of a great variety of services. The tool supports flexible configuration, response validation, and reporting tools.

CruiseControl: A tool that is used to test the availability of the ALOE frontend and backend twice every hour and provides the data required to analyse the reliability of the ALOE services.

Availability

The CruiseControl component conducts a test every 30 minutes, which results in approximately 1,400 tests each month. Figure 6.10 shows the respective results gathered for MACE in 2009 and until April 2010.

Figure 6.11 and Figure 6.12 show availability reports for ALOE-MACE and ALOE@KM created with the infrastructure monitoring system Nagios[25].

Performance

Besides general scalability efforts (see below), the performance of ALOE is achieved by elaborating a database design that supports fast queries also in large amounts of data. For instance tag search is accelerated by making use of MySQL's full text indexes. Furthermore, much effort has been put into keeping the algorithms as lean as possible in order limit the computational complexity of our algorithms. In case Lucene is integrated in ALOE instances, the performance is significantly faster due to the qualities of the framework regarding any kind of text based search even on very large corpora.

---

[25]See `http://www.nagios.org`
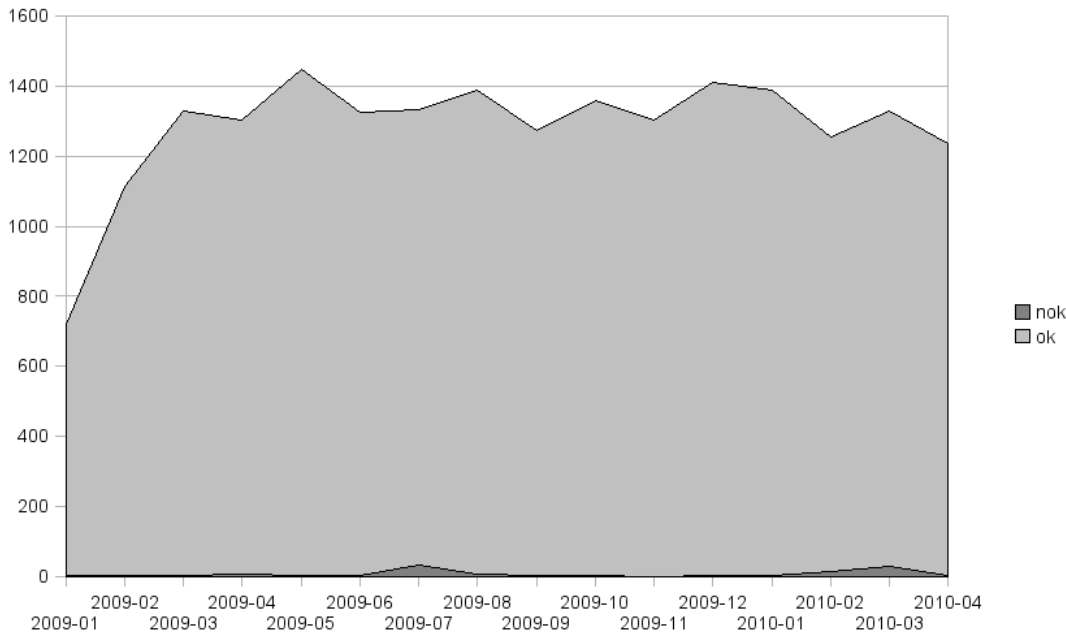
Figure 6.10: MACE cruise control results for ALOE services

Load Tests   For an evaluation of this aspect, results gathered within the context of the MACE project are now presented[26]. Load tests on the ALOE services (except for the *ALOE Multimedia* servlet) where conducted using soapUI . For each method, the number of simultaneous threads chosen were 1 (to test the response time in the simplest case), 3, and 15. Each test was run for 120 seconds. In the following we present the resulting tables containing the number of threads, minimum, maximum and average response time in milliseconds, the number of performed requests (Cnt), and the number of errors occurred.

The following setups were used for the load tests:

getTagCloud: No special setup was required, as this method does not expect any input.

getResourceComments: A representative sample set of five resource ids was provided, from which in each run one user was randomly chosen.

getResourceMetadata: A representative sample set of five resource ids was provided, from which in each run one user was randomly chosen.

---

[26]See [SWM+09] and [MWC+10] for a complete overview of the evaluation of the MACE infrastructure and services.

State	Type / Reason	Time	% Total Time	% Known Time
OK	Unscheduled	110d 23h 1m 7s	30.400%	99.926%
	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**110d 23h 1m 7s**	**30.400%**	**99.926%**
WARNING	Unscheduled	0d 0h 0m 0s	0.000%	0.000%
	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 0h 0m 0s**	**0.000%**	**0.000%**
UNKNOWN	Unscheduled	0d 0h 0m 0s	0.000%	0.000%
	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 0h 0m 0s**	**0.000%**	**0.000%**
CRITICAL	Unscheduled	0d 1h 58m 53s	0.023%	0.074%
	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 1h 58m 53s**	**0.023%**	**0.074%**
Undetermined	Nagios Not Running	0d 0h 0m 0s	0.000%	
	Insufficient Data	253d 23h 0m 0s	69.578%	
	Total	253d 23h 0m 0s	69.578%	
All	Total	365d 0h 0m 0s	100.000%	100.000%

Figure 6.11: Availability report for the ALOE-MACE instance

contributeBookmark, addComment: A sample set of five user nicknames was provided, from which in each run one user was randomly chosen. This user then contributed a URI that was randomly generated, and the user also added a comment. At the end of each run, the contributed resource was deleted. The two methods were tested in one run because ALOE does not offer a method to delete single comments.

addTags, deleteUserResourceTag: A sample set of five user nicknames and five representative resource ids was provided, from which in each run one user and one resource id was randomly chosen. The user then contributed a tag that was randomly generated. Afterwards, the respective tag was deleted.

searchSelectedFields: A representative sample set of five queries was provided, from which in each run one query was randomly chosen.

searchUsers: A representative sample set of five queries was provided, from which in each run one query was randomly chosen.

The results of the load tests clearly show that for one and three parallel threads the response times are always very fast. When testing 15 parallel threads, response times may go up to about 10 seconds in some cases, but these cases can be considered as exceptional (especially for writing access).

State	Type / Reason	Time	% Total Time	% Known Time
	Unscheduled	230d 2h 50m 27s	63.046%	99.600%
OK	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**230d 2h 50m 27s**	**63.046%**	**99.600%**
	Unscheduled	0d 14h 0m 0s	0.160%	0.252%
WARNING	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 14h 0m 0s**	**0.160%**	**0.252%**
	Unscheduled	0d 0h 0m 0s	0.000%	0.000%
UNKNOWN	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 0h 0m 0s**	**0.000%**	**0.000%**
	Unscheduled	0d 8h 9m 33s	0.093%	0.147%
CRITICAL	Scheduled	0d 0h 0m 0s	0.000%	0.000%
	Total	**0d 8h 9m 33s**	**0.093%**	**0.147%**
	Nagios Not Running	0d 0h 0m 0s	0.000%	
Undetermined	Insufficient Data	133d 23h 0m 0s	36.701%	
	Total	133d 23h 0m 0s	36.701%	
All	Total	365d 0h 0m 0s	100.000%	100.000%

Figure 6.12: Availability report for the ALOE@KM instance

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	749	948	772,54	59	0
3	720	1.850	1.025,33	118	0
15	929	8.574	5.086,47	121	0

Table 6.3: Load test results for *addComment*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	86	203	140,78	92	0
3	69	304	111,79	293	0
15	32	648	105,63	1527	0

Table 6.4: Load test results for *addTags*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	50	401	107,15	86	0
3	58	574	145,28	245	0
15	108	3.133	1.652,55	414	0

Table 6.5: Load test results for *contributeBookmark*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	90	203	147,72	92	0
3	70	402	113,2	293	0
15	32	674	108,03	1527	0

Table 6.6: Load test results for *deleteUserResourceTag*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	47	91	52,88	132	0
3	44	103	52,24	391	0
15	44	532	53,55	1966	0

Table 6.7: Load test results for *getResourceComments*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	734	1.229	774,07	76	0
3	814	2.058	1.306,94	169	0
15	994	10.476	9.613,51	180	0

Table 6.8: Load test results for *getResourceMetadata*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	22	78	62,58	134	0
3	22	200	66,54	333	0
15	24	430	87,45	1.921	0

Table 6.9: Load test results for *getTagCloud*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	52	736	318,51	103	0
3	18	1.477	490,36	275	0
15	17	6.508	2.463,92	518	0

Table 6.10: Load test results for *searchSelectedFields*

Threads	Min (ms)	Max (ms)	Avg (ms)	Cnt	Errors
1	50	118	65,92	129	0
3	50	370	69,42	387	0
15	14	1.961	78,3	1.912	0

Table 6.11: Load test results for *searchUsers*

To test the *ALOE Multimedia* servlet, which can be accessed using simple HTTP requests, JMeter was used. As a ramp-up period, 100 seconds was chosen for all tests. The maximum response time allowed was set to 60 seconds, everything above was considered as an error. In such cases, no thumbnail will be displayed. Only resourceThumbnailLarge was tested as this method is mostly used in ALOE and does use exactly the same mechanisms as the other methods offered by the *ALOE Multimedia* Servlet.

The average response times for the settings shown in the table are all within acceptable ranges. In rare cases, the maximum response time can be high, which could result in a missing thumbnail on a detail page or a search result overview. Still, this did not have a significant impact on the interaction with the MACE system.

Threads	Requests	Min (ms)	Max (ms)	Avg (ms)	Samples	Errors
25	50	44	1842	156	1250	0
50	50	44	30680	1150	2500	0
75	50	46	52431	4091	3750	0

Table 6.12: Load test results for *resourceThumbnailLarge*

Reliability

Several tactics are applied to ensure the reliability of ALOE services. First of all, several development versions of ALOE exist in addition to productive versions. These development versions ensure that extensive testing of new services is possible before deploying them on a productive system. Furthermore, they allow testing (e.g., stress tests) without interfering with a productive system. To ensure data integrity and availability, ALOE uses innoDB tables in MySQL, and backups are executed regularly (once a day). As ALOE is usually deployed on a virtual machine, the quick setup of a new ALOE instance is possible as a fallback solution in case of hardware problems.

Interoperability and Use of Standards

To ensure interoperability, ALOE uses several standards for content representation and delivery, e.g.:

- REST and JSON are used in the *ALOE Web Service*.

- An OAI target allows the harvesting of social metadata.

- A CAM service for usage metadata is provided.

- ALOE metadata uses DC elements wherever possible (dc:contributor, dc:date and dc:format (all created automatically when contributing a resource), dc:creator, dc:description, dc:rights and dc:title).

Modifiability and Extensibility

As ALOE is a service-based system, services can to a large extent be modified without a need to change respective infrastructures accessing the system. Concerning resource metadata, new types of categories (e.g., competencies) can easily be integrated by using the generic categories offered in ALOE.

Scalability

ALOE uses only standard technologies that have proven to scale also in professional environments. Among others, the system is based on Java, Apache Tomcat, MySQL with innoDB and myISAM tables, and Lucene. Thus scalability can be ensured from a software perspective.

From a hardware point of view, scalability is guaranteed up to a certain point as ALOE is deployed on a virtual machine that can easily be allocated with more storage or CPU.

6.6 ALOE as a Generic Infrastructure

The process of adapting and enhancing ALOE instances was already explained in Section 5.5. There are means to systematically add new data storage structures, to add new functionalities in the ALOE Backend Services, and especially to create variants of *ALOE View* that can be managed in branches. Furthermore, the ALOE categories provide generic means to allow adding arbitrary vocabularies and ontologies to allow for classifying contents.

The presented scenarios and use cases provide evidence that several ALOE variants have successfully been created and deployed, thus Hypothesis H3.2 can be considered as valid.

6.7 Supporting Access to Digital Resources with ALOE

In this section evidence is presented that a Social Resource and Metadata Hub can successfully tackle typical problems when accessing digital resources (Hypotheses 1). This concerns the problems of *abundance* (H1.1), *heterogeneous access* (H1.2), and *missing metadata* (H1.3). ALOE in general provides all functionalities identified to overcome these problems as presented in the system design checklist in Section 4.8. To provide results providing evidence about whether these problems could also be tackled in specific scenarios with ALOE, evaluations results from the use cases MACE and Mindpool are now presented, as both specifically tackled the mentioned access problems.

6.7.1 The Problem of Abundance

While MACE integrated more than 175,000 resources from various repositories, Mindpool integrated more than 35,000 press articles, photos, and resources with information about visits. In addition to the successful integration on a technical level, users were also asked about the quality of information access in Mindpool.

Figure 6.13 and Figure 6.14 show that 86% of the participants agreed that searching for contents in Mindpool is easy and intuitive, and that the means to browse and navigate within the contents in Mindpool are sufficient and easy to use.
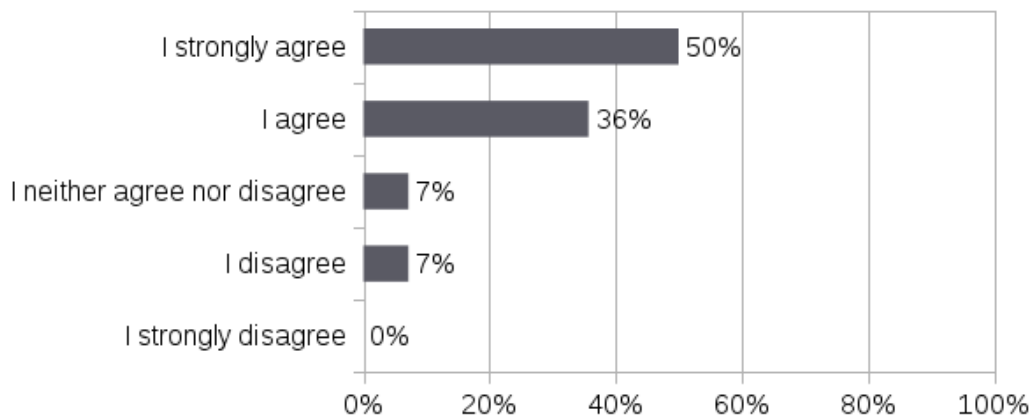
Figure 6.13: Survey results for *"Searching for contents in Mindpool is easy and intuitive"*
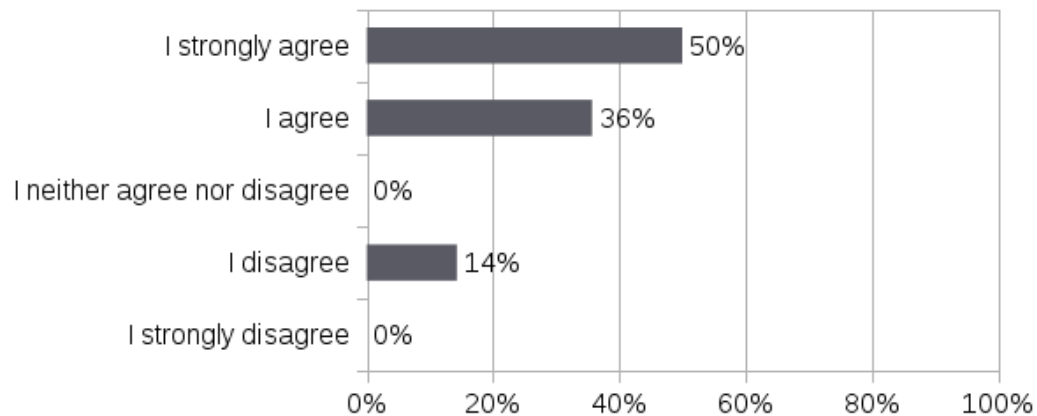
Figure 6.14: Survey results for *"The means to browse and navigate within the contents in Mindpool are sufficient and easy to use"*

6.7.2 The Problem of Heterogeneity

Heterogeneity in MACE In addition to the MACE frontend that was developed as an own component (not related to ALOE) from engineers at the University of Applied Sciences in Potsdam (FHP), uniform access to all resources was enabled through ALOE as a means to store and provide all social metadata in the MACE environment.

Heterogeneity in Mindpool In Mindpool, all resources were integrated by means of specific importers that harvested the existing data on HTML pages provided and maintained by the enterprise communication unit in DFKI.

In addition to the technical prerequisites that were met, the Mindpool questionnaire provided clear evidence about the added value of Mindpool in comparison to the situation before the system was introduced (see Figures 6.15, 6.16, 6.17, and 6.18).
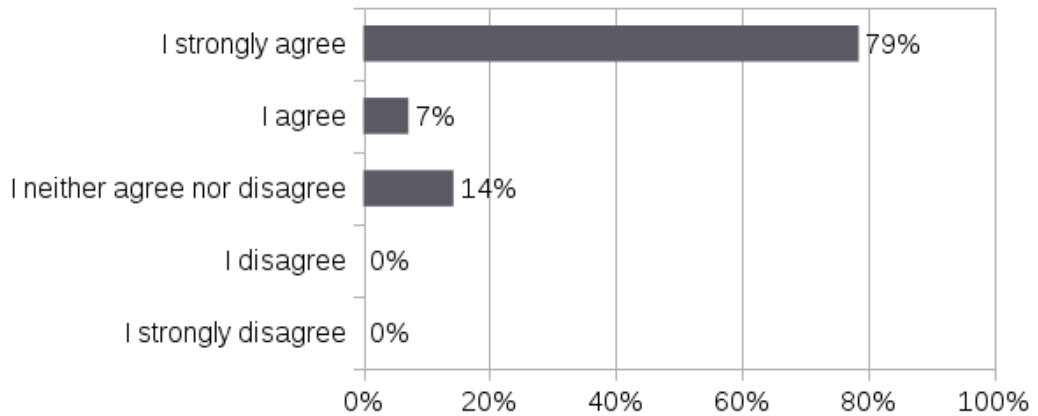
Figure 6.15: Survey results for *"With Mindpool I can get aware of DFKI related digital contents (especially press releases, visits, and photos) that I otherwise would not have found"*
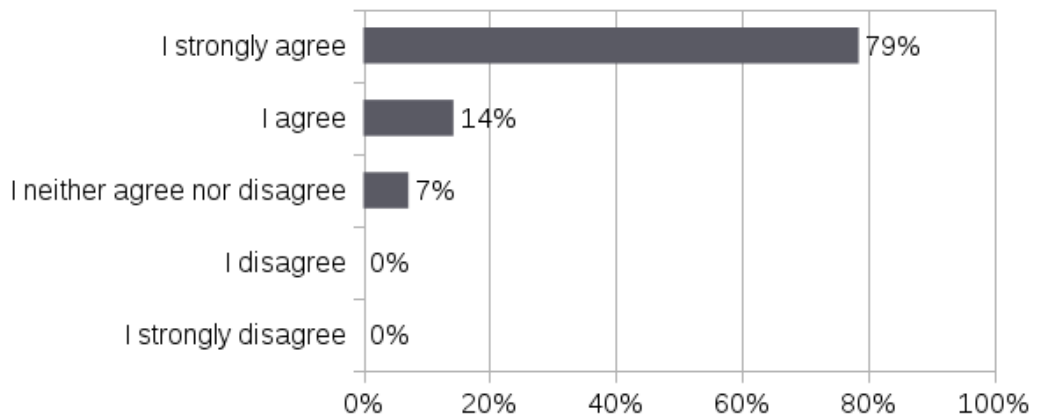
Figure 6.16: Survey results for *"Searching for press releases, visits, and photos has become easier and more intuitive with Mindpool"*
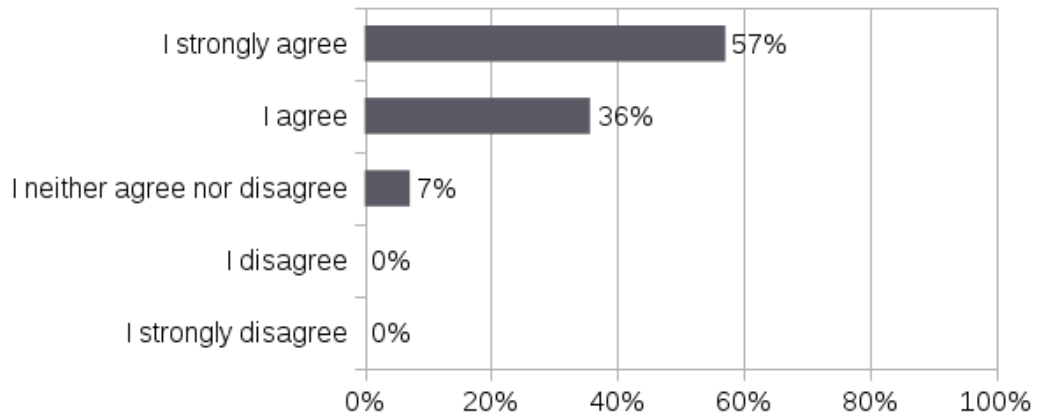
Figure 6.17: Survey results for *"There are more and better ways to browse and navigate within press releases, visits, and photos"*
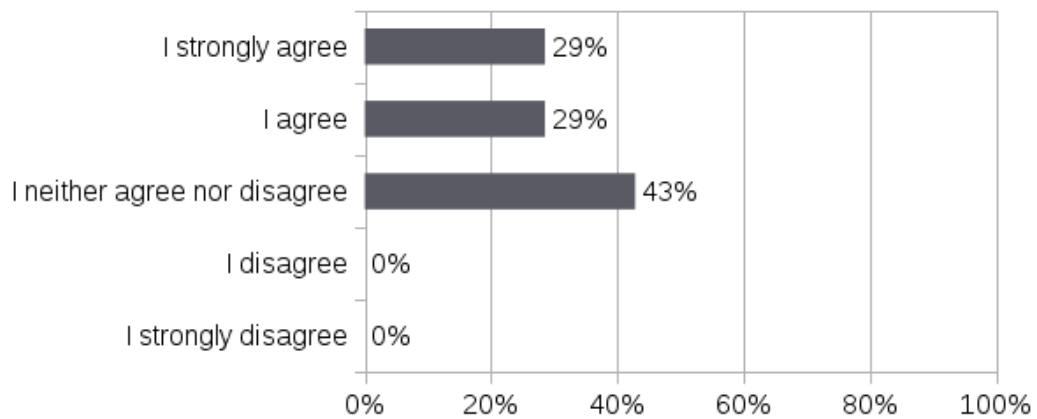
Figure 6.18: Survey results for *"Interaction and collaboration are fostered with Mindpool"*

6.7.3 The Problem of Missing Metadata for Digital Resources

Especially in Mindpool, most of the integrated contents were described with no or only few metadata elements. With the newly introduced resource types and respective interaction means, users were given the chance to provide new metadata. The Mindpool dashboard shows that users were making extensive use of this possibility and added more than 250,000 Tags for more than 35,000 resources (see Table D.4).

In the questionnaire, 93% of the users agreed that the metadata quality for the integrated contents in Mindpool has increased (see Figure 6.19).
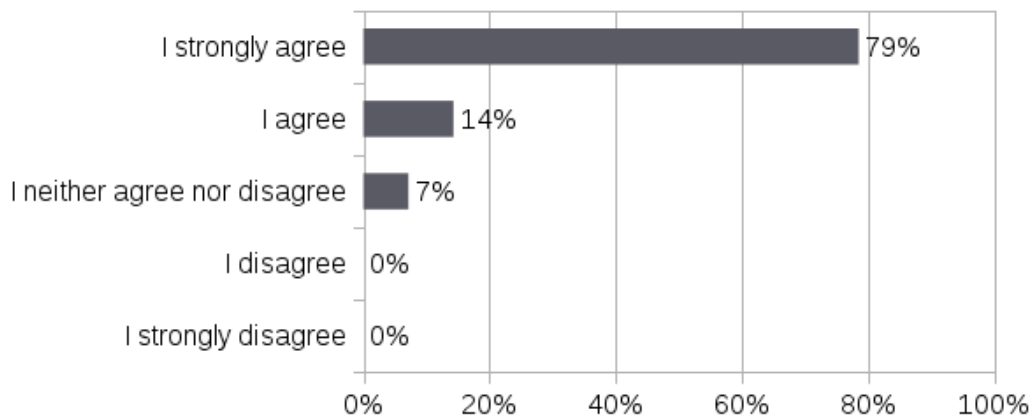
Figure 6.19: Survey results for *"The quality of information (metadata) about press releases, visits, and photos in Mindpool has increased"*

Information about the number of contributed metadata elements from other ALOE instances also support the validity of Hypothesis H1.3. E.g., users added more than 60,000 tags in MACE (see Table D.1).

Summary The feedback provided by the users and the statistical information allow to consider H1.1, H1.2, and H1.3 to be valid.

6.8 Offering Potentials to Exploit Social Web Phenomenons with ALOE

To investigate Hypothesis 2, results about how ALOE allowed for exploiting *interaction possibilities* (H2.1), the *Long Tail* phenomenon (H2.2), *Collective Intelligence* (H2.3), and *Crowdsourcing* (H2.4) will now be provided.

6.8.1 Interaction Possibilities

It was already shown that the required interaction possibilities exist and technically work properly. The usage statistics for the ALOE instances also show that they were used extensively (see Appendix D). E.g., more than 35,000 activities were tracked in Mindpool, carried out by more than 100 users on more than 45,000 different objects (see Table D.5). For ALOE@KM, more than 58,000 activities of 59 users on more than 3,300 objects were tracked.

Additionally, the feedback from the questionnaires clearly shows that users know and appreciate the offered interaction means, and that enough of them are offered (see Figures 6.20, 6.21, and 6.22 for the results from the Mindpool survey and Figures 6.23, 6.24, and 6.25 for the results regarding ALOE@KM).
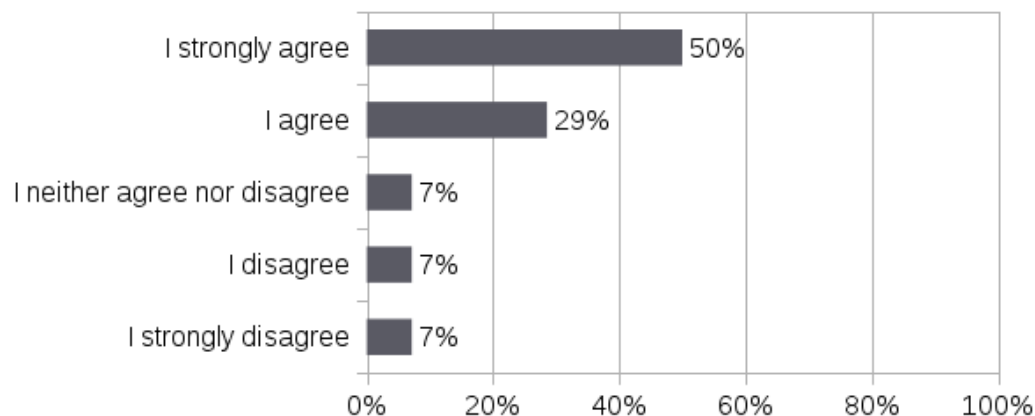
Figure 6.20: Survey results for *"I am aware of the interaction possibilities that Mindpool offers"*
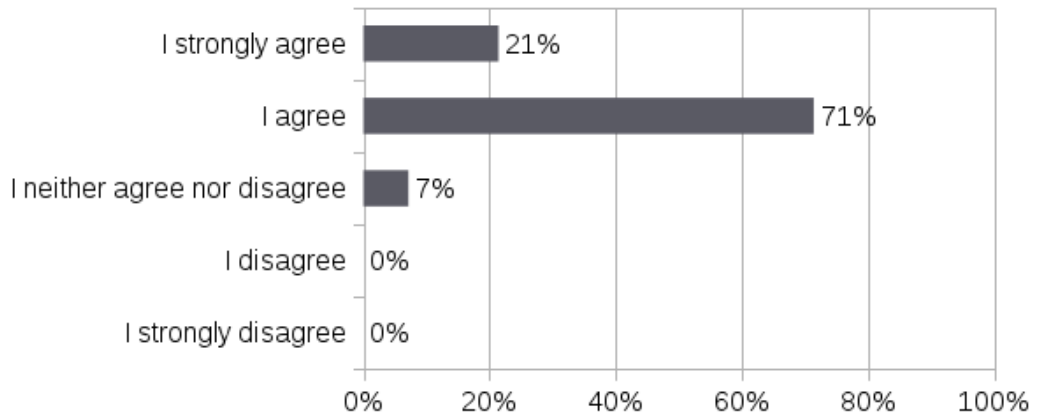
Figure 6.21: Survey results for *"The interaction possibilities offered by Mindpool are important for me"*
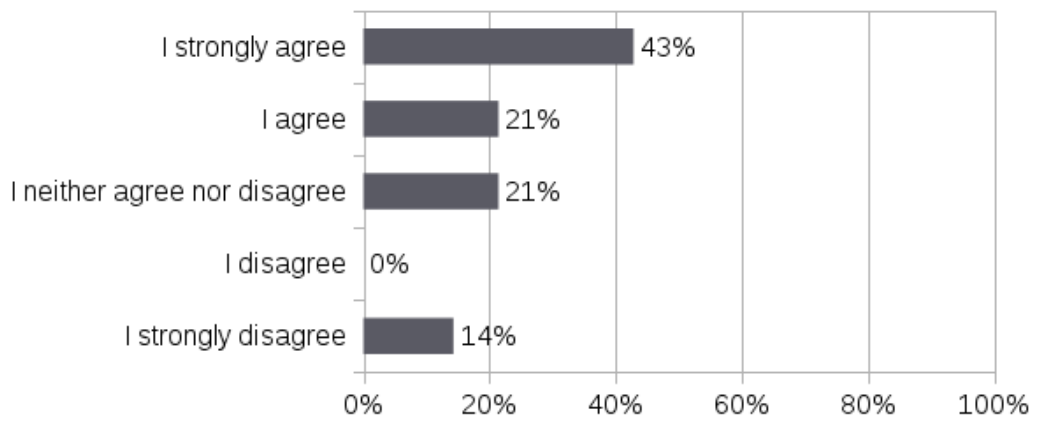
Figure 6.22: Survey results for *"I do not miss any interaction possibilities in Mindpool"*

