Personalized Access to Distributed Learning Repositories – PADLR – 2nd year continuation proposal (based on the 2001/2002 proposal)

Coordination: Wolfgang Nejdl

June 5, 2002

Abstract

The driving vision for this project is a distributed "Learning Web Infrastructure", which makes it possible to exchange / author / annotate / organize / market and personalize / navigate / use / reuse modular learning objects, supporting a variety of courses, disciplines and universities. Each of the PADLR subprojects deals with a specific problem on the way towards this vision. Infrastructure, tools, course materials and archives will be designed / developed in accordance with international standards for modularization and metadata, and will be compatible across the PADLR project. We will specify how course modules are built and described (both from a technical and from an educational point of view), how they are organized and how they are exchanged and reused. Several testbeds at our universities will be the test and application area of our infrastructure, tools and courselets, providing a rich source of requirements, feedback and evaluation results for steering our project into the right direction.

This proposal has been based on the 2001/2002 two-year proposal. It therefore describes work for two years (2001/2002 and 2002/2003), includes a progress report for each submodule describing its achievements for the first part of the 2001/2002 period, and - for some aspects / submodules - revisions and extensions for the second year.

Principal Investigators for the 2nd year submodules.

- Donald Broady, Department of Teacher Education, Uppsala University
- Stefan Fischer, Institute for Operating Systems and Computer Networks, TU Braunschweig
- Christian Floto, Göttingen IWF gGmbH Knowledge and Media, and Institute for Social Sciences, TU Braunschweig
- Carl Gustaf Jansson, Knowledge and Communication Laboratory, KTH Stockholm
- Klaus Jobmann, Institute for Communications Technology, University of Hannover
- Christopher Manning, Infolab, Computer Science Department, Stanford University
- Ambjörn Naeve, Center for User-Oriented Technology Design, KTH Stockholm
- Wolfgang Nejdl, Institute for Information Systems, University of Hannover
- Helmut Pralle, Computing Center Lower Saxony, University of Hannover
- Bernd Rebe, Institute for Social Sciences, TU Braunschweig
- Ulrich Reimers, Institute for Communications Technology, TU Braunschweig
- Tore Risch, Database Laboratory, Uppsala University
- Rudi Studer, Institute AIFB, University of Karlsruhe

1 Project Overview

The driving vision for this project is a distributed "Learning Web Infrastructure", which makes it possible to exchange / author / annotate / organize / market and personalize / navigate / use / reuse modular learning objects, supporting a variety of courses, disciplines and universities. The challenges for reaching this vision are manifold, and the PADLR Project Proposal therefore consists of quite a few modules and subprojects, each dealing with a specific problem on the way towards our goal, involving researchers from various areas of computer science, and from other disciplines. Together we are working on the necessary infrastructure, middleware, tools, courselets and distributed archives (to be designed/developed in accordance with international standards for modularization and metadata), and make them compatible across the whole PADLR project. We specify how courselets are built (both from a technical and from an educational point of view), how they are organized and how they are exchanged and reused, and how distributed content archives can be queried and navigated.

We use several testbeds at our universities for providing a rich source of requirements, feedback and evaluation results, and as test and application areas of our infrastructure, tools and course materials, helping us to steer our project into the right direction.

The PADLR subprojects are grouped into three different modules, including descriptions of testbeds at the participating sites. The proposal consists of three intertwined modules, with several submodules each:

- The module "Infrastructure and Intelligent Services" includes work on exchange facilities and basic infrastructure, personalized queries and views over distributed learning materials, automatic extraction of metadata and ontological information, as well as courseware discovery and annotation.
- The module "Server and Client Side Tools" includes work on modular content archives and video/audio capturing and metadata annotation tools.
- The module "Shared and Personalized Access to Educational Media" includes work on personalized learning sequences, interfaces, and guidance, personalized access to large text archives and personalized and shared mathematics courselets, as well as WebService-oriented courseware, marketing efforts, and copyright issues.

1.1 Proposal Structure

The proposal is structured as a framework + modules document. In the overview/framework pages we have given an overview of the project and its overall research agenda, and briefly listed the individual PADLR modules. Module descriptions written by the subgroups responsible describe the research agenda in each submodule in more detail. Each module description includes explicit information about how it is integrated into the total framework and describe the interaction with other modules and groups (focusing especially on cross-disciplinary and cross-institutional collaboration). The descriptions further specify deliverables, timelines, testbeds and budget information.

The core of the proposal corresponds to the original two-year proposal prepared for the 2001/2002 WGLN Call for Proposals. It additionally includes the two submodules Courseware Watchdog and CORI-PADLR, which were included as part of PADLR in the L3S grant application to the BMBF, and integrated as official PADLR submodules at the September 2001 WGLN Meeting. It further includes the L3SCPD submodule (started as a separate local project last year), with the goal of integrating it further with the other PADLR activities, and it includes an extension to the Edutella submodule (Large-Scale Edutella and Interfaces), which was started in January 2002 as local L3S addition to PADLR, adding one additional research position for this infrastructure submodule. Lastly, it includes the new submodule WebServices, to be started in the current year.

All submodules (except WebServices and Extract, which are to be started in the 2002/2003 funding period) have been running since fall/winter 2001, and include therefore a (separate) progress report and descriptions and lists of deliverables / publications (item 3 of the 2002/2003 WGLN Call for Proposals). We have also included short biographical sketches of the heads of the respective research groups (item 6 of the CfP) in the appendix of this proposal.

Compatible with the original 2001/2002 proposal, budget has been listed for two years in most cases and (in some cases) for a third year to give an impression for possible further extensions, though, for the 2002/2003 WGLN proposal process, only the second year budget (for continuing submodules) and the first year budget (for new submodules) is relevant. Consequently, only the 2002/2003 budget (grouped by research groups/PIs) is listed in the PADLR budget spreadsheet.

1.2 Introduction to the 2001/2002 Progress Reports

Projects have started in fall/winter 2001, so submodules have been running for 5 to 8 months until now. As it is obviously impossible to do justice to each submodule by including one joint progress report for the large number of submodules integrated within the PADLR framework, we have included a progress report for each submodule for the first funding period, which describes the work done in this submodule, including reports and publications (demonstrating the scientific quality of the work done so far), research visits (not including the numerous visits of participating institutes in the same country) as well as related work.

As documented by these progress reports, research collaboration within PADLR has been widespread and fruitful, resulting in many jointly authored publications, with more being worked upon right now, using short and longer term research visits and other forms of collaboration. Results of the PADLR project have already been included as a basis for several research consortia, including the open source project Edutella (which already includes about 45 participants from research institutions in Europe and the US), EU funded projects on semantic web technology and e-learning (5th framework), national consortia in Germany and Sweden, as well as several ongoing initiatives within these contexts as well as the forthcoming EU 6th framework program. Furthermore, PADLR participants have been active in various standardization committees for metadata (IMS, DIN, etc.), networks of excellence (OntoWeb etc.), working groups and program committees for conferences and workshops in the areas relevant for the PADLR project.

In addition, the PADLR project has been holding regular half-yearly workshops since its inception. The first (startup) workshop was held in Hannover (September 10 and 11, 2001), the second one in Uppsala (March 11 and 12, 2002), and the third one is scheduled for September 26 and 27, 2002, again in Hannover. Furthermore, we use the PADLR BSCW server both as internal common workspace (for workshop notes, discussion items, etc.) and as public presentation space[1].

2 Module: Infrastructure and Intelligent Services.

2.1 Exchange Facilities / Basic Infrastructure

2.1.1 Submodule Description

Working Title. Edutella: An Infrastructure for the Exchange of Educational Media

Contributing Research Groups and PIs. Stanford Infolab (Manning/Decker), Hannover KBS (Nejdl), Stockholm CID/KMR (Naeve)

Problem Description. Every single university usually has already a large pool of educational resources distributed over the institutions. These are under control of the single entities, and it is unlikely that these entities will give up this control. Thus central-server approaches for the distribution of educational media are unlikely to happen. To facilitate exchange of educational media, we propose to develop an exchange network for educational media, based on an approach based on peer-to-peer (P2P) networks. These P2P networks have already been quite successful for exchanging data in heterogeneous environments, and have been brought into focus with services like Napster and Gnutella, providing access to distributed resources like MP3 coded audio data.

However, pure Napster and Gnutella like approaches are not suitable for the exchange of educational media. For example, the metadata in Gnutella is limited to a file name and a path. While this might work for files with titles like “Madonna - like a Virgin”, it certainly does not work for “Introduction to Algebra - Lecture 23”. The educational domain, we thus see, is in need of a much richer metadata markup of resources, a markup that is often highly domain and resource type specific. In order to facilitate interoperability and reusability of educational resources, we need to build a system supporting a wide range of such resources. This places high demands on the interchange protocols and metadata schemata used in such a system, as well as on the overall technical structure.

The “Open Archives Initiative”, which has recently defined a HTTP-based protocol for retrieving metadata information about digital documents from library servers, is an interesting initiative, whose further development might also

[1] http://www.learninglab.de/workspace/padlr/index.html

be of interest to the issues discussed in this proposal. Currently, though, it is server-retrieval protocol only, and just only Dublin Core as its basic schema.

Research Plan and Deliverables. To solve these problems we propose to use a metadata based peer to peer system for educational resources, including Edutella nodes providing different client/server functionalities based on common and interoperable metadata and P2P conventions. Constructing such a system will pose two concrete problems: 1) define the metadata infrastructure and functionalities for Edutella and 2) define the peer-to-peer infrastructure and functionalities. For both problems we have to define the basic (abstract) model and to design a sound implementation.

For defining and implementing the metadata infrastructure and functionalities (1), we have the tackle the following tasks:

1. Metadata and metadata schemata for learning resources. We need to specify a basic model (RDF, RDF(S), [Las98]) to express metadata, to identify / describe metadata schemata (standard schemata like IEEE-LOM (P1484.12) to describe general aspects of learner resources, MPEG-7 for specific aspects of audio-visual content, and area/subject specific metadata like the ACM classification scheme for computer science topics, etc.) and to describe how additional kinds of metadata schemata can be added (e.g. learner models (IEEE P1484.2), which characterize learners and their knowledge/abilities to enable personalized instruction and allow creating and building personal learner models utilizized throughout their education and work life). Using RDF and RDF(S) will be strategically important in order to design our system to be compliant with the next generation (= semantic) web [DvHB+00, BKD+01, NWC00]. Interoperability with existing web standards and metadata schemata are crucial requirements to lower the entry level into the system. This is the fundamental part of defining the metadata information handled by the Edutella system (KBS, CID, IfN).

2. Defining the metadata handling capabilities (API, query language, updates) and identify required functionalities for the access to the metadata. Edutella will bring not only multiple and distibuted content, but multiple and distributed meta-data as well (where several different schemata for the same content will be possible and several authors provide metadata for the same content). Metadata therefore have to be identifiable and reproducible to ensure that different meta-data sets are consistently separable, metadata will be seen as a resource on its own (see also the next point on schema interoperabilitiy). This is the fundamental part of accessing and working with Edutella metadata irrespective of particular storage representations. See also [Nae99, NP99] (Infolab, CID, KBS).

3. Define and set up metadata schema bureaus, i.e. servers that store metadata schemata and classifications used in the Edutella network (see also [DEFS99]. Enabling access to such explicit descriptions of schemata is crucial in order to allow sharing of metadata schemata within user communities [HH00]. RDF schemata already support this mechanism, RDF metada explicitly reference machine readable descriptions of the schemata they are instantiated from. These schemata will then be available as resources within the Edutella network. (Infolab)

4. Implement ontology/schema mapping capabilities to be used within the Edutella network in order to allow exchange of metadata based on different, but related metadata schemata. (Infolab)

For the peer-to-peer infrastructure and functionalities (2), the following tasks are important:

1. Define a set of well-designed interfaces to the Edutella system and protocols between Edutella nodes, that allow access to and interchange between Edutella resources. Again we can build upon emerging Web standards (like the SOAP messaging protocol which uses HTTP to carry messages that are formatted with XML and thus provides a lightweight standard object invocation protocol built on Internet standards [BEK+00], or the new JXTA open source initiative proposed by SUN). This makes it possible to implement Edutella facilities e.g. as plugin modules to SOAP enabled servers (such as Apache). Even more lightweight basic servers can be imagined, for example using simple WebDAV over HTTP, or a read-only servers serving metadata as static XML files. Integrating such lightweight server capabilities into arbitrary Edutella nodes leads to a P2P network where every node can offer services to other nodes (possibly with different sophistification). This approach satisfies the very important demand that setting up a basic Edutella server should be a very simple procedure, and that different layers of added functionality can be added step by step later. This is the fundamental part of accessing Edutella services over the Internet/Web (see also [DB01]. (CID, KBS)

2. Implement an open-source base library and set of plugin modules with support for the basic metadata model for Edutella nodes and the peer-to-peer networking details between Edutella nodes. These libraries/modules will provide basic access and update functionalities for Edutella nodes, and will also be used in other submodules (like "modular content archives" and "lightweight tools") (KBS)

3. Implement a set of extended functionalities on top of this base library in order to support different types of extensions (e.g. an ontology inference engine or filtering [Pet00]). Functionalities provided by other submodules like "personalization" can also be implemented as such extension plugin modules plus appropriate metadata/metadata schemata. Important further aspects include loadbalancing and efficient querying in the Edutella network, influencing both protocol design and dynamic reconfiguration of network (KBS, Infolab).

Dissemination, Testbeds and Evaluation Dissemination of results will be done through reports and scientific publications on the different aspects outlined in the research plan. A set of prototype implementations at the participating sites as described above will be available after the first year, which will be refined and extended during the second year based on a evaluation and feedback from these implementations. We will use several specific courses as well as existing intra- and inter-university project cooperations as resources for our requirements analysis and as testbeds for our implementations.

In Germany for example, our testbed context will be courses for the ULI project, a distributed computer science program, where 8 universities and 17 CS professors will work together for 3 years to create a distributed computer science study program as well as course content for these courses (financed with about 3 Mill. Euro for the next three years). KBS is one of these partners (responsible for the area of artificial intelligence), and will use this project as a testbed for the infrastructure and tools developed, both in the requirements, refinement and evaluation phase. In Sweden, our testbeds will consist of the modular content archives for humanities and of personalized and shared mathematical courselets.

Collaboration and Scholarly Exchange. Strong interaction with all modules building tools und functionalities, in order to define common standards, strong interaction with all modules working with testbeds in order to define requirements, and to use evaluations to drive development. Use research visits (2 weeks up to 3 months) (Hannover, Stanford, Stockholm) in order to integrate design and development within this module and with other modules .