Figure 6.23: Survey results for *"I am aware of the interaction possibilities that ALOE@KM offers"*
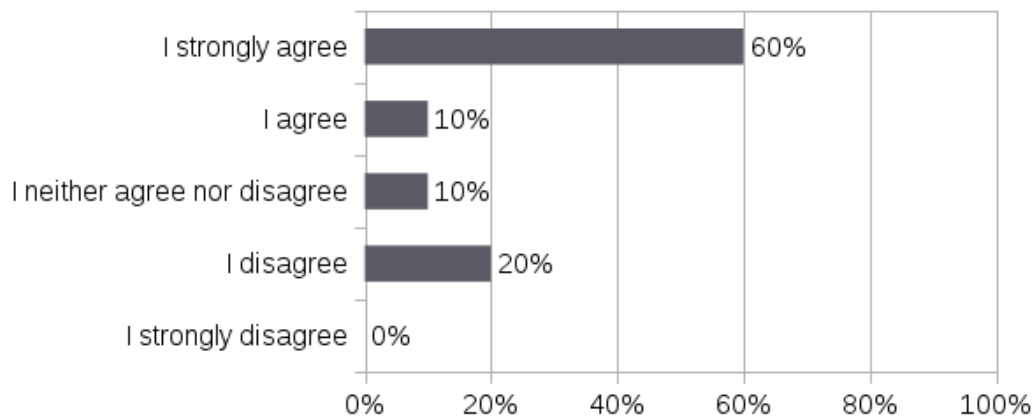
Figure 6.24: Survey results for *"The interaction possibilities offered by ALOE@KM are important for me"*
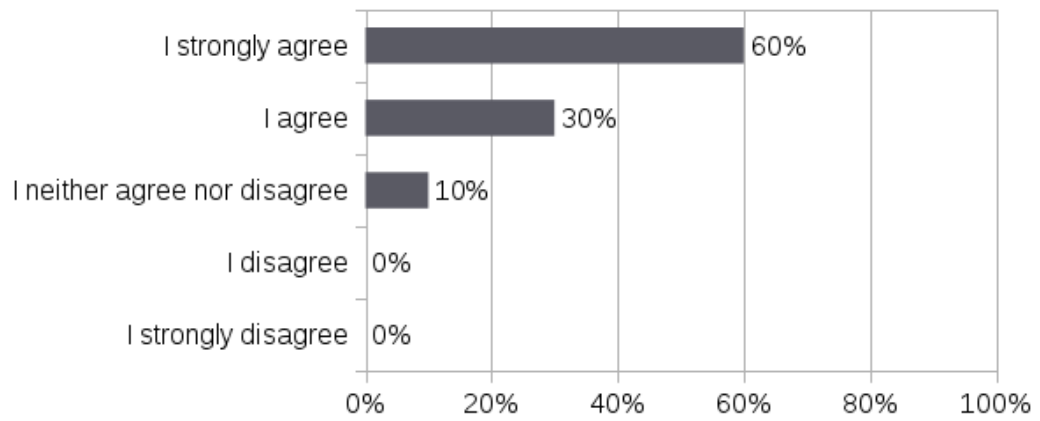
Figure 6.25: Survey results for *"I do not miss any interaction possibilities in ALOE@KM"*

6.8.2 The Long Tail Phenomenon

There are numerous search and navigation means in ALOE that allow for accessing resources, and users expressed their satisfaction with the provided means as already shown in Figure 6.13 or Figure 6.14. Furthermore, 93% of Mindpool users agreed that they can easily find contents regardless of their popularity, while 80% agreed in the survey for ALOE@KM (see Figure 6.26 and Figure 6.27).
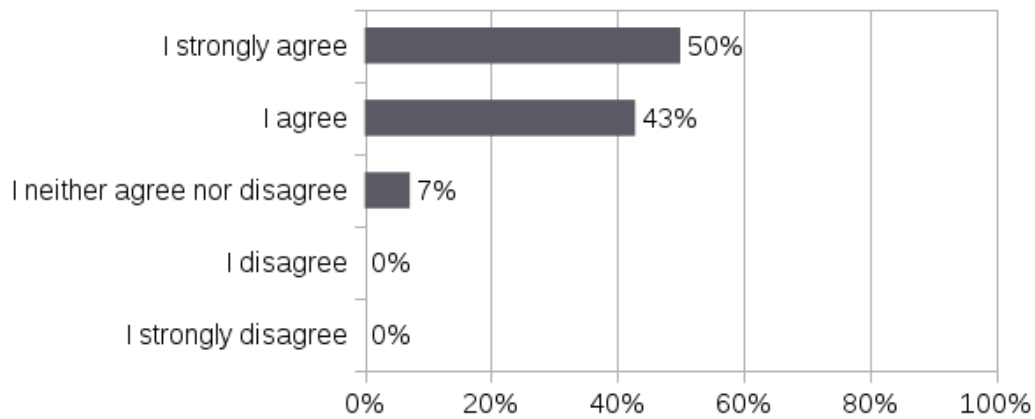
Figure 6.26: Survey results for *"I can easily find contents in Mindpool regardless of their popularity"*
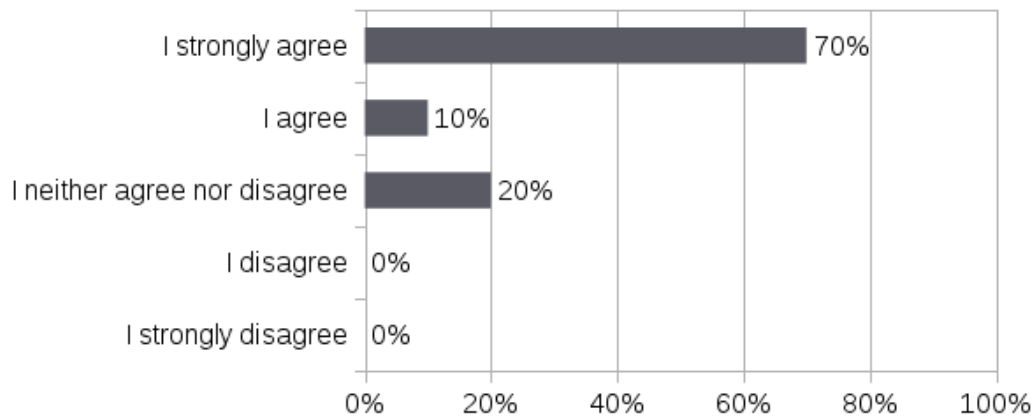
Figure 6.27: Survey results for *"I can easily find contents in ALOE@KM regardless of their popularity"*

6.8.3 Collective Intelligence

As a basis for the potential manifestation of Collective Intelligence, it can be stated that a significant number of users participated, and contributed a large number of artefacts in the variety of scenarios where ALOE was used. The survey results shown in Figure 6.28 and Figure 6.29 provide evidence that most users consider the metadata created in this way as being of good or even high quality.
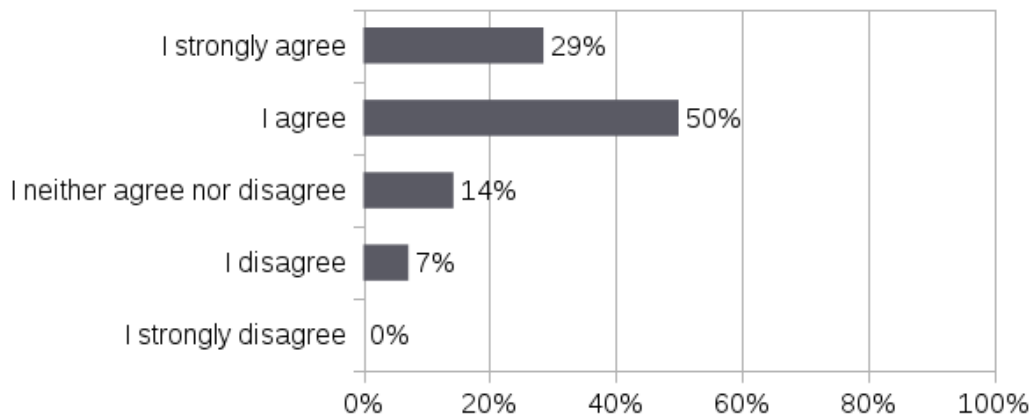
Figure 6.28: Survey results for *"The quality of information (metadata) about contents in Mindpool is high"*
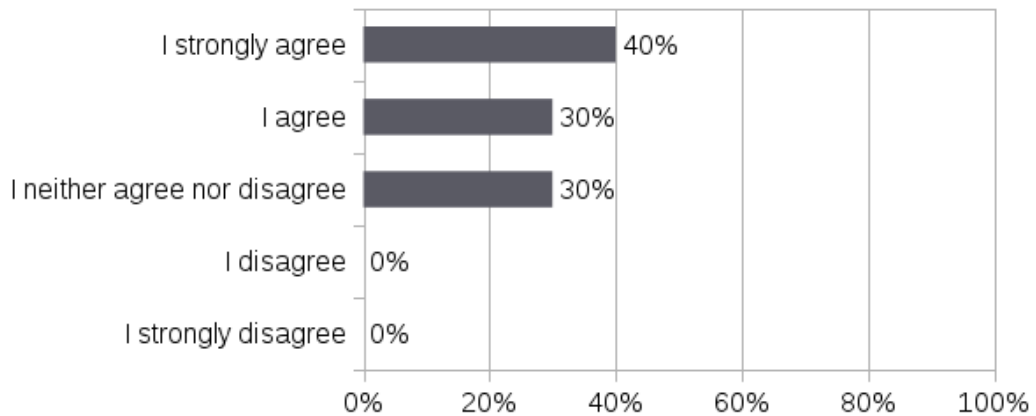
Figure 6.29: Survey results for *"The quality of information (metadata) about contents in ALOE@KM is high"*

To provide further information about whether contributions in ALOE instances can be considered as intelligent, results about the use of social data in

MACE are now presented.

Social Data as a Basis for Searching, Browsing, and Relevance Estimation in MACE

Usage Data   For the following calculations, stored CAM instances concerning search activities were used. Thereby, search sessions instead of single search action were considered. That is to say, a search (e.g., clicking on a classification term in the classification based search or entering the name of a country in the location based search) and its refinements (e.g., further clicking in the classification hierarchy or zooming into the map) are regarded as one action. A search session ends when the user accesses a learning resource or when he/she leaves the search page. This approach was chosen to achieve the comparability of the different searches in MACE, as for some actions, like the filtered search, more fine-grained CAM instances were collected as for others, like the location based search were zooming is not monitored.

Figure 6.30 depicts the relative frequency of the different searching and browsing facilities as already presented.
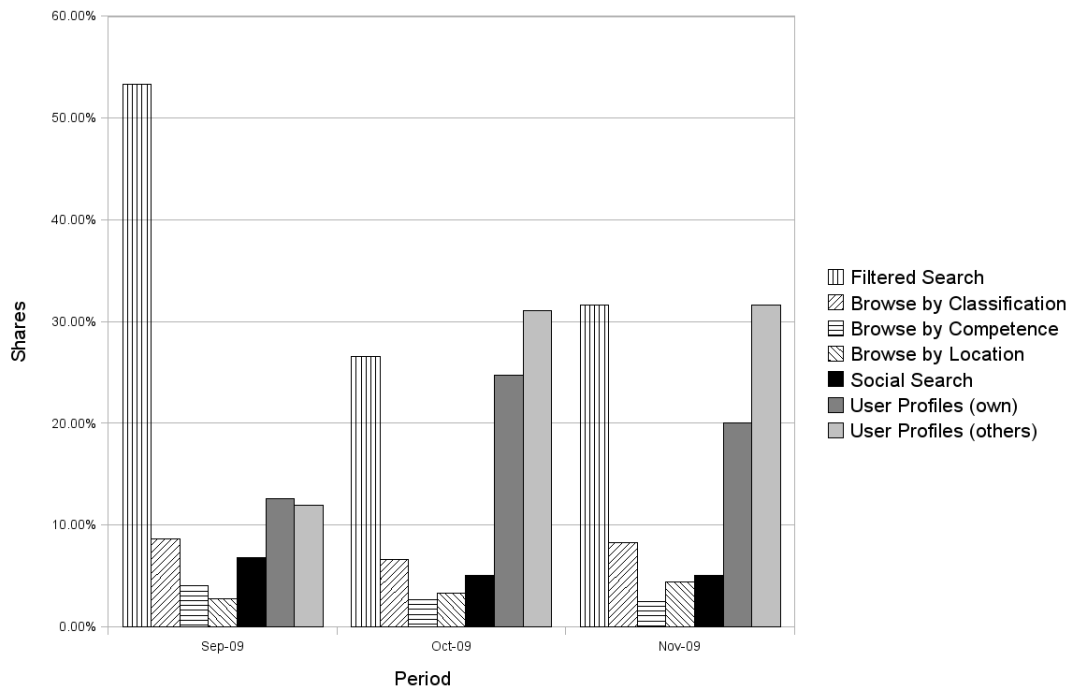
Figure 6.30: Relative frequency of search and browse types in MACE

The data clearly shows that *Filtered Search* (subsuming the simple keyword search) has been the most used tool during the examined period. Browsing in the own user portfolios and in user portfolios of other users is also used significantly more often than other facilities that each made out less than 10% of all conducted actions.

Survey Data   Besides this quantitative data, information was also collected about the end users' perception of the usefulness of different services and metadata types. Therefore, a survey was conducted with 14 participants of a MACE competition carried out in August and September 2009. The competition was carried out with teams of 20 Urban Design students from the University of Kaiserslautern. Their task was to gather and annotate resources about culturally relevant buildings in the areas Rhineland-Palatinate and Saar-Lor-Lux.

For the survey, a questionnaire was used that was made up of a rating scale based on 2 items. Utilising a 5-point scale, the users had to rate the importance of

- different search and browse facilities within MACE to find contents related to the impromptu, and

- different kinds of resource metadata to judge the relevance of a resource.

Figure 6.31 summarises the average value of each feature, ranging from 1 to 5 corresponding to a scale ranging from "rather not important" to "very important".
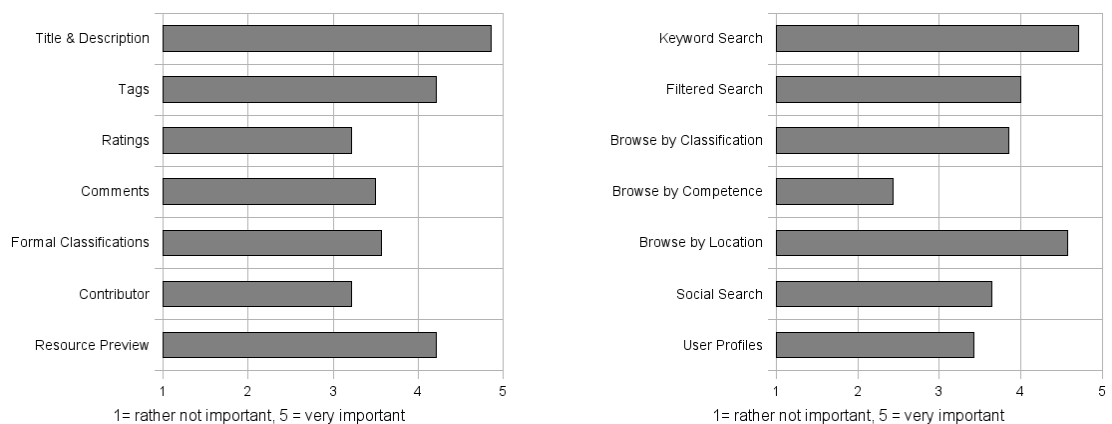
Figure 6.31: Survey results for the importance of information types and different search and browse types in MACE

Concerning the importance of search and browse facilities, *Keyword Search* and *Browse by Location* were rated as very important, and they were clearly considered as the most important facilities. Still, except for *Browse by Competence*, all other facilities were also rated as important, including *Social Search* and *User Profiles* that are based on social metadata.

To judge the relevance of a resource, *Title & Description*, *Tags*, and the *Resource Preview* were considered as the most important kinds of metadata, while all other types were rated similarly as less, but still clearly important.

From these evaluation results, it can be concluded that all social metadata is considered as important (especially tags associated with resources) to judge the relevance of a resource, and that the services based on social metadata are also considered as important means to search and browse contents in MACE.

Using Social Tags to Maintain the MACE Application Profile

An architectural project such as MACE constitutes a great syntheses effort, where different knowledge fields are involved – may they be connected to the poetic-artistic side (ideas, cultural and social message of a project) or to the technical one (functionality, living wellness, building ease). To find a coherent strategy to approach the knowledge organisation system of such a heterogeneous subject, several studies about semiotic interpretation of architectural knowledge were used as a basis to have the possibility to classify each aspect of the domain with semantic metadata (a detailed description can be found in [SCVN08]).

Thus, the MACE Application Profile (AP) was developed to organise all the architectural knowledge of MACE in a meaningful way. It contains a taxonomy whose categories are motivated and supported by the studies and the semiotic model previously mentioned. The AP is the base for the classification of all the contents harvested by MACE, and it is fundamental for several searching and browsing tools.

This profile enables to harmonise metadata descriptions of architectural contents. The AP is based on the Learning Object Metadata standard (LOM) with adaptations and extensions optimised for architecture and engineering.[27] The LOM standard organises metadata about learning objects in categories (e.g., identification, lifecycle, or classification), through which it is possible to express both media features and content features. Moreover, in the classification category, additional attributes from architectural taxonomies and classification systems were included.

---

[27]In LOM the metadata describing the learning objects are stored. Additionally, other metadata such as contextual, competency, and usage metadata is stored in special databases. See [APN+07] for a more detailed description.

Through this taxonomy (see Table 6.13) it was possible to classify and find all the LOs according to the principle described before.

Identification	Intervention type, project type, functional typology, form typology
Context	Location, geographic context, urban context
Technical design	Materials, construction form, building element, technological profile, structure profile, systems and equipments, technical performance, maintenance and conservation
Constructing	Construction management, construction phase, construction activity, machinery and equipments
Theories and concepts	Styles, periods and trends, theoretical concepts
Conceptual design	Project cues, project actions, form characteristics, perceptive qualities, relation with the context

Table 6.13: The categories featured in the MACE taxonomy grouped in six main facets

Besides the effort to initially create a wide, complete, and shared taxonomy (which was realised by combining several existing thesauri), the partners in the MACE consortium agreed that it must not be considered fixed and closed. In fact only with its use by the end users (teachers, students, etc.) it will be possible to understand and evaluate its soundness. For this reason, regular AP updates were scheduled every 6 months, the AP. To do this, the Protégé system[28] was used to modify or delete existing terms, but moreover to add new terms.

Consequently, an operational procedure was developed to manage the AP maintenance based on the communities' feedback. This feedback is not a direct suggestion coming from each user, but derived from the analysis of social metadata. Using the respective enriching interfaces, users are able to add freely chosen keywords (i.e., tags) to a specific resource, even if these keywords do not yet exist in the AP. During the maintenance work of the taxonomy, the AP panel (composed by experts of each specific domain) consulted the list of commonly

---

[28]See `http://protege.stanford.edu`

used tags to extract those that deserve to be inserted in the official taxonomy. The selected keywords tags were therefore inserted in the appropriate face of the taxonomy using the Protégé tool.

During the last MACE review operated in the final period of the project, these "unconscious suggestions" of expert users (adding metadata and keywords to contents during the scheduled enriching work) and students users (enriching contents during the evaluation phase of the project), the MACE AP taxonomy was extended with about 20 new terms. This could be regarded as a non-relevant number of terms, but if one considers the extent of the taxonomy (consisting of 2,850 terms) and its purpose (to be as much complete as possible), it is easy to understand that the possibility to find missing terms is not so high.

This hybrid approach combining a pre-defined top-down hierarchy and a bottom-up folksonomy allows us to utilise the wisdom of the crowds in a controlled manner to profit from existing personal knowledge. In this quality assured way, the taxonomy can be extended and improved over time, thus having the flexibility to adapt to emerging changes and arising innovations [SCVN08].

6.8.4 Crowdsourcing

As already argued in Section 6.8.3, the dashboard information as well as the Web analytics information available for the numerous ALOE instances clearly shows that it was possible to attract a significant number of users, and that these users contributed a large number of artefacts. Especially in the context of Mindpool these efforts can be considered as a successful crowdsourcing approach (see the dashboard information in Table D.4):

- More than 40 users tagged at least once (here, mainly users from the enterprise communication unit situated in different DFKI sites were involved).

- Resources were tagged more than 235,000 times (with more than 7,000 unique tags).

- More than 37,000 resources were tagged.

Summary   The feedback provided by the users and the statistical information allow to consider H2.1, H2.2, H2.3, and H2.4 to be valid.

6.8.5 ALOE Metadata as Basis for Further Applications

The metadata created within ALOE can also be used as a basis to realise advanced functionalities. These approaches are not in the focus of this work, but

are now briefly introduced to illustrate the potential added values beyond the benefits presented so far.

User profile extraction In [SBMD10], an approach is presented to extract contextualised user profiles according to the users' different topics of interest in ALOE. The approach analysed the social annotations as well as the preferred resources of each user and identifies thematic groups. For every group a weighted term vector was derived that represents the respective topic of interest. Each user profile consists of several such vectors that way enabling recommendation lists with a high degree of inter-topic diversity as well as targeted context-sensitive recommendations.

The proposed approach has been tested in ALOE@KM. Evaluations have shown that the method is likely to identify reasonable user interest topics, and that resource recommendations for these topics are widely appreciated by the users.

Resource recommenders A topic-based recommender system with the aim to provide personalised information was proposed in [SBMD11]. The approach applies algorithms from the domain of topic detection and tracking on the metadata profiles of the users' preferred resources to identify their interest topics. An evaluation of the approach has shown that the approach retrieved on-topic resources with a high precision.

Multi Source Tag Recommender In [MKS08] and [MKS09], a prototypical tag recommender system for ALOE was presented. The interactive system allows users to control the generation of the recommendations by selecting the sources to be used as well as their impact. The component was introduced at DFKI, and an evaluation showed that the recommender component was considered as helpful by a majority of users.

CHAPTER 7
Outlook and Conclusion

The previous chapters examined how the access to digital resources can be supported by means of a generic infrastructure that makes use of social media technologies, referred to as a Social Resource and Metadata Hub. The ALOE system was presented that realises such a hub. Several use cases where ALOE was deployed and used were presented, and the evaluation based on quantitative and qualitative data from the use cases provided evidence about the validity and usefulness of the approach.

To sum up the results, the major contributions of the thesis will be now presented. The thesis ends with an outlook and final remarks.

7.1 Conclusion

The **contributions** of this work are now presented along with their relations to the *research questions* introduced in Chapter 1 and the *hypotheses* summed up in Section 6.1.

In order to systematically identify what benefits social media technologies can provide for different aspects regarding the access to digital resources (*RQ1*), a **holistic architecture of information systems** and a **general access process for digital resources** in such systems were developed. The benefits and related challenges when aiming to realise them are summarised in Table 3.5, grouped by different social media phenomenons. Mainly the creation of social resources and social metadata has to be mentioned here.

RQ1

Together with the description of typical problems that can occur when accessing digital resources, these challenges and potentials were used as the basis to identify the **requirements for generic infrastructures that support access to digital resources by means of social media technologies, and that allow the introduction of social media technologies in existing digital environments** (*RQ2*).

RQ2

This leads to the **definition of a *Social Resource and Metadata Hub*** – an infrastructure that provides social media functionalities and allows to integrate existing resources and metadata for the specific needs of various application scenarios. The requirements for building such a system are assembled in a **system design checklist** in Section 4.8.

RQ3 The ALOE system was developed according to this checklist (*RQ3*). The system has been successfully deployed in a variety of projects and scenarios, and users have created hundreds of thousands of artefacts in the respective ALOE instances. The evaluation results from MACE and Mindpool as well as usage statistics provided from various other instances provide evidence that **ALOE technically realises a Social Resource and Metadata Hub** (*H3.1*). Information about the system design, how ALOE can be adapted and enhanced, and information about a variety of adapted instanced showed that **the approach is generic** (*H3.2*).

Especially the evaluations of the use cases MACE and Mindpool have proven that **a Social Resource and Metadata Hub can successfully tackle the problems of *abundance of digital resources*, *heterogeneous access to digital resources*, and *missing metadata for digital resources*** (*H1.1*, *H1.2*, and *H1.3*). These use cases together with the usage statistics from other ALOE instances have also served as the basis to show that **a Social Resource and Metadata Hub allows for exploiting social web phenomenons** (*H2.1*, *H2.2*, *H2.3*, and *H2.4*).

Benefits can not be guaranteed It is important to note that the simple deployment of ALOE – or any other information system that relies or even depends on the involvement and participation of users – is never enough to ensure that it provides the hoped-for effects that are technically possible. The wrong idea "If you build it, they will come"[1] is often the reason for the missing success of information systems, even if they technically work properly. As already stated, certain preconditions need to be fulfilled that cannot be enforced by means of a technical setup. Whether users participate or not depends on several factors influenced by the focus of a system, the context in which the application is deployed, the users that are involved, means to encourage participation, and so on. These aspects are extremely important, but they are beyond the scope of this thesis. In case of ALOE, the instances that were presented in more detail (e.g., Mindpool, MACE, and ALOE@KM) could be deployed under circumstances where a lot of users could be encouraged to participate. In MACE this was partly due to the fact that a lot of teaching institutions were involved, in Mindpool and ALOE@KM the user groups could directly be targeted, and the systems were advertised con-

---

[1]This adapted quote from the movie *Field of Dreams* is often used in the context of information systems

stantly. Yet, there are also ALOE instances where the participation of users was not reaching a significant level.

7.2 Outlook and Final Remarks

The ALOE system as a realisation of a Social Resource and Metadata Hub has proven to be a rich and powerful infrastructure that can be used to support the access to digital resources in a wide range of scenarios. It can be applied to integrate existing resources in contexts where they are distributed in heterogeneous sources with a lack of interaction possibilities, and also as an initially empty system that serves as a starting point for building up a "social repository". Due to the large number of projects where ALOE was employed as a basis for prototype and system development, continuous improvement and refactoring was possible, and the ecosystem of tools supporting the installation, maintenance, and adaptation of the environment has been continually growing. These are key factors when aiming at a sustainable infrastructure. Usually they are hard to achieve in a research context, even when the focus is put on application oriented research. In case of ALOE, it was only possible because the system was always intended as a generic system with the aim of potentially supporting any kind of digital resource. But this for sure was not enough – the most important part is that great people were involved. A special thanks to Rafael Schirru and Heinz Kirchmann that did an amazing job and have dedicated a lot of energy and passion to ALOE.

In the future, it will especially be interesting to see how the first approaches to add formal semantics into ALOE can be further developed. Automatically finding concepts in integrated resources, and the possibility to use inference mechanisms for finding related concepts has the potential of creating new and very interesting means to search and navigate resources.

IV
Appendices

A.1 Learning Object Metadata

The LOM format is intended to allow the description of digital resources that can be used to support learning. LOM defines a hierarchy of elements with nine top-level categories: *general*, *life cycle*, *meta-metadata* (information about the metadata), *technical*, *educational*, *rights*, *relation*, *annotation* and *classification*. The following table describes all LOM elements.

Nr.	Element name	Explanation
1	**General**	This category groups the general information that describes this learning object as a whole.
1.1	Identifier	A globally unique label that identifies this learning object.
1.1.1	Catalog	The name or designator of the identification or cataloging schema for this entry.
1.1.2	Entry	The value of the identifier within the identification or cataloging schema that designates or identifies this learning object.
1.2	Title	Name given to this learning object.
1.3	Language	The primary human language or languages used within this learning object to communicate to the intended user.
1.4	Description	A textual description of the content of this learning object.
1.5	Keyword	A keyword or phrase describing the topic of this learning object.

Nr.	Element name	Explanation
1.6	Coverage	The time, culture, geography or region to which this learning object applies.
1.7	Structure	Underlying organisational structure of this learning object.
1.8	Aggregation Level	The functional granularity of this learning object.
2	**Life Cycle**	This category describes the history and current state of this learning object and those entities that have affected this learning object during its evolution.
2.1	Version	The edition of this learning object.
2.2	Status	The completion status or condition of this learning object.
2.3	Contribute	Those entities (i.e., people, organisations) that have contributed to the state of this learning object during its life cycle (e.g., creation, edits, publication).
2.3.1	Role	Kind of contribution.
2.3.2	Entity	The identification of and information about entities (i.e., people, organisations) contributing to this learning object.
2.3.3	Date	The date of the contribution.
3	**Meta-Metadata**	This category describes this metadata record itself (rather than the learning object that this record describes).
3.1	Identifier	A globally unique label that identifies this metadata record.
3.1.1	Catalog	The name or designator of the identification or cataloging schema for this entry. A namespace schema.
3.1.2	Entry	The value of the identifier within the identification or cataloging schema that designates or identifies this metadata record. A namespace specific string.
3.2	Contribute	Those entities (i.e., people or organisations) that have affected the state of this metadata instance during its life cycle (e.g., creation, validation).
3.2.1	Role	Kind of contribution.