Budget Overview (including overhead costs).

KBS: 70K first year, 40K second year. Budget will pay for one / part of one Ph.D. Student, L3S infrastructure costs, travel and exchange.

IfN: 20K first year. Budget will pay for part of a Ph.D. student, L3S infrastructure costs, travel and exchange.

Infolab: 70K first year. Budget will pay for one Postdoc, overhead costs, travel and exchange.

CID/KMR: 30K first year, 30K second year. Budget will pay for part time Ph.D. Student, overhead costs, travel and exchange.

2.1.2 Progress Report (KBS/Nejdl, AIFB/Studer, Infolab CS/Manning/Decker, CID/KMR/Naeve, DBL/Risch).

Participants. Due to the importance of this common infrastructure module, we have allocated even more resources as planned originally to this submodule, including additional participation by the AIFB Institute (Karlsruhe) and the DBL Lab (Uppsala). This intensive collaboration has been very fruitful, and work has progressed fastly in the areas of infrastructure specification, design and implementation of the Edutella infrastructure as an open source project, as an intense collaborative effort between the two groups in Germany (KBS Hannover, AIFB Karlsruhe), the group in Stanford (Infolab CS), and two groups in Sweden (CID/KMR Stockholm, DBL Uppsala).

During the first months we have defined the Edutella P2P infrastructure in more detail, including a family of query exchange languages (QEL) which allows heterogeneous and distributed systems to be asked for their metadata [NWQ+02]. This work has been complemented by implementations of adapters which translate QEL to query languages of other systems. Adapters to SQL, RDQL, XPath and O-Telos are available. A first step towards definition

of a modification language has been done [NSST02]. This will enable annotation and replication of metadata using a standardized interface.

As underlying peer-to-peer framework the open source project JXTA has been chosen, and a working infrastructure implementation has been built on this foundation. We have set up Edutella as as part of the JXTA project (see the Edutella Project Home Page[2]), which includes all sources and libraries to be downloaded as open source. As of end of May, the Edutella project encompasses 45 project members, contributing to or testing the Edutella infrastructure (see the Edutella Project Member Page[3]).

Several application prototypes and metadata providers have been implemented to show the validity of our approach. Integration with other submodules has started, especially with Modular Content Archives: Metadata describing the content of the Open Learning Repository, developed at KBS, Hannover, can already be accessed via an Edutella provider. Conzilla, developed at CID/KMR, Stockholm, can act as Edutella peer. Edutella has also been integrated with the KAON framework, developed at AIFB, to create an annotation service for Edutella [NWST02]. Also, Edutella will be used as basic infrastructure in an EU 5th framework project which will be starting this summer (ELENA), [KSV+02], and has already been adopted by several German research institutes (including two Fraunhofer institutes) as underlying P2P infrastructure. Work has also started to combine the Edutella approach with the Open Archive Initiative standards [ANS02]. For an overview over Edutella and related projects/submodules, see [Nej02].

To complete the set of available RDF query languages in EDUTELLA a new query language for querying RDF was developed [SD02a, SD02b]. The language, TRIPLE, is based on Horn logic and designed for querying and transforming RDF models. TRIPLE does not have a fixed built-in semantics for object-oriented features like classes and inheritance, but the modular architecture allows such features to be easily defined for different object-oriented and other data models like UML, Topic Maps, or RDF Schema. Recent ontology language approaches like DAML+OIL [DAM01] that cannot be fully handled by Horn logic are provided as modules that interact with a description logic classifier, e.g. FaCT [Hor01], resulting in a hybrid rule language. Since EDUTELLA connects heterogeneous peers with metadata descriptions of their content based on various standards, TRIPLE enables the effective integration and joint querying of these peers.

Peer-to-peer networks like EDUTELLA are envisioned to be deployed in a wide range of applications. However, unorganized P2P networks exhibit characteristics that severely hamper their scalability to a large number of nodes, among them long search times, network traffic overload and traffic hotspots and pushing search times to unacceptable limits. To address these problems within EDUTELLA, we developed a novel approach to selforganizing P2P networks, HyperCuP [SSDN02b]: We propose a Hypercube-based network topology which enable efficient broadcast and search, and we describe a broadcast algorithm that exploits the topology to reach all nodes in the network with the minimum amount of messages possible. We provide an efficient topology construction and maintenance algorithm which, crucial to symmetric peer-to-peer networks, does not require any central server nor super nodes in the network. Nodes can join and leave the self-organizing network at any time, and the network is resilient against failure.

To further optimize the EDUTELLA networks we developed a novel organization approach for P2P networks based on ontologies [SSDN02a, SSDN02c]: Usually, each peer in a P2P network can provide content that is associated with particular topics. These topics can be arranged as concepts in a global ontology, hierarchically classified by is-a links. It is desirable to restrict the broadcast of a query message to peers that can potentially provide information related to the concepts asked for in the query. We extended the Hypercube routing algorithms with ontology-based routing methods which can be exploited for directed search and broadcast on HyperCuP-based P2P networks.

Coordinated by the CID/KMR-group (Naeve et al) and together with partners from L3S, from the University of Economy, Vienna (UNIVERSAL project) and other, we have produced an RDF binding for the IMS Metadata information model. This work has been done mainly within IMS. Moreover, a parallel track has recently been launched within IEEE to make it an official standard which will build upon the work done for our IMS binding. During the 2nd year the KMR-group will hopefully be able to finalize this work and start to work on an RDF binding for the IMS Content Packaging information model to enable expression and search for course structures (and more).

Also, we have been involved in separating and defining the different levels of query language and their expression in RDF. The KMR-group has contributed especially in two ways in this process. First the represention in RDF is done very carefully, e.g. it doesn't break the basic assertional semantics of RDF (mainly via using a reified syntax which solves several problems, e.g. allowing queries to be addressed and searched for just like any other metadata in the

[2] http://edutella.jxta.org

[3] http://edutella.jxta.org/servlets/ProjectMemberList

system). Secondly, a special user-friendly flavor has been designed and made compatible with QEL level 2 (it may be extended to higher levels later) and a user-interface supporting this expressiveness is being developed in Conzilla [NNP01, Nae01]. This graphical query language has been documented and presented e.g. at WWW2002 (see also [NWQ+02]). This work will continue during the second year.

The KMR-group is further designing and leading the development of the SCAM (Standardized Content Archive Management) system - a data/metadata archiving managment system which in version two (under development) will act as an Edutella provider. Since SCAM is a small but extendable server-side system based on semantic web techniques, it fits well into the picture of a distributed network of locally-controlled providers. SCAM is actually (and will continue to be) a separately funded project that utilizes the techniques of Edutella and semantic web in general, and some of the design is based on compatibility with the Edutella network. SCAM is already used in several of the testbeds to avoid reinvention of the wheel. Also the Swedish national educational television and radio (UR) has choosen SCAM as a base for their future system for providing their resources on the web, which will enrich the Edutella network with high quality resources. During the second year, KMR will further work on implementations of Edutella peers (clients and providers) and their interfaces. A more detailed exposition of the connections between SCAM, Edutella and Conzilla is available at http://kmr.nada.kth.se/papers/SemanticWeb/Edutella-2nd-year-CID.pdf[4].

Research visits

- M. Palmer (Stockholm) has spent a week at L3S to work on RDF bindings for educational metadata standards and integration of Edutella and Conzilla (September 2001)
- T. Snackerstrom (Uppsala) has spent a week at L3S to work on integration of AMOS as Edutella hub (March 2002).
- W. Siberski (L3S) has been at Stanford to work on an enhanced peer-to-peer implementation and plan integration of TRIPLE and Edutella (April 2002).
- W. Nejdl (L3S) is spending a sabbatical at Stanford, CS Department, Infolab, working on various aspects of the PADLR project (March to August 2002).

2.2 Exchange Facilities / Basic Infrastructure II

2.2.1 Submodule Description

Working Title. Large-Scale Edutella and Edutella Interfaces

Contributing Research Groups and PIs. University of Hanover, L3S (Nejdl)

Problem Description. With simple distribution approaches (e.g. broadcast of queries to all peers) a peer-to-peer network like Edutella is unable to scale for more than hundreds of peers. Therefore more sophisticated techniques have to be introduced. But scalability not only has a technical side; there is also the community aspect. By building communities of peers with similar contexts a large network becomes feasible. How to structure these communities best for educational purposes has to be investigated.

Besides these architectural issues, interfacing of existing systems with Edutella is important. We will therefore extend our previously planned work by implementing additional wrappers for various backend systems (learning repositories, RDF query languages and XML-based systems). We also need to define and implement additional mappings between XML and RDF representations for educational metadata to let XML based systems become peers easily.

[4] http://kmr.nada.kth.se/papers/SemanticWeb/Edutella-2nd-year-CID.pdf

Research Plan and Deliverables. Our research will cover the following aspects:

Replication: One way to enhance performance of a large network is replication of data. Therefore we plan to introduce means to replicate metadata between repositories. This part will cover specification of a suitable replication (or general modification) protocol for RDF repositories (along the lines of the proposed query protocol). We also want to investigate which kind of replication strategy is optimal in the Edutella context (source peer vs. target peer controlled, point-to-point vs. multicast, etc.).

Optimized Request Handling: To reduce overhead we can exploit techniques to find the nearest peer which provides the metadata to answer ones request. There are already approaches proposed for peer-to-peer networks which allow redirecting requests to suitable peers (e.g. by maintaining local indices, see [CGM02]). However, these approaches are targeted to P2P networks for simple file exchange, which have other properties and constraints than an educational peer to peer network. Therefore it is necessary to build an infrastructure supporting performance experiments, which will enable us to evaluate and compare different approaches in an environment similar to the "real-world" network.

Community Management: Another way to let the network scale is by logically partitioning it. Requests by a peer will be handled at his (logical) context first and distributed further only if the request can't be met (regardless of technical settings like subnets, etc.). The primary purpose of such a partition is to allow peers to join groups most suitable for their interests. Of course, this will typically also reduce the need for distribution of messages and therefore decrease performance requirements (bandwidth, etc).

Typical learner communities will be participants of courses, study programs, or students at the same university. For instructors partitioning by research topics or advertised interests might be appropriate.

Advancing Educational Metadata Standards: To allow interoperability between Edutella and SCORM-conformant systems we will specify a RDFS binding for SCORM 1.2 and implement libraries and tools for conversion between XML and RDF representations of educational metadata. Our learning system OLR will incorporate the instructional roles approach [ADN02a]. We will expand the existing standards with schema information to capture metadata about these roles.

Mapping RDF metadata and Java Objects for educational systems: Mapping RDF metadata and Java Objects for educational systems: In metadata driven educational systems we often find two different models for educational material. For storage and exchange purposes the material properties are captured as statements about these materials, according to a metadata schema. For modification and presentation purposes they are held as objects, according to a class model. Currently there is a gap between these models. There are a lot of similar concepts (e.g. rdf:type vs. java interface/class, rdf:property vs. java attributes), but no seamless translation between them. The necessary mapping between metadata and object model has to be developed manually for each system. Our goal is to automate this mapping as much as possible.

Wrappers for Medatada Stores: Wrappers for metadata stores: To integrate existing metadata stores we plan to implement an extended set of wrappers: As a native XML database, dbXML can be used to store meta-data application profiles (XML/RDF) in their original XML format. At present a dbXML-based learning resource repository constructed using Dublin Core and its RDF binding has already been integrated into the Edutella network. Other repositories based on some higher level learning resource specifications, e.g., LOM and SCORM, will be available very soon. All these repositories will share a minimal interoperable level (Dublin Core) in the Edutella network. RDQL is a RDF query language implemented as part of the Jena framework. The implementation allows querying RDF files directly. The wrapper for RDQL we will implement is targeted to enable lightweight provider peers, using simple files to store metadata.

Furthermore by implementing wrappers for systems that provide services in addition to data and metadata storage we provide these services to the Edutella network. As an example for this kind of services we will develop a wrapper for the ConceptBase system, which is a deductive object manager for meta databases, as well as another wrapper to Prolog. All wrappers accept queries formulated in RDF-QEL, answers are stated using simple RDF.

Deliverables.

- Specification of replication/modification protocol for Edutella
- Reference implementation for Edutella replication module
- Test infrastructure for Edutella performance experiments
- Community management module for Edutella
- RDF-Binding for SCORM, extended with instructional role information
- Wrapper modules for RDQL, dbXML, Conceptbase and Prolog as Edutella modules
- Mapping framework between RDF data and java objects as Edutella module

Dissemination, Testbeds and Evaluation. Dissemination of results will be done through reports and scientific publications on the aspects outlined in the research plan. One of our most important means for dissemination is the open source project Edutella. All specifications and implemented modules will be available as part of this project. For the replication specification and implementation we will be able to use the UNIVERSAL platform [BEG+01] as productive testbed.

Collaboration and Scholarly Exchange.

- Wirtschaftsuniversität Wien: Replication
- P2P Group Stanford: Scalability (technical aspects)
- AIFB, Karlsruhe: Edutella interfaces

Budget Overview (including overhead costs).

KBS: 70K second year. Budget will pay for one Ph.D. Assistant, L3S infrastructure costs, travel and exchange.

2.2.2 Progress Report (KBS, Nejdl).

Some deliverables have already been achieved in the first months of this year (especially in the areas of replication/update, metadata standards, Edutella wrappers, and P2P topologies), which are described in Section 2.1.2 and Section 3.1.2. Furthermore, within the collaboration started within this submodule, we have successfully proposed the EU/IST 5th framework project ELENA (Creating a Smart Space for Learning), under the topic "ambient intelligence", where the L3S is focussing on P2P- and personalization issues in a smart learning space [KSV+02].

2.3 Personalized Queries and Views over Distributed Learning Materials

2.3.1 Submodule Description

The module will provide an infrastructure for advanced personalized query facilities on top of the Edutella meta-data and peer-to-peer infrastructure. This would complement the deliverables in the module 'Infrastructure and Intelligent Services'.

Working Title. PSELO: A personalized search engine for learning objects.

Contributing Research Groups and PIs. Uppsala University (Risch)

Problem Description. In a distributed learning environment there will be large numbers of learning objects and courselets stored in many distributed and differing data stores on the Internet. Without guidance, students will have great difficulties to find the learning objects relevant for a particular learning task. The meta-data descriptions provide information about properties of the learning objects, but the meta-data by itself does not provide for handling reconciliation of differences between different objects nor does it provide for customized and efficient queries over views of reconciled learning objects.

Research Plan and Deliverables. The purpose of this module is to provide technology for defining personalized learning views of relevant learning objects / courselets for each subject, student, or task. These personalized views focus the data primarily seen by a user to a particular set of relevant objects. The student can then explore the relevant learning objects through a powerful and *subject-oriented* query language. Here subject-orientation means that the operators used in queries are specialized for a particular learning subject. Such a query language provides a tool for the student for advanced exploration of a subject. It should be possible to dynamically change and adapt the personalized view for each student, task etc., as new knowledge is increasingly deeper explored.

The personalized views are defined from the meta-data model. In our approach these views are defined for the user as a set of object-oriented (OO) data views inferred from meta-data. Subject-oriented queries are then specified in terms of the personalized data views using an extensible OO query language.

For defining and implementing such a query infrastructure the following tasks need to be solved:

1. The subject-oriented queries should be expressed in terms of the meta-data model of learning objects. The system must therefore have the ability to interpret meta-data definitions of Eduella. The learning object often have complex structures that need to be viewed as object structures. The query language must thus have the ability to express queries and views over complex object structures.

2. To support subject-oriented queries it must be possible to easily extend the query engine with knowledge modules implementing the subject-oriented query operators. This includes plug-in facilities for implementing both query operators as well as optimization rules for subject-oriented queries.

3. Since the learning objects are stored in many different places and in many different formats the query engine must be able to deal with queries that span many heterogeneous and distributed data sources.

   A major challenge is here to efficiently process subject-oriented queries that access and process many distributed and heterogeneous knowledge objects. The query execution should thereby utilize the peer-to-peer infrastructure of Eduella.

The project enhances data integration technology for support of learning applications and is expected to enhance the state-of-the-art in the areas of database and data integration techniques. It will produce reports and scientific publications.

The work will extend on research developed at Uppsala Database Laboratory (http://www.dis.uu.se/~udbl) on OO queries over distributed and heterogeneous data [VT99a, VT99b, RJ01]. Research from other institutes on wrapping and searching heterogeneous, distributed, and semistructured data, e.g. [GMPQ$^+$97, LC97, TRV98], is also applicable. Other important supplemental components are user interface modules for subject-oriented visual query specifications and for visualizing retrieved learning objects, which can be added as plugins.

Dissemination, Testbeds and Evaluation In the project we will implement a prototype search engine, PSELO, that provides personalized OO views of the distributed learning objects. The prototype system will demonstrate the feasibility of the approach and serve as a platform for further experimentation and evaluations.

It is envisioned that this technology will be an important component of the Edutella infrastructure. It would utilize the meta-data protocols of Edutella for providing the terminology in which the personalized views and queries are expressed. It will use the Edutella peer-to-peer infrastructure for efficient access to learning objects from the queries. With the proposed system it will furthermore be possible to incorporate algorithms for a particular subject as system plug-ins that define subject-oriented query operators.

Collaboration and Scholarly Exchange. The work on this module is closely related to other modules building tools and functionalities. Research visits to other participants are expected to be complemented with dayly e-mail contacts.

Budget Overview (including overhead costs).

Uppsala: 35K first year, 35K second year. Budget will pay for half a Ph.D. Student, infrastructure costs, travel and exchange.

2.3.2 Progress Report (DBL, Risch).

A general query engine was developed for queries to RDF and RDFSchema meta-data descriptions. It is fully implemented and can be downloaded from http://www.csd.uu.se/~udbl/amos/rdf[5]. The query engine can import general RDFSchema descriptions from the web, which are then converted to an object-relational semantic data model. General object-relational queries to the imported meta-data can then be issued both in terms of RDFSchema objects and RDF binary relations. This is the core of the PSELO engine. It is an extension of the Amos II mediator system.

The mediator system Amos II was interfaced with the JXTA peer-to-peer infrastructure. The interface allows Amos II peers to be called through JXTA. With the interface general object-relational mediator queries can be submitted through JXTA to an Amos II peer for evaluation. It can also be downloaded from http://www.csd.uu.se/~udbl/amos/rdf (soon).

The PSELO architecture, its RDFSchema meta-data query engine, and the JXTA interface to Amos II mediators were presented at the PADLR workshop in Uppsala, March 11, 2002. PSELO provides an infrastructure with which arbitrary web sources described through RDFSchema can be integrated and queried. The JXTA-interface allows Amos II to be called from other peers and provides the basis for using PSELO as an Edutella peer. In our contribution to the paper on Edutella presented at the WWW-2002 conference in Honolulu [NWQ+02] we showed how Amos II could be used as a peer in the Edutella infrastructure allowing integration of different data sources both from the web (through PSELO) or from other sources (e.g. relational databases).

In [LRK02] we show a general method to import XML data into an object-relational mediator system allowing to query and integrate general XML data using an object-relational data model. The object-relational schemas are inferred from the DTDs when available. However, for DTDs with unspecified properties, for XML documents without DTDs, or for XML documents with errorneous DTDs the inferred object-relational schema is adapted as new XML documents are added.

2.4 Automatic extraction of metadata and ontological information

2.4.1 Submodule Description

Contributor. Chris Manning (Infolab/CS, Stanford)

Working Title. Extract: Automatic extraction of metadata and ontological information.

Problem description. The basic infrastructure portion of the proposal depends on the availability of accurate meta-data for learning resources, built to fit within a loose ontology. While a metadata framework provides a lightweight solution to many of the problems to be addressed in this project, it leaves the question of where the metadata is to come from. For core content, hand-annotation is possible. However, even for core content, this will be difficult: busy educators will not appreciate having to carry out an additional task beyond content creation. A reason why many CBL projects fail is because the materials change more slowly than educational needs. *Any tool that can reduce this burden by semi-automatically providing metadata will be useful*. The problem is even more acute for the enormous world of information (e.g., on the Web) outside the core resources. There is already a vast array of useful teaching resources on the web, and students could often get value from making use of advanced or complementary materials at other institutions. However, connected to the personalization issues of adaptation comes the difficult and very time-consuming tasks of finding appropriate materials, and determining their prerequisites, etc. In this context, standard keyword search is of very limited effectiveness, because it cannot filter for: (i) the *type* of information (tutorial, applet or demo, review questions, etc.), (ii) the *level* of the information (aimed at secondary school students, the general public, or graduate students?), (iii) the *prerequisites* for understanding the information, or (iv) the *quality* of the information. Moreover, there are all the usual problems of keyword-based information retrieval, such as problems with synonymy, polysemy,

[5] http://www.csd.uu.se/~udbl/amos/rdf

and so on. *Any method which automatically annotates information from other sources so that it can be easily accessed within our content repository will be of enormous value*. Such additional resources, while of less consistent overall quality, will magnify the value of our content repository.

Research plan and deliverables. The starting point is the use of statistical information extraction and natural language parsing techniques to automatically derive classifcatory and metadata information from primarily textual data (web pages, Word, postscript or similar documents, etc.). While still challenging for large ontologies, text classification methods which semantically categorize an entire document are now relatively well-understood, and provide a good level of performance. A central research challenge is how to extend these methods to address issues like trying to determine the prerequisites for understanding a topic, or the quality of the information presented. We believe that two of the most important sources of information here will be linkage information (in the case of hypertext sources like the Web), and natural language processing terminology extraction (to identify terms that are assumed without definition within the text). Determining the level of information is also little explored, beyond crude measures of reading level. Exploring NLP and machine learning approaches to such problems will be the main focus of the research.

The research will lead to a prototype system which will classify educational resources from outside the repository with the metadata attributes defined in the central Edutella infrastructure.

Dissemination and Evaluation. The results of the research will be disseminated through scientific publications. The prototype will be integrated with the Edutella infrastructure in the second year, and evaluated for its success in finding and classifying suitable information. In the first instance, this success will be measured using information retrieval evaluation methods. A more task-specific evaluation would involve demonstrating the additional value over existing resources provided by returned results, as seen in user studies. This evaluation can be an aspect of the more general evaluation of the testbeds.

Collaborations. On the one hand, this module fits in strongly with aspects of the Edutella infrastructure, in particular the issue of ontology mapping (InfoLab), which requires similar techniques. On the other hand, resource identification and accurate classification, in particular identifying approaches and prerequisites are important enabling technologies for effective personalization, and this project will work closely with KBS and Uppsala groups on personalized content delivery and computational linguistics methods. Research visits will assist in integrating the design and development between these two groups.

Budget Overview (including overhead costs).

InfoLab: 70K second year. Budget will pay for one PhD student, overhead costs, and travel.

2.4.2 Progress Report (Infolab, Manning).

This subproject will start in the second year.

2.5 Module: Courseware Watchdog

2.5.1 Submodule Descriptions.

Topics in education are changing with an ever faster pace. Especially in the field of life-long learning the aspects that need to be taught by information providers must keep up to date with the process in its field. The courseware watchdog is a comprehensive module which allows users to focus on existing subfields of a discipline, but thereby be aware of important drifts and tendencies in the field. The courseware watchdog consists of four major components:

1. A focused crawler that gathers data from relevant educational media sources
2. A subjective clustering algorithm that allows to group educational media with similar contents together following different types of ontology similarity criteria

3. An intelligent browsing capability, which allows to understand similarities and differences of educational media sources

4. A mechanism for updating the ontology in order to reflect a drift of courseware topics

These four components are elaborated in the following submodules. The significant difference to modules like the automatic extraction of metadata and the personal search engine is the attempt at using unsupervised learning methodology and intelligent browsing facilities in order to detect trends and tendencies rather than in classifying or retrieving according to existing structures. Thus, the courseware watchdog nicely complements these other work packages.

2.5.2 Focused Crawler

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer,Stumme)

Working Title. Focused Crawler.

Problem Description. In order to detect related material and topics it is necessary to observe related sites and check them for congruency with one's own interests (cf. [HMSS02] on focused crawling for a Human Ressource Management problem). For this purpose, it is necessary to have a sophisticated crawler that takes advantage of one's focus such as expressed in the ontology and in one's documents.

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Crawler: Re-use existing crawlers to collect metadata and relevant documents from "seed" sites. Because existing crawlers typically "only" deal with gathering but rarely with selection of links according to content, the existing crawler has to be adapted in a twofold way: 1) Adapt crawler to focus on ontology-congruent documents. 2) Adapt crawler to focus on documents similar to the ones a user has given as preferable.

Dissemination, Testbeds and Evaluation. Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia. The prototype will be integrated with the Edutella infrasturcture at the end of the first year.

Collaboration and Scholarly Exchange. Interactions with other modules:

1. Edutella module (Exchange Facilities / Basic Infrastructure)

2. Automatic extraction of metadata and ontological information

3. Personal Search Engine

Use research visits (2 weeks up to 3 months) (Braunschweig, Hannover, Stanford, Stockholm, Uppsala) in order to integrate design and development with other modules. In particular, we expect fruitful interaction with the workpackage on metadata extraction, because based on additional metadata accuracy of crawling may be increased.

Budget Overview (including overhead costs).

AIFB: 70K first year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

2.5.3 Subjective Clustering

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer, Stumme)

Working Title. Clustering with multiple, ontology-based views.

Problem Description. In order to detect trends and tendencies it is necessary to find outliers or groups of outliers. Because users typically do not want to exactly specify their complete profile and because users' profiles tend to change rather often (which is a major problem for recommender systems), we want to give the user views onto existing educational media. For this purpose, there have been developed clustering algorithms. Typically, however, clustering was used to give a single, "optimal" view on (learning) components. This is not suitable to account for the plurality of views that exist when looking at educational media. E.g. the same or similar slides on the internet might be used for introducing social effects of the internet or its technical foundations. We have recently developed preliminary clustering mechanisms that allow to provide *subjective views* onto documents [HMS01a], which are based on an underlying ontology. For instance, one view may concentrate on differences and similarities along the ontology parts that deal with social effects and another view may concentrate on ontology parts that focus on technology. The objective of this workpackage lies in the elaboration of these techniques and their realization within a framework that makes them applicable within Edutella.

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Adapt ontology for clustering: The ontology determines the view onto the resources. If there are too few concepts and relations in the ontology there may be the necessity to remodel parts of the ontology or — preferably — to provide methods to derive a slightly changed ontology from the given one such that it is suitable for clustering.

Implement Clustering: The clustering mechanism must be implemented and integrated in the overall framework. In particular, it must build on the crawled information and the users' own data.

Dissemination, Testbeds and Evaluation Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia.

Collaboration and Scholarly Exchange Interactions with other modules:

1. Edutella module (Exchange Facilities / Basic Infrastructure)
2. Focused Crawler

Budget Overview (including overhead costs).

AIFB: 35K first year, 35K second year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

2.5.4 Browsing of Watchdog Data.

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer, Stumme)

Working Title. Browsing of Watchdog Data

Problem Description. Results from crawling and clustering need to be visualized. We plan to blend techniques from formal concept analysis and open conceptual hypermedia [GBC+01] in order to visualize the similarities and differences between documents and clustering results through lattices [SW00].

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Analysis of Underlying Data. This is necessary in order to understand underlying properties of data and metadata, such as "which data lies on a scale", etc. This analysis serves as input to the FCA-based browsing.

FCA-based Browsing. FCA-based browsing allows the user to explore similarities and differences between educational media.

Dissemination, Testbeds and Evaluation Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia. At the end of this workpackage it will be integrated into the overall environment.

Collaboration and Scholarly Exchange Interactions with other modules:

1. Edutella module (Exchange Facilities / Basic Infrastructure)
2. Clustering with multiple, ontology-based views

Use research visits (2 weeks up to 3 months) (Braunschweig, Hannover, Stanford, Stockholm, Uppsala) in order to integrate design and development within this module and with other modules. Workshops will be suitable to promote collaboration as well.

Budget Overview (including overhead costs).

AIFB: 35K second year, 35K third year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

2.5.5 Ontology Evolution

Contributing Research Groups and PIs. Univ. Karlsruhe, Institute AIFB (Studer)

Working Title. Ontology evolution.

Problem Description. The watchdog depends on having a reasonable ontology of current topics in order to distinguish upcoming topics and in order to cluster and present educational material. AIFB will extend its experiences from ontology learning and detection of new conceptual relationships [MS01, MS00] to ontology evolution. Of particular interest in the latter case, are the tendencies of drift in usage of terms, while "conventional" ontology learning rather considers text corpora to be static. Thus, we may extend the given ontology and adapt the overall results of the courseware watchdog.