Nr.	Element name	Explanation
3.2.2	Entity	The identification of and information about entities (i.e., people, organisations) contributing to this metadata instance.
3.2.3	Date	The date of the contribution.
3.3	Metadata Schema	The name and version of the authoritative specification used to create this metadata instance.
3.4	Language	Language of this metadata instance.
4	**Technical**	This category describes the technical requirements and characteristics of this learning object.
4.1	Format	Technical datatype(s) of (all the components of) this learning object.
4.2	Size	The size of the digital learning object in bytes.
4.3	Location	A string that is used to access this learning object. It may be a location (e.g., Universal Resource Locator), or a method that resolves to a location (e.g., Universal Resource Identifier).
4.4	Requirement	The technical capabilities necessary for using this learning object.
4.4.1	OrComposite	Grouping of multiple requirements.
4.4.1.1	Type	The technology required to use this learning object, e.g., hardware, software, network, etc.
4.4.1.2	Name	Name of the required technology to use this learning object.
4.4.1.3	Minimum Version	Lowest possible version of the required technology to use this learning object.
4.4.1.4	Maximum Version	Highest possible version of the required technology to use this learning object.
4.5	Installation Remarks	Description of how to install this learning object.
4.6	Other Platform Requirements	Information about other software and hardware requirements.
4.7	Duration	Time a continuous learning object takes when played at intended speed.
5	**Educational**	This category describes the key educational or pedagogic characteristics of this learning object.

Nr.	Element name	Explanation
5.1	Interactivity Type	Predominant mode of learning supported by this learning object.
5.2	Learning Resource Type	Specific kind of learning object.
5.3	Interactivity Level	The degree of interactivity characterising this learning object. Interactivity in this context refers to the degree to which the learner can influence the aspect or behaviour of the learning object.
5.4	Semantic Density	The degree of conciseness of a learning object. The semantic density of a learning object may be estimated in terms of its size, span, or – in the case of self-timed resources such as audio or video – duration.
5.5	Intended End User Role	Principal user(s) for which this learning object was designed.
5.6	Context	The principal environment within which the learning and use of this learning object is intended to take place.
5.7	Typical Age Range	Age of the typical intended user.
5.8	Difficulty	How hard it is to work with or through this learning object for the typical intended target audience.
5.9	Typical Learning Time	Approximate or typical time it takes to work with or through this learning object for the typical intended target audience.
5.10	Description	Comments on how this learning object is to be used.
5.11	Language	The human language used by the typical intended user of this learning object.
6	**Rights**	This category describes the intellectual property rights and conditions of use for this learning object.
6.1	Cost	Whether use of this learning object requires payment.
6.2	Copyright and Other Restrictions	Whether copyright or other restrictions apply to the use of this learning object.

Nr.	Element name	Explanation
6.3	Description	Comments on the conditions of use of this learning object.
7	**Relation**	This category defines the relationship between this learning object and other learning objects, if any.
7.1	Kind	Nature of the relationship between this learning object and the target learning object, identified by 7.2:Relation.Resource.
7.2	Resource	The target learning object that this relationship references.
7.2.1	Identifier	A globally unique label that identifies the target learning object.
7.2.1.1	Catalog	The name or designator of the identification or cataloging schema for this entry.
7.2.1.2	Entry	The value of the identifier within the identification or cataloging schema that designates or identifies the target learning object.
7.2.2	Description	Description of the target learning object.
8	**Annotation**	This category provides comments on the educational use of this learning object, and information on when and by whom the comments were created.
8.1	Entity	Entity (i.e., people, organisation) that created this annotation.
8.2	Date	Date that this annotation was created.
8.3	Description	The content of this annotation.
9	**Classification**	This category describes where this learning object falls within a particular classification system.
9.1	Purpose	The purpose of classifying this learning object.
9.2	Taxon Path	A taxonomic path in a specific classification system.
9.2.1	Source	The name of the classification system.
9.2.2	Taxon	A particular term within a taxonomy. A taxon is a node that has a defined label or term.
9.2.2.1	Id	The identifier of the taxon, such as a number or letter combination provided by the source of the taxonomy.
9.2.2.2	Entry	The textual label of the taxon.

Nr.	Element name	Explanation
9.3	Description	Description of the learning object relative to the stated 9.1:Classification.Purpose of this specific classification, such as discipline, idea, skill level, educational objective, etc.
9.4	Keyword	Keywords and phrases descriptive of the learning object relative to the stated 9.1:Classification.Purpose of this specific classification, such as accessibility, security level, etc., most relevant first.

Table A.1: The LOM v1.0 base schema [oEL02]

A.2 DaMiT User Metadata

In DaMiT, the metadata schema used to represent users in the system was a slightly adapted subset of the IMS LIP schema. The following categories were defined:

General: General information about the user.

Identification: Information allowing to identify and contact a user.

Accessibility: This part contains settings defined by a user, language information, technical information etc.

Security Key: Here, information such as passwords and security keys are stored.

Competency: This category offer information about competencies acquired by a user.

Goal: A user's goals.

The single categories will now be presented in the following tables.

A.2.1 General

name	meaning	domain	role
idtype	id-Typ (no primary key!)	matrikel social identity DB-generated	
id	user id	\<text\>	
registration_time	time of registration	DB-timestamp	
user_role[1]	user role	admin content_provider learner_anonymous learner_pseudonymous learner_standard_academic learner_commercial learner_educational	

---

[1]defined which functionalities (e.g., chat, forum) will be offered to the user

name	meaning	domain	role
		manager teacher	
learner_group[2]	user group	al02 ...	

<p align="center">**Table A.2:** DaMiT user metadadata, category *General*</p>

A.2.2 Identification

name	meaning	domain	role
[address]*			
address_type	address type	bill home work	
street	street	<text>	
street_number	house number	<text>	
postcode	post code	<text>	
city	city	<text>	
country	country	ISO 3166	
address_add	additional address information	<text>	
[contact]*			
contact_type	type of contact information	bill home work	
email	email address	<text>	
phone	phone number	<text>	
mobile	mobile number	<text>	
[name]			
first_name	first name	<text>	
last_name	last name	<text>	
[demographics]			
birthday	birthday	dd.mm.yyyy	

---

[2]defined which content may be accessed

name	meaning	domain	role
gender	gender	male female	

Table A.3: DaMiT user metadadata, category *Identification*

A.2.3 Accessibility

name	meaning	domain	role
[language]			
language	chosen language	*de* en	
[preference]			
difficulty	preferred level of difficulty	*basic* advanced	
presentation	presentation type	*embedded* illustrated	
[technical]			
connection	connection type	slow *medium* fast	
formula_mode	representation of formulas	*gif* math_ml	
resolution	screen resolution	? (640x480 etc.)	1024x768
[disability]			
disability*	disabilities	*none* deaf blind	

Table A.4: DaMiT user metadadata, category *Accessibility*

A.2.4 Security

name	meaning	domain	role
login	unique login	<text>	
password	encrypted password	<text>	
certificate	certificate	X.509	

Table A.5: DaMiT user metadadata, category *Security*

A.2.5 Competency

name	meaning	domain	role
knowledge	state of knowledge	basic intermediate advanced	

Table A.6: DaMiT user metadadata, category *Competency*

A.2.6 Goal

name	meaning	domain	role
date	when shall the goal be achieved?	dd.mm.yyyy	
status	status	inactive active complete	
type	goal type	content difficulty	
priority	goal priority	dd.mm.yyyy	
description	goal description	<text>	

Table A.7: DaMiT user metadadata, category *Goal*

A.3 ALOE Resource Metadata

On the following pages, a detailed description of all ALOE resource metadata elements (including occurrences, value spaces, and data types) is provided.

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
1	basicMetadata	basic resource metadata	1	-	-	-
1.1	dc:identifier	URI identifying the resource uniquely in ALOE[3]	1	ALOE	URI	Character String
1.2	uri	URI identifying the resource uniquely	1	ALOE/ User[4]	URI	Character String
1.3	visibility	the resource visibility	1	User	*public* (default), *private*, or the identifier of a closed ALOE group	Character String
1.4	dc:format	either MIME type or a proprietary format	0..1	ALOE	-	Character String
1.5	resourceThumbnail	an image file associated with the resource	0..1	ALOE	-	PNG
1.6	individualMetadata	a set of metadata elements that can be specified by each contributing user	1..∞	-	-	-

[3]This identifier will automatically be generated by ALOE. Every combination of *uri* and *visibility* is a unique resource within ALOE.
[4]In case of a file upload, the URI is automatically generated by ALOE.

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
1.6.1	contributor	ALOE identifier of the user who contributed the additional meta-data (and added the resource to his/her ALOE portfolio)	1	ALOE	ALOE userId	Character String
1.6.2	contributionDate	contribution date	1	ALOE	ISO8601	DateTime
1.6.3	dc:publisher	entity responsible for making the resource available	0..1	User	-	Character String
1.6.4	dc:title	title of the resource	1	User	-	Character String
1.6.5	dc:description	a free-text description of the resource	0..1	User	-	Character String
1.6.6	dc:date	creation date	1	ALOE	ISO8601	DateTime
1.6.7	dc:creator	author(s) of the resource (not to be confused with the publisher)	0..∞	User	-	Character String
1.6.8	dc:language	language of the resource content	0..1	User	ISO639	Character String

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
1.6.9	dc:rights	license which is associated with the resource.	0..1	User	Any Creative Commons licenses as well as *unknown* (default), *none* (*all rights reserved*) and *public domain*	Character String
1.6.10	rightsholder	rightsholder of the resource	0..1	User	-	Character String
1.6.11	dc:source	origin of the resource	0..1	User	-	Character String
2	tag	tag associated with the resource	0..∞	-	-	-
2.1	dc:creator	ALOE identifier of the user who contributed the tag	1	ALOE	ALOE userId	Character String
2.2	dc:date	contribution date	1	ALOE	ISO8601	DateTime
2.3	dc:subject	the tag associated with the resource	1	User	-	Character String
3	comment	comment associated with the resource	0..∞	-	-	-
3.1	dc:creator	ALOE identifier of the user who contributed the comment	1	ALOE	ALOE userId	Character String

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
3.2	dc:date	contribution date	1	ALOE	ISO8601	DateTime
3.3	commentText	the comment text associated with the resource	1	User	-	Character String
4	rating	rating associated with the resource	0..∞	-	-	-
4.1	dc:creator	ALOE identifier of the user who contributed the rating	1	ALOE	ALOE userId	Character String
4.2	dc:date	contribution date	1	ALOE	ISO8601	DateTime
4.3	ratingValue	rating value associated with the resource	1	User	1..5	Integer
5	usageMetadata	usage metadata associated with the resource	1	-	-	-
5.1	averageRating	the average rating of the resource	1	ALOE	-	Float
5.2	timesBookmarked	the number of users that have the resource in their portfolios	1	ALOE	-	Integer
5.3	timesCommented	the number of comments left on the resource	1	ALOE	-	Integer

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
5.4	timesViewed	the number of times the resource was viewed (views by contributors are not counted in ALOE)	1	ALOE	-	Integer
6	group	group the resource is part of	0..∞	User	-	-
6.1	groupId	the identifier of the group	1	User	ALOE identifier of a group	Character String
6.2	sharingDate	date the resource was shared to the group	1	ALOE	ISO8601	DateTime
7	collection	collection the resource belongs to	0..∞	User	-	Character String
7.1	collectionId	collection the resource belongs to	1	User	ALOE identifier of a collection	Character String
7.2	addingDate	date the resource was added to the collection	1	ALOE	ISO8601	DateTime
8	associatedMetadata	additional metadata set associated with the resource[5]	0..∞	User	ALOE identifier of a metadata set	Character String

[5]Currently, no license information can be provided here. Yet, enhancements basing on approaches such as Open Data Commons (http://www.opendatacommons.org) are possible.

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
8.1	dc:identifier	URI identifying the metadata set uniquely	1	ALOE	-	Character String
8.2	dc:publisher	ALOE identifier of the user who added the metadata set	1	ALOE	ALOE userId	Character String
8.3	dc:creator	author(s) of the metadata set (not to be confused with the publisher)	0..∞	ALOE	-	Character String
8.4	dc:date	contribution date	1	ALOE	ISO8601	DateTime
8.5	dc:description	a free-text description of the metadata set	0..1	User	-	Character String
8.6	dc:format	metadata format	0..1	User	-	Character String
8.7	relationType	the type of relation between the resource and the metadata set	0..1	ALOE/ User	-	Character String
8.8	provenance	provenance information	0..1	ALOE/ User	-	Character String
9	category	generic resource classification	0..∞	User	-	-

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
9.1	taxonomyId	the identifier of a taxonomy the category refers to	1	ALOE/ User	-	Character String
9.2	categoryId	the identifier of a category in the specified taxonomy	1	ALOE/ User	-	Character String
9.3	name	the display name of a category	1	ALOE/ User	-	Character String
9.4	description	the description of a category	1	ALOE/ User	-	Character String
9.5	resourceCategory Relation	relation to category	$0..\infty$	ALOE/ User	-	-
9.5.1	contributor	ALOE identifier of the user who contributed the relation	1	ALOE	ALOE userId	Character String
9.5.2	addingDate	relation contribution date	1	ALOE	ISO8601	DateTime
9.5.3	relationType	the type of relation between the resource and the specified category	0..1	ALOE/ User	-	Character String
9.5.4	weight	a weight associated with the relation	0..1	ALOE/ User	-	Integer

Nr	Element Name	Explanation	Occurrence	Creator	Value space	Data type
9.5.5	provenance	provenance infor-mation	0..1	ALOE/ User	-	Character String

Table A.8: ALOE resource metadata

The ALOE Database Design

In the following, information about the ALOE database design is provided for user histories and different entity types.

B.1 ALOE Database Schema – User Histories
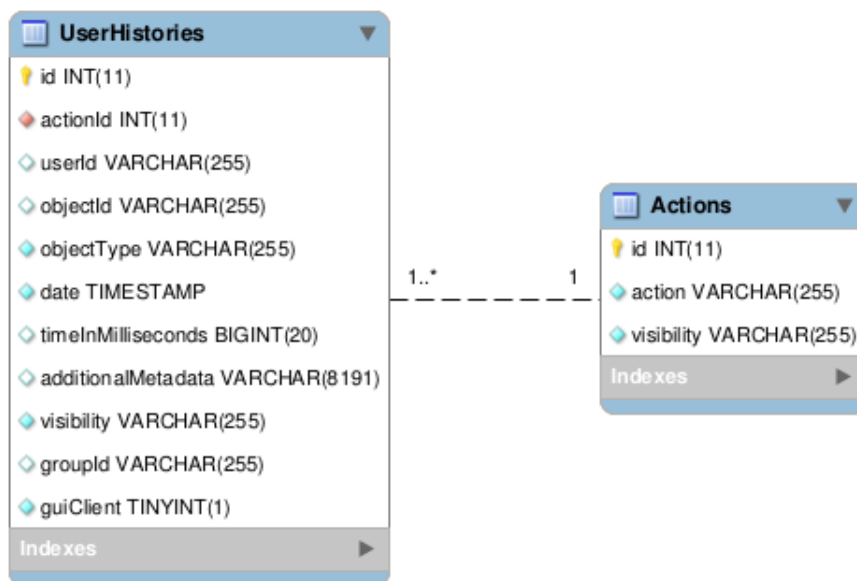
Figure B.1: Excerpt from the ALOE database schema for user histories

B.2 ALOE Database Schema – Resources
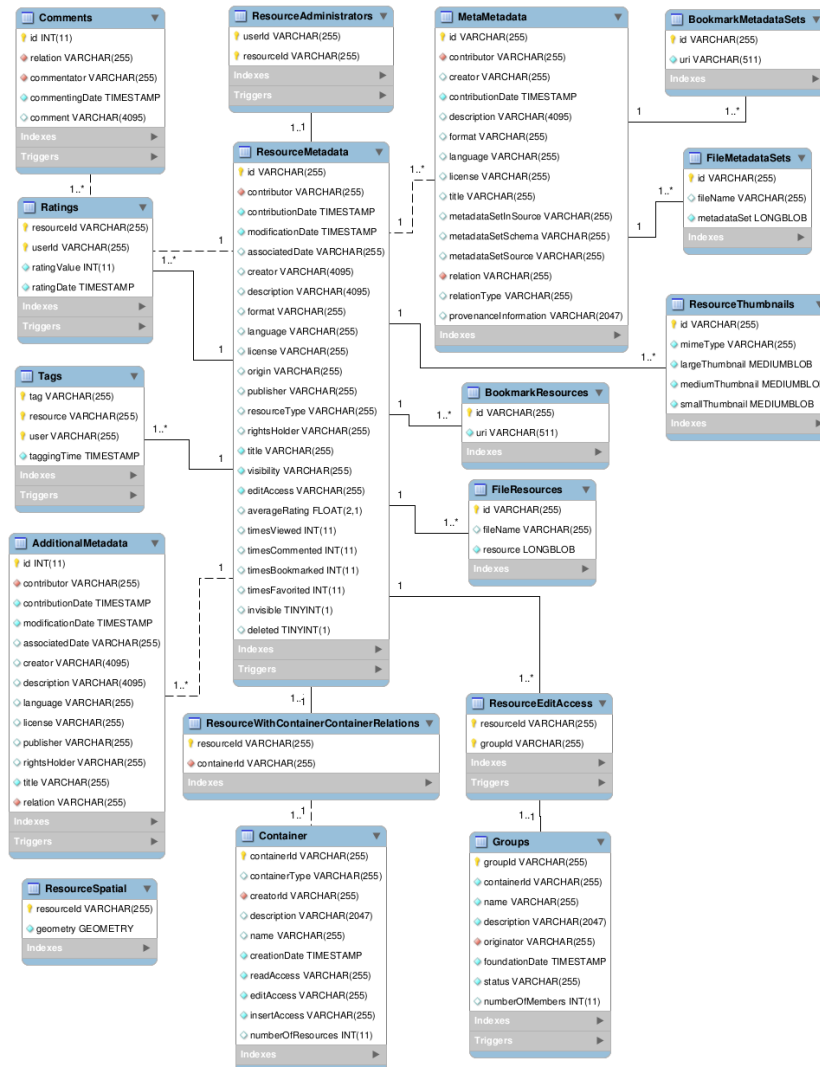
Figure B.2: Excerpt from the ALOE database schema for resources

B.3 ALOE Database Schema – Resource Types
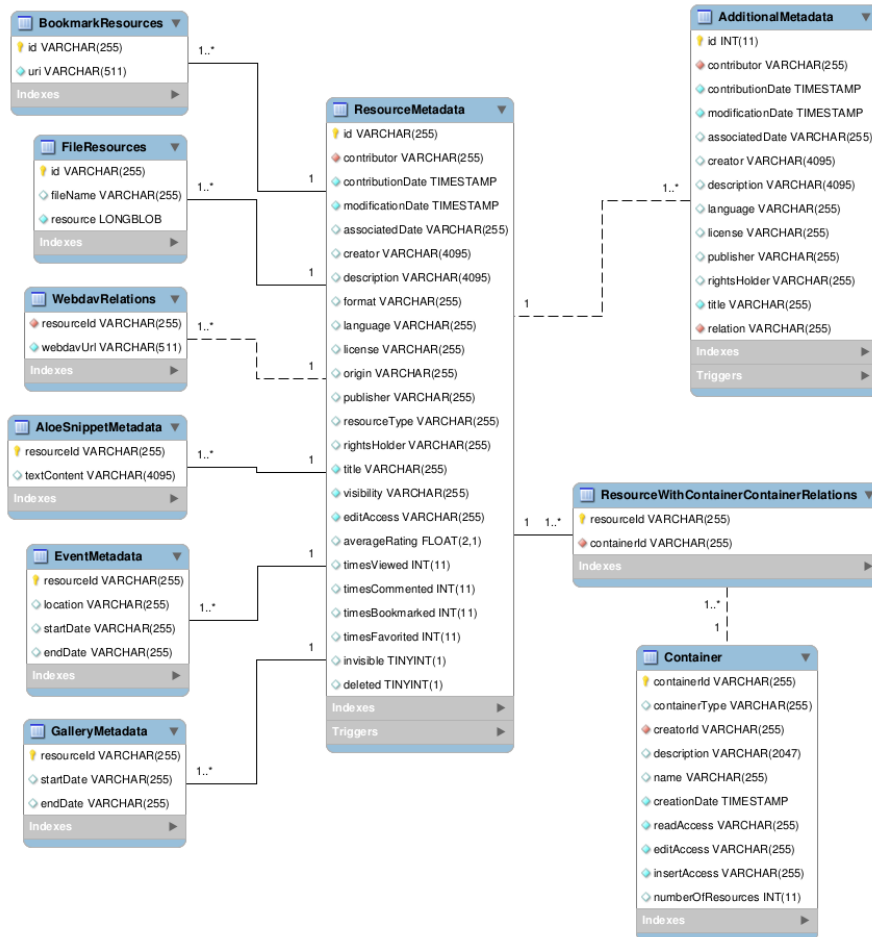
Figure B.3: Excerpt from the ALOE database schema for resource types

B.4 ALOE Database Schema – Containers
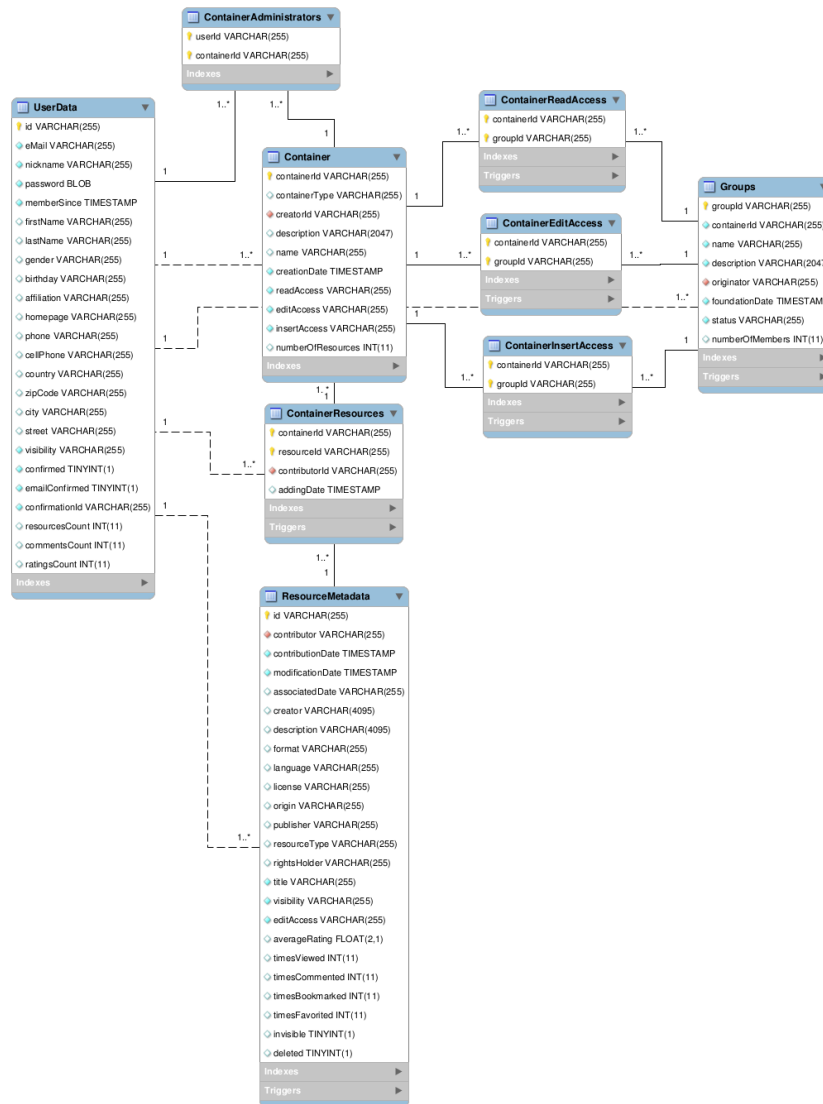
Figure B.4: Excerpt from the ALOE database schema for containers

B.5 ALOE Database Schema – Categories
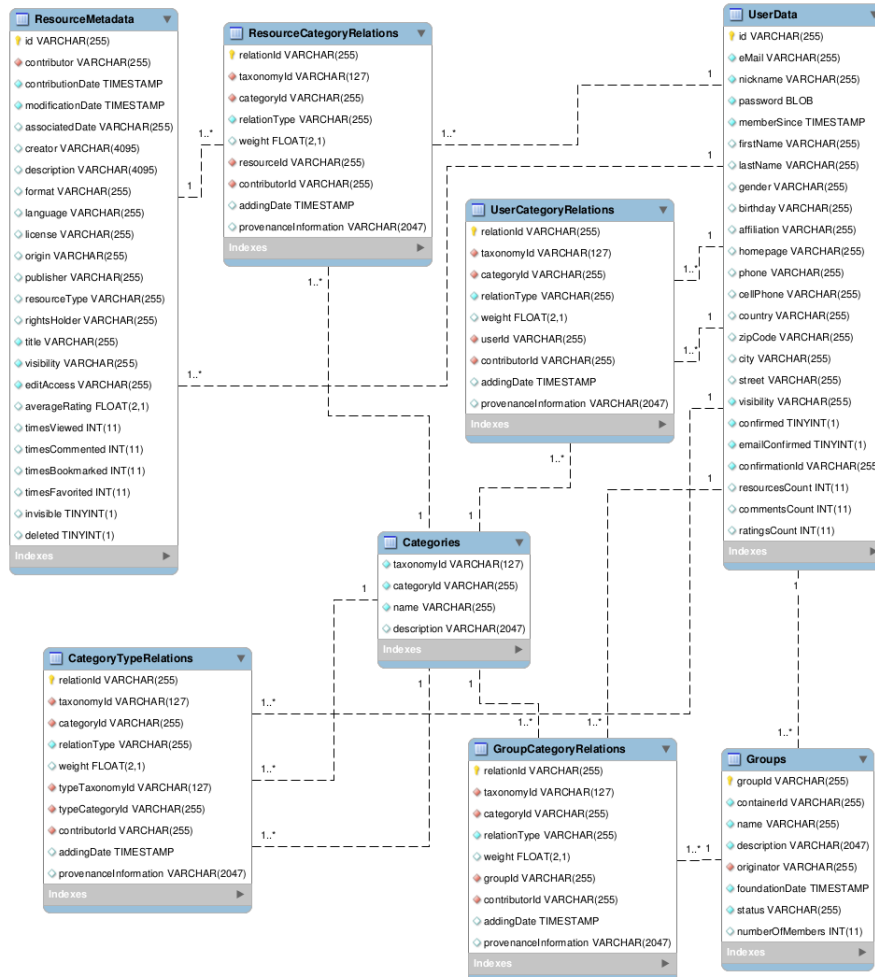
Figure B.5: Excerpt from the ALOE database schema for categories

B.6 ALOE Database Schema – Users
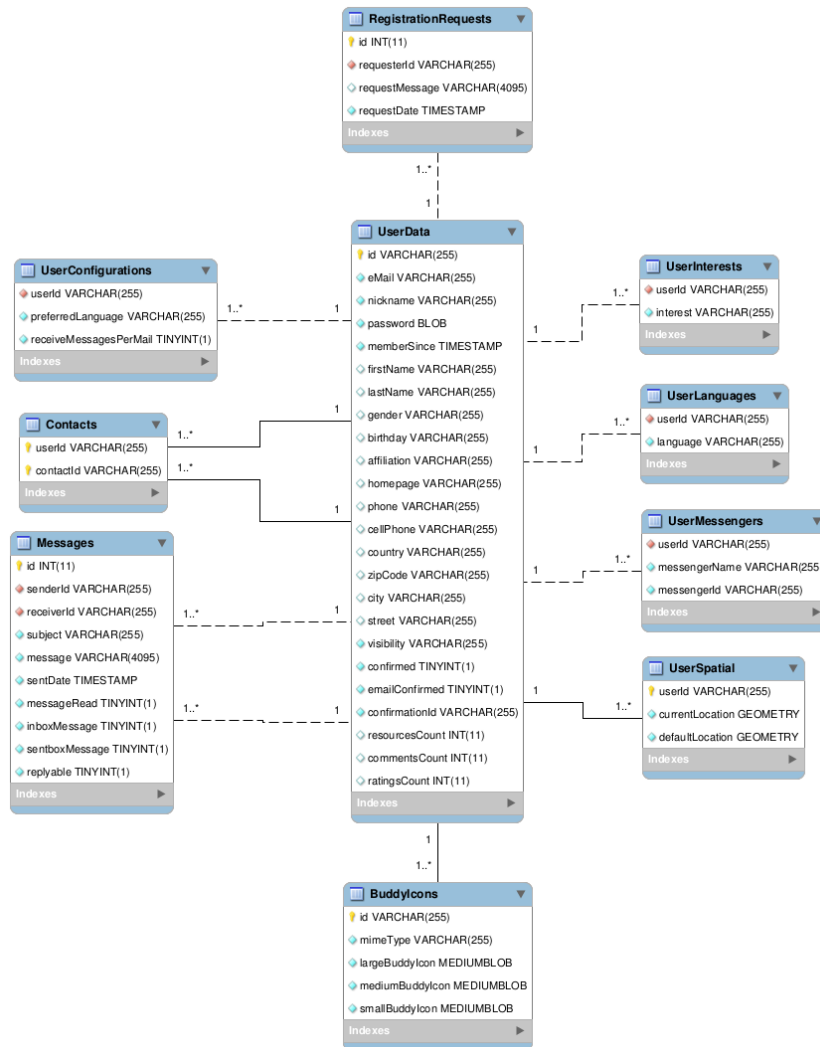
Figure B.6: Excerpt from the ALOE database schema for users

B.7 ALOE Database Schema – Groups
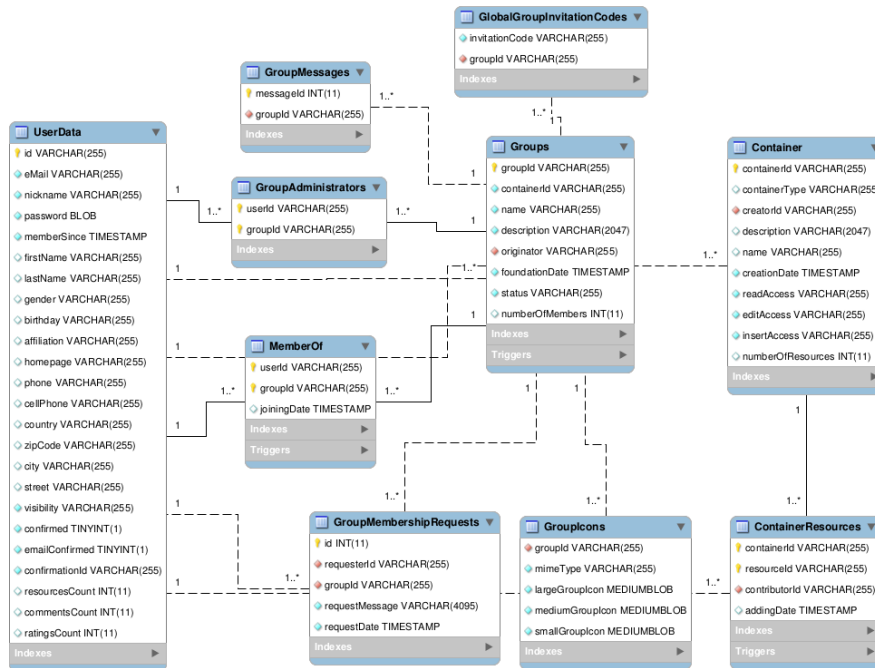
Figure B.7: Excerpt from the ALOE database schema for groups

ALOE Terms of Service and ALOE Privacy Policy

When users want to register in an ALOE instance, they need to agree to the ALOE Terms of Service and the ALOE Privacy Policy. The creation of these texts were coordinated with the DFKI management board, DFKI's legal department, and DFKI's data protection officer. In different ALOE instaces, the texts usually only differ regarding the name of the instance. We will here as a sample provide the texts used in the ALOE-public instance.