Research Plan and Deliverables. To solve the described problems the following tasks have to be accomplished:

Adapt existing learning environment: AIFB already has a set of tools which may be used for this step. However, it is necessary to augment and adapt them in order to consider the different nature of ontology evolution (in constrast to ontology extraction), taking into account temporal and spatial heterogenety of educational media.

Dissemination, Testbeds and Evaluation. Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia.

Collaboration and Scholarly Exchange Interactions with other modules:

1. Automatic extraction of metadata and ontological information

Use research visits (2 weeks up to 3 months) (Stanford) in order to integrate design and development with other modules. AIFB has been organizing two workshops on ontology learning and is planning to continue this "tradition".

Budget Overview (including overhead costs).

AIFB: 70K second year, 35K third year. Budget will pay for one Ph.D. assistant, L3S infrastructure costs, travel and exchange.

2.5.6 Progress Report (AIFB, Studer).

Focused Crawler A focused crawler [MEH+02] has been developed that computes ontology-based metrics on web pages and hyperlinks in order to restrict the crawling process to pages relevant to the user's needs. The crawler is integrated with the Karlsruhe Ontology Framework KAON [HMSV], so that it can benefit from the available infrastructure.

Subjective Clustering We have devised algorithms that employ ontology-based techniques to improve text clustering [HMS01b]. These algorithms reduce the large number of dimensions by generating views onto the ontology and thereby making use of background knowledge. A user-defined set of concepts can be used for these views, so that the granularity of the clustering process can be steered.

Browsing of Watchdog Data Formal Concept Analysis (FCA) is a means to provide structured views on complex data sets. The FCA-based browsing component of the Watchdog is under development, based on our experiences gained with the development of the Conceptual Email Manager (CEM) ([CS00],[CES02]) that we have created in cooperation with Richard Cole (Griffith University, Australia). The visualization will be based on the FCA-based open-source tool ToscanaJ[6] which we are currently integrating with the KAON framework.

Ontology Evolution As ontology evolution relies on the results of the components described above, we plan to tackle this problem in the second year. Preliminary work on ontology evolution has already been done [SS02].

Edutella Integration *[Not mentioned in original proposal]*

As the ontology-based tools of the Courseware Watchdog operate on ontologies expressed in RDF, they can take advantage of the Edutella peer-to-peer network in a natural way. We have provided capabilities for exchanging data [NSST02, NWST02] between KAON applications such as the Courseware Watchdog via the Edutella network.

3 Module: Server and Client Side Tools

3.1 Modular Content Archives

3.1.1 Module Description.

Contributing Research Groups and PIs. IfN Braunschweig (Ulrich Reimers), KBS Hannover (Wolfgang Nejdl), Uppsala (Broady), CID/KMR (Naeve)

Working Title. Archives: Intuitive Storage, Use and Retrieval of Archived Educational Media.

Problem Description. At every university usually plenty of material is available or being produced for the planning, preparation and execution of curricular activities like lectures, seminars and project work. This material is an inhomogeneous amount of content of various type. These could be lecture slides, presentations, scientific graphics and texts, description of experiments, software, animations, simulations.

Existing distributed learning environments often demand content to be molded into proprietary and application dependant formats. Therefore it is difficult to re-use content for different purposes and different audiences. Hence, teachers today are prone to produce archives and curricula based on redundant information, proprietary applications and formats, and non-modularized solutions.

An intelligent and flexible archiving, management, allocation and distribution of this modular content is an intense problem. Several commercial products like Hyperwave or Lotus offer certain functionality providing a step into the right direction but there is still the lack of a sophisticated archive system to optimally support curricular activities. As mentioned earlier in this proposal a central server approach is due to the distributed nature not the optimum type of architecture, and - building on the Edutella infrastructure - one of the distinguishing feature to other projects (e.g.

[6] http://souceforge.net/projects/toscanaj

the German "Teachware-on-Demand" project[7]) is exactly this distributed peer-to-peer infrastructure, with plugins for different kinds of peers participating within the Edutella network.

Intelligent archiving and flexible and intuitive access to modularized content come along with several problems: Content modules are often detached from context so that interrelated content modules can hardly be found and are therefore of no use. This for instance will prevent an instructor to answer a students question with appropriate instructive material if this material is not intended for that particular course. Content modules can be of different file types. Locally distributed creation leads to the need for version lists. Offline archiving of created content modules is difficult and tedious. Instructors have no fast access to content modules due to the lack of appropriate search strategies and missing context information. Students have no access to many interesting content modules. Especially colleagues from Sweden and Germany will produce same content in different languages that should be linked.

Another problem is that in many educational settings teachers and students are not able to profit from the international development of agreed markup schemes that are evolving within research communities (mainly SGML-based or XML-based, such as the TEI encoding guidelines among scholars in the humanities). Since the tools and practices used in courses often do not keep pace with such international de facto standards, the well-structured content of existing and emerging digital research archives are in many cases not easily available to teachers and students.

Yet another problem concerns the rapidly increasing volume of scientific results/deliverables available at the WWW. In some rapidly evolving fields, such as bioinformatics, there are hardly any course books. This means that curricular development will be dependent on teachers' ability to overview, navigate, identify and re-use appropriate selections of existing digital archives. There is a need for better principles for the design of modularized content repositories.

There are also a series of more narrow technical problems. Since existing materials are often stored in various file formats and developer versions along with non-valid filenames, it is difficult and sometimes impossible to re-use information.

Research Plan and Deliverables To solve the described problems the following tasks have to be accomplished:

Metadata and Windows Applications. A strong focus lies on the integration of (Windows) de-facto standard file formats like Word, PowerPoint, JPG, GIF, HTML, Mathcad, Windows executables etc. into Edutella archives. We will investigate, which kind of constraints this leads to, how Edutella metadata are used especially in this environment.

To access this content very fast efficient and flexible text/keyword search facilities are necessary. One situation that occurs during a lecture is the student having a question that deals with topic outside the primary scope of the lecture. Using these sophisticated search facilities instructors can access appropriate content very fast and answer the students question instructively. Another important aspect is context information for content modules. If every single module can have appended context information, it will be possible to provide fast access to connected modules via visual presentations (context maps).

The described tasks will lead to intensive cooperation with the "Edutella" and the "Personalized Interfaces" modules. (IfN)

Repository Architecture and Tools. We propose a local database system with a client server architecture to support instructors in an optimum way. This single database functions as a university wide repository that can exchange metadata and content with other distributed repositories as described in the "Edutella" module. Commercial products like Hyperwave will be examined if these can be extended to suit our needs. XML/SOAP et al can function as a standardized protocol for the exchange of metadata and data between repositories or a repository and a user. Similar developments at other institutions will also be taken into account (for previous work see e.g. [NW99]).

An important issue is the linking of personalized electronic student portfolios with such a repository. Students will be able to download interesting material to their own electronic portfolio and manage their content. These issues will be discussed in cooperation with members of the METAFOLIO group. (IfN)

Another task is to develop and explore methods and tools for the design and creation of modularized content archives, as well as develop and explore methods and tools that allow teachers and students to access and use

[7] Teachware-on-Demand has recently adopted Edutella as P2P infrastructure.

already existing repositories on the web, with a strong focus on modularity and re-usability in different context, based on the content archive systems described in the previous paragraphs. They should support the teachers and students to work with these archives. For a specific course the teacher might propose the students certain paths through the archive and certain subsets of content modules to be used by the students and in some cases added to their portfolios. (IfN, KBS, CID, Uppsala)

The knowledge management tool Conzilla already supports the organization of annotated content into personalized portfolios in a way that is compliant with the emerging standards for automated information exchange (RDF) and metadata handling (IMS). We will modify Conzilla to enhance this support in various ways, based on the structure of this repository, and on user-feedback from our activities in other modules. (CID)

Metadata-Based Learning Repositories. Based on the Edutella submodule we will implement a second server implementation called an open learning repository (OLR), which just stores (RDF) metadata (both for classification/annotation and for structure) in a central database, but no content (which is accessible via URLs/HTTP).

Resources potentially distributed all over the Internet can be combined to entities we call "Courses", logical documents integrating content from different sources. While the content stays at its original location our database holds all information about the structure of each course and metadata about the content. The responsibility for maintaining content stays with the people providing it. When browsing a course all HTML-pages are generated dynamicall.

The system provides a platform for testing out different navigation schemes including traditional concepts such as hierarchical tree structures and tutorial-like course-trails as well as highly adaptive approaches (personalized navigation) – see the personalized access submodules. (KBS)

A second system will explore an alternative architecture based on XML, XML Schema and JSP, again compatible with the Edutella functionalities. It will include a WebDAV-based courseware authoring module which enables geographically dispersed courseware authors to collaboratively work on the course contents, and a standalone, JSP-based courseware publishing engine which can achieve flexible, dynamic courseware presentation. The system can be aware of any changes of the course contents and automatically reflect the changes on the Web at any time, again customizable by metadata. (KBS, CID)

A third system will provide further reasoning capabilities for the Edutella network. Using the ConceptBase metadata object manager it will provide Datalog$\neg$ reasoning functionality for courseware metadata (e.g. RDF metadata) possibly stored elsewhere. This enhanced reasoning facility will enable the generation of additional information for courseware authors, and will simplify metadata handling within the network, e.g. by supporting the integration of various metadata schemas. (KBS)

Dissemination, Testbeds and Evaluation Dissemination of the achieved results will be accomplished by scientific publications in appropriate journals and by demonstrations at symposia.

In several testbeds we will show the impact of the developments in this module on teaching and learning. In Braunschweig, we will use several courses in the field of communications, encouraging exchange to other interested faculties from other universities, in Hannover we will use several computer science courses and the ULI project. At Uppsala University several courses in humanties and social sciences and bioinformatics will provide suitable testbeds.

Collaboration and Scholarly Exchange Interactions with other modules:

1. Edutella module (Exchange Facilities / Basic Infrastructure)
2. Automatic extraction of metadata and ontological information
3. Personal Search Engine

There are also close links between student portfolios and content archives. The students should be offered (and themselves able to create) content in portable modularized formats, suitable to be incorporated into their portfolios and to be re-used for various purposes that may not be foreseen by the teacher. Therefore the development of portfolio practices presupposes the creation of archives of content modules.

Use research visits (2 weeks up to 3 months) (Braunschweig, Hannover, Stanford, Stockholm, Uppsala) in order to integrate design and development within this module and with other modules. Workshops will be suitable to promote collaboration as well.

Budget Overview (including overhead costs).

IfN: 50K first year, 70K second year. Budget will pay for one Ph.D. assistant (part-time first year), L3S infrastructure costs, travel and exchange.

KBS: 20K first year, 40K second year. Budget will pay for a part-time Ph.D. student, L3S infrastructure costs, travel and exchange.

Uppsala: 15K first year, 15K second year. Budget will pay for a part-time Ph.D. student, infrastructure costs, travel and exchange.

CID/KMR: 15K first year, 15K second year. Budget will pay for a part-time Ph.D. student, infrastructure costs, travel and exchange.

3.1.2 Progress Report MoCa-OLR/CB (KBS, Nejdl).

The focus of the project Modular Content Archives (MoCA for short) in its first year was to design and implement tools and methods for intelligent and flexible archiving, management, allocation and distribution of modular content and to explore methods for the design and creation of modularized content archives with a strong focus on the reusability of the existing learning materials in different contexts [DNWW01].

For this purpose we implemented a server-side course portal, called Open Learning Repository (OLR for short), which structure and connect modularised course materials over the Web. The modular content can be distributed anywhere on the internet, and is integrated by explicit metadata information in order to build courses and connected sets of learning materials [KBN02]. The metadata used to structure, annotate, and classify the learning objects/materials are expressed in the semantic web standard RDF (Resource Description Framework) and stored in a relational database. In order to annotate and classify the learning objects we used existing metadata standards, e.g. the IEEE LOM draft standard for learning object metadata. In particular we have discussed the educational category of this standard and tried to define new pedagogical dimensions to the metadata elements in this category and suggested an extension of this category with new elements [ADN02a].

In close collaboration with the PerINaG team we have designed and implemented different navigation interfaces. Different views on a given course are possible, realized by different queries and page designs. A user has a choice between a classic hierarchical tree-like navigation, a trail navigation, where he can move forward and backward on a trail, and semantic (context) net navigation. We have experimented with these different navigation schemes in order to build an idea, which navigation the learner prefers.

In the current phase of MoCA we focused on scenarios-based design for the interface of OLR. The results of the evaluation of the first prototypes of our Open Learning Repositories motivated us to address the general conditions that influence the use of an e-learning system. We are analysing different instructional models and interfaces to structure and visualize learning materials based on different scenarios from an interdisciplinary (pedagogical, psychological and technical (computer science)) point of view [ADK+02].

We are currently using the OLR system in the context of three courses, one in “Introduction to Artificial Intelligence”, one in “Artificial Intelligence II”, and one in “Software Engineering”. In these test beds we will show the impact of the methods, interfaces and tools in this module on teaching and learning. OLR is also a peer in Edutella, which aims at storing, querying, exchanging and processing any kind of RDF metadata.

In continuation of previous work, we also finished designing and implementing a SCORM-conformant Computer Science courseware based on the latest ADL-SCORM specification 1.2 released on Oct. 1, 2001. All learning resources contained in an existing Java course are re-designed according to the SCORM Content Model and further described with corresponding Assets-, SCO-, and Content Aggregation meta-data. Also the course structure is re-constructed taking advantage of SCORM Content Packaging, which provides a standardized way to represent the aggregation of learning resources and further exchange them between different learning management systems [QN02a, QN02b].

Within the Modular Content Archives subproject of PADLR the CB-peer subproject aims at providing a peer for the Edutella network that enables the storage of data and furthermore also reasoning capabilities about such data. Therefore we developed a prototype of a provider peer with such reasoning capabilities within the last year. The provider peer uses the object-oriented database ConceptBase which implements the language O-Telos, a dialect of Telos which amalgamates properties of deductive and object-oriented languages. In order to represent metadata described in RDF we proposed and implemented a mechanism that translates property-centered data representations (e.g. RDF) to and from object-centred data representations (e.g. O-Telos). Employing this translation enables us to use ConceptBase for the management of RDF data. Beyond that we can use ConceptBase to reason about the RDF data [NDW01b, NDW01a, WBN02].

Martin Wolpers will spend 5 weeks at Stanford (August 2002), to work on advanced reasoning and inference capabilities as well as schema/data integration issues as part of the Edutella infrastructure.