C.1 Terms of Service

Standard Terms of Use

Preamble

The following Standard Terms of Use shall apply to the relationship between DFKI GmbH, represented by its managing directors Dr. Walter Olthoff and Prof. Dr. Wolfgang Wahlster, Trippstadter Str. 122, 67663 Kaiserslautern, Germany, as the provider of the web based platform "ALOE" (hereinafter referred to as ALOE) and registered users (hereinafter referred to as the user) and shall apply to all offerings of ALOE and to the services offered on the associated sub-pages and additional pages.

§1 Description of the ALOE Service

1.1 ALOE makes a platform available to users who wish to save and manage digital resources and links in a personal account. Users not only have access to their own saved content but also to the content saved and managed by other users. Every user can decide whether to save a resource or link for him or herself, for members of selected groups or publicly for other users. ALOE makes it possible to exchange such resources

and links with other users, to evaluate them, and to provide them with personal comments.

1.2 If a user does not consent to the amended Terms of Use, he can terminate his ALOE membership by cancellation or serving notice. The user is committed to abide by the amended Terms of Use if he continues to use the ALOE website after one month following announcement of the amendments.

1.3 The provider of ALOE reserves the right to change its offerings as and when required.

1.4 The provider of ALOE also reserves the right to delete or privatize members' names and user accounts, keywords (tags), URLs (links), comments, resources, collections and groups if they:

- have racist, sexist, insulting, slanderous, inciting, religiously injurious, or threatening content

- have content that is criminally relevant in any way or violate legal regulations

- may be regarded as spam (see §8 Anti-Spam Guidelines).

- improperly use the name of a natural person or legal entity or any other legally protected names

- misuse the function for sending messages to other users for the purpose of distributing spam, advertising, or bulk messages.

- affect or compromise the functionalities offered by ALOE, e.g., by automatically generating information such as tags and comments.

§2 Registration; Conclusion of a Contract

2.1 Part of the ALOE platform is accessible without registration as a user. However, in order to save and manage personal content on the ALOE platform it is necessary to register as a user.

2.2 A contract concerning use of the ALOE website comes into being upon submission of registration after prior acceptance of the present Standard Terms and Conditions.

2.3 The User can cancel his registration with ALOE at any time in text form (e.g. letter, fax, email) without having to state reasons. Cancellation must be sent to:

DFKI GmbH
Chief Executive Officers:

Dr. Walter Olthoff, Prof. Dr. Wolfgang Wahlster

Address:
Trippstadter Str. 122
67663 Kaiserslautern

Telefax:
0631-20575-5030

E-Mail:
aloe@dfki.uni-kl.de

§3 Users' Rights and Duties

3.1 Registered users shall be entitled to use all ALOE services in full within the scope of these Terms of Use.

3.2 When ALOE services are being used, general laws must be observed. In particular, users are not allowed to use the ALOE website

- in order to disseminate material that is either illegal or in any other way defamatory or offensive

- in order to threaten or harass third parties or to violate third-party rights (especially the general right of personal privacy, copyrights and any other rights involving industrial rights protection)

- in order to upload to the ALOE website any data that is either infected or harmful in any other way

- in order to make incorrect or erroneous entries concerning their person upon registration, especially registering with a false identity.

3.3 Notwithstanding any consequences in common law and/or criminal law, non-observance of the above code of conduct shall entitle the provider of ALOE to terminate the relevant user's access without notice and to permanently exclude the same from use for the future, or to restrict the visibilty of the information he provided. The method of sanctioning is at the discretion of the ALOE provider.

§4 Rights and Duties of the ALOE provider

4.1 The provider of ALOE is not obliged to make its website and services available at all times or to ensure that it can be used without errors. This particularly applies if access to the ALOE website and services is affected by disruptions that are beyond the

influence of the provider of ALOE. Furthermore, the provider of ALOE shall be entitled, at all times and without prior notice, to temporarily completely discontinue or restrict the availability of its website for maintenance or modernization work.

4.2 The provider of ALOE shall be entitled, at all times and without prior notice, to limit extent of its website and services, or to completely shut down its website and services.

4.3 The provider of ALOE shall be entitled, in response to such a request, to surrender the data saved by the user in the event of any criminally relevant actions on the part of the user (especially such concerning pornographic or anti-alien content) and any other common-law violations, especially with respect to trademark law, competition law or copyright law concerning third parties (e.g. criminal prosecution authorities, parties involved or injured directly or indirectly).

4.4 The provider of ALOE shall be entitled but not obliged to check the user's bookmarks, links and any other content (hereinafter referred to as "content") with regard to compliance with general laws and these Terms of Use and to delete such content at its own discretion if a violation is revealed.

§5 Guarantee and Liability

5.1 The provider of ALOE does not give any guarantee that the website and services made available by it is available at all times without any errors or that it is suitable for certain purpose. This particularly applies if access to the ALOE website is affected or unavailable on account of disruptions, the cause of which is beyond the influence of the provider of ALOE.

5.2 The provider of ALOE gives no guarantee for content provided by users and in particular it is not obliged to check such content for legal violations.

5.3 If third parties claim the provider of ALOE because of illegal contents by users, the responsible user is obliged to exempt the provider of ALOE from any liability, and to compensate the provider of ALOE for all costs caused by a possible pursuit of legal claims.

5.4 The provider of ALOE is liable

- for any damage caused by harm to the body, life or health that is due to willful intent or negligent violation of duties on the part of the provider of ALOE, a legal representative of the provider of ALOE, or a universal agent of the provider of ALOE;

- for any damage that is covered by liability under the German Product Liability Act;

- for any other damage that is due to willful or grossly negligent violation of duties on the part of the provider of ALOE, a legal representative, or a universal agent of the provider of ALOE;

- for any damage that is caused by simple negligence on the part of the provider of ALOE, inasmuch as the negligence involves violation of material contractual duties (cardinal duties). Cardinal duties are such duties as the contract has to grant the contracting party according to the spirit and meaning of the contract, or ones whose fulfillment enables proper performance of the contract and on compliance of which the contracting party may rely. In such cases, however, the liability of ALOE shall be restricted to damage that is typically associated with the contract and is foreseeable.

5.5 The provision in §5.4 covers all contractual and legal claims that result from these Terms of Use and use of the service provided by ALOE. In all other cases the provider of ALOE explicitly accepts no liability. This particularly applies to any loss of data that has not been caused by willful intent on the part of the provider of ALOE and its employees.

§6 Data Protection

The provider of ALOE respects the privacy of its users. The provider of ALOE treats the personal data of its users confidentially and only makes such data available to third parties if this conforms to applicable law or the user has consented to data being passed on. Details concerning data protection at ALOE are contained in the Data Protection Regulations.

§7 Term of Contract and Termination

7.1 The contract of use is concluded for an unlimited period of time.

7.2 Either party may terminate this contract at any time without having to state reasons.

7.3 The provider of ALOE is also entitled to cancel membership at any time for a good reason. A good reason in particular is if the user has violated these Terms of Use or applicable law.

§8 Anti-Spam Guidelines

According to §1 para. 5 the provider of ALOE reserves the right, in the case of spamming, to perform permanent privatization (or deletion) of an account, a URL or a whole

domain (blacklist); these will then no longer appear publicly on the ALOE website.

Spam is defined by ALOE as follows:

- Use of a number of accounts in order to manipulate the popularity of a bookmark

- Saving pure affiliate links or partner links

- Made for AdSense websites which only serve the purpose of presenting AdSense, banners and affiliate links and do not have far-reaching personal content

- Get rich quick websites ('Get rich with only 5 minutes of work a day')

- Saving websites under construction and domain sale offers

- Saving an usually large number of bookmarks pointing to just one domain

- Use of advertising in the description of the bookmarks

- Unsuitable and/or erroneous descriptions of bookmarks that are not associated with the website proper

- Misuse in order to develop link popularity

- Links to pornographic, right-wing extremist, violence-glorifying, illegal, morally objectionable and/or unreasonable content

- Links to dialer sites, phishing sites or sites with other damaging content

- Automatically redirecting websites and websites that open a large number of advertising windows

- Saving bookmarks for advertising purposes

§9 Referrals and Links

The provider of ALOE has no influence whatsoever on the current or future design, content, or authorship of the linked sites. For this reason the provider of ALOE explicitly dissociates itself from all content on all linked sites. This shall apply to all links and references placed on its own website and to any third-party entries. With regard to any illegal, erroneous or incomplete content, and particularly any damage or loss that may result from use or any information not disclosed in such a way, liability is held solely by the owner of the site to which the linked pointed, and not the party that merely points to the respective publication in links.

§10 Copyright and Distinctive Mark Law

In all publications the provider of ALOE endeavors to observe the copyrights of the pictures, graphics, sound documents, video sequences, and texts used, to use pictures, graphics, sound documents, video sequences, and texts created by the provider of ALOE itself, or to resort to license-free graphics, sound documents, video sequences, and texts.

All brand names and trademarks indicated on the website and possibly protected by third parties are entirely subject to the provisions of distinctive mark law as amended and the property rights of the various registered owners. On the basis of a mere naming it is not possible to infer that brand names are not protected by third-party rights!

The copyright for published items created by the provider of ALOE itself remains solely with the provider of ALOE. Duplication and use of such graphics, sound documents, video sequences, and texts in other electronic or printed publications is not permitted without the explicit consent of the provider of ALOE.

§11 Final Provisions

11.1 In the event of disputes in connection with the ALOE website German law shall apply and the UN Convention on Contracts for the International Sale of Goods shall not apply.

11.2 Merchants, public law entities and special public corporations shall be subject to the jurisdiction of Kaiserslautern courts.

11.3 If individual clauses of these Terms of Use should be or become entirely or partially invalid, this shall not affect the validity of the other Terms of Use. If any provision should be invalid, a provision shall apply that commercially comes as close as possible to the spirit and meaning of the invalid provision. The same shall apply to any provisions omitted.

C.2 Privacy Policy

Data Protection

Data protection policy at ALOE conforms to the general data protection regulations in the German Federal Data Protection Act ('BDSG') and the German Telemedia Act ('TMG') for the individual use of combinable data such as characters, images and sounds and for the offering and use of information and communication services directed at the public comprising text, sound and images.

The collection, processing and use of your personal data is based on the strict provisions of German data protection law, the policies being data avoidance, data transparency, and data security. The data collected is processed and used by ALOE and its cooperating partners only to the extent that this is necessary, legally permissible, and desired by yourself for the fulfillment of contracts concluded with you and maintenance of customer relations resulting therefrom.

Inasmuch as the website provides an opportunity to enter personal or business data (email addresses, names, postal addresses) the disclosure of such data by the user is on an explicitly voluntary basis. Entries made must be truthful. Use of all services offered is also permitted without entering such data or by entering anonymized data or a pseudonym, provided it is technically possible and reasonable. The misuse of names belonging to a natural person or legal entity or any other legally protected names is not permitted and can lead to closure of the user account.

Use of the contact data published as site owner information or comparable details, e.g. postal addresses, telephone numbers, fax numbers, and email addresses by third parties for mailing information that has not been explicitly requested is not allowed. We reserve the right to take legal steps against senders of so-called spam mail if this prohibitive clause is violated.

Whenever the ALOE website is accessed, a data record is stored on the web server. This data record consists of data that is required in order to handle the user session properly. In the case of unregistered users these data records are evaluated for internal, statistical purposes only. No data concerning unregistered users is passed on to external third parties. With each page impression access data is also stored in a log file. These log files provide information about your IP address, the remote host, the time of day, the status, any data quantity transferred, and the website from which the user arrived on the ALOE website, as well as product/version information concerning the browser being used.

In addition we hereby inform all users of our website that ALOE uses so-called cookies to manage user sessions. Cookies are small files that are stored on the respective user's computer for the duration of the session. Session management is not possible without cookies.

This website uses Google Analytics, a web analytics service provided by Google, Inc. ('Google'). Google Analytics uses 'cookies', which are text files placed on your computer, to help the website analyze how users use the site. The information generated by the cookie about your use of the website will be transmitted to and stored by Google on servers in the United States.

In case IP-anonymisation is activated on this website, your IP address will be truncated within the area of Member States of the European Union or other parties to the Agreement on the European Economic Area. Only in exceptional cases the whole IP address will be first transfered to a Google server in the USA and truncated there. The IP-anonymisation is active on this website.

Google will use this information on behalf of the operator of this website for the purpose of evaluating your use of the website, compiling reports on website activity for website operators and providing them other services relating to website activity and internet usage.

The IP-address, that your Browser conveys within the scope of Google Analytics, will not be associated with any other data held by Google. You may refuse the use of cookies by selecting the appropriate settings on your browser, however please note that if you do this you may not be able to use the full functionality of this website. You can also opt-out from being tracked by Google Analytics with effect for the future by downloading and installing Google Analytics Opt-out Browser Addon for your current web browser: `http://tools.google.com/dlpage/gaoptout?hl=en`.

Personal data is only collected with your knowledge and your consent having been given by means of checkbox confirmation upon registration as a user. Upon informal request you can obtain free information about the personal data stored concerning you and you can also request deletion of your data at any time, either in writing or in the form of an email. For this purpose please contact our data protection officer: dsb@dfki.de.

D.1 Usage Data From ALOE-MACE

D.1.1 Created Artefacts

Type	Number
Users	
Registered users	1425
Resources	
Resources contributed	189022
Unique resources contributed	182147
File resources contributed	0
Users that contributed at least once	507
Tags	
Tags added	63435
Unique tags added	14585
Tagged resources	9968
Users that tagged at least once	435
Comments	
Comments	453
Commented resources	337
Users that commented at least once	78
Ratings	
Ratings	1856
Rated resources	1755
Users that rated at least once	131

Table D.1: Created entities and usage information for ALOE-MACE (without users 'Martin', 'zeus'), retrieved December 13th, 2014

D.1.2 Tracked Activities

The first action on this instance was tracked on May 5th, 2009.

Table D.2 was created by querying the UserHistories table in ALOE-MACE and shows the total number of tracked activities, the number of tracked gui activities, the number of users with tracked activities, and number of objects affected by the tracked activities. Table D.3 shows information about the ten most tracked activities.

all actions	gui actions	users	objects
2256777	526654	1135	158543

Table D.2: Information about activities tracked for ALOE-MACE (retrieved December 13th, 2014)

Action	Freq (gui)	Freq (all)
.../resource/viewed/	292549	1006583
.../resource/deleted/	129657	147644
.../user/login/	49738	150173
.../resource/tagged/	30993	31006
.../resource/contributed/bookmark/	9691	323388
.../resource/contributed/existingBookmark/	7176	7176
.../resource/tagDeleted/	3861	3861
.../user/register/	1005	1005
.../resource/rated/	706	706
.../user/addedToContacts/	551	551

Table D.3: Information about the ten most tracked activities for ALOE-MACE (retrieved December 13th, 2014)

D.1.3 Web Analytics Data
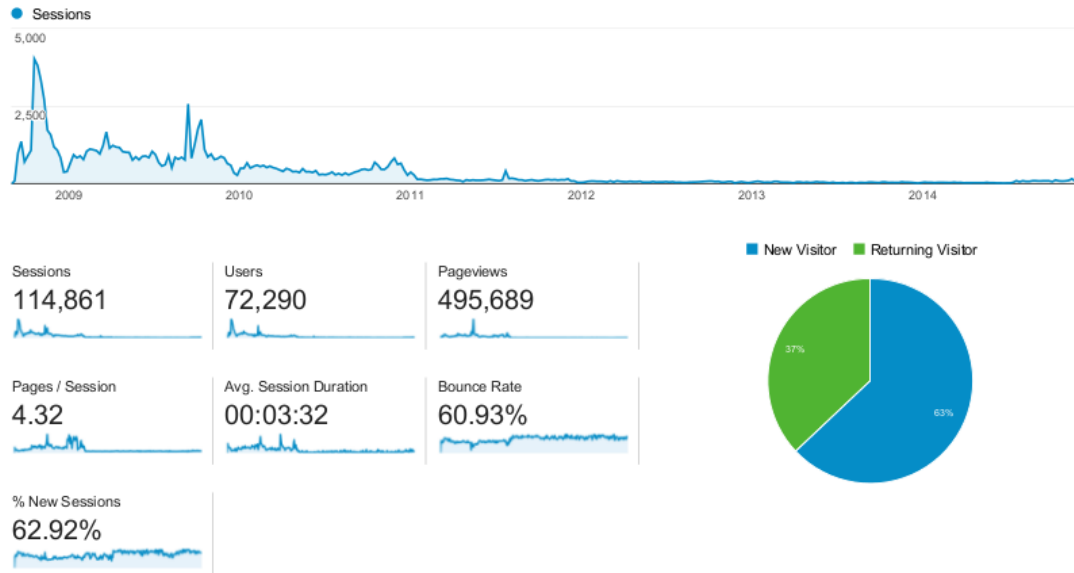
Figure D.1: Web Analytics data gathered for MACE (date range: 2008/09/01 - 2014/12/12)

Please note that the information gathered here concerns the complete MACE portal and not only the pages with contents and interaction means from the ALOE-MACE instance as part of this infrastructure.

D.2 Usage Data From Mindpool

D.2.1 Created Artefacts

Type	Number
Users	
Registered users	101
Sets	
Collections created	7
Resources added to collections	11
Groups	
Open groups created	7
Closed-public groups created	0
Closed-private groups created	2
Resources shared to open groups	26
Resources shared to closed-public groups	0
Resources shared to closed-private groups	14
Open group memberships	27
Closed-public group memberships	2
Closed-private group memberships	33
Resources	
Resources added by end users	16789
Unique resources contributed by end users	16789
File resources contributed by end users	333
Users that contributed at least once	39
Resource Types	
Resources added by end users (File)	333
Resources added by end users (Snippet)	1
Resources added by end users (Bookmark)	14244
Resources added by end users (Gallery)	187
Resources added by end users (DFKI Visit)	139
Resources added by end users (DFKI press coverage item)	1885
Tags	
Tags	236727
Unique tags	7187
Tagged resources	37717
Users that tagged at least once	41
Comments	
Comments	18
Commented resources	15
Users that commented at least once	9

Type	Number
Ratings	
Ratings	75
Rated resources	69
Users that rated at least once	14
Categories	
Categories	587
Relations between categories and resources	560
Unique categories for resources	84
Categorised resources	481
Users that categorised at least once	17
Persons	
Persons	15992
Relations between persons and resources	2522
Unique persons for resources	1039
Resources annotated with persons	1782
Users that added a person at least once	13

Table D.4: Created entities and usage information for ALOE-DFKI (without users 'Martin', 'zeus'), retrieved December 13th, 2014

D.2.2 Tracked Activities

The first action on this instance was tracked on January 24th, 2010.

Table D.5 was created by querying the UserHistories table in Mindpool and shows the total number of tracked activities, the number of tracked gui activities, the number of users with tracked activities, and the number of objects affected by the tracked activities. Table D.6 shows information about the ten most tracked activities.

all actions	gui actions	users	objects
350019	231400	102	42451

Table D.5: Information about activities tracked in Mindpool (retrieved December 13th, 2014)

Action	Freq (gui)	Freq (all)
.../search/	61157	68728
.../resource/metadataEdited/	48175	48175
.../resource/tagged/	44700	44700
.../resource/viewed/	31491	32132
.../resource/contributed/bookmark/	15125	35748
.../resource/addedToGallery/	14221	34844
.../user/login/	2850	3621
.../resource/metadataEdited/dfkiMediaContent/	2118	2118
.../resource/contributed/dfkiMediaContent/	1970	8536
.../resource/deleted/	1511	3482

Table D.6: Information about the ten most tracked activities in Mindpool (retrieved December 13th, 2014)

D.2.3 Web Analytics Data
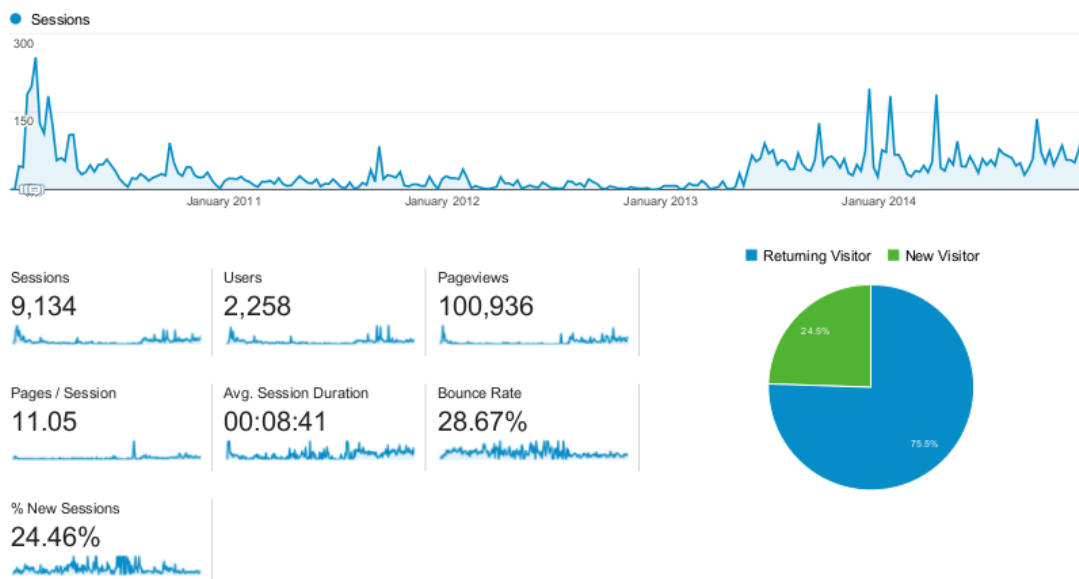
Figure D.2: Web Analytics data gathered for Mindpool (date range: 2010/01/10 - 2014/12/12)

D.3 Usage Data From ALOE@KM

D.3.1 Created Artefacts

Type	Number
Users	
Registered users	52
Sets	
Collections created	8
Resources added to collections	50
Groups	
Open groups created	26
Closed-public groups created	0
Closed-private groups created	5
Resources shared to open groups	1061
Resources shared to closed-public groups	0
Resources shared to closed-private groups	374
Open group memberships	145
Closed-public group memberships	4
Closed-private group memberships	32
Resources	
Resources added by end users	1951
Unique resources contributed by end users	1774
File resources contributed by end users	211
Users that contributed at least once	38
Resource Types	
Resources added by end users (File)	211
Resources added by end users (Bookmark)	1562
Resources added by end users (Snippet)	0
Resources added by end users (Event)	1
Tags	
Tags	7756
Unique tags	3275
Tagged resources	1830
Users that tagged at least once	38
Comments	
Comments	224
Commented resources	192
Users that commented at least once	24
Ratings	
Ratings	304

Type	Number
Rated resources	264
Users that rated at least once	28
Categories	
Categories	0
Relations between categories and resources	0
Unique categories for resources	0
Categorised resources	0
Users that categorised at least once	0
Persons	
Persons	0
Relations between persons and resources	0
Unique persons for resources	0
Resources annotated with persons	0
Users that added a person at least once	0

Table D.7: Created entities and usage information for ALOE@KM (without user 'Martin'), retrieved December 13th, 2014

D.3.2 Tracked Activities

The first action on this instance was tracked on February 12th, 2009.

Table D.8 was created by querying the UserHistories table in ALOE@KM and shows the total number of tracked activities, the number of tracked gui activities, the number of users with tracked activities, and number of objects affected by the tracked activities. Table D.9 shows information about the ten most tracked activities.

all actions	gui actions	users	objects
58597	44789	59	3242

Table D.8: Information about activities tracked in ALOE@KM (retrieved December 13th, 2014)

Action	Freq (gui)	Freq (all)
.../resource/viewed/	19694	24212
.../search/	13860	13877

Action	Freq (gui)	Freq (all)
.../resource/contributed/bookmark/	2826	2961
.../user/login/	2346	6863
.../resource/sharedToGroup/	1755	1840
.../resource/tagged/	895	895
.../resource/rated/	509	509
.../resource/commented/	427	427
.../resource/contributed/existingBookmark/	319	319
.../user/viewed/	256	256

Table D.9: Information about the ten most tracked activities in ALOE@KM (retrieved December 13th, 2014)

D.3.3 Web Analytics Data
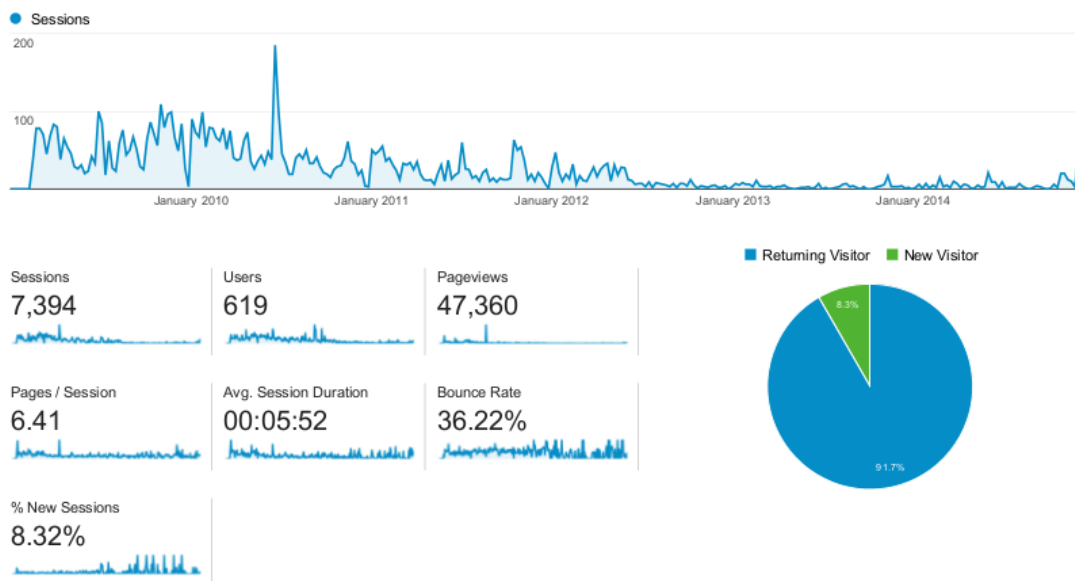
Figure D.3: Web Analytics data gathered for ALOE@KM (date range: 2009/01/01 - 2014/12/12)

D.4 Usage Data From ALOE-Public

D.4.1 Created Artefacts

Type	Number
Time span	
Users	
Registered users	52
Sets	
Collections created	8
Resources added to collections	50
Groups	
Open groups created	26
Closed-public groups created	0
Closed-private groups created	5
Resources shared to open groups	1061
Resources shared to closed-public groups	0
Resources shared to closed-private groups	374
Open group memberships	145
Closed-public group memberships	4
Closed-private group memberships	32
Resources	
Resources added by end users	1951
Unique resources contributed by end users	1774
File resources contributed by end users	211
Users that contributed at least once	38
Resource Types	
Resources added by end users (File)	211
Resources added by end users (Bookmark)	1562
Resources added by end users (Snippet)	0
Resources added by end users (Event)	1
Tags	
Tags	7756
Unique tags	3275
Tagged resources	1830
Users that tagged at least once	38
Comments	
Comments	224
Commented resources	192
Users that commented at least once	24
Ratings	

Type	Number
Ratings	304
Rated resources	264
Users that rated at least once	28
Categories	
Categories	0
Relations between categories and resources	0
Unique categories for resources	0
Categorised resources	0
Users that categorised at least once	0
Persons	
Persons	0
Relations between persons and resources	0
Unique persons for resources	0
Resources annotated with persons	0
Users that added a person at least once	0

Table D.10: Created entities and usage information for ALOE-Public (without user 'Martin'), retrieved December 13th, 2014

D.4.2 Tracked Activities

The first action on this instance was tracked on February 12th, 2009.

Table D.11 was created by querying the UserHistories table in ALOE-Public and shows the total number of tracked activities, the number of tracked gui activities, the number of users with tracked activities, and the number of objects affected by the tracked activities. Table D.12 shows information about the ten most tracked activities.

all actions	gui actions	users	objects
1164618	1153700	96	2841

Table D.11: Information about activities tracked in ALOE-Public (retrieved December 13th, 2014)

Action	Freq (gui)	Freq (all)
.../search/	832336	832501

Action	Freq (gui)	Freq (all)
.../resource/viewed/	282653	284834
.../searchInGroup/	33112	33112
.../user/login/	2038	6342
.../resource/contributed/bookmark/	1513	1516
.../user/logout/	1321	5586
.../user/viewed/	149	149
.../searchAdvanced/	134	134
.../resource/sharedToGroup/	99	99
.../resource/tagged/	97	97

Table D.12: Information about the ten most tracked activities in ALOE-Public (retrieved December 13th, 2014)

D.4.3 Web Analytics Data
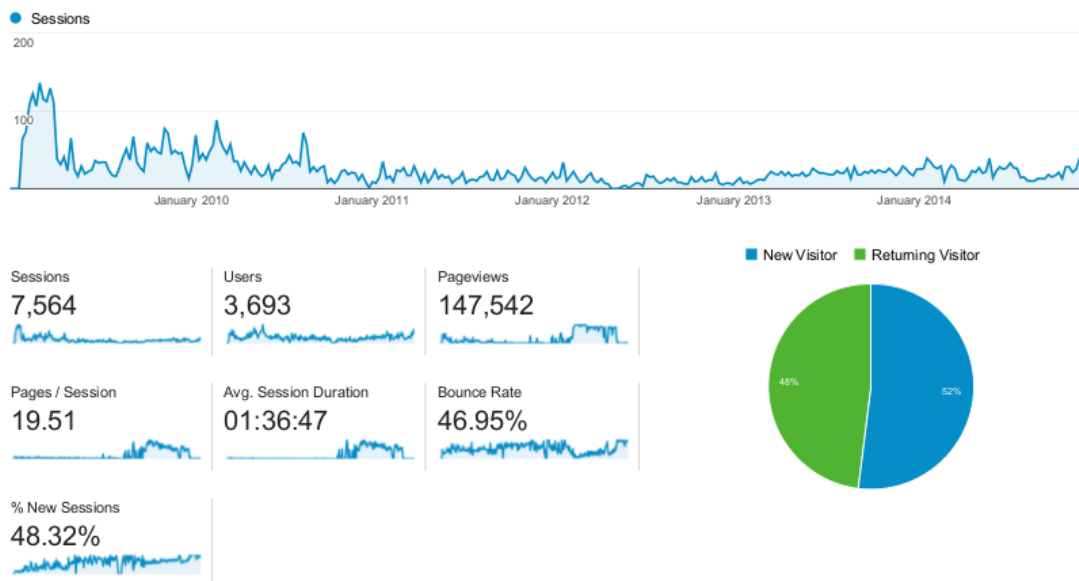
Figure D.4: Web Analytics data gathered for ALOE-Public (date range: 2009/01/01 - 2014/12/12)

D.5 Usage Data From ALOE-CMCM

D.5.1 Created Artefacts

Type	Number
Users	
Registered users	25
Sets	
Collections created	0
Resources added to collections	0
Groups	
Open groups created	0
Closed-public groups created	0
Closed-private groups created	0
Resources shared to open groups	0
Resources shared to closed-public groups	0
Resources shared to closed-private groups	0
Open group memberships	1
Closed-public group memberships	4
Closed-private group memberships	18
Resources	
Resources added by end users	11
Unique resources contributed by end users	11
File resources contributed by end users	0
Users that contributed at least once	4
Resource Types	
Resources added by end users (Mathematical Model)	11
Tags	
Tags	10
Unique tags	8
Tagged resources	5
Users that tagged at least once	1
Comments	
Comments	1
Commented resources	1
Users that commented at least once	1
Ratings	
Ratings	3
Rated resources	3
Users that rated at least once	3
Categories	

Type	Number
Categories	6653
Relations between categories and resources	7
Unique categories for resources	7
Categorised resources	1
Users that categorised at least once	1
Persons	
Persons	35
Relations between persons and resources	21
Unique persons for resources	14
Resources annotated with persons	10
Users that added a person at least once	3

Table D.13: Created entities and usage information for ALOE-CMCM (without user 'Martin'), retrieved December 13th, 2014

D.5.2 Tracked Activities

The first action on this instance was tracked on August 3rd, 2010.

Table D.14 was created by querying the UserHistories table in ALOE@CMCM and shows the total number of tracked activities, the number of tracked gui activities, the number of users with tracked activities, and the number of objects affected by the tracked activities. Table D.15 shows information about the ten most tracked activities.

all actions	gui actions	users	objects
27241	27219	29	114

Table D.14: Information about activities tracked in ALOE-CMCM (retrieved December 13th, 2014)

Action	Freq (gui)	Freq (all)
.../search/	18667	18668
.../resource/viewed/	6423	6442
.../searchAdvanced/	1122	1122
.../user/login/	488	489
.../resource/metadataEdited/womModel/	150	150

Action	Freq (gui)	Freq (all)
.../user/logout/	89	90
.../resource/contributed/bookmark/	49	49
.../resource/deleted/	29	58
.../resource/contributed/womModel/	26	26
.../group/requestMembership/	25	25

Table D.15: Information about the ten most tracked activities in ALOE-CMCM (retrieved December 13th, 2014)

D.6 Usage Data From RADAR-Kaiserslautern

D.6.1 Created Artefacts

Type	Number
Users	
Registered users	95
Sets	
Collections created	5
Resources added to collections	0
Groups	
Open groups created	7
Closed-public groups created	6
Closed-private groups created	3
Resources shared to open groups	16
Resources shared to closed-public groups	579
Resources shared to closed-private groups	1
Open group memberships	9
Closed-public group memberships	76
Closed-private group memberships	3
Resources	
Resources added by end users	1935
Unique resources contributed by end users	1933
File resources contributed by end users	1019
Users that contributed at least once	51
Resource Types	
Resources added by end users (File)	1019
Resources added by end users (Snippet)	0
Resources added by end users (Bookmark)	193
Resources added by end users (Venue)	97
Resources added by end users (Geocontent)	78
Resources added by end users (Event)	546
Tags	
Tags	5150
Unique tags	1295
Tagged resources	1739
Users that tagged at least once	52
Comments	
Comments	9
Commented resources	8
Users that commented at least once	6

Type	Number
Ratings	
Ratings	63
Rated resources	49
Users that rated at least once	22
Categories	
Categories	35
Relations between categories and resources	16
Unique categories for resources	3
Categorised resources	14
Users that categorised at least once	1
Persons	
Persons	0
Relations between persons and resources	0
Unique persons for resources	0
Resources annotated with persons	0
Users that added a person at least once	0

Table D.16: Created entities and usage information for RADAR-Kaiserslautern (without user 'Martin'), retrieved December 13th, 2014

D.6.2 Tracked Activities

The first action on this instance was tracked on June 11th, 2010.

Table D.17 was created by querying the UserHistories table in RADAR-Kaiserslautern and shows the total number of tracked activities, the number of tracked gui activities, the number of users with tracked activities, and the number of objects affected by the tracked activities. Table D.18 shows information about the ten most tracked activities.

all actions	gui actions	users	objects
541525	540781	110	3094

Table D.17: Information about activities tracked in RADAR-Kaiserslautern (retrieved December 13th, 2014)

Action	Freq (gui)	Freq (all)
.../resource/viewed/	238757	238894
.../search/	237393	237395
.../searchInGroup/	51296	51296
.../resource/relationAdded/	1596	1596
.../user/login/	1400	1449
.../resource/contributed/file/	1388	1388
.../resource/sharedToGroup/	1161	1161
.../resource/addedToEvent/	890	890
.../resource/metadataEdited/	850	850
.../resource/metadataEdited/event/	847	847

Table D.18: Information about the ten most tracked activities in RADAR-Kaiserslautern (retrieved 2014/12/13)

D.6.3 Web Analytics Data
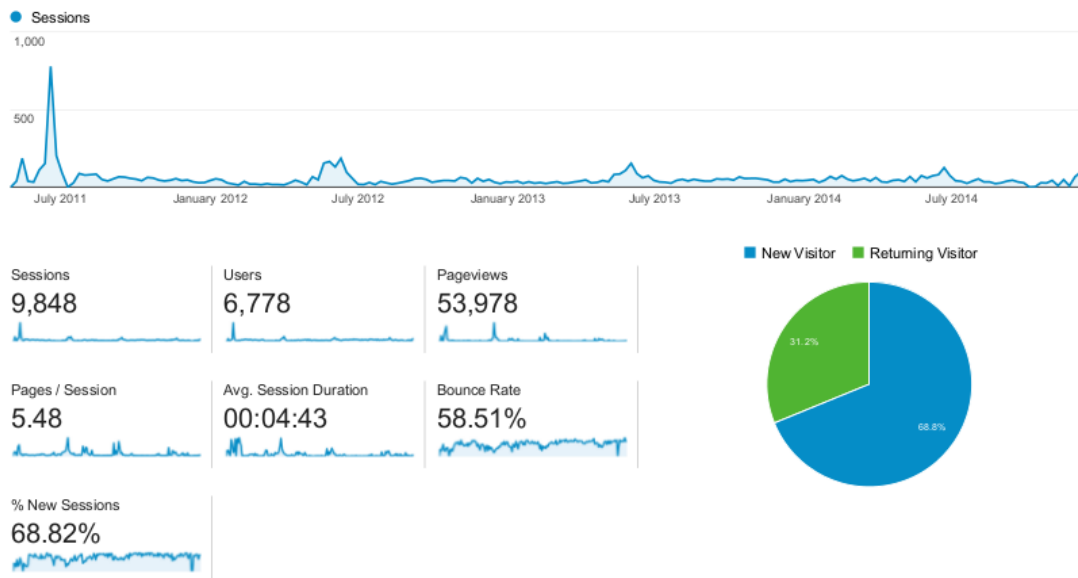
Figure D.5: Web Analytics data gathered for RADAR-Kaiserslautern (date range: 2011/05/01 - 2014/12/12)

D.7 Usage Data From ALOE-UNIFARM

D.7.1 Created Artefacts

Type	Number
Users	
Registered users	14
Sets	
Collections created	3
Resources added to collections	20
Groups	
Open groups created	1
Closed-public groups created	1
Closed-private groups created	1
Resources shared to open groups	13
Resources shared to closed-public groups	1
Resources shared to closed-private groups	0
Open group memberships	1
Closed-public group memberships	1
Closed-private group memberships	1
Resources	
Resources added by end users	107
Unique resources contributed by end users	107
File resources contributed by end users	2
Users that contributed at least once	9
Resource Types	
Resources added by end users (File)	2
Resources added by end users (Bookmark)	11
Resources added by end users (InfoItem)	94
Tags	
Tags	204
Unique tags	61
Tagged resources	83
Users that tagged at least once	7
Comments	
Comments	5
Commented resources	5
Users that commented at least once	1
Ratings	
Ratings	1
Rated resources	1

Type	Number
Users that rated at least once	1
Categories	
Categories	16
Relations between categories and resources	121
Unique categories for resources	13
Categorised resources	89
Users that categorised at least once	6
Persons	
Persons	0
Relations between persons and resources	0
Unique persons for resources	0
Resources annotated with persons	0
Users that added a person at least once	0

Table D.19: Created entities and usage information for ALOE-UNIFARM (without user 'MartinM'), retrieved December 13th, 2014

Table D.20 was created by querying the UserHistories table in ALOE@UNIFARM and shows the total number of tracked activities, the number of tracked gui activities, the number of users with tracked activities, and the number of objects affected by the tracked activities. Table D.21 shows information about the ten most tracked activities.

all actions	gui actions	users	objects
1767	1767	14	146

Table D.20: Information about activities tracked in ALOE-UNIFARM (retrieved 2014/12/13)

Action	Freq (gui)	Freq (all)
.../resource/viewed/	658	658
.../search/	513	513
.../resource/contributed/aloesnippet/	112	112
.../resource/metadataEdited/aloesnippet/	110	110
.../user/login/	87	87
.../resource/addedToAloeSnippet/	71	71

Action	Freq (gui)	Freq (all)
.../resource/tagged/	68	68
.../user/viewed/	26	26
.../resource/addedToCollection/	21	21
.../user/logout/	14	14

Table D.21: Information about the ten most tracked activities in ALOE-UNIFARM (retrieved December 13th, 2014)

For Mindpool as well as for ALOE@KM, an online questionnaire was published. As the questionnaires were almost identical, only the Mindpool questionnaire is shown here. The only differences for ALOE@KM were:

- the name *Mindpool* was exchanged with *ALOE@KM*,

- the *Impact*-question was left out because it was specific for Mindpool, and

- the questionnaire for ALOE@KM contained a last page where the participants were informed about how they can take part in a contest.

Mindpool Survey

Mindpool Survey
Welcome to this Mindpool survey.

Thanks a lot in advance for participating - this helps us a lot!

There are 10 questions in this survey

Some general questions

In this part of the questionnaire, we want to ask for some general information.

Your gender is...

Please choose **only one** of the following:

○ Female
○ Male

Figure E.1: Mindpool questionnaire – page 1

What is your role in DFKI? If you want you can provide further explanations in the comment field.

Please choose **only one** of the following:

○ Student

○ Administrative Staff

○ Software Development

○ Research

○ Project Management

○ Senior Management

Make a comment on your choice here:

Figure E.2: Mindpool questionnaire – page 2

How do you use Mindpool?

In this part we would like to ask you about how you use Mindpool.

How frequently do you use Mindpool? *

Please choose **only one** of the following:

○ never
○ seldom (about once a week)
○ not daily, but several times each week
○ daily, but not more than 10minutes
○ more than 10minutes each day

Mindpool and other tools: I do use Mindpool... *

Please choose the appropriate response for each item:

	very frequently	frequently	sometimes	rarely	never
...to substitute existing tools with a similar purpose	○	○	○	○	○
...in addition to other tools with a similar purpose	○	○	○	○	○
...as the first tool for such purposes	○	○	○	○	○

I use Mindpool as a tool... *

Please choose the appropriate response for each item:

	very frequently	frequently	sometimes	rarely	never
...to organize and manage information for myself	○	○	○	○	○
...to contribute and share information with others	○	○	○	○	○
...to get information from others	○	○	○	○	○
...to get informed about activities from others	○	○	○	○	○

Figure E.3: Mindpool questionnaire – page 3

I intend to use the Mindpool system in the next 6 months. *

Please choose **only one** of the following:

- ○ I strongly disagree
- ○ I disagree
- ○ I neither agree nor disagree
- ○ I agree
- ○ I strongly agree

Figure E.4: Mindpool questionnaire – page 4

Accessing digital contents with Mindpool

The following questions concern the quality of information access in Mindpool in general. *

Please choose the appropriate response for each item:

	I strongly disagree	I disagree	I neither agree nor disagree	I agree	I strongly agree
Searching for digital contents in Mindpool is easy and intuitive	○	○	○	○	○
I can easily find contents in Mindpool regardless of their popularity	○	○	○	○	○
The quality of information (metadata) about contents in Mindpool is high	○	○	○	○	○
The means to browse and navigate within the digital contents in Mindpool are sufficient and easy to use	○	○	○	○	○
Detail pages for contents in Mindpool allow to easily estimate the relevance of the contents	○	○	○	○	○

Figure E.5: Mindpool questionnaire – page 5

Interaction Possibilities & Social Media

Mindpool not only offers means to search for information, but also a variety of social media functionalities, e.g., to contribute own information. This includes, among others:

- Contribution of arbitrary bookmarks and files
- Search for contents using full text and metadata search and filters for a variety of attributes
- Browse contents based on their popularity (ratings, number of views, number of comments, etc.)
- Organize contents in favorite lists
- Join or initiate open and closed groups for contens and other users - to share information about specific topics, projects, etc.
- Tag, rate, and comment on contents
- Social Networking functionalities (user profiles, contact lists, messaging)

Interaction possibilites in Mindpool *

Please choose the appropriate response for each item:

	I strongly disagree	I disagree	I neither agree nor disagree	I agree	I strongly agree
I am aware of the interaction possibilities that Mindpool offers	○	○	○	○	○
The interaction possibilities offered by Mindpool are important for me	○	○	○	○	○
I do not miss any interaction possibilities in Mindpool	○	○	○	○	○
I feel encouraged to contribute information in Mindpool	○	○	○	○	○

Figure E.6: Mindpool questionnaire – page 6

Impact of Mindpool

These questions concern the impact of Mindpool on your work with digital contents.

One of the aims of Mindpool is to support the access to digital contents within DFKI. A lot of these contens (e.g., information about press releases, visits, and photo galleries) were also available before on special pages in the DFKI intranet. With the following questions we want to find out if Mindpool helps users in accessing these contents. *

Please choose the appropriate response for each item:

	I strongly disagree	I disagree	I neither agree nor disagree	I agree	I strongly agree
With Mindpool I can get aware of DFKI related digital contents (especially press releases, visits, and photos) that I otherwise would not have found	○	○	○	○	○
Searching for press releases, visits, and photos has become easier and more intuitive with Mindpool	○	○	○	○	○
There are more and better ways to browse and navigate within press releases, visits, and photos	○	○	○	○	○
The quality of information (metadata) about press releases, visits, and photos in Mindpool has increased	○	○	○	○	○
Interaction and collaboration are fostered with Mindpool	○	○	○	○	○

Figure E.7: Mindpool questionnaire – page 7

Suggestions for improvement

If you want, you can finally provide suggestions for improving Mindpool here!

Please write your answer here:

Figure E.8: Mindpool questionnaire – page 8

In this part, selected methods from the ALOE Web Service are presented. For a complete overview of all methods, please refer to `http://aloe-project.de/AloeWebServiceDocs`.

- **addAssociatedResource**
  `public void` **addAssociatedResource**`(String` **sessionId,** `String` **resourceWithContainerId,** `String` **resourceId**`)`

    - **Description**

      Adds an associated resource to a specified resourceWithContainer.

    - **Parameters**

        * `sessionId` – The user's current session ID.

        * `resourceWithContainerId` – Identifier of the resourceWithContainer to which a resource shall be added.

        * `resourceId` – Identifier of the resource.

- **addCategoryToGroup**
  `public String` **addCategoryToGroup**`(String` **sessionId,** `String` **groupId,** `String` **taxonomyId,** `String` **categoryId,** `String` **relationType,** `float` **weight,** `String` **provenanceInformation**`)`

    - **Description**

      Adds a relation between a category and a group to the system.

    - **Parameters**

        * `sessionId` – The user's current session ID.

        * `groupId` – ID of the group.

        * `taxonomyId` – ID of the taxonomy.

        * `categoryId` – ID of the category.

        * `relationType` – Relation type that has to be added.

        * `weight` – Weight of the new relation

* provenanceInformation – Where does this information come from? (optional).
  – **Returns** – ALOE identifier of the relation.

- **addCategoryToResource**
  public String **addCategoryToResource**(String **sessionId,** String **resourceId,** String **taxonomyId,** String **categoryId,** String **relationType,** float **weight,** String **provenanceInformation**)
  – **Description**

    Adds a relation between a category and a resource to the system.

  – **Parameters**

    * sessionId – The user's current session ID.
    * resourceId – ID of the resource.
    * taxonomyId – ID of the taxonomy.
    * categoryId – ID of the category.
    * relationType – Relation type that has to be added.
    * weight – Weight of the new relation
    * provenanceInformation – Where does this information come from? (optional).

  – **Returns** – ALOE identifier of the relation.

- **addCategoryToUser**
  public String **addCategoryToUser**(String **sessionId,** String **userId,** String **taxonomyId,** String **categoryId,** String **relationType,** float **weight,** String **provenanceInformation**)
  – **Description**

    Adds a relation between a category and a user to the system.

  – **Parameters**

    * sessionId – The user's current session ID.
    * userId – ID of the user.
    * taxonomyId – ID of the taxonomy.
    * categoryId – ID of the category.
    * relationType – Relation type that has to be added.
    * weight – Weight of the new relation
    * provenanceInformation – Where does this information come from? (optional).

- **Returns** – ALOE identifier of the relation.

- **addComment**
  public CommentBean **addComment**(String **sessionId,** String **relation,** String **comment**)

  – **Description**

  Adds a comment to a resource.

  – **Parameters**

  * sessionId – The user's current session ID.

  * relation – Identifier of the resource for which the comment has to be added.

  * comment – Text of the comment.

  – **Returns** – Text and metadata about the comment that has been added.

- **addContact**
  public void **addContact**(String **sessionId,** String **contactId,** boolean **sendNotification**)

  – **Description**

  Adds a user to the current user's list of contacts.

  – **Parameters**

  * sessionId – The user's current session ID.

  * contactId – Identifier of the user who has to be added to the list of a user's contacts.

  * sendNotification – True if a notification has to be sent to the user who is added to the contact list, false otherwise.

- **addGroup**
  public String **addGroup**(String **sessionId,** String **name,** String **description,** String **status,** byte[] **groupIcon**)

  – **Description**

  Opens a new group and expells the current user as an administrator of the group. Further the current user is registered as a member of the group.

  – **Parameters**

  * sessionId – The user's current session ID.

  * name – Identifier of the group that has to be opened.

  * description – Description of the group that has to be opened.

    ∗ `status` – Status of the group that has to be opened. Allowed values are open, closed-private and closed-public.

    ∗ `groupIcon` – File whose thumbnails have to be set as group icons.

  – **Returns** – ALOE identifier of the group.

- **addResourceAdministrator**
  public void **addResourceAdministrator**(String **sessionId,** String **resourceId,** String **userId**)

  – **Description**

  Adds an administrator for a resource.

  – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `resourceId` – Identifier of the resource for which the administrator has to be added.

      ∗ `userId` – Identifier of the user that has to be added as administrator for the resource.

- **addResourceToFavorites**
  public void **addResourceToFavorites**(String **sessionId,** String **resourceId**)

  – **Description**

  Adds a resource to a user's favorites.

  – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `resourceId` – Identifier of the resource that has to be added to the favorites.

- **addTags**
  public void **addTags**(String **sessionId,** String **tags,** String **resourceId**)

  – **Description**

  Adds tags to a resource. Duplicate entries by a user are ignored.

  – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `tags` – Keywords which classify the resource (space separated).

      ∗ `resourceId` – Identifier of the resource that has to be tagged.

- **contributeAloeSnippet**
  `public String` **contributeAloeSnippet**`(String` **sessionId,** `String` **associatedDate,** `String` **creator,** `String` **description,** `String` **language,** `String` **license,** `String` **publisher,** `String` **rightsHolder,** `String` **tags,** `String` **textContent,** `String` **title,** `String` **visibility,** `String` **groupId**`)`

  – **Description**

    Inserts an AloeSnippet and its associated metadata into the system.

  – **Parameters**

    * `sessionId` – The user's current session ID.

    * `associatedDate` – A point in time associated with the AloeSnippet. Has to match the pattern AD YYYY-MM-DD hh:mm:ss with AD = 0 if the date is before anno domine and 1 otherwise.

    * `creator` – Author of the AloeSnippet.

    * `description` – Description of the AloeSnippet.

    * `language` – Language of the AloeSnippet content.

    * `license` – URL of the license which is associated with the AloeSnippet. Shortcuts are provided as follows:
      0: None (All rights reserved)
      1: Attribution License
      2: Attribution-NoDerivs License
      3: Attribution-NonCommercial License
      4: Attribution-ShareAlike License
      5: Attribution-NonCommercial-NoDerivs License
      6: Attribution-NonCommercial-ShareAlike License
      7: Public Domain Mark

    * `publisher` – Name of the publishing instance (optional).

    * `rightsHolder` – Rightsholder of the AloeSnippet (optional).

    * `tags` – Tags which classify the AloeSnippet.

    * `textContent` – The textual content of the AloeSnippet.

    * `title` – Title of the AloeSnippet (mandatory).

    * `visibility` – Visibility of the AloeSnippet. Allowed values are public, group and private. If the parameter is omitted the visibility is set to public per default.

    * `groupId` – Identifier of the group to which the AloeSnippet has to be published (only if parameter visibility is group).

> – **Returns** – ALOE identifier of the AloeSnippet.

- **contributeBookmark**
  public String **contributeBookmark**(String **sessionId,** String **uri,**
  String **associatedDate,** String **creator,** String **description,** String
  language, String **license,** String **publisher,** String **resourceType,**
  String **rightsHolder,** String **tags,** String **title,** String **visibility,**
  String **groupId**)

  > – **Description**
  >
  > Inserts a bookmark and its associated metadata into the system.
  >
  > – **Parameters**
  >
  > > * `sessionId` – The user's current session ID.
  > >
  > > * `uri` – URI of the bookmark that has to be inserted into the system.
  > >
  > > * `associatedDate` – A point in time associated with the resource. Has to match the pattern AD YYYY-MM-DD hh:mm:ss with AD = 0 if the date is before anno domine and 1 otherwise.
  > >
  > > * `creator` – Author of the resource.
  > >
  > > * `description` – Description of the resource.
  > >
  > > * `language` – Language of the resource content.
  > >
  > > * `license` – URL of the license which is associated with the resource. Shortcuts are provided as follows:
  > >   0: None (All rights reserved)
  > >   1: Attribution License
  > >   2: Attribution-NoDerivs License
  > >   3: Attribution-NonCommercial License
  > >   4: Attribution-ShareAlike License
  > >   5: Attribution-NonCommercial-NoDerivs License
  > >   6: Attribution-NonCommercial-ShareAlike License
  > >   7: Public Domain Mark
  > >
  > > * `publisher` – Name of the publishing instance (optional).
  > >
  > > * `resourceType` – Type of the resource (optional).
  > >
  > > * `rightsHolder` – Rightsholder of the resource (optional).
  > >
  > > * `tags` – Tags which classify the resource.
  > >
  > > * `title` – Title of the resource (mandatory).
  > >
  > > * `visibility` – Visibility of the resource. Allowed values are public, group and private. If the parameter is omitted the visibility is set to public per default.

* groupId – Identifier of the group to which the resource has to be published (only if parameter visibility is group).

– **Returns** – ALOE identifier of the resource.

● **contributeBookmarkMetadataSet**
public String **contributeBookmarkMetadataSet**(String **sessionId,** String **uri,** String **creator,** String **description,** String **language,** String **license,** String **title,** String **metadataSetIdInSource,** String **metadataSetSchema,** String **metadataSetSource,** String **relation,** String **relationType,** String **provenanceInformation**)

– **Description**

Inserts a user-defined metadata set and its associated metadata into the system.

– **Parameters**

* sessionId – The user's current session ID.

* uri – URI of the metadata set.

* creator – Creator of the metadata set. (optional)

* description – Description of the metadata set. (optional)

* language – Language of the metadata set. (optional)

* license – (optional) URL of the license which is associated with the metadata set. Shortcuts are provided as follows:
0: None (All rights reserved)
1: Attribution License
2: Attribution-NoDerivs License
3: Attribution-NonCommercial License
4: Attribution-ShareAlike License
5: Attribution-NonCommercial-NoDerivs License
6: Attribution-NonCommercial-ShareAlike License
7: Public Domain Mark

* title – Title of the metadata set. (optional)

* metadataSetIdInSource – Id of the metadata set in its source (i.e., a foreign in outside ALOE). (optional)

* metadataSetSchema – Format of the metadata which is contained in the metadata set. (optional)

* metadataSetSource – Source of the metadata set. (optional)

* relation – Identifier of the resource which is described by the metadata set.

∗ `relationType` – Relation type (optional).

∗ `provenanceInformation` – Where does this information come from? (optional).

   – **Returns** – Identifier of the metadata set.

- **contributeFile**
  `public String` **contributeFile**`(String` **sessionId,** `byte[]` **resource,** `String` **associatedDate,** `String` **creator,** `String` **description,** `String` **language,** `String` **license,** `String` **publisher,** `String` **resourceType,** `String` **rightsHolder,** `String` **tags,** `String` **title,** `String` **visibility,** `String` **groupId,** `String` **fileName**`)`

     – **Description**

     Inserts a file and its associated metadata into the system.

     – **Parameters**

  ∗ `sessionId` – The user's current session ID.

  ∗ `resource` – File that has to be inserted into the system.

  ∗ `associatedDate` – A point in time associated with the resource. Has to match the pattern AD YYYY-MM-DD hh:mm:ss with AD = 0 if the date is before anno domine and 1 otherwise.

  ∗ `creator` – Author of the resource.

  ∗ `description` – Description of the resource.

  ∗ `language` – Language of the resource content.

  ∗ `license` – URL of the license which is associated with the resource. Shortcuts are provided as follows:
  0: None (All rights reserved)
  1: Attribution License
  2: Attribution-NoDerivs License
  3: Attribution-NonCommercial License
  4: Attribution-ShareAlike License
  5: Attribution-NonCommercial-NoDerivs License
  6: Attribution-NonCommercial-ShareAlike License
  7: Public Domain Mark

  ∗ `publisher` – Name of the publishing instance (optional).

  ∗ `resourceType` – Type of the resource (optional).

  ∗ `rightsHolder` – Rightsholder of the resource (optional).

  ∗ `tags` – Tags which classify the resource (mandatory).

    ∗ `title` – Title of the resource (mandatory).

    ∗ `visibility` – Visibility of the resource. Allowed values are public, group and private. If the parameter is omitted the visibility is set to public per default.

    ∗ `groupId` – Identifier of the group to which the resource has to be published (only if parameter visibility is group).

    ∗ `fileName` – The file name of the resource may be specified in order to support the MIME type detection. Only the file name, not the path to the file should be provided.

  – Returns – ALOE identifier of the resource.

- **contributeFileMetadataSet**
  `public String` **contributeFileMetadataSet**(`String` **sessionId,** `byte[]` **metadataSet,** `String` **creator,** `String` **description,** `String` **language,** `String` **license,** `String` **title,** `String` **metadataSetIdInSource,** `String` **metadataSetSchema,** `String` **metadataSetSource,** `String` **relation,** `String` **relationType,** `String` **provenanceInformation**) `String` **fileName,**

   – Description

    Inserts a user-defined metadata set and its associated metadata into the system.