3.1.3 Progress Report MoCa - ReFlAcT (IfN, Reimers).

Research during the stated period started with the creation of a first technical report [Pai02] describing goals and plans of our development, the ReFlAcT Content Browser helping instructors with archiving and organising modular content modules. Many conceptual ideas have been outlined in this document. Several of these ideas are a result of a one-week exchange where I went to Uppsala and Stockholm (December 2001). Discussion with staff of Uppsala University and of CID, Stockholm, lead to new ideas that found their way into the technical report.

At CID/KMR a Metadata Editor (the ImseVimse Metadata Editor) has been developed in the past as an open source project which provides modules that can be re-used for the ReFlAcT Content Browser. At IfN this Metadata Editor has been updated to IMS 1.2 compliant. These alterations are now being used at CID/KMR for the development of an RDF backend for ImseVimse and will be used at IfN from now on.

For the testbed "Digital Communications"[8] several interesting and interactive modules have been created to be used during a lecture. Metadata for these modules has been appended, still there has to be done work for the preparation of this testbed. The ReFlAcT Content Browser will function desirably as an Edutella client. This feature has to be planned more thoroughly in the next few weeks.

3.1.4 Progress Report (CID/KMR, Naeve)

SCAM has been designed by the KMR group to be small and extendable allowing it to be used in different settings. The Modular Content Archives (MoCA) module has provided several scenarios resulting in four distinct requirements on SCAM. 1) SCAM should allow personal expression which has been solved by letting each user have their own pool of metadata (the only limit of which metadata that can be stored is the user interface). 2) SCAM should support collaboration around resources. This isn't solved in version 1, and during the next year we will work on this. 3) SCAM deals with interoperability and portability issues allowing an archive to be used, extracted and moved between systems. This has been tackled by using IMS Content-Packaging expressed in RDF. 4) Several ways to organize and present the archived resources should be possible at the same time (both structural and presentational allowing different purposes and work processes/methodologies to live side by side). In SCAM this is solved on a representational level by using IMS Content-Packaging. However, it remains to provide a graphical user interface that is flexible enough to encompass this (SCAM version 1 provides a generic GUI that has limited adaptation capabilities). These requirements have been identified during an iterative process with the Uppsala team. In this collaboration a user interface for SCAM version 1 has also been developed, tested and evaluated.

Conzilla has been further developed by the KMR group to allow for metadata expressed in RDF rather than in XML as originally suggested. Together with Braunschweig the metadata editor in Conzilla, ImseVimse, have been updated to compensate for the changes in the LOM standard. During year two this work will continue and also provide an early implementation of an RDF-binding developed as a standard in IEEE [PNN01, NPN02].

There is also an ongoing work intending to present a generic archive as a set of surfable maps possibly displaying contextual relations that are hard to see in non-graph like tools. SCAM will probably be one of the easiest archiving systems to present. The KMR group will aim for this during year two, however it is uncertain at this stage if there will be enough time.

[8] http://www.ifn.ing.tu-bs.de/dnue/

3.2 Video/Audio Capturing and Metadata

3.2.1 Submodule Description

Working Title. VACE: Video/Audio Capturing and Embedding

Contributing Research Groups and PIs. Hannover RVS (Pralle)

Problem Description. Currently we see the need to integrate captured audiovisual material from actual seminars, courses, lectures etc. into on-line (web-based) and off-line (e. g. CD-ROM-based) learning systems.

These A/V recordings can help students in understanding the content, especially when they are embedded in other kinds of learning material, as a simple example, the script of the course. In addition this, a larger group of learners can be reached and students can get access to previous material.

But nevertheless, it does not help to just do a recording of a course an publish this on the web, a CD-ROM or a video tape. The key for a useful solution is the integration in a multi-media learning system. The idea is not to produce a "TV-Show" but to use captured media as a supplement to other learning material.

There are some approaches that all have several advantages and disadvantages.

1. One way is to put all information (the A/V-recording of the lecture and the slides) in a TV-quality audio/video stream, like it is done in a project called "Uni-TV" (http://www.lrz-muenchen.de/wir/gigabit-vpp/demo-unitv/).

   There are certain drawbacks in this "single-media" approach: The resolution of the slides is restricted to the TV standard (about 720 576 pixels). A high bandwidth is needed to deliver the media stream (at least four Megabits per second for a MPEG-2 stream). Apart from a VCR-like control (pause, play, fast forward, rewind) the user cannot navigate the content. Finally, there is no association to other media types possible.

2. There are certain presentation tools commercially available (Real Presenter, Microsoft Media Tools) that allow to record an actual presentation and generate an on-line representation of this. Two media streams are produced, one containing the slides, the other one containing the A/V recording of the instructor.

   As the produced media streams are vendor-dependent, they cannot be easily integrated in own learning applications. The re-use of the generated material is difficult or even not possible because it the produced data depends on the format used by the tool. The quality of the streams is often poor because the tools are mainly developed for low-bandwidth transport.

3. One general approach is to record the presentation on video, convert this to a format suitable for online delivery and link it to web pages containing the slides of the presentation.

   Especially if different media types are used much effort has to be spent on the post-production. The slides (e.g. in MS Powerpoint format) have to be converted to browseable web pages and the recorded audio and video data has to be encoded to different streaming formats.

4. There is some development going on in building a system for the "authoring-on-the-fly" of lectures. Several media streams are captured (audio, video, whiteboard) an can be published right after the presentation.

   This LINUX-based systems needs special software on both, the instructors and the students computers. Common applications, file- and streaming-formats cannot be used with this system. Although a platform-independence is not urgently necessary, the applications needed to obtain the media should at least run on common PCs in order to achieve a wide acceptance.

As we see, although it seems to be a common problem there is no ideal solution yet.

Research Plan and Deliverables. To come closer to a solution the features of ideal capturing systems has be seen from different kinds of view.

1. The authors (i. e. the instructors or professors) view. In the terms of "light-weight-authoring" the system has to be easy to use for the producer. In an ideal case, the capturing tool should run in the background so that the professor can focus on the actual presentation. Common presentation tools like MS PowerPoint, PDF viewers or web browsers should be supported and associations to the captured A/V content should be generated automatically. The post-production should be reduced to a minimum. The content should be reusable for different kinds of media (e. g. on-line/web, off-line/CD).

2. The users (students) view. The viewer should have the possibility to navigate and browse the content. Different ways of using the content have to be offered to the learner to satisfy the individual preferences: Watching the course like a video ("timeline-based") or reading, browsing and searching it like a book. The slides and the A/V streams should be synchronized to each other so that the viewer can choose between watching the AV stream or browsing through the web pages without losing access to the associated other data type. The integration of other systems like discussion-forums or chat could be helpful.

In order to construct a flexible, open system it would be useful to think of several modules with defined interfaces. There has to be modules for capturing meta-data, interfaces to A/V-systems and publishing. The different modules make use of different existing applications for presenting, A/V-codecs and streaming-systems but itself will be vendor-independent.

Example scenario: The author uses PowerPoint to present his slides in the actual presentation. The lecture is recorded by a PC with capture card. A browseable web presentation and a CD-ROM with additional learning material should be created.

In this case we need a module that captures the events in PowerPoint (changing slides) and extracts the content of the slides for web pages. Another module has to generate metadata for A/V streaming (e. g. RealMedia) by associating the time-stamps and keywords from the slides. A third module generates an HTML framework for the web presentation that contains the content from the slides and the links to the A/V streaming media. Finally a module that extracts the information for the CD-ROM production is needed.

The concept of modules also allows the integration of forthcoming techniques for media delivery. Metadata will be recorded in a way compatible to the Edutella submodule.

The generated content should be accessible by the use of common applications like web-browsers an streaming players like Real or Windows-Media so that no special programs have to be installed on the students computer.

The content can be use for different purposes, e. g. for archives, portfolios and can be integrated in existing or forthcoming learning systems. The system can be seen as a framework. Extensions as chat, forums or other communication systems can be integrated when using on-line publishing.

One could think of on-demand delivery in web-based on-line systems as well as the use in a real-time learning scenarios or for stored media (on CD, DVD).

Collaboration / Scholarly Exchange. Collaboration especially with the Edutella module, based on direct collaboration and exchange visits to other participating labs. Use of this tool in the different testbeds (e.g. ULI).

Budget Overview (including overhead costs).

RVS: 70K for first and second year, which pays for one full Ph.D. student, including L3S overhead costs, travel and exchange.

3.2.2 Progress Report (RVS, Pralle).

The VACE submodule consists of two working packages: the development of a distributed audio/video capturing system and the design of the technical infrastructure for capturing live presentations that is evidently needed as a basic platform for the system.

Infrastructure. For capturing live presentations additional functionalities are needed. As a prototype the "L3S Learning Space" - a multifunctional lecture room - at L3S has been set up. It contains equipment for several different kind of applications with a focus on audio/visual recording and presentation facilities. E. g. video-conferencing,

video-streaming and -recording. The main aspects of the concept were to build an easy-to-use system and to achieve medium (near-TV-like) A/V-quality.

The installation is completely working but can be considered as a moving target - several improvements have been made due to user feedback. Side effects of this projects include the set-up of a multi-point video conferencing infrastructure that is used by the learning-labs, the evaluation and introduction of high quality video over IP transmission systems and integration of video editing systems.

A description and examples for utilization can be found at: http://www.mml.uni-hannover.de/vace/mminfra/[9].

Capturing System. For the VACE capturing framework the architecture has been defined as a distributed system. The VACE server as a core acts as a central information gathering system. Several client systems ("capturer") deliver information about events to the server in real time but also can work independently - the gathered data can be synchronized later. The architecture offers possibilities for the flexible integration of existing and forthcoming presentation and audio/video systems.

4 Module: Shared and Personalized Access to Educational Media

4.1 Personalized Learning Sequences

4.1.1 Submodule Description

Working Title. PLeaSe: Personalized Learning Sequences

Contributing Research Groups and PIs. Hannover KBS (Nejdl/Henze)

Problem Description. It is generally agreed that it is a desirable goal in any educational setting to be able to tailor courses and course materials to individual students' needs, as determined by such factors as their previous knowledge of the subject, their learning style, their general attitude, as well as their cultural and linguistic background. A skilled teacher will be able to achieve this goal in one-to-one interactions with learners, but not as a rule in the lecture hall/classroom settings more typical of higher education. Nor will a developer of traditional course materials be able to cater for individual learner needs to any great extent.

ICT-based educational materials can potentially be much more flexible than traditional course materials, however. They offer a unique opportunity to achieve the mentioned goal through *personalization*, where both the selection and presentation sequence of the units of educational material making up a courselet or course are determined by a *dynamically updated learner model*, which takes into account at least the learner's previous knowledge and her progress in mastering the material.

The aim of the personalization module is to provide individualized access to learning materials, to courses, courselets, etc. A "one-size-fits-all" learning path through these materials (or even through parts of it) would neglect individual needs, knowledge or preferences of the users [HNW99]. To maximize the students' success with the open learning repository it is necessary to provide a quick, user-focussed access to those entities in the learning repository which correspond to the user's actual information needs, to her/his knowledge, current situation and preferences.

Research Plan. Research in adapting information systems to the individual users is conducted e.g. in the area of adaptive hypermedia systems. Adaptive hypermedia systems use a model of the user to collect information about her/his knowledge, goals, experience, etc., in order to adapt the content and the navigational structure. There are two main adaptation strategies in adapting hypermedia systems [Bru96]: link-level adaptation calculates useful navigational strategies for the individual user, content-level adaptation individually places content-chunks on the information entities. We assume that our repositories will provide lots of information on any one topic: different explanations, examples, use-cases, etc., therefore in PLeaSe module we will concentrate on selecting and presenting the most appropriate material for each user into personalized learning sequences [Bru99]. In module PALaTe we will explore the

[9] http://www.mml.uni-hannover.de/vace/mminfra/

issue of creating new hypertext pages out of existing text, particularly in connection with the large text archives testbed [SS00].

The overall problem setting for our adaptation functionalities is the following: We will have access to a large set of learning materials. Many different authors can modify, delete or add new content to the OLR, the learning objects might be distributed, and we can expect that we will often find more than one learning object on the same topic.

Adaptive hypermedia systems normally combine learning materials with reading sequences, didactical rules, pre- and post-knowledge or pre- and post-learning objects [HN99]. As we want to adapt materials from different courses and from different contexts, we need a flexible approach to adaptation [Hen00, HN00], where information about learning objects is read out from metadata, but also inferred from the objects themselves. Using information retrieval, information extraction and natural language processing methods for obtaining automatic or semi-automatic indexing is of advantage (see the submodule "Automatic extraction of metadata and ontological information"), even crucial in the case of large textual resources as used e.g. in the Languages/Humanities (see the section on the "PALaTe submodule/testbed").

Deliverables.

1. *User orientation:* The adaptation component will focus on orientation guides: Provide access to relevant information, select and show case studies, give hints, show examples, show context, generate reading sequences.
   - How to deal with different teaching strategies in one single course? Different authors of courses/courselets will follow different teaching strategies.
   - How to deal with information oversupply? Present "best", find best fitting tour.
   - How to generate personalized learning sequences based on different learning theories (see module PerI-NaG)?
   - How to deal with structured materials? In our repositories we will find lots of structured materials. How can we make use of these structures (for presentation, visualization, selection of materials?).
2. *User Model / Knowledge Model:* The user model stores individual data like name, overall preferences, abilities, etc. The knowledge model expresses the domain knowledge of some application domain. In defining these models and implementing tools for using them, we will build upon work done by IEEE LTSC Working and Study Groups, especially P1484.2 (Learner Model WG), P1484.4 (Task Model WG), P1484.6 (Course Sequencing WG), and P1484.20 (Competency Definitions SG).
   - Employ Ontologies for knowledge modeling in open learning environment
     - Ontologies provide a well-founded plattform for knowledge modeling
     - Different Ontologies for different course materials can be mapped to one common ontology by ontology mapping
     - Ontologies allow the exchange of information about knowledge throughout different course materials
   - Learning Metadata for adaptation
     - Use Machine Learning techniques for text categorization
     - How much information can we obtain automatically or semi-automatically, is there information that can not be learnt?

Dissemination and Evaluation Dissemination of results will be done through reports and scientific publications on the different aspects outlined in the research plan. A set of prototype implementations at the participating sites as described above will be available after the first year, which will be refined and extended during the second year based on a evaluation and feedback from these implementations. We will use several specific courses as well as existing intra- and inter-university project cooperations as resources for our requirements analysis and as testbeds for our implementations.

Collaboration and Scholarly Exchange. Interaction with the Module "Infrastructure and Intelligent Services": incorporating requirements from the adaptation component: content and usage (including context of use) of each learning object; indexing (semi-automatic), information retrieval methods in the meta-data definition. In addition, users should be enabled to store and manipulate retrieved learning materials. These learning materials might be structured (hierarchical, sequential, concept maps, etc.).