   – Parameters

      ∗ `sessionId` – The user's current session ID.

      ∗ `metadataSet` – Metadata set which has to be inserted.

      ∗ `creator` – Creator of the metadata set. (optional)

      ∗ `description` – Description of the metadata set. (optional)

      ∗ `language` – Language of the metadata set. (optional)

      ∗ `license` – (optional) URL of the license which is associated with the metadata set. Shortcuts are provided as follows:
      0: None (All rights reserved)
      1: Attribution License
      2: Attribution-NoDerivs License
      3: Attribution-NonCommercial License
      4: Attribution-ShareAlike License
      5: Attribution-NonCommercial-NoDerivs License
      6: Attribution-NonCommercial-ShareAlike License
      7: Public Domain Mark

* title – Title of the metadata set. (optional)

* metadataSetIdInSource – Id of the metadata set in its source (i.e., a foreign in outside ALOE). (optional)

* metadataSetSchema – Format of the metadata which is contained in the metadata set. (optional)

* metadataSetSource – Source of the metadata set. (optional)

* relation – Identifier of the resource which is described by the metadata set.

* relationType – Relation type (optional).

* provenanceInformation – Where does this information come from? (optional).

* fileName – The file name of the metadata set may be specified in order to support the MIME type detection. Only the file name, not the path to the file should be provided.

– **Returns** – Identifier of the metadata set.

* **deleteGroup**
  public void **deleteGroup**(String **sessionId,** String **groupId**)

  – **Description**

    Deletes a group from the system.

  – **Parameters**

    * sessionId – The user's current session ID.

    * groupId – Identifier of the group that has to be deleted.

* **deleteResource**
  public void **deleteResource**(String **sessionId,** String **resourceId,** String **userId,** boolean **allOccurrences**)

  – **Description**

    Deletes a resource from the system.

  – **Parameters**

    * sessionId – The user's current session ID.

    * resourceId – Identifier of the resource which has to be deleted.

    * userId – Identifier of the user whose resource has to be deleted. May only be specified by system administrators, ordinary users should leave this parameter empty.

    * allOccurrences – True if all occurrences of the resource have to be deleted, false otherwise. May only be set to true by administrators.

- **deleteTag**
  public void **deleteTag**(String **sessionId,** String **resourceId,** String **tagName**)
  - **Description**

    Deletes a tag that is associated with a resource. This method may only be executed by system administrators.
  - **Parameters**
    - ∗ sessionId – The user's current session ID.
    - ∗ resourceId – Identifier of the resource of which the tag shall be deleted.
    - ∗ tagName – Name of the tag.

- **deleteUserAccount**
  public void **deleteUserAccount**(String **sessionId,** String **userId**)
  - **Description**

    Deletes the account of the current user. The contacts of the user will be informed that the user account has been deleted.
  - **Parameters**
    - ∗ sessionId – The user's current session ID.
    - ∗ userId – Identifier of the user whose account has to be deleted. May only be specified by system administrators, ordinary users should leave this parameter empty.

- **deleteUserResourceTag**
  public void **deleteUserResourceTag**(String **sessionId,** String **resourceId,** String **tagName**)
  - **Description**

    Deletes a user's tag that is associated with a resource.
  - **Parameters**
    - ∗ sessionId – The user's current session ID.
    - ∗ resourceId – Identifier of the resource of which the tag shall be deleted.
    - ∗ tagName – Name of the tag.

- **deleteUserTag**
  public void **deleteUserTag**(String **sessionId,** String **tagName**)
  - **Description**

    Deletes a user's tag (all occurences).

– **Parameters**

* `sessionId` – The user's current session ID.

* `tagName` – Name of the tag.

- **getAllMetaMetadata**
  `public MetaMetadataBean[]` **getAllMetaMetadata**`(String `**sessionId,**
  `String `**resourceId**`)`

  – **Description**

  Determines metadata about all user-defined metadata sets which have been annotated for a certain resource.

  – **Parameters**

  * `sessionId` – The user's current session ID.

  * `resourceId` – Identifier of the resource for which metadata about all the associated user-defined metadata sets has to be determined.

  – **Returns** – Metadata about all user-defined metadata sets which have been annotated for the specified resource.

- **getAllUserTags**
  `public TagBean[]` **getAllUserTags**`(String `**sessionId,** `String `**userId**`)`

  – **Description**

  Determines the tags that have been used by the specified user as well as their usage frequency. (MACE)

  – **Parameters**

  * `sessionId` – The user's current session ID.

  * `userId` – Identifier of the user whose used tags and their usage frequency have to be determined.

  – **Returns** – Tags which the specified user has used as well as their usage frequency.

- **getAloeSnippetMetadata**
  `public GenericResourceMetadataBean`
  getAloeSnippetMetadata`(String `**sessionId,** `String `**resourceId**`)`

  – **Description**

  Determines metadata about the specified AloeSnippet.

  – **Parameters**

  * `sessionId` – The user's current session ID.

* `*` `resourceId` – Identifier of the AloeSnippet for which metadata has to be determined.

 – **Returns** – Metadata about the specified AloeSnippet.

* **getAssociatedResources**
 public ResourcesResultBean **getAssociatedResources**(String **sessionId,** String **resourceWithContainerId,** String **order,** String **orderDirection,** int **lowerIndex,** int **numberOfResources**)

 – **Description**

 Determines metadata about the associated resources of a specified resourceWithContainer.

 – **Parameters**

 * `*` `sessionId` – The user's current session ID.

 * `*` `resourceWithContainerId` – Identifier of the resourceWithContainer for which metadata about the associated resources has to be determined.

 * `*` `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfComments, orderByNumberOfViews, orderByDate.

 * `*` `orderDirection` – Determines whether the results should be ordered in ascending (ASC) or descending (DESC) direction. If not specified, a default value will be chosen for the given order.

 * `*` `lowerIndex` – Index of first resource in the result set that has to be returned.

 * `*` `numberOfResources` – Number of resources that have to be returned (max number 100).

 – **Returns** – Metadata about the associated resources of the specified resourceWithContainer.

* **getCategory**
 public CategoryMetadataBean **getCategory**(String **sessionId,** String **taxonomyId,** String **categoryId**)

 – **Description**

 Determines information about a category.

 – **Parameters**

 * `*` `sessionId` – The user's current session ID.

 * `*` `taxonomyId` – ID of the taxonomy.

 * `*` `categoryId` – ID of the category.

 – **Returns** – Metadata about a category.

- **getCategoryRelationsForGroups**
  `public CategoryObjectRelationBean[]`
  getCategoryRelationsForGroups(`String` **sessionId,** `String[]` **groupIds,** `String` **relationType,** `String` **contributorId**)

  – **Description**

     Determines information about category relations for an array of groups.

  – **Parameters**

     * `sessionId` – The user's current session ID.

     * `groupIds` – Array of group identifiers for which information about category relations has to be determined (mandatory).

     * `relationType` – Relation type (optional)

     * `contributorId` – Identifier of the user who added the related categories for the groups. (optional)

  – **Returns** – Information about category relations of groups.

- **getCategoryRelationsForResources**
  `public CategoryObjectRelationBean[]`
  getCategoryRelationsForResources(`String` **sessionId,** `String[]` **resourceIds,** `String` **relationType,** `String` **contributorId**)

  – **Description**

     Determines information about category relations for an array of resources.

  – **Parameters**

     * `sessionId` – The user's current session ID.

     * `resourceIds` – Array of resource identifiers for which information about category relations has to be determined (mandatory).

     * `relationType` – Relation type (optional)

     * `contributorId` – Identifier of the user who added the related categories for the resources. (optional)

  – **Returns** – Information about category relations of resources.

- **getCategoryRelationsForUsers**
  `public CategoryObjectRelationBean[]`
  getCategoryRelationsForUsers(`String` **sessionId,** `String[]` **userIds,** `String` **relationType,** `String` **contributorId**)

  – **Description**

     Determines information about category relations for an array of users.

– **Parameters**

* `sessionId` – The user's current session ID.

* `userIds` – Array of user identifiers for which information about category relations has to be determined (mandatory).

* `relationType` – Relation type (optional)

* `contributorId` – Identifier of the user who added the related categories for the users. (optional)

– **Returns** – Information about category relations of users.

- **getFavoritesMetadata**
  `public FavoriteMetadataResultBean` **getFavoritesMetadata**(`String` **sessionId,** `String` **resourceId,** `int` **lowerIndex,** `int` **numberOfMetadataSets**)

  – **Description**

  Determines the favorite metadata of all users that added the specified resource to their favorites.

  – **Parameters**

  * `sessionId` – The user's current session ID.

  * `resourceId` – Identifier of the resource for which the metadata has to be determined.

  * `lowerIndex` – Index of first favorite metadata in the result set that has to be returned.

  * `numberOfMetadataSets` – Number of favorite metadata sets that have to be returned (max number 100).

  – **Returns** – Metadata of all users that added the specified resource to their favorites.

- **getGroup**
  `public GroupBean` **getGroup**(`String` **sessionId,** `String` **groupId**)

  – **Description**

  Determines metadata about a specified group.

  – **Parameters**

  * `sessionId` – The user's current session ID.

  * `groupId` – Identifier of the group for which metadata has to be specified.

  – **Returns** – Metadata about the specified group.

- **getGroupMembers**
  `public UsersResultBean` **getGroupMembers**`(String `**sessionId,**
  `String `**groupId,** `String `**order,** `int `**lowerIndex,** `int`
  numberOfUsers`)`

  - **Description**

    Determines user data of the members of a specified group.

  - **Parameters**

    * `sessionId` – The user's current session ID.

    * `groupId` – Identifier of the group of whose members user data has to be determined.

    * `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByDate.

    * `lowerIndex` – Index of the first user in the result set that has to be returned.

    * `numberOfUsers` – Number of users that have to be returned (max number 100).

  - **Returns** – User data of the members of the specified group.

- **getGroupResources**
  `public ResourcesResultBean` **getGroupResources**`(String `**sessionId,**
  `String[] `**groupIds,** `String `**resourceTypes,** `String `**order,** `int`
  lowerIndex, `int `**numberOfResources**`)`

  - **Description**

    Determines metadata of the resources in a specified list of groups.

  - **Parameters**

    * `sessionId` – The user's current session ID.

    * `groupIds` – List of group identifiers for which metadata of the contained resources has to be determined.

    * `resourceTypes` – Resource types of the searched resources (optional).

    * `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfBookmarks/orderByNumberOfFavorites, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderBySharingDate.

    * `lowerIndex` – Index of first resource in the result set that has to be returned.

* numberOfResources – Number of resources that have to be returned
(max number 100).

– **Returns** – Metadata of the resources in a specified group.

* **getGroups**
public GroupsResultBean **getGroups**(String **sessionId,** String
order, int **lowerIndex,** int **numberOfGroups**)

– **Description**

Determines metadata about all groups in the system.

– **Parameters**

* sessionId – The user's current session ID.

* order – Determines how the result should be ordered. Valid values
are: orderByAlphabet, orderByDate, orderByNumberOfMembers and
orderByNumberOfResources.

* lowerIndex – Index of first group in the result set that has to be
returned.

* numberOfGroups – Number of groups that have to be returned (max
number 100).

– **Returns** – Metadata about all groups in the system.

* **getMessage**
public MessageBean **getMessage**(String **sessionId,** int **messageId**)

– **Description**

Determines the text of a specified message and its associated metadata.

– **Parameters**

* sessionId – The user's current session ID.

* messageId – Identifier of the message for which the text and its
metadata has to be determined.

– **Returns** – Text and metadata of the specified message.

* **getMetadataSet**
public ByteArrayWithMimeTypeBean **getMetadataSet**(String
sessionId, String **metadataSetId**)

– **Description**

Determines the specified file metadata set.

– **Parameters**

* sessionId – The user's current session ID.

* ∗ `metadataSetId` – Identifier of the file metadata set.
    – **Returns** – File metadata set content.

* **getMetaMetadata**
  `public MetaMetadataBean` **getMetaMetadata**`(String `**sessionId,**
  `String `**metadataSetId**`)`
    – **Description**

      Determines metadata about the specific user-defined metadata set which
      has been annotated for a certain resource.

    – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `metadataSetId` – Identifier of the user-defined metadata set for
        which metadata has to be determined.

    – **Returns** – Metadata about all user-defined metadata sets which have been
      annotated for the specified resource.

* **getResource**
  `public ByteArrayWithMimeTypeBean` **getResource**`(String `**sessionId,**
  `String `**resourceId**`)`
    – **Description**

      Determines the specified (file) resource.

    – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `resourceId` – Identifier of the resource.

    – **Returns** – Resource content.

* **getResourceAdministrators**
  `public UsersResultBean` **getResourceAdministrators**`(String`
  sessionId, `String `**resourceId,** `String `**order,** `int `**lowerIndex,** `int`
  numberOfUsers`)`
    – **Description**

      Determines user data about resource administrators.

    – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `resourceId` – Identifier of the resource for which administrators
        have to be retrieved

      ∗ `order` – Determines how the result should be ordered. Valid values
        are: orderByAlphabet, orderByDate.

* `lowerIndex` – Index of the first user in the result set that has to be returned.

* `numberOfUsers` – Number of users that have to be returned (max number 100).

– **Returns** – User data about the resource administrators

• **getResourceComments**
`public CommentBean[]` **getResourceComments**`(String` **sessionId,**
`String` **resourceId**`)`

– **Description**

Determines the comments that have been posted for a certain resource.

– **Parameters**

* `sessionId` – The user's current session ID.

* `resourceId` – Identifier of the resource for which the comments have to be determined.

– **Returns** – Comments that have been posted for the specified resource.

• **getResourceMetadata**
`public GenericResourceMetadataBean` **getResourceMetadata**`(String`
sessionId, `String` **resourceId**`)`

– **Description**

Determines metadata about the specified resource. If the current user has bookmarked the specified resource then his metadata for the resource is determined otherwise the metadata of the first resource contributor is determined per default.

– **Parameters**

* `sessionId` – The user's current session ID.

* `resourceId` – Identifier of the resource for which metadata has to be determined.

– **Returns** – Metadata about the specified resource.

• **getResourceRelatedCategories**
`public CategoryMetadataBean[]`
getResourceRelatedCategories`(String` **sessionId,** `String` **resourceId,**
`String` **relationType,** `String` **contributorId**`)`

– **Description**

Determines related categories for a resource.

– **Parameters**

* sessionId – The user's current session ID.

* resourceId – Identifier of the resource for which related categories have to be determined (mandatory).

* relationType – Relation type (optional)

* contributorId – Identifier of the user who added the related categories for a resource. (optional)

– **Returns** – Related categories of a resource.

- **getResourcesByOrderAndResourceTypes**
  public GenericResourceMetadataBean[]
  getResourcesByOrderAndResourceTypes(String **sessionId,** String **resourceTypes,** String **order,** int **lowerIndex,** int **numberOfResources**)

  – **Description**

    Determines metadata of resources retrieved by order and resourceTypes.

  – **Parameters**

    * sessionId – The user's current session ID.

    * resourceTypes – Resource types of the searched resources.

    * order – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderByNumberOfBookmarks/orderByNumberOfFavorites.

    * lowerIndex – Index of first resource in the result set that has to be returned.

    * numberOfResources – Number of resources that have to be returned (max number 100).

  – **Returns** – Metadata of the resources found.

- **getResourceTags**
  public TagBean[] **getResourceTags**(String **sessionId,** String **resourceId**)

  – **Description**

    Determines all tags and their frequencies that are associated with a resource.

  – **Parameters**

    * sessionId – The user's current session ID.

    * resourceId – Identifier of the resource of which the tags have to be determined.

– **Returns** – Tags and their frequencies that are associated with the resource.

- **getUserConfiguration**
  `public UserConfigurationBean` **getUserConfiguration**(`String` **sessionId**)
  - **Description**

    Determines the configuration of the current user.
  - **Parameters**
    - ∗ `sessionId` – The user's current session ID.
  - **Returns** – Configuration of the current user.

- **getUserContacts**
  `public UsersResultBean` **getUserContacts**(`String` **sessionId, **`String` **userId,** `String` **order,** `int` **lowerIndex,** `int` **numberOfUsers**)
  - **Description**

    Determines user data about a user's contacts.
  - **Parameters**
    - ∗ `sessionId` – The user's current session ID.
    - ∗ `userId` – Identifier of the user whose contacts have to be determined. (Only admin users may specify this parameter!)
    - ∗ `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByDate.
    - ∗ `lowerIndex` – Index of the first user in the result set that has to be returned.
    - ∗ `numberOfUsers` – Number of users that have to be returned (max number 100).
  - **Returns** – User data about the current user's contacts.

- **getUserData**
  `public UserDataBean` **getUserData**(`String` **sessionId,** `String` **profileOwnerId**)
  - **Description**

    Determines the user data of a specified user.
  - **Parameters**
    - ∗ `sessionId` – The user's current session ID.
    - ∗ `profileOwnerId` – Identifier of the user whose user data has to be determined.

- **Returns** – User data of the specified user.

- **getUserFavorites**
  public ResourcesResultBean **getUserFavorites**(String **sessionId,** String **userId,** String **order,** int **lowerIndex,** int **numberOfResources**)

  - **Description**

    Determines metadata about a user's favorites

  - **Parameters**

    * sessionId – The user's current session ID.

    * userId – Identifier of the user for whose favorites metadata has to be determined. If the parameter is omitted, metadata about the current user's favorites will be determined.

    * order – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderByNumberOfFavorites.

    * lowerIndex – Index of first resource in the result set that has to be returned.

    * numberOfResources – Number of resources that have to be returned (max number 100).

  - **Returns** – Metadata about a user's favorites

- **getUserRelatedCategories**
  public CategoryMetadataBean[] **getUserRelatedCategories**(String **sessionId,** String **userId,** String **relationType,** String **contributorId**)

  - **Description**

    Determines related categories for a user.

  - **Parameters**

    * sessionId – The user's current session ID.

    * userId – Identifier of the user for which related categories have to be determined (mandatory).

    * relationType – Relation type (mandatory)

    * contributorId – Identifier of the user who added the related categories for a user. (optional)

  - **Returns** – Related categories of a user.

- **getUserResources**
  `public ResourcesResultBean` **getUserResources**`(String` **sessionId,**
  `String` **userId,** `String` **order,** `int` **lowerIndex,** `int`
  numberOfResources`)`

  - **Description**

    Determines metadata about a user's resources.

  - **Parameters**

    - `* sessionId` – The user's current session ID.

    - `* userId` – Identifier of the user for whose resources metadata has to be determined. If the parameter is omitted, metadata about the current user's resources will be determined.

    - `* order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderByNumberOfBookmarks/orderByNumberOfFavorites.

    - `* lowerIndex` – Index of first resource in the result set that has to be returned.

    - `* numberOfResources` – Number of resources that have to be returned (max number 100).

  - **Returns** – Metadata about a user's resources.

- **getUserTags**
  `public String[]` **getUserTags**`(String` **sessionId,** `String` **resourceId**`)`

  - **Description**

    Returns the tags which the current user has annotated for a certain resource.

  - **Parameters**

    - `* sessionId` – The user's current session ID.

    - `* resourceId` – Identifier of the resource for which the user's tags have to be determined.

  - **Returns** – Tags which the current user has annotated for the specified resource.

- **hasAccessRightForResource**
  `public java.lang.Boolean` **hasAccessRightForResource**`(String`
  sessionId, `String` **resourceId,** `String` **accessType**`)`

– **Description**

Checks whether a session id is valid either for an access to a certain resource. Note that the user id can be extracted from the session id.

– **Parameters**

* `sessionId` – Session identifier which has to be checked.

* `resourceId` – Identifier of the resource for which access right has to be checked

* `accessType` – Specifies the access type. Allowed values are edit and read.

– **Returns** – True if the session id is valid for an access to resourceId.

- **isExistingBookmark**
  `public String` **isExistingBookmark**`(String` **sessionId,** `String` **uri,** `String` **visibility,** `String` **groupId**`)`

  – **Description**

  Determines whether a bookmark is already published with the specified visibility in the system.

  – **Parameters**

  * `sessionId` – The user's current session ID.

  * `uri` – URI of the bookmark.

  * `visibility` – Searched visibility under which the bookmark has been published. Valid values are public, group and private. If the parameter is omitted then public is used per default.

  * `groupId` – If parameter visibility is set to group then the identifier of the group has to be specified here.

  – **Returns** – Identifier of the bookmark with the specified visibility. Null if no such bookmark is available.

- **isGroupAdministrator**
  `public java.lang.Boolean` **isGroupAdministrator**`(String` **sessionId,** `String` **groupId**`)`

  – **Description**

  Determines whether the current user is an administrator of a specified group.

  – **Parameters**

  * `sessionId` – The user's current session ID.

* groupId – Identifier of the group for which has to be checked whether the current user is an administrator of it.

– **Returns** – True if the user is an administrator of the specified group, false otherwise.

* **joinGroup**
  public void **joinGroup**(String **sessionId,** String **groupId**)

  – **Description**

  Subscribes the current user to a group.

  – **Parameters**

  * sessionId – The user's current session ID.

  * groupId – Identifier of the group which the current user wants to join.

* **login**
  public String[] **login**(String **nickname,** String **password,** boolean **guiClient**)

  – **Description**

  Performs a login to ALOE.

  – **Parameters**

  * nickname – Nickname of the user that has to be logged in.

  * password – Password of the user that has to be logged in.

  * guiClient – True in case that the invoking client is a GUI, false otherwise.

  – **Returns** – At position 0 a session ID of a logged in session is returned. At position 1 the identifier of the current user is returned. At position 2 the information is stored whether the current user is a system administrator.

* **rate**
  public void **rate**(String **sessionId,** String **resourceId,** int **ratingValue**)

  – **Description**

  Performs a rating.

  – **Parameters**

  * sessionId – The user's current session ID.

  * resourceId – Identifier of the resource that has to be rated.

  * ratingValue – Value with which the resource should be rated. Valid values are integers from one to five.

- **registerUser**
  `public String` **registerUser**`(String[]` **userData,** `String` **interests,** `String` **languages,** `String` **messengers,** `byte[]` **buddyIcon,** `String` **confirmationMailLanguage**)
  - **Description**

    Registers a user in the system and sends a confirmation mail to provided eMail address.
  - **Parameters**
    * `userData` – User data in the following order: eMail address (required), password (required), nickname (required), first name, last name, gender, birthday, affiliation, homepage, phone, cell phone, country, zip code, city, street and the visibility of the user data (public, contacts or private). If parameter visibility is omitted the visibility of the profile data is set to public per default.
    * `interests` – The user's interests (separated by ,).
    * `languages` – The languages a user speaks (separated by ,).
    * `messengers` – Messenger name and the user's ID for the messenger have to be provided in alternating order (separated by ,).
    * `buddyIcon` – The user's buddy icon.
    * `confirmationMailLanguage` – The language in which the confirmation mail has to be sent (otional) If not provided, the default language from the configuration properties will be used
  - **Returns** – Identifier of the user that has been registered.

- **removeAssociatedResource**
  `public void` **removeAssociatedResource**`(String` **sessionId,** `String` **resourceWithContainerId,** `String` **resourceId**)
  - **Description**

    Removes an associated resource from a specified resourceWithContainer.
  - **Parameters**
    * `sessionId` – The user's current session ID.
    * `resourceWithContainerId` – Identifier of the resourceWithContainer from which the resource should be removed.
    * `resourceId` – Identifier of the resource.

- **removeCategoryFromGroup**
  `public void` **removeCategoryFromGroup**`(String` **sessionId,** `String` **groupId,** `String` **taxonomyId,** `String` **categoryId,** `String`

relationType, String **provenanceInformation,** String **contributorId,**
`boolean` **allOccurrences**)

- **Description**

  Removes a relation between a category and a group in the system.

- **Parameters**

  * `sessionId` – The user's current session ID.

  * `groupId` – ID of the group.

  * `taxonomyId` – ID of the taxonomy.

  * `categoryId` – ID of the category.

  * `relationType` – Relation type that has to be removed.

  * `provenanceInformation` – Provenance information (optional).

  * `contributorId` – Identifier of the user whose relation has to be deleted. May only be specified by system administrators, ordinary users should leave this parameter empty.

  * `allOccurrences` – True if all relations of the specified type between the category and the group have to be deleted, false otherwise. May only be set to true by administrators.

- **removeCategoryFromResource**
  `public void` **removeCategoryFromResource**(`String` **sessionId,** `String`
  resourceId, `String` **taxonomyId,** `String` **categoryId,** `String`
  relationType, `String` **provenanceInformation,** `String` **contributorId,**
  `boolean` **allOccurrences**)

  - **Description**

    Removes a relation between a category and a resource in the system.

  - **Parameters**

    * `sessionId` – The user's current session ID.

    * `resourceId` – ID of the resource.

    * `taxonomyId` – ID of the taxonomy.

    * `categoryId` – ID of the category.

    * `relationType` – Relation type that has to be removed.

    * `provenanceInformation` – Provenance information (optional).

    * `contributorId` – Identifier of the user whose relation has to be deleted. May only be specified by system administrators, ordinary users should leave this parameter empty.

         ∗ `allOccurrences` – True if all relations of the specified type between the category and the resource have to be deleted, false otherwise. May only be set to true by administrators.

- **removeCategoryFromUser**
  public void **removeCategoryFromUser**(`String` **sessionId,** `String` **userId,** `String` **taxonomyId,** `String` **categoryId,** `String` **relationType,** `String` **provenanceInformation,** `String` **contributorId,** `boolean` **allOccurrences**)

  – **Description**

  Removes a relation between a category and a user in the system.

  – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `userId` – ID of the user.

      ∗ `taxonomyId` – ID of the taxonomy.

      ∗ `categoryId` – ID of the category.

      ∗ `relationType` – Relation type that has to be removed.

      ∗ `provenanceInformation` – Provenance information (optional).

      ∗ `contributorId` – Identifier of the user whose relation has to be deleted. May only be specified by system administrators, ordinary users should leave this parameter empty.

      ∗ `allOccurrences` – True if all relations of the specified type between the category and the user have to be deleted, false otherwise. May only be set to true by administrators.

- **removeResourceFromFavorites**
  public void **removeResourceFromFavorites**(`String` **sessionId,** `String` **resourceId**)

  – **Description**

  Removes a resource from a user's favorites.

  – **Parameters**

      ∗ `sessionId` – The user's current session ID.

      ∗ `resourceId` – Identifier of the resource that has to be removed from the favorites.

- **removeResourceFromUserSet**
  public void **removeResourceFromUserSet**(`String` **sessionId,** `String` **userSetId,** `String` **resourceId**)

– **Description**

Removes a resource from a specified user set.

– **Parameters**

* `sessionId` – The user's current session ID.

* `userSetId` – Identifier of the user set from which a resource should be removed.

* `resourceId` – Identifier of the resource.

• **renameUserTag**
`public void` **renameUserTag**`(String` **sessionId,** `String` **tagNameOld,** `String` **tagNameNew**`)`

– **Description**

Renames a tag of a user (all occurences)

– **Parameters**

* `sessionId` – The user's current session ID.

* `tagNameOld` – Name of the tag which has to be renamed.

* `tagNameNew` – New name for the tag.

• **requestGroupMembership**
`public void` **requestGroupMembership**`(String` **sessionId,** `String` **groupId,** `String` **requestMessage**`)`

– **Description**

Performs an authorization request for the membership in a closed group.

– **Parameters**

* `sessionId` – The user's current session ID.

* `groupId` – Identifier of the group for which the authorization request for membership has to be performed.

* `requestMessage` – Explanatory statement of the user why he wants to be a member of the specified group.

• **resignResourceFromGroup**
`public void` **resignResourceFromGroup**`(String` **sessionId,** `String` **resourceId,** `String` **groupId**`)`

– **Description**

Resigns a resource from a specified group.

– **Parameters**

* `sessionId` – The user's current session ID.

      * `resourceId` – Identifier of the resource which has to be resigned from a group.

      * `groupId` – Identifier of the group from which the resource has to be resigned.

- **resignUserFromGroup**
  `public void` **resignUserFromGroup**`(String` **sessionId,** `String` **groupId,** `String` **memberId**`)`

  – **Description**

  Resigns a user from a specified group.

  – **Parameters**

      * `sessionId` – The user's current session ID.

      * `groupId` – Identifier of the group from which the user has to be resigned.

      * `memberId` – Identifier of the user which has to be resigned from the specified group. `null` if the current user has to be resigend from the specified group. Different group members may only be resigned from groups by group administrators.

- **search**
  `public ResourcesResultBean` **search**`(String` **sessionId,** `String` **searchString,** `String` **order,** `int` **lowerIndex,** `int` **numberOfResources**`)`

  – **Description**

  Determines metadata about the resources that match the specified search string in their metadata fields author, description, tags or title.

  – **Parameters**

      * `sessionId` – The user's current session ID.

      * `searchString` – Keywords which have to be contained in the metadata fields author, description, tags or title.

      * `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderByNumberOfBookmarks/orderByNumberOfFavorites.

      * `lowerIndex` – Index of first resource in the result set that has to be returned.

      * `numberOfResources` – Number of resources that have to be returned (max number 100).

– **Returns** – Metadata about the found resources.

- **searchAdvanced**
  `public ResourcesResultBean` **searchAdvanced**`(String` **sessionId,**
  `String` **creator,** `String` **description,** `String` **licenses,** `String`
  mimeTypes, `String` **resourceTypes,** `String` **tags,** `String` **title,**
  `String` **fullText,** `boolean` **bookmarkResources,** `boolean`
  fileResources, `boolean` **ownResources,** `String` **contributionDateFrom,**
  `String` **contributionDateTo,** `String` **dateFrom,** `String` **dateTo,**
  `String[]` **categories,** `String` **order,** `int` **lowerIndex,** `int`
  numberOfResources`)`

  – **Description**

  Determines metadata about the resources that match the search strings in
  the specified fields.

  – **Parameters**

  * `sessionId` – The user's current session ID.

  * `creator` – Terms that have to be contained in the creator metadata
    element.

  * `description` – Terms that have to be contained in the description
    metadata element.

  * `licenses` – Licenses under which the searched resources have to be
    published. Licenses have to be separated by ";". Shortcuts are
    provided as follows:
    0: None (All rights reserved)
    1: Attribution License
    2: Attribution-NoDerivs License
    3: Attribution-NonCommercial License
    4: Attribution-ShareAlike License
    5: Attribution-NonCommercial-NoDerivs License
    6: Attribution-NonCommercial-ShareAlike License
    7: Public Domain Mark

  * `mimeTypes` – MIME types of the searched resources. Valid values are:
    audio, document, image and video.

  * `resourceTypes` – Resource types of the searched resources - this is
    used as a filter, so only one of them must fit.

  * `tags` – Tags that have to be annotated for the searched resources.

  * `title` – Terms that have to be contained in the title metadata element

  * `fullText` – Terms that have to be contained in the resource's full text.

* `bookmarkResources` – True, when bookmark resources are allowed in the result set. False otherwise.

* `fileResources` – True, when file resources are allowed in the result set. False otherwise.

* `ownResources` – Restricts the search to the current user's own resources. Will be ignored for anonymous users.

* `contributionDateFrom` – Resources in the result may not be contributed before contributionDateFrom. Please provide contributionDateFrom in the form YYYY-MM-DD.

* `contributionDateTo` – Resources in the result may not be contributed after contributionDateTo. Please provide contributionDateTo in the from YYYY-MM-DD.

* `dateFrom` – Resources in the result may not be older than dateFrom. Please provide dataFrom in the form YYYY-MM-DD.

* `dateTo` – Resources in the result may not be newer than dateTo. Please provide dateTo in the from YYYY-MM-DD.

* `categories` – Array of category (and optionally taxonomy) identifiers in which the search has to be performed. Each entry can consist of only a categoryId or a taxonomyId and a categoryId, separated with a blank (optional)

* `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderByNumberOfBookmarks/orderByNumberOfFavorites.

* `lowerIndex` – Index of first resource in the result set that has to be returned.

* `numberOfResources` – Number of resources that have to be returned (max number 100).

– **Returns** – Metadata about the found resources.

* **searchCategories**
  `public CategoryMetadataBean[]` **searchCategories(**`String` **sessionId,**
  `String` **searchString,** `String` **taxonomyId**`)`

  – **Description**

    Determines information about categories which match the specified search string in their name / categoryId.

  – **Parameters**

    * `sessionId` – The user's current session ID.

* searchString – Search string of which the contained words have to be found in the category fields.

* taxonomyId – Identifier of the taxonomy in which categories are searched (optional)

– **Returns** – Information about categories which match the specified search string.

• **searchGroups**
  public GroupsResultBean **searchGroups**(String **sessionId,** String **searchString,** String **order,** int **lowerIndex,** int **numberOfGroups**)

  – **Description**

    Determines the groups in the system which match the specified search string in their name or description.

  – **Parameters**

    * sessionId – The user's current session ID.

    * searchString – Search string of which the contained words have to be found in the groups' names or descriptions.

    * order – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByDate, orderByNumberOfMembers and orderByNumberOfResources.

    * lowerIndex – Index of first group in the result set that has to be returned.

    * numberOfGroups – Number of groups that have to be returned (max number 100).

  – **Returns** – Metadata of the groups which match the specified search string in their name or description.

• **searchInGroups**
  public ResourcesResultBean **searchInGroups**(String **sessionId,** String **searchString,** String[] **groupIds,** String **order,** int **lowerIndex,** int **numberOfResources**)

  – **Description**

    Determines metadata about resources in a specified list of groups that match a specified search filter.

  – **Parameters**

    * sessionId – The user's current session ID.

    * searchString – Keywords which have to be contained in the title, description, creator or tags of a matching resource.

* `groupIds` – List of group identifiers in which the search has to be performed.

* `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfBookmarks/orderByNumberOfFavorites, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderBySharingDate.

* `lowerIndex` – Index of first resource in the result set that has to be returned.

* `numberOfResources` – Number of resources that have to be returned (max number 100).

– **Returns** – Metadata about resources in the specified group that match the specified search filter.

- **searchSelectedFields**
  `public ResourcesResultBean` **searchSelectedFields**`(String` **sessionId,** `String` **searchString,** `String` **order,** `int` **lowerIndex,** `int` **numberOfResources,** `boolean` **title,** `boolean` **tags,** `boolean` **description,** `boolean` **creator**`)`

  – **Description**

    Determines metadata about the resources that match the specified search string in selected metadata fields (author, description, tags or title).

  – **Parameters**

    * `sessionId` – The user's current session ID.

    * `searchString` – Keywords which have to be contained in the selected metadata fields.

    * `order` – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByAverageRating, orderByNumberOfComments, orderByNumberOfViews, orderByDate, orderByNumberOfBookmarks/orderByNumberOfFavorites.

    * `lowerIndex` – Index of first resource in the result set that has to be returned.

    * `numberOfResources` – Number of resources that have to be returned (max number 100).

    * `title` – Indicates whether metadata field title shall be included in the search.

    * `tags` – Indicates whether the tags of a resource shall be included in the search.

* description – Indicates whether metadata field description shall be included in the search.

* creator – Indicates whether metadata field creator shall be included in the search.

– **Returns** – Metadata about the found resources.

● **searchUsers**
public UsersResultBean **searchUsers**(String **sessionId,** String **searchString,** String **order,** int **lowerIndex,** int **numberOfUsers**)

  – **Description**

  Determines user data of the members in the system which match the specified search string in their nickname or country. Logged in users automatically also search in the user data fields affiliation, first name, last name and city.

  – **Parameters**

    * sessionId – The user's current session ID.

    * searchString – Search string of which the contained words have to be found in the user data fields.

    * order – Determines how the result should be ordered. Valid values are: orderByAlphabet, orderByDate.

    * lowerIndex – Index of the first user in the result set that has to be returned.

    * numberOfUsers – Number of users that have to be returned (max number 100).

  – **Returns** – User data of the members which match the specified search string.

● **sendGroupMessage**
public void **sendGroupMessage**(String **sessionId,** String **receiverGroup,** String **subject,** String **message**)

  – **Description**

  Sends a message to all members of a specified group.

  – **Parameters**

    * sessionId – The user's current session ID.

    * receiverGroup – Identifier of the group to whose members the message has to be sent.

    * subject – Subject of the message.

∗ `message` – The message text.

- **sendMessage**
  public void **sendMessage**(String **sessionId,** String **receiverId,** String **subject,** String **message**)

  – **Description**

  Sends a message to a user of the system.

  – **Parameters**

  ∗ `sessionId` – The user's current session ID.

  ∗ `receiverId` – Identifier of the receiver of the message.

  ∗ `subject` – Subject of the message.

  ∗ `message` – The message text.

- **setResourceEditAccessRight**
  public void **setResourceEditAccessRight**(String **sessionId,** String **resourceId,** String **editAccessRight,** String[] **groupIds**)

  – **Description**

  Sets the edit access right for a resource

  – **Parameters**

  ∗ `sessionId` – The user's current session ID.

  ∗ `resourceId` – Identifier of the resource for which the edit access right has to be set

  ∗ `editAccessRight` – Edit access right of the resource. Allowed values are public, group and private.

  ∗ `groupIds` – Identifiers of the closed groups that have to be added to the edit access right if the chosen edit access right is group.

- **setResourceWithContainerInsertAccessRight**
  public void **setResourceWithContainerInsertAccessRight**(String **sessionId,** String **resourceId,** String **insertAccessRight,** String[] **groupIds**)

  – **Description**

  Sets the insert access right for a resource with container

  – **Parameters**

  ∗ `sessionId` – The user's current session ID.

  ∗ `resourceId` – Identifier of the resource for which the insert access right has to be set

* insertAccessRight – Insert access right of the resource. Allowed values are public, group and private.

* groupIds – Identifiers of the closed groups that have to be added to the insert access right if the chosen insert access right is group.

- **setUserConfiguration**
  public void **setUserConfiguration**(String **sessionId,** String **preferredLanguage,** boolean **receiveMessagesPerMail**)

    – **Description**

    Sets the configuration for the current user.

    – **Parameters**

    * sessionId – The user's current session ID.

    * preferredLanguage – The preferred language of the user that will be used in the ALOE components.

    * receiveMessagesPerMail – Determines whether ALOE messages should be sent to the user via eMail.

- **shareResourceToGroups**
  public void **shareResourceToGroups**(String **sessionId,** String **resourceId,** String[] **groupIds**)

    – **Description**

    Makes a public resource visible in the list of group resources of the specified groups.

    – **Parameters**

    * sessionId – The user's current session ID.

    * resourceId – Identifier of the resource which has to be shared to the specified groups.

    * groupIds – Identifiers of the groups to which the resource has to be shared.

- **subscribeToInfoMail**
  public void **subscribeToInfoMail**(String **sessionId,** int **mailTypeId,** String **objectId,** String **frequency,** String **additionalMetadata**)

    – **Description**

    Subscribes a user to an ALOE info mail.

    – **Parameters**

    * sessionId – The user's current session ID.

         * `mailTypeId` – Identifier of the info mail type to which the user subscribes.

         * `objectId` – Identifier of the object for which the info mail is subscribed.

         * `frequency` – Determines how often the info mail is delivered. Valid values are: daily and weekly, default: weekly.

         * `additionalMetadata` – Further data describing the info mail.

- **unsubscribeFromInfoMail**
  public void **unsubscribeFromInfoMail**(String **sessionId,** int **subscriptionId**)

  – **Description**

  Unsubscribes the current user from the specified info mail.

  – **Parameters**

      * `sessionId` – The user's current session ID.

      * `subscriptionId` – Identifier of the info mail from which the user intends to be unsubscribed.

- **updateAloeSnippetMetadata**
  public void **updateAloeSnippetMetadata**(String **sessionId,** String **aloeSnippetId,** String **creator,** String **description,** String **textContent,** String **title**)

  – **Description**

  Updates the metadata of an AloeSnippet.

  – **Parameters**

      * `sessionId` – The user's current session ID.

      * `aloeSnippetId` – The AloeSnippet ID.

      * `creator` – Author of the resource.

      * `description` – Description of the resource.

      * `textContent` – Textual content of the resource.

      * `title` – Title of the resource (mandatory).

- **updateGroupMetadata**
  public void **updateGroupMetadata**(String **sessionId,** String **groupId,** String **name,** String **description**)

  – **Description**

  Updates the metadata of the specified group.

– **Parameters**

* `sessionId` – The user's current session ID.

* `groupId` – Identifier of the group for which the metadata has to be updated.

* `name` – New name of the group.

* `description` – New description of the group.

- **updateResourceMetadata**
  public void **updateResourceMetadata**(String **sessionId,** String **resourceId,** String **associatedDate,** String **creator,** String **description,** String **language,** String **license,** String **publisher,** String **rightsHolder,** String **title**)

  – **Description**

  Updates the metadata of a resource.

  – **Parameters**

  * `sessionId` – The user's current session ID.

  * `resourceId` – Identifier of the resource for which the metadata has to be updated.

  * `associatedDate` – A point in time associated with the resource. Has to match the pattern AD YYYY-MM-DD hh:mm:ss with AD = 0 if the date is before anno domine and 1 otherwise.

  * `creator` – Author of the resource.

  * `description` – Description of the resource.

  * `language` – Language of the resource content.

  * `license` – URL of the license which is associated with the resource. Shortcuts are provided as follows:
    0: None (All rights reserved)
    1: Attribution License
    2: Attribution-NoDerivs License
    3: Attribution-NonCommercial License
    4: Attribution-ShareAlike License
    5: Attribution-NonCommercial-NoDerivs License
    6: Attribution-NonCommercial-ShareAlike License
    7: Public Domain Mark

  * `publisher` – Name of the publishing instance.

  * `rightsHolder` – Rightsholder of the resource.

* title – Title of the resource (mandatory).

* **updateUserData**
  public void **updateUserData**(String **sessionId,** String **firstName,** String **lastName,** String **gender,** String **birthday,** String **affiliation,** String **homepage,** String **phone,** String **cellPhone,** String **country,** String **zipCode,** String **city,** String **street,** String **visibility,** String **interests,** String **languages,** String **messengers**)

  – **Description**

    Updates the user data of the current user.

  – **Parameters**

    * sessionId – The user's current session ID.

    * firstName – The user's first name.

    * lastName – The user's last name.

    * gender – The user's gender.

    * birthday – The user's birthday.

    * affiliation – The user's affiliation.

    * homepage – URL of the user's homepage.

    * phone – The user's phone number.

    * cellPhone – The user's cell phone number.

    * country – The country the user lives in.

    * zipCode – The zip code of the city the user lives in.

    * city – The name of the city the user lives in.

    * street – The name of the street the user lives in.

    * visibility – Visibility of the user data (public, contacts or private).

    * interests – The user's interests (separated by ,).

    * languages – The languages a user speaks (separated by ,).

    * messengers – Messenger name and the user's ID for the messenger have to be provided in alternating order (separated by ,).

Curriculum Vitae

Persönliche Daten		Martin Memmel * 1975 in Speyer
Schulbildung		
1981 – 1985		Grundschule Zeppelinschule Speyer
1985 – 1994		Hans–Purrmann–Gymnasium Speyer Allgemeine Hochschulreife
Studium		
Okt. 1994 – Apr. 2001		Studium der Mathematik an der Universität Kaiserslautern mit Nebenfach Informatik
Beruflicher Werdegang		
		Arbeitsgruppe "Datenverwaltungssysteme" *(Univ. Kaiserslautern)*
April 1997 – Juli 1997		Wissenschaftliche Hilfskraft
		Arbeitsgruppe "Effiziente Algorithmen" *(Univ. Kaiserslautern)*
Okt. 1997 – Juli 1998		Wissenschaftliche Hilfskraft
		Arbeitsgruppe "Algorithmisches Lernen" *(Univ. Kaiserslautern)*
Okt. 1998 – April 2001		Wissenschaftliche Hilfskraft
Mai 2001 – Sep. 2003		Wissenschaftlicher Mitarbeiter
2001		*Gründung der Quertex GmbH*
		Deutsches Forschungszentrum für Künstliche *Intelligenz GmbH*
Seit Oktober 2003		Researcher

Bibliography

[AAB⁺03]   James Allan, Jay Aslam, Nicholas Belkin, Chris Buckley, Jamie
             Callan, Bruce Croft, Sue Dumais, Norbert Fuhr, Donna Harman,
             David J. Harper, Djoerd Hiemstra, Thomas Hofmann, Eduard
             Hovy, Wessel Kraaij, John Lafferty, Victor Lavrenko, David Lewis,
             Liz Liddy, R. Manmatha, Andrew McCallum, Jay Ponte, John
             Prager, Dragomir Radev, Philip Resnik, Stephen Robertson, Roni
             Rosenfeld, Salim Roukos, Mark Sanderson, Rich Schwartz, Amit
             Singhal, Alan Smeaton, Howard Turtle, Ellen Voorhees, Ralph
             Weischedel, Jinxi Xu, and ChengXiang Zhai.   Challenges in in-
             formation retrieval and language modeling: report of a workshop
             held at the center for intelligent information retrieval, university of
             massachusetts amherst, september 2002. *SIGIR Forum*, 37(1):31–47,
             2003.

[AB08]   Ben Adida and Mark Birbeck. Rdf primer, bridging the human and
             data webs. Technical report, W3C, 2008.

[Acc09]   Accenture.   Millenials at the gates: Results from accenture's
             high performance it research.   Report., 2009.   Date of pub-
             lication: February 2009. Retrieved February 23, 2009, from
             `http://www.scribd.com/docinfo/12707647?access_`
             `key=key-v4yyzelegzzya3n5e7q`.

[All84]   Thomas J. Allen. *Managing the flow of technology: technology trans-
             fer and the dissemination of technological information within the R&D
             organization*. MIT Press, Cambridge, Massachusetts, USA, 1984.

[AMZS11]   Benjamin Allbach, Martin Memmel, Peter Zeile, and Bernd Streich.
             Mobile augmented city – new methods for urban analysis and ur-
             ban design processes by using mobile augmented reality services.
             In Manfred Schrenk, Vasily V. Popovich, and Peter Zeile, editors,
             Proceedings of REAL CORP 2011, pages 633–641, Essen, May 2011.
             ISBN 978-3-9503110-1-3.

[And04]   Chris Anderson. The long tail. *Wired magazine*, 12(10), 2004.

[And06]   Chris Anderson. *The Long Tail*. Random House Business Books, 2006.

[And07]   Paul Anderson. What is web 2.0? ideas, technologies and implications for education. *JISC Technology and Standards Watch*, 2007.

[APN+07]   Stefan Apelt, Christian Prause, Till Nagel, Martin Wolpers, and Marcus Eisenhauser. Enriching e-learning contents for architecture in the mace project – activities and outlook. In Paolo Nesi Jaime Delgado and Pierfrancesco Bellini, editors, *Proceedings Variazioni Workshop held in conjunction with the 3rd International Conference on Automated Production of Cross Media Content for Mulit-channel Distribution (Axemidis 2007)*, pages 94–101, 2007. ISBN = 978-88-8453-677-8.

[ASPJ+03]   Christoph Altenhofen, Mirjana Stanisic-Petrovic, Markus Junker, Thomas Kieninger, and Haigo Hofmann. Empirische studie zum werkzeugeinsatz im umfeld der dokumentenverwaltung, 2003.

[Atl04]   Tom Atlee. Defining "Collective Intelligence", 2004. Blog of Collective Intelligence. Electronic document. Date of publication: August 09, 2004. Retrieved October 31, 2008, from `http://www.community-intelligence.com/blogs/public/2004/08/defining_collective_intelligen.html`.

[Bar94]   Carol L. Barry. User-defined relevance criteria: An exploratory study. *JASIS*, 45(3):149–159, 1994.

[BBC+07]   Mark Baillie, Gloria Bordogna, Fabio Crestani, Monica Landoni, and Gabriella Pasi. The peng system: Integrating push and pull for information access. In Dion Hoe-Lian Goh, Tru Hoang Cao, Ingeborg SÃ¸lvberg, and Edie M. Rasmussen, editors, *ICADL*, volume 4822 of *Lecture Notes in Computer Science*, pages 351–360. Springer, 2007. ISBN: 978-3-540-77093-0.

[BC92]   Nicholas J. Belkin and Bruce B. Croft. Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM*, 35(12):29–38, December 1992.

[BCK03]   Len Bass, Paul Clements, and Rick Kazman. *Software Architecture in Practice*. Addison-Wesley, Boston, MA, 2. edition, 2003. ISBN 0-321-15495-9.

[Bel08]   Michael Bell. *Service-Oriented Modeling: Service Analysis, Design, and Architecture*. Wiley & Sons, 2008. ISBN 978-0-470-14111-3.

[BEM+13a]   Benjamin Bergner, Jan Exner, Martin Memmel, Rania Raslan, Dina Taha, Manar Talal, and Peter Zeile. Human sensory assessment linked with geo- and mobile-data processing methods in urban planning exemplified on different planning cultures in germany and egypt. In Stan Geertman, John Stillwell, and Fred Toppen, editors, *Proceedings of CUPUM 2013*, Utrecht, the Netherlands, July 2013. ISBN 978-90-816960-2-9.

[BEM+13b]   Benjamin Bergner, Jan Exner, Martin Memmel, Rania Raslan, Dina Taha, Manar Talal, and Peter Zeile. Human sensory assessment methods in urban planning – a case study in alexandria. In Manfred Schrenk, Vasily V. Popovich, Peter Zeile, and Pietro Elisei, editors, *Proceedings of REAL CORP 2013*, pages 407–417, Rome, Italy, May 2013. ISBN 978-3-9503110-5-1.

[BH04]   Thomas R. Bruce and Diane I. Hillmann. *Metadata in Practice*, chapter The Continuum of Metadata Quality: Defining, Expressing, Exploiting, pages 238–256. American Library Association, 2004.

[BL94]   Tim Berners-Lee. Universal resource identifiers in www. Informational memo RFC1630, Internet Engineering Task Force, June 1994.

[Bla05]   Simon Blackburn. *Wahrheit – Ein Wegweiser für Skeptiker*. Primus Verlag, Darmstadt, 2005.

[BLFIM98]   Tim Berners-Lee, R. Fielding, U.C. Irvine, and L. Masinter. Uniform Resource Identifiers (URI): Generic Syntax. rfc 2396, IETF, 1998.

[Bor03]   Pia Borlund. The concept of relevance in ir. *Journal of the American Society for Information Science and Technology*, 54(10):913–925, May 2003. ISSN 1532-2890.

[Bra08]   Daren C. Brabham. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1):75–90, February 2008.

[BS04]   Steven Bird and Gary Simons. *Metadata in Practice*, chapter Building an Open Language Archive Community on the DC foundation, pages 203–222. American Library Association, 2004.

[Bus45]   Vannevar Bush. As we may think. *The Atlantic Monthly*, 176(1):101–108, 1945.

[BYRN99]   Ricardo A. Baeza-Yates and Berthier A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

[Car95]   Martin Carrier. *Enzyklopädie Philosophie und Wissenschaftstheorie*, volume 3, chapter Relativismus, page 564f. Metzler, Stuttgart; Weimar, 1995.

[CDO06]   Kris Cardinaels, Erik Duval, and Henk Olivié. 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.

[Cha74]   K. Alok Chakravarti. The role of champion in product innovation. *California Management Review*, 17(2):58–62, 1974.

[Coa03]   Tom Coates. My working definition of social software..., 2003. Electronic document. Date of publication: May 8, 2003. Retrieved October 27, 2008, from `http://www.plasticbag.org/archives/2003/05/my_working_definition_of_social_software/`.

[Com]   Creative Commons. Choosing a license. Electronic document. Retrieved October 23, 2006, from `http://creativecommons.org/about/licenses`.

[com07]   comScore. Social networking goes global – major social networking sites substantially expanded their global visitor base during past year. Press Release., July 2007. Date of publication: July 31, 2007. Retrieved November 09, 2008, from `http://www.comscore.com/press/release.asp?press=1555`.

[Com09]   UNESCO OER Community. Access2oer/issues, 2009. Electronic document. Retrieved March 24, 2009, from `http://oerwiki.iiep-unesco.org/index.php?title=Access2OER/Issues`.

[com10]   comScore. The state of social networks in asia pacific, with a focus on singapore. White Paper., January 2010. Date of publication: January 15, 2010. Retrieved August 10, 2010, from `http://www.comscore.com/Press_Events/Presentations_Whitepapers/2010/The_State_of_Social_Networks_in_Asia_Pacific_with_a_Focus_on_Singapore`.

[CPPB01]   Rob Cross, Andrew Parker, Laurence Prusak, and Stephen P. Borgatti. Knowing what we know: Supporting knowledge creation and sharing in social networks. *Organizational Dynamics*, 30(2):100–120, November 2001.

[CS01]   Keith Cheverst and Gareth Smith. Exploring the notion of information push and pull with respect to the user intention and disruption. In *International workshop on Distributed and Disappearing User Interfaces in Ubiquitous Computing*, pages 67–72, 2001.

[Dav86]   D. Fred Davis. *A technology acceptance model for empirically testing new end-user information systems: Theory and results: PhD Thesis.* PhD thesis, Massachusetts Institute of Technology, Boston, 1986.

[DC003]   Iso 15836:2003(e) – information and documentation - the dublin core metadata element set. International standard published, International Organization for Standardization (ISO), 2003.

[Den06]   Andreas Dengel. Six thousand words about multi-perspective personal document management. 2006.

[DHSW02]   Erik Duval, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel. Metadata principles and practicalities. *D-lib Magazine*, 8(4):1–16, April 2002.

[Dim05]   Garrett Dimon. Microformats primer. *Digital Web Magazine*, 2005.

[Doc01]   Cory Doctorow. Metacrap: Putting the torch to seven strawmen of the meta-utopia, 2001. Electronic document. Date of publication: August 26,2001. Retrieved March 14, 2007, from http://www.well.com/ doctorow/metacrap.htm.

[Dow04]   Stephen Downes. Resource profiles. *Journal of Interactive Media in Education*, 5, 2004. ISSN:1365-893X.

[DUBqW09] Manni Duan, Adrian Ulges, Thomas M. Breuel, and Xiu qing Wu. Style modeling for tagging personal photo collections. In *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8, New York, NY, USA, 2009. ACM.

[Enc08] Encyclopaedia Britannica Online. Writing, 2008. David R. Olson (editor). Electronic document. Retrieved August 04, 2008, from `http://www.britannica.com/EBchecked/topic/649670/writing`.

[Eng62] Douglas C. Engelbart. Augmenting human intellect: A conceptual framework. Technical report, Stanford Research Institute, Palo Alto, October 1962.

[FG08] Martin Fisch and Christoph Gscheidle. Mitmachnetz web 2.0: Rege beteiligung nur in communitys. *Media Perspektiven*, 7:356–364, 2008.

[FM06] Ichiro Fuinaga and Daniel McEnnis. On-demand metadata extraction network (omen). In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pages 346–346, New York, NY, USA, 2006. ACM.

[FM08] Andrew J. Flanagin and Miriam J. Metzger. *Digital Media, Youth, and Credibility*, chapter Digital Media and Youth: Unparalleled Opportunity and Unprecedented Responsibility, pages 5–28. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. MIT Press, Cambridge, MA, 2008.

[GAB09] Marie-Luise Groš, Benjamin Adrian, and Gerhard Budin. Study: The web 2.0 - a high capacity research landscape for professional translators? In Klaus Tochtermann and Hermann Maurer, editors, *Proceedings of I-KNOW '09 and I-SEMANTICS '09.*, pages 348–353. Know-Center, Graz, Journal of Universal Computer Science, September 2009. ISBN 978-3-85125060-2.

[Gar08] Gartner. Gartner identifies the top 10 strategic technologies for 2008, 2008. Electronic document. Date of publication: October 09, 2007. Retrieved November 04, 2008, from `http://www.gartner.com/it/page.jsp?id=530109`.

[Ger91]   Anne Geraci. *IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries*. Institute of Electrical and Electronics Engineers, 1991.

[GGM⁺11]   Jean-Marie Gaillourdet, Thomas Grundmann, Martin Memmel, Karsten Schmidt, Arnd Poetzsch-Heffter, and Stefan Dessloch. Wom: An open interactive platform for describing, exploring, and sharing mathematical models. In Andreas König, Andreas Dengel, Knut Hinkelmann, Koichi Kise, Robert J. Howlett, and Lakhmi C. Jain, editors, *Knowledge-Based and Intelligent Information and Engineering Systems, Part IV*, volume 6884 of *Lecture Notes in Computer Science*, pages 126–135. Springer, September 2011.

[GGS⁺11]   Thomas Grundmann, Jean-Marie Gaillourdet, Karsten Schmidt, Arnd Poetzsch-Heffter, Stefan Dessloch, and Martin Memmel. The web of mathematical models: A schema-based, wiki-like, interactive platform. In Christoph Lange and Josef Urban, editors, *Proceedings of the ITP 2011 Workshop on Mathematical Wikis (MathWikis 2011)*, volume 767 of *CEUR Workshop Proceedings*, pages 19–27. CEUR-WS.org, August 2011.

[GH05]   Scott Golder and Bernardo A. Huberman. The structure of collaborative tagging systems, 2005.

[Gil05]   Jim Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, December 2005. ISSN: 0028-0836.

[GLM03]   Gunter Grieser, Steffen Lange, and Martin Memmel. DaMiT: Ein adaptives Tutorsystem für Data-Mining. In Klaus P. Jantke, Jörg Herrmann, and Wolfgang S. Wittig, editors, *Von e-Learning bis e-Payment. Das Internet als sicherer Marktplatz*, pages 192–203. Akademische Verlagsgesellschaft Aka, 2003.

[GM03]   Rebecca Guenther and Sally McCallum. New metadata standards for digital resources: Mods and mets. In *Bulletin of the American society for Information Science and Technology*, volume 29, pages 12–15. Richard B. Hill, 2003.

[Gol01]   Goldcorp. Us $575.000 goldcorp challenge awards world's first 6 million ounce internet gold rush yields high grade results! Press Release, March 2001.

[Goo08]   Google. We knew the web was big..., July 2008. Electronic document. Date of publication: July 25, 2008. Retrieved November 09, 2008, from `http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html`.

[Gre03]   Jane Greenberg. Metadata generation: Processes, people and tools. In *Bulletin of the American society for Information Science and Technology*, volume 29, pages 16–19. Richard B. Hill, 2003.

[Gro06]   Lev Grossman. Time's person of the year: You. *Time Magazine*, 2006.

[GS96]   Ralph Grishman and Beth Sundheim. Message understanding conference – a brief history. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, pages 466–471, Kopenhagen, Denmark, 1996.

[GSH07]   G. Hans Gemünden, Sören Salomo, and Katharina Hölzle. Role models for radical innovations in times of open innovation. *Creativity and Innovation Management*, 16(4):408–421, 2007.

[HJSS06]   Andreas Hotho, Robert Jaschke, Christoph Schmitz, and Gerd Stumme. Trend detection in folksonomies. In *First International Conference on Semantics and Digital Media Technologies (SAMT)*, pages 56–70, Athens, Greece, 2006.

[Hol02]   Harald Holz. *Process-Based Knowledge Management Support for Software Engineering*. PhD thesis, Technische Universität Kaiserslautern, 2002.

[How06a]   Jeff Howe. Crowdsourcing: A definition, 2006. Electronic document. Date of publication: June 02, 2006. Retrieved November 04, 2008, from `http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html`.