Interaction with the Module "Interfaces and Navigation": Outcome of the adaptation component (reading sequences, access to examples, alternative views, explanations) - how to integrate it in a smart user interface? Interaction with "Personalized Access to Large Text Archive"s: How can textual resources be effectively used for personalized learning? Use research visits (2 weeks up to 3 months) (Hannover, Stanford, Uppsala) in order to integrate design and development within this module and with other modules.

Budget Overview (including overhead costs).

KBS: 20K first year, 30K second year. Budget will pay for one part-time Ph.D. student, L3S infrastructure costs, travel and exchange.

L3S central: 20K first year, 20 K second year, for a part-time Ph.D. student in media and design education, L3S infrastructure costs, travel and exchange.

4.1.2 Progress Report (KBS, Nejdl).

Research in the first project phase has concentrated on adaptation in open corpus hypermedia environment. The PADLR project aims to develop an open infrastructure for the exchange, distribution and access to learning objects. To personalize the access to the individual needs of a user, methods and techniques from adaptive hypermedia are very usefull and settle on well-founded research. However, so far developed techniques can not be applied one-to-one to the PADLR project. Adaptive Hypermedia has so far considered the space of documents under consideration to be closed. Adaptation information is therefor context - dependent, where the context is defined by the overall (closed) set of documents. To apply these techniques to open learning environments, we need a generalized approach to open adaptive hypermedia (OAH).

In the current state of our research we have analyzed possible solution strategies for OAH. Based on this analysis we have defined a set of metadata entities which are at minimum required for open adaptive hypermedia. We are currently investigating whether these metadata can be learned (semi-)automatically by employing text classification algorithms. This is done in cooperation with the subproject "Automatic extraction of metadata and ontological information" (Chris Manning , Infolab/CS, Stanford). Further we have developed a knowledge modeling approach for OAH, see [HN01, Hen01, HN02].

4.2 Personalized Access to Large Text Archives (Submodule/Testbed)

4.2.1 Submodule Description

Working Title. PALaTe: Personalized Access to Large Text Archives

Contributing Research Groups and PIs. Uppsala/Linguistics (Borin), Uppsala/Teacher Education – KTH/NADA/CID (Broady)

Problem Description. Text is still important in the teaching of almost any subject, viz. in the form of textbooks and other course texts. In Languages and Humanities education, (large) textual resources are also quite often objects of study in themselves. Arguably, their effective deployment as study objects in the context of ICT-based personalized learning demands some kind of language understanding. Hence, personalized access and navigation among such resources should – almost by definition – make use of Computational Linguistics (CL) / Natural Language Processing (NLP) techniques, to complement the more general personalization tools which will be developed in the submodule "PLeaSe: Personalized Learning Sequences".

In this submodule/testbed, we thus consider the issue of *personalized access to large text archives* in Languages and Humanities education. In order to make the fruits of our labor in the proposed project useable also in other subject areas, we will focus on certain aspects of this issue, namely how (aspects of the) content and difficulty of texts or parts of texts can be inferred and utilized for creating personalized access to text material.

Research plan and deliverables, 1st year (shortened). We will consider the use of two fairly different kinds of large text archives:

1. In language education and linguistics, large text archives are important mainly (but not only!) because of their (linguistic) *form*. Here, the so-called *text corpus* has become an important educational (and research) resource. The uses of text corpora in language education are manifold:

   - as a data source for the preparation of (monolingual or bilingual) word lists, grammars [Axe00, Bor00, Bor a], test items (e.g. for diagnostic tests such as the Didax system being developed in the Swedish Learning Lab (SweLL) APE-DRHum project [BkSBa , BkSBZa]), etc.
   - as a source of empirical examples in 'data-driven learning' [BD99]. The English Department at Uppsala uses the British National Corpus in this way, and other language departments are getting ready to do the same, e.g. the Slavic Department for use in their Russian courses.
   - as a source of reading matter, user-adapted as to its level of difficulty and subject area (where content obviously becomes important, too)

2. On the other hand, in such Humanities subjects as History, Literature Studies, History of Science, Teacher Training, etc., large text archives are important mainly because of their *content*, i.e. because of the information contained in the texts (and, as a rule, the range of languages dealt with will be much smaller; see below).

The work with large text archives will proceed along two interconnected lines of research:

1. We will explore the issue of using partial parsing and information extraction techniques for marking text portions for persons, places, and times, and carry out formative evaluation of these techniques in an educational setting. This work will be pursued in collaboration with the work in the submodules "Automatic extraction of metadata and ontological information" and "PLeaSe: Personalized learning sequences".

2. We will pursue the issue of how to (operationally) define and determine the level of difficulty (or "level of information"; see above) of a text or a portion of a text (for language education purposes it would be useful to be able to determine this even for small linguistic units such as phrases or clauses), and carry out formative evaluation of this definition in an educational setting. This work, too, will be a collaboration with the work in the submodules "Automatic extraction of metadata and ontological information" and "PLeaSe: Personalized learning sequences".

Research plan and deliverables (2nd year, revised). The proposed second year PALaTe submodule of PADLR represents the merger of the first-year Swedish PLeaSe and PALaTe submodules. For year two, the work with large text archives will continue along the following interconnected lines of research:

1. The prototype text segmenter developed in the project will be integrated in the Edutella infrastructure and the SCAM portfolio application, for experimental deployment in History education.

2. We will explore the issue of using partial parsing and information extraction techniques for marking text portions for persons, places, and times, and carry out formative evaluation of these techniques in an educational setting.

3. We will further pursue the issue of how to refine the definition of 'level of difficulty' beyond the preliminary definition (readability plus subject matter) used in [NB02], of a text or a portion of a text for second language learners, and carry out formative evaluation of this definition in an educational setting. The aim here is to utilize text difficulty level classification in creating successively more difficult reading sequences as well as corpus-based exercise sequences.

4. Markup tools for teachers and students will be evaluated and implemented in educational testbed settings.

5. Tools and methods for browsing, retrieval, sharing and management of existing, primarely TEI-encoded, digital text resources will be explored and employed in testbeds, i.a. in teacher education..

Dissemination, Testbeds and Evaluation Dissemination of results will be done through reports and scientific publications on the different aspects outlined in the research plan. In general, we plan to do research/development and evaluation in parallel (i.e., formative evaluation). We will use existing courses in the departments of the Faculty of Languages, in the History Department and in the Department of Teacher Education as resources for our requirements analysis and as testbeds for our implementations.

Collaboration and Scholarly Exchange. Interactions with the submodules "MoCA: Modular Content Archives", "Courseware Watchdog", and "Automatic extraction of metadata and ontological information".

Budget Overview (including overhead costs).

Uppsala/Linguistics: 25K first year, 45K second year. Budget will pay for one part-time Postdoc, for systems development, and for faculty involvement in testbed integration in regular Languages/Humanities curricula, overhead costs, travel and exchange.

Uppsala/Teacher Education: 30K first year, 10K second year, Budget will pay for a part-time PhD student and faculty involvement in testbed integration in regular Teacher Education/Social Sciences curricula, overhead costs, travel and exchange.

CID: 10K first year, 10K second year. Budget will pay for a part-time PhD student and faculty involvement in testbed integration in regular Teacher Education/Social Sciences curricula, overhead costs, travel and exchange.

4.2.2 Progress Report (Uppsala, Borin, Broady, Langert-Zettermann).

In this report, as in our day-to-day work in the projects, we treat the subprojects MoCA (Modular Content Archives), PLeaSe (Personalized Learning Sequences), PALaTe (Personalized Access to Large Text Archives) as a unit. They were closely related already at their conception, and we have decided to make them even more so, in the name of efficiency and effectiveness, particularly in view of the fact that the original project budget was to be reduced by approximately one half for the last version of the project proposals. For the same reasons, the focus of the subprojects had to be narrowed somewhat, so that out of the testbeds originally planned, we have concentrated on two, History and Teacher Education, and left most of the work on Language Education for the second project year (although we have managed some work on language learning resources even now, see below). Note, that up until now the projects has been a little behind schedule because of initial difficulties to recruit collaborators with the adequate qualifications.

Accomplished and ongoing work.

- Development of the APE portfolio on the SCAM platform which is built on international standards for metadata and learning technology such as RDF, Dublin Core, IMS and IEEE/LOM. The development of this system for distributed archiving and shared use of educational components is a main effort - coordinated by MoCA - by several of the PADLR sub-projects (and also for some Swedish testbeds within the ongoing WGLN project "Personal Learning Portfolios - Folio Thinking"). The development is co-funded by Skolverket (the Swedish Board of Education), Utbildningsradion, and Nationellt Centrum för Flexibelt Lärande (NCFL). The aim is to enable students and teachers to create and share personal collections of material to be used within, e.g., the Edutella framework. The APE/SCAM application is currently being employed in testbeds at Uppsala University and KTH. SCAM is an open source project; documentation available at the SCAM Project Page at CID/KMR[10] and the code at the SCAM Project Homepage[11].

[10] http://kmr.nada.kth.se/scam/index.html
[11] http://sourceforge.net/projects/scam/

- Extensive search for existing software in the domains language recognition, text segmentation, and named entity recognition; a report is in preparation. A simple web interface for trying out and evaluating existing such software has been developed by Lilian Karlsson.
- Kristina Nilsson and Lars Borin have worked on the problem of locating reading matter for language learners of Nordic languages. The language, the subject matter, and the difficulty were considered. This research was undertaken jointly with the project on corpus-based computer-assisted language learning of Nordic languages, funded by NorFA. The work has resulted in a prototype application [NB02, Bor02].
- A number of existing text materials have been explored and made available to the project: the national edition of the collected works of Swedish 19th century author August Strindberg, in a TEI marked-up version; 17th century Stockholm court records, in a TEI marked-up version; interviews with Hungarian Holocaust escapees, in HTML format.
- We are negotiating about the acquisition and use of other text materials: an 18th century diary, from the Department of Scandinavian Languages at Uppsala University; a digital version of Axel Oxenstierna's letters, from the Swedish National Archive; Old Swedish texts, from the Bank of Swedish, University of Göteborg.
- Donald Broady, Monica Langerth Zetterman and Jan Sjunnesson are working with creation of educational text collections using markup compliant to the TEIXLITE DTD (the Text Encoding Initiative Document Type Definition, light XML version) [Bro01].

Testbeds at Uppsala University In different case studies we have engaged a number of both undergraduate and graduate students from several disciplines. At Uppsala University approximately 80-90 students/participants have been or are involved in the testbeds. We are currently testing the usefulness of some tools, such as the APE-portfolio/SCAM application along with tests of existing web-based tools for content organization, sharing and annotation. The focus is to explore how students and researchers use these different tools to organize content, annotate and share information. The tests are conducted in natural settings i.e. in physically located or web-based courses that serve as testbeds. We use various methods to collect data such as questionnaires and group interviews.

- An Undergraduate courses at the Dept. for Teacher Education, "Learning and ICT!" (35 ECTS), a course in Technology, Nature and Society (35 ECTS), and a course in Methodology given during the spring term (January-June 2002). 65 students are using the APE-portfolio/SCAM and the user study/evaluation will be done during the first week in June 2002, which means one week before the students complete the courses.
- A graduate course in "Social Science Classics" (7,5 ECTS) October 2001- January 2002 where the platform SenSei was tested. 20 participants.
- A course "Correspondence analysis", November 2001 followed by seminars during Spring 2002, where the aim was to support the sharing of and personalized views on extensive data sets. 17 participants from several Scandinavian universities. The same course will be given during Autumn 2002 which gives the opportunity for more observations.
- Studies on collaboration within an interdisciplinary research program "Formation for the public sphere", where a standard platform (BSCW) is used.

4.3 Personalized and Shared Mathematics Courselets

4.3.1 Submodule Description

Working Title. MathViz: Personalized and Shared Mathematics Courselets.

Contributing research groups and PIs. Stockholm KTH/DSV (Jansson), Stockholm CID/KMR (Naeve).

Problem Description. This submodule attacks two major difficulties for teachers and learners: the difficulty to share and reuse learning material among students and teachers and across geographical and organizational boundaries and the difficulty to personalize and adapt existing learning material to a particular learning situation. Adaptation is relevant both with respect to the students characteristics and the context of learning Our particular concern is mathematics education. By courselets we mean fragments of courses composed from multimedia explanation modules or content modules in electronic form.

The work will extend on research at DSV on personalized support for learning conceptual modeling [TKRR99, RTK99, TRJ+99] and research at CID, where the idea of a concept browser has been developed by Naeve and his team over the past 4 years [Nae97, Nae99]. A first prototype of a concept browser, called Conzilla, has already been implemented [NP99, Nil00]. While conforming to evolving international e-learning standards (such as RDF and IEEE LOM), Conzilla combines conceptual modeling with annotated access to multi-media based archives in a novel way. This makes Conzilla a powerful basic platform for the kind of problems that our project aims to address. Moreover, since Conzilla is being developed as an open source project, it has the potential to evolve into a widely used tool that provides support for students and teachers in handling multimedia-based archives of digital information.

Using Conzilla, Naeve's group has started to create a Virtual Mathematics Exploratorium (VME)[12]. This work has been done partly with funding from WGLN (within the APE project)[13], and partly within the National Research School for Mathematics Didactics. The aim is to open up the construction work of the VME and get other teachers involved, thereby effectively disseminating the tools as well as the methodology involved in using them efficiently.

Within the Mathemagic project, in cooperation with the Advanced Media Technology (AMT) laboratory at KTH, CID is presently coordinating an effort to introduce new teaching methodology into mathematics courses at both the university and the school level[14].

Cybermath[15] is a shared virtual reality environment for the interactive exploration of mathematics. It has been developed at CID with partial support from WGLN within the APE-project, track C[16].

The Cybermath environment allows: teaching of both elementary, intermediate and advanced mathematics and geometry, the teacher to teach in a direct manner, teachers to present material that is hard to visualize using standard teaching tools, students to work together in groups, global sharing of resources. Cybermath has received widespread attention and has been presented at several international conferences, including SIGGRAPH-2001.

Research Plan. We articulate seven main concerns for this research: the commitment to shared standards, languages and tools to make sharing and reuse possible on a technical level (relations to the Module on Infrastructure and intelligent services), the management of incrementally growing multimedia content archives, in particular version handling and the handling of problems of structure and navigation (relations to the module on Infrastructures and intelligent services), the sharing of simple models for courselet structures, the modeling of domain knowledge, tasks and user competence as well as personal user preferences in such a way that it can form a basis for course structuring and personalization, the personalization of content modules through annotations expressed as meta-data, the personalization of courses or courselets through different ways to configure and modify structures of modules and the adaption and combination of user interface metaphors for authoring and use of courselets.