[How06b]   Jeff Howe. The rise of crowdsourcing. *Wired*, 14(6), 2006. Electronic document. Date of publication: June, 2006. Retrieved May 11, 2008, from `http://www.wired.com/wired/archive/14.06/crowds.html`.

[HP00]   Rachel Heery and Manjula Patel. Application profiles: mixing and matching metadata schemas. In *Ariadne Issue 25*. Ariadne, 2000.

© *Martin Memmel, 2015*

[HR08]   Brian Hilligoss and Soo Young Rieh. Developing a unifying frame-
            work of credibility assessment: Construct, heuristics, and interac-
            tion in context. *Information Processing & Management*, 44(4):1467–
            1484, 2008.

[HSHR09]   Keith Hampton, Lauren Sessions, Eun Ja Her, and Lee Rainie. So-
            cial isolation and new technology - how the internet and mobile
            phones impact americans'Â social networks. Technical report, Pew
            Internet & American Life Project, November 2009.

[HSS01]   Uri Hanani, Bracha Shapira, and Peretz Shoval. Information filter-
            ing: Overview of issues, research and systems. *User Modeling and
            User-Adapted Interaction*, 11(3):203–259, 2001.

[Hun98]   Jane Hunter. The application of metadata standards to video in-
            dexing. In *ECDL '98: Proceedings of the Second European Conference
            on Research and Advanced Technology for Digital Libraries*, pages 135–
            156, London, UK, 1998. Springer-Verlag.

[IHB⁺08]   Mizuko Ito, Heather A. Horst, Matteo Bittanti, danah boyd, Becky
            Herr-Stephenson, Patricia G. Lange, C.J. Pascoe, Laura Robinson
            (with Sonja Baumer, Rachel Cody, Dilan Mahendran, Katynka
            MartÃnez, Dan Perkel, Christo Sims, and Lisa Tripp). Living and
            learning with new media: Summary of findings from the digital
            youth project. White paper. the john d. and catherine t. macarthur
            foundation reports on digital media and learning., The John D. and
            Catherine T. MacArthur Foundation, November 2008.

[Int97]   Lucas D. Introna. Privacy and the computer: Why we need privacy
            in the information society. *Metaphilosophy*, 28(3):259–275, 1997.

[JDG⁺04]   Klaus P. Jantke, Gerhard Degel, Gunter Grieser, Martin Memmel,
            Oleg Rostanin, and Bernd Tschiedel. Technology enhanced dimen-
            sions in e-learning. In Michael E. Auer and Ursula Auer, editors,
            *International Conference on Interactive Computer Aided Learning, ICL
            2004, Sept. 29 – Oct. 1, 2004, Villach, Austria (CD-ROM)*, 2004. ISBN
            3-89958089-3.

[JGLM04]   Klaus P. Jantke, Gunter Grieser, Steffen Lange, and Martin Mem-
            mel. DaMiT: Data Mining lernen und lehren. In A. Abecker,
            S. Bickel, U. Brefeld, I. Drost, N. Henze, O. Herden, M. Minor,
            T. Scheffer, L. Stojanovic, and S. Weibelzahl, editors, *LWA 2004,*

Lernen − Wissensentdeckung − Adaptivität, 4.-6.Oktober 2004, pages 171–179. Humboldt-Universität zu Berlin, October 2004.

[Koc08]   Michael Kockler. Conception & implementation of tagging support mechanisms in a web 2.0 social media sharing platform. Master's thesis, Technische Universität Kaiserslautern, 2008.

[Kön92]   Gert König. *Relativismus*, volume 8, chapter Relativismus, pages 619–622. Wissenschaftliche Buchgesellschaft Darmstadt, Basel, 1992.

[Kos03]   Ilpo Koskinen. User-generated content in mobile multimedia: Empirical evidence from user studies. In *ICME '03: Proceedings of the 2003 International Conference on Multimedia and Expo*, pages 645–648, Washington, DC, USA, July 2003. IEEE Computer Society.

[KZ01]   Yu-Hwan Kim and Byoung-Tak Zhang. Document indexing using independent topic extraction. In *Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2001.

[LDP08]   Daniel Lemire, Stephen Downes, and Sébastien Paquet. Diversity in open social networks, October 2008. Electronic document. Date of publication: October 2008. Retrieved October 27, 2008, from http://www.daniel-lemire.com/fr/abstracts/DIVERSITY2008.html.

[Len09]   Amanda Lenhart. Adults and social network websites. Technical report, Pew Internet & American Life Project, January 2009.

[LGZ08]   Xin Li, Lei Guo, and Yihong Eric Zhao. Tag-based social interest discovery. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 675–684, New York, NY, USA, 2008. ACM.

[LHDL04]   Scott Lederer, Jason I. Hong, Anind K. Dey, and James A. Landay. Personal privacy through understanding and action: five pitfalls for designers. *Personal and Ubiquitous Computing*, 8(6):440–454, 2004.

[LJ07]   Kristina Lerman and Laurie Jones. Social browsing on flickr. In *Proceedings of International Conference on Weblogs and Social Media (ICWSM)*, March 2007.

[LMMS07]   Amanda Lenhart, Mary Madden, Alexandra Rankin Macgill, and Aaron Smith. Teens and social media – the use of social media gains a greater foothold in teen life as they embrace the conversational nature of interactive online media. Technical report, Pew Internet & American Life Project, December 2007.

[Luh61]   Hans Peter Luhn. Selective dissemination of new scientific information with the aid of electronic processing equipment. *American Documentation*, 12:131–138, 1961.

[LV09]   Peter Lyman and Hal R. Varian. How much information, 2009. Electronic document. Retrieved October 11, 2008, from `http://www.sims.berkeley.edu/how-much-info-2003`.

[LW91]   Jean Lave and Etienne Wenger. *Situated learning. Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, 1991.

[Mar65]   M. E. Maron. Mechanized documentation: The logic behind a probabilistic interpretation. In M.E. Stevens et al., editor, *Statistical Association Methods for Mechanized Documentation Symposium Proceedings 1964*, volume 269, pages 9–13, Washinton, 1965. National Bureau of Standards.

[Mar06]   Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006. ISSN = 0001-0782.

[MD07]   Martin Memmel and Andreas Dengel. Sharing contextualized attention metadata to support personalized information retrieval. In Erik Duval, Jehad Najjar, and Martin Wolpers, editors, *Proceedings of the 2nd International ACM/IEEE Workshop on Contextualized Attention Metadata: Personalized Access to Digital Resources, CAMA2007*, volume 266, pages 19–26. CEUR workshop proceedings, 2007. ISSN 1613-0073.

[MG11]   Martin Memmel and Florian Gross. Radar – potentials for supporting urban development with a social geocontent hub. In Manfred Schrenk, Vasily V. Popovich, and Peter Zeile, editors, *Proceedings of REAL CORP 2011*, pages 777–784, Essen, May 2011. ISBN 978-3-9503110-1-3.

[MKS08]   Martin Memmel, Michael Kockler, and Rafael Schirru. Providing multi source tag recommendations in a social resource sharing platform. In Hermann Maurer, Frank Kappe, Werner Haas, and Klaus Tochtermann, editors, *Proceedings of I-MEDIA '08*, pages 226–233. Know-Center, Graz, Journal of Universal Computer Science, September 2008. ISSN 0948-695x.

[MKS09]   Martin Memmel, Michael Kockler, and Rafael Schirru. Providing multi source tag recommendations in a social resource sharing platform. *Journal of Universal Computer Science*, 15(3):678–691, 2009. `http://www.jucs.org/jucs_15_3/providing_multi_source_tag`.

[MM04]   Frank Manola and Eric Miller. Rdf primer, w3c recommendation. Technical report, W3C, 2004.

[MNBD06]   Cameron Marlow, Mor Naaman, Danah Boyd, and Marc Davis. Ht06, tagging paper, taxonomy, flickr, academic article, toread. In *Proceedings of Hypertext 2006*, New York, 2006. ACM, New York: ACM Press.

[MOD07]   Michael Meire, Xavier Ochoa, and Erik Duval. Samgi: Automatic metadata generation v2.0. In C. Montgomerie and J. Seale, editors, *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, pages 1195–1204. AACE, October 2007.

[Mog07]   Bill Moggridge. From the desk to the palm – interview with jeff hawkins, 2007. Electronic document. Retrieved March 8, 2010, from `http://www.designinginteractions.com/interviews/JeffHawkins`.

[MPS99]   Christos J. Moschovitis, Hilary Poole, and Theresa M. Senf. *History of the Internet: A Chronology, 1843 to the Present*. A B C-CLIO, Inc., 1999.

[MRW$^+$06]   Martin Memmel, Eric Ras, Stephan Weibelzahl, Daniel Burgos, Daniel Olmedilla, and Martin Wolpers, editors. *Proceedings of the Joint International Workshop on Professional Learning, Competence Development and Knowledge Management*, Heraklion, Crete, Greece, October 2006. CEUR workshop proceedings.

[MRWA07]   Martin Memmel, Eric Ras, Martin Wolpers, and Frans Van Aasche, editors. *Proceedings of the 3rd Workshop on Learner-Oriented Knowledge Management & KM-Oriented E-Learning (LOKMOL 2007)*, Sissi, Lassithi, Crete, Greece, October 2007. CEUR workshop proceedings.

[MRZ05]   Martin Memmel, Eric Ras, and Lars Zapf. Tutorial "Setting Up Your Own Learning Object Environment", affiliated to the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn 2005), 24 October - 28 October 2005, Vancouver, Canada., October 2005.

[MSWT08]   Martin Memmel, Rafael Schirru, Martin Wolpers, and Elia Tomadaki. Towards the combined use of metadata to improve the learning experience. In Paloma DÃaz, Kinshuk, Ignacio Aedo, and Eduardo Mora, editors, *Proceedings of the The 8th IEEE International Conference on Advanced Learning Technologies*, pages 930–932. IEEE, IEEE Conference Publishing Services, July 2008. ISBN 978-0-7695-3167-0.

[MTW08]   Martin Memmel, Elia Tomadaki, and Martin Wolpers. An approach to enable collective intelligence in digital repositories. In Joseph Luca and Edgar R. Weippl, editors, *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications, ED-MEDIA 2008, Vienna*, pages 1803–1811. AACE, 2008. ISBN: 1-880094-65-7.

[MWC$^+$10]   Martin Memmel, Martin Wolpers, Massimiliano Condotta, Katja Niemann, and Rafael Schirru. Introducing a social backbone to support access to digital resources. In Martin Wolpers, Paul A. Kirschner, Maren Scheffel, Stefanie Lindstaedt, and Vania Dimitrova, editors, *EC-TEL 2010 - Sustaining TEL. From Innovation to Learning in Practice*, volume 6383 of *Lecture Notes in Computer Science*, pages 560–565. Springer, 2010. ISBN 978-3-642-16019-6.

[Net04]   Media Awareness Network. Young canadians in a wired world. phase ii. focus groups. Technical report, Media Awareness Network, Ottawa, Ontario, Canada, February 2004.

[Nie00]   Jakob Nielsen. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, 2000.

[NND$^+$08] Mikael Nilsson, Ambjörn Naeve, Erik Duval, Pete Johnston, and David Massart. Harmonization of metadata standards. Technical report, ProLearn: Network of Excellence in Professional Learning, February 2008.

[NPN02] Mikael Nilsson, Matthias Palmér, and Ambjörn Naeve. Semantic web metadata for e-learning - some architectural guidelines. In *Proceedings of the 11th World Wide Web Conference*, 2002.

[NS05] M. Nottingham and R. Sayre. The atom syndication format. RFC 4287 (Proposed Standard), December 2005.

[OD08] Xavier Ochoa and Erik Duval. Quantitative analysis of user-generated content on the web. In *Proceedings of webevolve2008: web science workshop at WWW2008*, 2008.

[OEC03] OECD. Quality framework for oecd statistical activities. Technical Report 2003/1, Organisation for Economic Co-operation and Development (OECD), October 2003.

[oEL02] Institute of Electrical and Electronics Engineers Learning Technology Standards Committee (IEEE LTSC). Draft standard for learning object metadata, 2002.

[O'R05a] Tim O'Reilly. Web 2.0: Compact Definition?, 2005. O'Reilly Radar. Electronic document. Date of publication: October 01, 2005. Retrieved February 12, 2007, from `http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html`.

[O'R05b] Tim O'Reilly. 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`.

[Org04] National Information Standards Organization. Understanding metadata. Technical report, National Information Standards Organization, Bethesda, MD, USA, 2004.

[otFRfBR98] IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records: Final Report*, volume 19 of *UBCIM Publications-New Series*. K.G.Saur, München, 1998.

[PBMW98]   Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, January 1998.

[PD03]   Leysia Palen and Paul Dourish. Unpacking privacy for a networked world. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 136. ACM, 2003.

[Por07]   Joshua Porter. The Danger of Aggregate Displays in Social Software, 2007. Electronic document. Date of publication: September 07, 2007. Retrieved November 05, 2008, from `http://bokardo.com/archives/the-danger-of-aggregate-displays-in-social-software/`.

[PTdJ⁺07]   Christian Prause, Stefaan Ternier, Tim de Jong, Stefan Apelt, Marius Scholten, Martin Wolpers, Markus Eisenhauer, Bram Vandeputte, Marcus Specht, and Erik Duval. Unifying learning object repositories in mace. In David Massart, Jean-Noel Colin, and Frans Van Assche, editors, *LODE*, volume 311 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2007.

[Rai07]   Lee Rainie. 28% of online americans have used the internet to tag content. Technical report, Pew Internet & American Life Project, January 2007.

[Ray94]   W. Boyd Rayward. Visions of xanadu: Paul otlet (1868-1944) and hypertext. *Journal of the American Society for Information Science*, 45:235–250, 1994.

[RD07]   Soo Young Rieh and David R. Danielson. Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology*, 41:307–364, 2007.

[Ree66]   A. M. Rees. The relevance of relevance to the testing and evaluation of document retrieval systems. In *Aslib Proceedings*, volume 18, pages 316–324, 1966.

[Ree04]   Inga Elisabeth Reeps. Joy-of-use: eine neue qualität für interaktive produkte. Master's thesis, University of Konstanz, 2004.

[Ric83]   Elaine Rich. Users are individuals: individualising user models. *International Journal of Man-Machine Studies*, 18:199–214, 1983.

[RMW05]   Eric Ras, Martin Memmel, and Stephan Weibelzahl. *Professional Knowledge Management*, volume 3782 of *Lecture Notes in Artificial Intelligence LNAI*, chapter Integration of E-Learning and Knowledge Management - Barriers, Solutions and Future Issues, pages 155–164. Springer-Verlag, Berlin, 2005.

[Rob81]   Stephen E. Robertson. The methodology of information retrieval experiment. In Karen Sparck Jones, editor, *Information Retrieval Experiment*, pages 9–31. Butterworth-Heinemann, Newton, MA, 1981. ISBN: 978-0408106481.

[Rut05]   Ian Ruthven. *Integrating Approaches to Relevance*, volume 19 of *Information Retrieval Series*, chapter 4, pages 61–80. Springer, 2005.

[SA88]   Ralph R. Swick and Mark S. Ackerman. The x toolkit: More bricks for building user-interfaces or widgets for hire. In *USENIX Winter*, pages 221–228, 1988.

[Sal68]   Gerard M. Salton. *Automatic Information Organization and Retrieval*. McGraw Hill Text, 1968. isbn = 0070544859.

[Sar96]   Tefko Saracevic. Relevance reconsidered. In *Proceedings of the Second Conference on Conceptions of Library and Information Science (Co-LIS 2)*, pages 201–218, 1996.

[SBMD10]   Rafael Schirru, Stephan Baumann, Martin Memmel, and Andreas Dengel. Extraction of contextualized user interest profiles in social sharing platforms. *Journal of Universal Computer Science*, 16(16):2196–2213, 2010.

[SBMD11]   Rafael Schirru, Stephan Baumann, Martin Memmel, and Andreas Dengel. Topic-based recommendations for enterprise 2.0 resource sharing platforms. In Andreas König, Andreas Dengel, Knut Hinkelmann, Koichi Kise, Robert J. Howlett, and Lakhmi C. Jain, editors, *Knowledge-Based and Intelligent Information and Engineering Systems*, volume 6881 of *Lecture Notes in Computer Science*, pages 495–504. Springer, September 2011.

[Sch08]   Rolf Schulmeister. Gibt es eine "net generation"?, 2008. Electronic document. Retrieved January 9, 2009, from `http://www.zhw.uni-hamburg.de/pdfs/Schulmeister_Netzgeneration.pdf`.

[Sch10]   Sven Schwarz. *Context-Awareness and Context-Sensitive Interfaces for Knowledge Work*. Dissertation, Technische Universität Kaiserslautern, Fachbereich Informatik, March 2010.

[Sco04]   *Sharable Content Object Reference Model (SCORM©) 2004 2nd Edition Overview*. Advanced Distributed Learning (ADL), 2004.

[SCVN08]   Vittoria Spigai, Massimiliano Condotta, Elisa Dalla Vecchia, and Till Nagel. Semiotic based facetted classification to support browsing architectural contents in mace. In Marja Naaranoja, Ad den Otter, Matthijs Prins, Anu Karvonen, and Ville Raasakka, editors, *Proceedings of Joint CIB Conference: Performance and Knowledge Management*, pages 273–284, June 2008. ISBN 978-951-758-492-0.

[Seg07]   Toby Segaran. *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. O'Reilly Media, Inc., August 2007.

[SEN90]   Linda Schamber, Michael B. Eisenberg, and Michael Sanford Nilan. A re-examination of relevance: toward a dynamic, situational definition. *Information Process & Management*, 26(6):755–776, 1990.

[Sen04]   Arun Sen. Metadata management: past, present and future. *Decision Support Systems*, 37(1):151–173, April 2004.

[Six05]   Mario Sixtus. Das Web sind wir. Social Software und das neue Leben im Netz, 2005. Electronic document. Date of publication: July, 2005. Retrieved February 05, 2007, from `http://www.sixtus.net/article/614_0_2_0_C/`.

[SMA13]   Caroline Sabty, Martin Memmel, and Slim Abdennadher. Geoevents – an interactive tool to analyze and visualize spatial information from the social web. In Randall Bilof, editor, *Proceedings of the ASE/IEEE International Conference on Social Computing (SocialCom) 2013*, pages 803–808, Washington, D.C., USA, September 2013.

[SMP03]   Carol Sansone, Carolyn C. Morf, and A. T. Panter, editors. *The Sage Handbook of Methods in Social Psychology*. Sage Publications, Inc, 2003. ISBN-13: 978-0761925354.

[SNLW04]   Herbert Van De Sompel, Michael Nelson, Carl Lagoze, and Simeon Warner. Resource harvesting within the oai-pmh framework. *D-Lib Magazine*, 10(12), December 2004.

[Sur04]   James Surowiecki. *The Wisdom of Crowds*. Doubleday, 2004.

[SVC⁺07]   Moritz Stefaner, Elisa Dalla Vecchia, Massimiliano Condotta, Martin Wolpers, Marcus Specht, Stefan Apelt, and Erik Duval. Mace – enriching architectural learning objects for experience multiplication. In Erik Duval, Ralf Klamma, and Martin Wolpers, editors, *Proceedings of the 2nd European Conference on Technology Enhanced Learning (EC-TEL 2007)*, volume 4753 of *Lecture Notes in Computer Science*, pages 322–336. Springer, 2007. ISBN = 978-3-540-75194-6.

[SWM⁺09]   Moritz Stefaner, Martin Wolpers, Martin Memmel, Erik Duval, Marcus Specht, Dirk Börner, Marion Gruber, Tim De Jong, Alberto Giretti, and Roland Klemke. Mace: Joint deliverable "evaluation of the mace system". Deliverable, The MACE consortium, 2009.

[T08]   AT & T. Social networking in the workplace increases efficiency. Press Release., November 2008. Date of publication: November 11, 2008. Retrieved November 25, 2008, from `http://www.corp.att.com/emea/insights/pr/eng/social_111108.html`.

[Tay62]   Robert S. Taylor. The process of asking questions. *American Documentation*, 13(4):391–396, 1962.

[Til04]   Barbara Tillett. What is frbr? a conceptual model for the bibliographic universe. Technical report, Library of Congress, 2004.

[Tor08]   Nat Torkington. Radar Roundup: Collective Intelligence, 2008. Electronic document. Date of publication: February 26, 2008. Retrieved October 15, 2008, from `http://radar.oreilly.com/2008/02/radar-roundup-collective-intel.html`.

[TW06]   Don Tapscott and Anthony D. Williams. *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio, 2006.

[UKBB09]   Adrian Ulges, Markus Koch, Damian Borth, and Thomas M. Breuel. Tubetagger - youtube-based concept detection. In *ICDMW '09: Proceedings of the 2009 IEEE International Conference on Data Mining Workshops*, pages 190–195, Washington, DC, USA, 2009. IEEE Computer Society.

[vAD04]   Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, USA, 2004. ACM Press.

© *Martin Memmel, 2015*

[vAGK$^+$06] Luis von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum. Improving accessibility of the web with a computer game. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 79–82, New York, USA, 2006. ACM Press.

[VWV07] Graham Vickery and Sacha Wunsch-Vincent. *Participative Web and User-Created Content: Web 2.0, Wikis and Social Networking*. OECD, 2007.

[Wal07] Thomas Vander Wal. Folksonomy Coinage and Definition, February 2007. Electronic document. Date of publication: February 02, 2007. Retrieved October 15, 2008, from `http://vanderwal.net/folksonomy.html`.

[WD06] Wolfgang Wahlster and Andreas Dengel. Web 3.0: Convergence of web 2.0 and the semantic web. *Technology Radar, Feature Paper, Edition II*, pages pp. 1–23, 2006.

[Wei02] David Weinberger. *Small Pieces Loosely Joined: A Unified Theory of the Web*. Perseus Books Group, April 2002. ISBN 0-7382-0543-5.

[Wei07] David Weinberger. *Everything is Miscellaneous: The Power of New Digital Disorder*. Times Books, New York, 2007. ISBN-13 978-0-8050-8043-8.

[Wik08a] Wikipedia. Collective intelligence — wikipedia, the free encyclopedia, 2008. Electronic document. Retrieved October 30, 2008, from `http://en.wikipedia.org/w/index.php?title=Collective_intelligence&oldid=248588214`.

[Wik08b] Wikipedia. Computer — wikipedia, the free encyclopedia, 2008. Electronic document. Retrieved August 4, 2008, from `http://en.wikipedia.org/w/index.php?title=Computer&oldid=230963828`.

[Wik08c] Wikipedia. History of computing hardware — wikipedia, the free encyclopedia, 2008. Electronic document. Retrieved August 4, 2008, from `http://en.wikipedia.org/w/index.php?title=History_of_computing_hardware&oldid=232747794`.

[Wik08d]   Wikipedia.   History of printing — wikipedia, the free en-
            cyclopedia, 2008.   Electronic document. Retrieved August
            4, 2008, from `http://en.wikipedia.org/w/index.php?`
            `title=History_of_printing&oldid=231937873`.

[Wik08e]   Wikipedia.   History of radio — wikipedia, the free en-
            cyclopedia, 2008.   Electronic document. Retrieved August
            4, 2008, from `http://en.wikipedia.org/w/index.php?`
            `title=History_of_radio&oldid=232677701`.

[Wik08f]   Wikipedia.   History of telecommunication — wikipedia, the
            free encyclopedia, 2008.   Electronic document. Retrieved Au-
            gust 4, 2008, from `http://en.wikipedia.org/w/index.`
            `php?title=History_of_telecommunication&oldid=`
            `228934633`.

[Wik08g]   Wikipedia.   History of television — wikipedia, the free en-
            cyclopedia, 2008.   Electronic document. Retrieved August
            4, 2008, from `http://en.wikipedia.org/w/index.php?`
            `title=History_of_television&oldid=232738848`.

[Wik08h]   Wikipedia.   History of writing — wikipedia, the free en-
            cyclopedia, 2008.   Electronic document. Retrieved August
            4, 2008, from `http://en.wikipedia.org/w/index.php?`
            `title=History_of_writing&oldid=232277266`.

[Wik08i]   Wikipedia.   Printing — wikipedia, the free encyclope-
            dia, 2008.   Electronic document. Retrieved August 4, 2008,
            from   `http://en.wikipedia.org/w/index.php?title=`
            `Printing&oldid=231498898`.

[Wik08j]   Wikipedia.   Radio — wikipedia, the free encyclopedia,
            2008.   Electronic document. Retrieved August 4, 2008,
            from   `http://en.wikipedia.org/w/index.php?title=`
            `Radio&oldid=232654061`.

[Wik08k]   Wikipedia.   Social media — wikipedia, the free ency-
            clopedia, 2008.   Electronic document. Retrieved August
            4, 2008, from `http://en.wikipedia.org/w/index.php?`
            `title=Social_media&oldid=229282906`.

[Wik08l]   Wikipedia.   Telecommunication — wikipedia, the free en-
            cyclopedia, 2008.   Electronic document. Retrieved August

© *Martin Memmel, 2015*

4, 2008, from `http://en.wikipedia.org/w/index.php?`
`title=Telecommunication&oldid=231864483`.

[Wik08m] Wikipedia. Television — wikipedia, the free encyclope-
dia, 2008. Electronic document. Retrieved August 4, 2008,
from `http://en.wikipedia.org/w/index.php?title=`
`Television&oldid=232678756`.

[Wik08n] Wikipedia. Writing — wikipedia, the free encyclope-
dia, 2008. Electronic document. Retrieved August 4, 2008,
from `http://en.wikipedia.org/w/index.php?title=`
`Writing&oldid=229464337`.

[Wik09] Wikipedia. White-label product — wikipedia, the free
encyclopedia, 2009. Electronic document. Retrieved May
25, 2009, from `http://en.wikipedia.org/w/index.php?`
`title=White-label_product&oldid=289642140`.

[Wil81] T.D. Wilson. On user studies and information needs. *Journal of
Librarianship*, 37(1):3–15, 1981.

[Wit73] Eberhard Witte. *Organisation für Innovationsentscheidungen: Das
Promotoren-Modell*, volume 2 of *Schriften der Kommission für
Wirtschaftlichen und Sozialen Wandel*. Schwartz, Göttingen, 1973.

[Wit01] Ludwig Wittgenstein. *Philosophische Untersuchungen*. Wis-
senschaftliche Buchgesellschaft, Frankfurt am Main, kritisch-
genetische edition edition, 2001.

[WMG+12] Martin Wolpers, Martin Memmel, Alberto Giretti, Miquel Casals,
Katja Niemann, and Marcus Specht. *Collaborative Learning 2.0:
Open Educational Resources*, chapter Supporting Collaborative
Learning in the Architectural Domain, pages 328–356. IGI Global,
2012.

[WMK+09] Martin Wolpers, Martin Memmel, Joris Klerkx, Gonzalo Parra,
Bram Vandeputte, Erik Duval, Rafael Schirru, and Katja Niemann.
Bridging repositories to form the mace experience. *New Review of
Information Networking*, 14(2):102–116, November 2009.

[WMN+11] Martin Wolpers, Martin Memmel, Katja Niemann, Joris Klerkx,
Marcus Specht, Alberto Giretti, and Erik Duval. Aggregating
metadata to improve access to resources. In Reda Alhajj, James

Joshi, and Mei-Ling Shyu, editors, *Proceedings of the 12th IEEE International Conference on Information Reuse and Integration (IRI 2011)*, number IEEE Catalog Number: CFP11IRI-ART, pages 187–192, Las Vegas, Nevada, USA, 2011. IEEE. ISBN 978-1-4577-0966-1.

[WMS⁺09]   Martin Wolpers, Martin Memmel, Hans-Christian Schmitz, Martin Friedrich, Marco Jahn, and Rafael Schirru. Usage metadata based support for learning activity reflection. In Klaus Tochtermann and Hermann Maurer, editors, *Proceedings of I-KNOW '09*, pages 354–359. Know-Center, Graz, Journal of Universal Computer Science, September 2009. ISBN 978-3-85125060-2.

[WMS10]   Martin Wolpers, Martin Memmel, and Moritz Stefaner. Supporting architecture education using the mace system. *International Journal of Technology Enhanced Learning*, 2(1/2):132–144, 2010. ISSN (Online) 1753-5263, ISSN (Print) 1753-5255.

[WNVD07]   Martin Wolpers, Jehad Najjar, Katrien Verbert, and Erik Duval. Tracking actual usage: the attention metadata approach. *Educational Technology & Society*, 10(3):106–121, 2007.

[Wol08]   Martin Wolpers. Contextualized attention metadata in learning environments. *Upgrade - The European Journal for the Informatics Professional*, IX(3):57–61, June 2008. ISSN 1684-5285.

[WRG00]   David A. Wiley, Mimi Recker, and Andrew S. Gibbons. 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`.

[YJNT07]   Yusuke Yanbe, Adam Jatowt, Satoshi Nakamura, and Katsumi Tanaka. Can social bookmarking enhance search in the web? In *JCDL '07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 107–116, New York, NY, USA, June 2007. ACM. ISBN 978-1-59593-644-8.

[ZD01]   Ingrid Zukerman and Albrecht W. Daven. Predictive statistical models for user modeling. In *User Modeling and User Adaptive Interaction*, pages 5–18. Kluwer Academic Publishers, 2001.

[ZJ06]   Jin Zhang and Iris Jastram. A study of the metadata creation be-
           havior of different user groups on the internet. *Information Process-
           ing & Management*, 42(4):1099–1122, July 2006.

[ZME12]   Peter Zeile, Martin Memmel, and Jan Exner. A new urban sensing
           and monitoring approach: Tagging the city with the radar sens-
           ing app. In Manfred Schrenk, Vasily V. Popovich, Peter Zeile, and
           Pietro Elisei, editors, *Proceedings of REAL CORP 2012*, pages 17–25,
           Vienna, May 2012.