In our approach the students will be stimulated to play with preauthored vizualisations and other multimedia explanation modules for mathematical concepts, create, reuse or modify such modules, create their own conceptual models of mathematical knowledge, annotate nodes in conceptual structures with personal information, create courselets based on sequences of explanations generated from the personal conceptual models, indirectly create courselets generated through knowledge-based techniques basing their inferences on meta-data coding relevant contextual information, browse courselet structures, exchange courselets. Conzilla will provide the basic platform for the these activities, and the Graphing Calculator (http://www.pacifict.com) will be the basic visualization tool.

Teachers will also be engaged in the same kind of activities with the purpose to create relevant courselets. Typically the learning situation will be partly teacher driven and partly student driven. The above activities need a set of tools both for authoring (for conceptual modeling and meta- data specification) and browsing. This set will include Conzilla

[12] http://kmr.nada.kth.se/math/conzilla-demo.html
[13] http://kmr.nada.kth.se/papers/MathematicsEducation/APE-A/APE-A-final-report.pdf
[14] http://www.amt.kth.se/projekt/matemagi.html
[15] http://www.nada.kth.se/~gustavt/cybermath
[16] http://kmr.nada.kth.se/proj/ape.html#apec

and the Graphing Calculator from the start, but will be incrementally revised during the course of the project. An important part of the project will be the user-testing of Conzilla in a realistic learning environment, and the features of the program will be modified according to the feedback of the participating teachers and students.

From a methodological and technical point of view, the research will combine methods and techniques from the areas of conceptual modeling, ie. design of ontologies, concepts and relationsships suitable for modeling mathematical knowledge, artificial intelligence, i.e. knowledge representation for personal information, human machine interaction, i.e. adaptive interfaces, cognitive science, i.e., cognitive aspects of concept learning.

Deliverables, Timelines, Testbeds The submodule will be focussed on the use of personalized courseware for a few courses in Mathematics on the Information Technology Program at KTH.

In the first half year, we will introduce the Conzilla/GraphingCalculator-based methodology in two mathematics courses, set up an archive of appropriate multimedia explanation modules. Teachers will add appropriate meta-data to explanation modules, following up on studies of students modeling their own mathematical knowledge. Teachers will create prototypical courselets manually using conceptual modeling and semiautomatically using knowledge-based techniques.

In the second half year, students will also use these pre-authored courselets, author their own explanatory module, and be able to browse archives of multimedia explanations. Students will generate personalized courselets based on conceptual models, will share courselets among each other on the same course and will add annotations to modules and experiment with knowledge-based generation of courselets.

Due to the acceptance of the methodology by first year mathematics students at KTH which we managed to establish during year 1, we will be able to scale up experiments during the second year, so that they will affect all mathematics courses during year one. We will still introduce the methodology as a complementary support for students who experience problems with certain mathematical concepts. For the last part of year two we will conduct a more substantial empirical study which not only will study the students' acceptance of the methodology, but also the students' ability to reuse, adapt and personalize modules within and among courses. The larger scale application will also make it possible to detect potential positive changes for the students' examination results. Another development during year 2 will be to utilize the Edutella tools more systematically in the applications. In the fourth half year, evaluations will be carried out, guidelines for the methodology will be developed and documented. During the second year of the MathViz project we also plan to:

- introduce Cybermath at the IT-program in mathematics. This will add new types of mathematical interactivity (compared to what is provided by the Graphing Calculator), which we expect to increase the motivation of the students to further explore the corresponding concepts.
- restructure the Virtual Mathematics Exploratorium in order to take advantage of the Edutella system (developed in the infrastructure module of PADLR) and the SCAM-system[17] developed under the coordination of the KMR-group. Since SCAM (version 2) will implement an Edutella peer, it will be possible to publish the content of the VME in the Edutella network.

Results from the submodule will be prototypes, empirical studies, courselet and content archives for particular subdomains of mathematics as well as general reports.

Collaboration/Scholarly Exchange. Strong collaboration between DSV/KTH and KBS/Hannover with respect to interfaces for personalized course material and knowledge-based techniques to generate coursematerial from course-modules as described in two of the other submodules.

Connections to the work with meta-data in module on Infrastructure and intelligent services. The empirical results of the meta-data activities of this module will serve as input to the development of new metadata-handling capabilities in the infrastructure module.

Strong collaboration with the Mathemagic project at the Advanced Media Technology (AMT) laboratory at KTH, where a multi-media based component archive with mathematical content is being developed under the coordination of Naeve.

Exchange of graduate students planned between KTH and Hannover.

[17] http://kmr.nada.kth.se/scam/index.html

Budget Overview (including overhead costs).

KTH/DSV: 30K first year, 30K second year. Budget will pay for one part-time Ph.D. student, infrastructure costs, travel and exchange.

CID/KMR: 10K first year, 10K second year. Budget will pay for one part-time Ph.D. student, infrastructure costs, travel and exchange.

4.3.2 Progress Report (DSV Jansson, CID/KMR Naeve).

Students at KTH in Stockholm have trouble with their mathematics courses and difficulties learning certain math concepts. The overall goal of this project has been to develop, launch and evaluate the use of computer-based support for math education on a university level. The point of this has been a pedagogical one, to make abstract concepts that cause problems for students more concrete and understandable through the use of the computer-based tools.

During the fall of 2001 an archive consisting of a number of mathematics courselets were developed which illustrate central math concepts concerning linear algebra and geometry. The courselets were developed using the Graphing Calculator software, which is a graph drawing tool that can be used for the visualization of and interaction with mathematical expressions. In the spring, students were given the opportunity to participate in weekly, student driven, visualization sessions where math problems were visualized. The courselets were made available to the students over the web. Students could download the courselets, run them on their own computers and manipulate them as they wanted. Students appreciated the sessions as a complement to regular lectures by offering more thorough and developed explanations of central concepts and a support for learning concepts which caused them learning difficulties. The visualizations made it possible to develop a more concrete understanding of the math concepts. The sessions also brought up and made explicit the mathematical terminology related to the domain. In sum, the students were very satisfied with the support that the visualization sessions provided.

The math visualization courselets and interactive visualization sessions used in this project have been successful and reached a high level of acceptance of methodology by the 1st year mathematics students at KTH. The sessions have been documented on video and further analysis is continued to develop so called “practitioner tracks” - support for students to understand and use the visualization courselets on their own in between the sessions. See the report “Results in the Personalized and Shared Mathematics Courselets (MathViz) Project” for a full report of the MathViz project.

Dissemination The Norwegian National Concerts Organisation (”Rikskonserter”) has decided to create a national tour of performaces based on the Cybermath environment. These performances will be geared towards inspiring interest in mathematics on all levels - something that is felt to be of strategic importance by the Norwegian ministry of education.

Gjövik college has decided to run a distance course in mathematics using the Cybermath environment. The course will be taught from Stockholm, and it will be specially tailored towards addressing the conceptual difficulties that the students at Gjövik are experiencing in the ordinary mathematics courses.

4.4 Personalized Interfaces and Access to Educational Media

4.4.1 Submodule Description

Working Title. Personalized Interfaces, Navigation and Guidance - PerINaG

Contributing Research Groups and PIs. Hannover KBS (Nejdl/Allert)

Problem Description. Using information spaces learners must integrate new chunks of information into a coherent mental representation. This coherence formation process makes great demand on learners’ cognitive and metacognitive skills. They must make many decisions and there is a huge number of possible routes which can be constructed and performed by the learner. They must orient themselves and build up connections between single concepts, learning objects, units and courselets. They have to relate important items of content. In hypertext navigation and navigation in

non-linear data bases learners suffer from conflicting and competing goal intentions as well as from cognitive overload if the navigational task consumes too much of their resources. There is a strong need of distraction and violation protection in learning and problem solving. The instructional design is therefore based on theories of working memory.

By using information spaces in project oriented learning, learners have to perform real tasks such as answering questions, solving problems, writing reports, etc. So they do not only have to find information but also scan, read, interpret, evaluate for utility, annotate, and form coherence. How can learners be assisted most effectively in orientation, access and navigation? What kinds of direct and indirect help/aid should be presented? How to aid navigation in information spaces? Does it help learners to be presented the underlying complex domain/knowledge model or do they need different access structure? For example: task-adaptive access/navigation structure facing the needs in different stages of a project (broad overview, goal setting, information seeking, performing the procedural tasks, reflection and analyses), depending on their prior knowledge. How to contextualize learning objects and units and to organize access?

State-of-the-art research and systems concentrate on presentation and structuring of content. Trails and access in the Open Learning Repository [DNWW01, KBN02, ADK+02] instead focus on skills students are to be able to perform. We call them competencies. Example 1: Project orientation qualifies students to organize, coordinate, and manage projects. Example 2: the trail "Introduction to Scientific Communities" introduces students to scientific work. Access is structured along these "trails of competence".

Using metadata and existing standards we state a lack of instructional information in metadata standards for eLearning. State-of-the-art concepts for integrating instructional information in standards focus on "instructional neutrality". Because of neutrality they lack important instructional information. We constitute an approach which is deduced from philosophy of science (especially postmodern and poststructuralistic approaches) and science of education. This approach of Instructional Roles supports any instructional context and any theory of cognition, learning theory and instructional principle. This concept has to be introduced to standardization initiatives. We also apply it to OLR.

Research Plan and Deliverables (shortened 1st year, with 2nd year revisions). This submodule will deal with course structuring, access, navigation, and orientation in non-linear information spaces, as well as instructional and cognitive impacts on the design of distributed learning repositories, and requirements for designing these systems with evaluation in mind from the beginning. These general issues address many important actual problems and questions. We will focus on the following:

The initiatives of Standardization in eLearning (IEEE LOM, ADL SCORM and others) are driven by a narrowed perspective on learning. We want to open the view on learning in standardization and propose an approach of Instructional Roles in eLearning Metadata Standards, which explicitely supports the idea of annotating different paradigms, models and principles in learning. Based on this general approach we will apply it to a concrete instructional principle as well as theory of cognition. Corresponding to the 'trails of competence' which will be implemented in the Open Learning Repository (OLR) we will define a set of metadata suitable for situated learning and the concept of 'Communities of practice' (CoP according to Lave and Wenger). Nevertheless, the approach of Instructional Roles is applicable to any instructional model.

The Open Learning Repository (submodule MoCa - OLR) is based on metadata and is flexible to different instructional models and instructional principles. In a scenario-based-design we integrate different access structure which are designed along different concepts such as 'Communities of Practice' and project-based-learning. The OLR will be used in CS courses at KBS (University of Hanover). Testbeds in Germany are described in the Edutella module.

The scenario-based-approach will support the implementation of the OLR into the practice of teaching and learning at the University as well as the redesign of the courses themselves. We will further develop the scenario-based-approach for specific purposes of the design of learning environments. The evaluation within the scenario-based-design of OLR focusses on scenarios written by all stackeholders of the design process as well as by teaching staff. These scenarios of intended use are contrasted by scenarios of current use.

Dissemination & Testbeds Testbeds: mainly ULI and CS courses at KBS (see Edutella), other testbeds in interaction with collaborating modules. Dissemination: Publications and Reports.

Collaboration and Scholarly Exchange Especially with Edutella, PleaSe, Modular Content Archives, in Sweden with Department of Teacher Education. Cooperation and research visits.

Budget Overview (including overhead costs).

KBS: 30K first year, 30K second year, for part-time Ph.D. student, incl. overhead costs, travel and exchange.

L3S central: 15K first year, 15K second year, for part-time Ph.D. student in media and design education, especially for evaluation, incl. overhead costs, travel and exchange.

4.4.2 Progress Report (KBS, Nejdl).

Research in the first project phase has concentrated on three main aspects:

1. We modelled different access structure to information spaces based on different learning theories and instructional principles, called trails of competencies. We modelled an instructional model specifing instructional metadata in order to be able to offer different forms of navigation in the Open Learning Repository (OLR3) based on the following approaches: Expository Teaching (cognitive approach), Project Based Learning and Communities of Practice (both situated approaches). We integrated the course Artificial Intelligence I and currently redesign the interface [ADK+02].

2. Evaluation of the former version of OLR in order to guide the design of OLR3. We used a scenario-based-approach which matches the key aspects of theory-based-evaluation proposed within the WGLN. We adopt the scenario-based-approach (eg. [Bod99, Car00] to the design of learning systems and develop it further on. The further development of the scenario-based-approach will focus on the differences between learning and work on the one hand and modelling the context of a learning situation on the other [RA02].

3. As we designed access structure and concepts of navigation based on different learning theories we stated a lack of instructional information in metadata standards such as LOM (IEEE) and SCORM (ADL). We therefore developed the approach of Instructional Roles which integrates information on different theories of cognition, learning theories and instructional principles to metadata standards [ADN02b, ADN02a].

Additionally, we were invited to give the following talks:

- Allert, Heidrun. Instructional Roles in eLearning Metadata Standards. "Open Source eLearning". Bauhaus University Weimar (Veranstaltungsreihe der Hochschulleitung, Fakultät Medien, Projekt medienquadrat sowie Fachschaft Medien). Weimar, Dec 19th, 2002.
- Allert, Heidrun. Standardisierung in didaktischer Perspektive. 5. Hagener MultimediaWerkstatt - ZFE Zentrum fuer Fernstudienentwicklung. Hagen, March 14th, 2002.
- Allert, Heidrun. Theoretischer Ansatz zur Rolle der Didaktik in Metadaten Standards. Workshop "Standardisierung im eLearning"[18]. Frankfurt 10/11 April 2002.
- We also currently contribute to German standardization initiative at DIN - Deutsches Institut für Normung e.V. (German Institute for Standardization).

We collaborate with PleaSe (Personalized Learning Sequences). Some of the Learning Sequences are designed along learning theories used in OLR3. We specify instructional metadata which serves both, the OLR3 as well as the Personalized Learning Sequences. The Learning Sequences are integrated in the interface designed currently. The OLR3 is open to Personalized Learning Sequences. We also collaborate with Modular Content Archieves (MoCA) in the scenario-based-design of the Open Learning Repository (OLR3). It is designed along the concept of Communities of Practice and Situated Learning. But it is also flexible to integrate models of Expository Teaching. We collaborate in the specification of metadata in order to annotate these concepts.

Research Visits within PerINaG:

[18] http://www.httc.de/nmb/

- November 2001: a research visit by Heidrun Allert and Hadhami Dhraief to the Learning Lab Uppsala, Sweden and to the Department of Computer and Systems Sciences / CID, University of Stockholm (two weeks)
- June 2002: a research visit by Heidrun Allert and Christoph Richter at Stanford (5 weeks)

4.5 Learning Platforms based on Web Services

4.5.1 Submodule Description

Working Title. WebServices: Learning Platforms based on Web Services

Contributing Research Groups and PIs. Institute for Operating Systems and Computer Networks), TU Braunschweig (Fischer)

Problem Description. Today's learning platforms are usually stand-alone platforms in that they basically provide only one service (teaching/learning) for a given set of learning units (courses) which are stored within this system. Course unit exchange with other platforms is often difficult, even when the units follow the SCORM standard, since there is not standardized online way of exchanging these units. In addition, integration of such platforms in huger contexts, including for example human capital management, as it would be desirable in professional educational systems, is at least difficult.

Research Plan and Deliverables. We propose to investigate the newly standardized approach of web services to be used as the communication infrastructure for e-Learning services. Web services basically provide an XML-based way of implementing and using services in the WWW. The approach consists of a number of simple and efficient protocols such as HTTP and SOAP, a language to describe services named WSDL, and a directory service that helps locating available services called UDDI. SOAP, WSDL and UDDI are all based on XML as basic data description language.

The big advantage we see in using web services lies in its potential for standardization. Once access to certain services (and we are interested in e-learning services here) is provided by the way of web services, this service can be very easily integrated into any application that also speaks the web services' languages. In addition, these applications can make use of other services in the Net, thus allowing the creation of highly-integrated applications as described above. Ideally, at the end, SCORM-based and thus standardized course descriptions would be exchanged over web-services-based and thus also standardized communication infrastructures, creating a huge set of easily to be integrated course units.

The plan described here covers the first year of the project. All in all, the project is expected to run for roughly 3 years. In the first year, we plan to do a feasibility study by first deriving the requirements on a solution and then implementing a first and simple prototype. In the second and third year, the approach will be refined and completely implemented, making use of all the features provided by web services that make sense in our context. In parallel, it is planned to create first versions of a number of courses based on the early prototype. The test bed for this will be outlined in the next section.

Within the first year, the following tasks have to be solved:

1. Current developments in the field of web services have to be surveyed. This is a rapidly evolving field, making it necessary to gain a current overview (expected time: 2 person months).
2. A framework for the use of web services in e-learning environments has to be designed. Issues to be discussed include:

   How can course units be accessed through web services? Are they web services themselves, or is access to them provided through web services?

   How could learning applications look like that are based on web services? Is there a standard way (a recipe) for building web-services based e-Learning applications?

   Does it make sense to build hierarchies of web services, i.e., web services that themselves act as clients to other web services. A purpose could be the integration of services on an even higher level.

Overall time requirement for this phase should be roughly 4 to 5 person months. The expected outcome is a detailed report that answers the questions posed above.

3. Create a first prototype for a web-services based educational application. The idea is to use one of the existing and freely (in source code) available learning platforms as a basis. A web services interface should be build and integrated into this platform, providing the server interface. On the other hand, a web-based client-side (client for the educational platform) application (for instance based on servlets or similar technologies) has to be built; access to this application by users is via the web. The new application needs to demonstrate the added value of using web services by also integrating other services available in the web. This phase should be doable within 5 to 6 person months. As a result, a working prototype can be expected.

Dissemination, Test Beds, and Evaluation. Dissemination of results will be done through reports and scientific publications on the different aspects outlined in the research plan. At the end of the first year, a prototypical implementation of our approach will be available and can be expected to demonstrate the most important aspects of web-services-based educational platforms.

The approach will be evaluated by applying the prototype within the "ELAN" environment. ELAN is a new e-Teaching/e-Learning project in Lower Saxony, providing resources of 3 million Euro for the next 2.5 years to the Universities of Braunschweig and Hannover. ELAN is both meant to use an existing well-established educational software platform and experimental platforms, making it an ideal test bed for our purposes, especially for comparative studies.

Collaboration and Scholarly Exchange. The work on this module is closely related to other modules building tools and functionalities. Research visits to other participants are expected to be complemented with daily e-mail contacts.

Budget Overview (including overhead costs).

IBR: 70K USD for one year. Budget will pay for one PhD student, L3S overhead cost, travel and exchange.

4.5.2 Progress Report (IBR, Fischer).

This subproject will start in the second year.

4.6 L3S Center for Professional Development (L3SCPD)

4.6.1 Submodule Description

Working Title. L3SCPD: L3S Center for Professional Development

Contributing Research Groups and Pls. Hannover IANT (Jobmann/Krüger), KBS (Steimann), GEML (Garbe), Schering Institut (Gockenbach)

Problem Description. L3S shall be enabled to market educational content, operating in two ways: 1) as profit center, which is self funding, and 2) as support activity to enhance the internal lecturing of the university of Hannover and possibly other participating universities.

Several approaches throughout the last two years have shown, that the market of education in Germany is so much different from the US-market, that the straight adoption of US-defined continuing education activities into the German environment are not self financing. The main reason for this is that in Germany university education is financed indirectly by the government, not directly by the students. The bachelor or master degree can be achieved as a fulltime student for an average fee of 100 Euro per half term at more than 30 Universities and more than 100 politechnical high schools or as parttime student at the Fernuniversität Hagen or the FHS-Verbund for the same fee.

The knowledge on corporate mass education in office support technologies like training wordprocessors, spreadsheets and others is not leading edge at universities and thus unlikely to be a L3SCPD market.

A third possible opportunity could be the direct transfer of research results within continuing education directly out of the labs of the university.

Internal teaching processes at German universities are still mainly based on traditional lectures with professor and students present at the lecture hall at certain times. First tests show that there is a demand from the students for tele-education. The methods to achieve similar or better results in tele-learning are still under development and need further enhancements.

To run the L3SCPD it is mandatory to test and install a learning platform including campus management.

Development plan and deliverables. A business plan shall be developed in order to show the business opportunities for self financing parts of L3SCPD. First draft ready in December 2002. Depending on the outcome further activities will be defined.

The discussion on the learning platform needs to be based on an evaluation of three examples, after studying five versions on paper basis. The paper studies and the actual experiences will result in a requirement portfolio which will be used to challenge the three test versions. Available studies from the market will be bought. The paper studies will be ready April 2003 and the test will be finalized December 2003.

The support of the internal activities in teleteaching will be in installing the tele-educational testbed at the L3S. Due date April 2003. Based on the testbed, teaching scenarios will be developed, implemented and tested with three types of lectures. Due date July 2004.

Dissemination, Testbeds and Evaluation. The business plan is to be presented to the board of management of L3S, further steps need to be agreed with them. The outcome of the platform evaluation is to be presented to the participants of L3S and to be agreed by the board of management. The evaluation of the testbed activities will be performed using questionairs, from users, trainers and technicians.

Collaboration and Scholarly Exchange. With infrastructure modules (especially basic infrastructure and WebServices) for exchange of technical evaluations and extensions of commercial platforms.

Budget Overview (including overhead costs).

L3S central: 70K USD per year, 1 MY first year, 1MY second year for a Ph.D. student in professional education who is engineer and teacher, L3S infrastucture costs, travel and exchange, external studies.

4.6.2 Progress Report (IANT, Jobmann).

During the first project phase about 60 training hours were recorded and placed on the streaming server at IANT. At the same time KBS recorded about 30 training hours. About 20 students in IANT and about 40 in KBS used these streams for their studies. The first round was done without formal evaluation due to the high learning progress of the teaching and support staff and the changes which were implemented permanently. Now in the first term 2002 (SS2002) the tool and methodological situation is stable and thus an evaluation process is under progress. Apart from the lecture a tutorial process was developed which is under evaluation as well.

Several attempts in marketing the recorded lectures and some lectures from Stanford CPD were taken. The business partners were interested in taking test lectures until we talked about prices for regular lectures. In parallel we did a quick analysis on two platforms (Blackboard, Hut) and decided to buy a test licence from Hut to run our first term 2002 with that. The platform will be included in the running evaluation.

4.7 Copyright in Personalized Access to Distributed Learning Repositories

4.7.1 Submodule Description

Working Title. Copyright in Personalized Access to Distributed Learning Repositories (CORI-PADLR)

Contributing Research Groups and PIs. Göttingen IWF gGmbH Knowledge and Media (Floto/Rebe/Sander/Dudzik)

Problem Description and Research Plan (1st year, shortened) .

Objective of the PADLR framework is to specify how courselets are build, how they are organized and how they are exchanged and reused, and how distributed content archives can be queried and navigated. While these are necessary and important educational and technical issues, they are not sufficient to determine the future of the exchange of Educational Media. Legal problems in conjunction with the use of Napster - an exchange platform for music - have clearly shown the need to address copyright issues when exchanging media. Especially in an educational environment, where public (universities) and private (publisher) partner collaborate, there is a strong demand for the protection of intellectual property rights and the availability of appropriate business models. Therefore, inside the PADRL framework there is a strong demand to investigate copyright issues for the exchange of media.

To solve this problems, several forms of protection of intellectual property right will be evaluated (watermark, server-based-only media, copy protection, fees like the German Gemma etc.). Second, business models are going to be evaluated (E-Commerce, Micropayment, Pay per View, Subscription, Syndication, Open Source etc.). Third, appropriate forms of intellectual property right protection will be combined with sound business models and tested.

Problem Description and Research Plan (2nd year, revised). As regards the methods of protection of intellectual property rights and business models, the main area of the research work in the second year will be the guideline 2001/29/EG on the harmonisation in special aspects of copyright in the European Union, which was passed by the European Comission on 22nd May 2001. This guideline is based on the WIPO contracts, referring to the WCT = WIPO Copyright Treaty and the WPPT = WIPO Phonograms and Performances Treaty, both from December 1996, which were signed from the EU itself, most EU countries as well as the United States. As such, the ratification and the transformation of these contracts and, in the EU, the transformation of the guideline could be the beginning of a worldwide standardisation of copyright codification. The guideline is to be accepted into national law by December 22nd 2002. In Germany, the Ministry of Justice passed their first draft on the modification of the UrhG (German copyright law) last month. However, due to the election of the German parliament in fall 2002 it is already certain that they will not be in time.

Nevertheless there will be a great deal of public discussion on the changes which have to be made, should be made, should not be made and must not be made to copyright law because of the guideline. As many different copyright traditions exist that are concerned by these multinational arrangements (e.g. the American, British, Swedish, German, and Italian legislations are all different), many people are afraid that this new codification could destroy the social balance between the copyright holders and the public. In Germany in particular, there is concern for the ”Privatkopie” - a right for every person to make one copy of nearly everything for private use without the consent of the copyright holder. This right is not recognised in the US and there is concern inin Germany that it will be cancelled with the modification of the UrhG (German copyright law). Chapter three of the guideline, which should support technical methods to improve copyright security states that this new UrhG would be the last copyright legislation made by parliament. It could thus be an authorization for every copyright holder to make his or her own law.

Indeed, what is meant by ”technical methods” in chapter three of the guideline must also be closely examined, as musts also the effect of these methods onthe social ”contract” mentioned above, If, for example, digital rights management systems would undermine fundamental public interests; or, on the other hand, if these technical methods could be used to improve copyright security. It must be examined whether these modifications will have an influence on the economic behaviour of consumers, copyright holders, authors and all others who are involved in copyright issues. As previously mentioned, the importance of digital rights management systems could increase and so the use of (juridical) metadata could also increase. On the other hand, the availability of free sources, for example, open source software could decease. This should be examined as well.

Finally article 5 of the guideline offers the possibility of implementing several limitations in national copyright legislation. This article is a conglomeration of the limitations which could be found in different EU member state legislations, and so it is interesting to compare the foreign limitations with the German ones, thus making suggestions for additions. Paragraph 3 a) of article 5 could make an interesting additional limitation for German copyright legislation. This would allow the copying and non-commercial use of copyright protected material, including sound and picture media, for all kinds of teaching and scientific research. Similar legislation already exists, for example, in Austria, France, and the Netherlands. A way has to be found to build this limitation into German legislation as well as there has to be found a way to compensate copyright holders for this. In this way, the contradiction between public pressure on universities to finance their research projects and current copyright law can be examined as well as its relationship

with the German privilege for non-commercial users to merely name the copyright holder.

Deliverables, Dissemination A document is planned (doctoral thesis from a faculty of law) which addresses the issues mentioned above, especially the problems and possibilities concerning the EU guideline and other international treaties. There will be also be several reports, papers, and scientific publications about these problems. Finally, the implementation of the schemes on copyright legislation into a software solution is planned.

Testbeds and Evaluation The schemes and the software solution will be used with several testbeds from the participating institutes. Reports on these tests will be collected, additionally an evaluation from the PADLR evaluation team could also be carried out. Feedback forms will be designed so that the papers and reports can be analyzed especially in order to obtain feedback from the participating institutes.

Collaboration and Scholarly Exchange. The modification of the national copyright laws is to be observed by all PADLR projects. On the one hand, the issue of whether open source projects are still useful should be considered. It is to be examined, whether digital rights management systems should be implemented in Edutella, and if so, which ones, and also whether they should be implemented into the IWF Cells Platform[19] and the IWFdigiclip platform. Finally the "technical methods" named in the EU guideline could cause existential problems for all peer-to-peer networks and a way should be found to solve these problems. All participating institutes who want to use their research results commercially are faced with these legal issues and so will all profit from research work in this field.

Budget Overview (including overhead costs).

IWF: 35 K USD for a PostDoc, including all travel costs and overhead, for two years.

4.7.2 Progress Report (IWF/TU-BS, Floto/Rebe).

The original proposal noted that the PADLR framework will raise several legal issues concerning the exchange of (educational) media, especially in an international context using the Internet, the collaboration of public and private partners and the availability of appropriate business models. Therefore, it was decided to investigate copyright issues with the transfer of media. The research plan was firstly, to evaluate various methods of protecting intellectual property rights, then to evaluate business models and lastly to combine the protection methods with sound business models. The investigations of recent months have clearly shown that this approach was not detailed enough, especially regarding the PADLR - projects.

The most important issue in the case of multimedia projects with international collaboration is not how the intellectual property rights can be protected and how to make any research results commercially available. The first and most important issue in such a project is the creation of an extremely clear structure. This structure should contain contracts between all partners and rights holders detailing their functions and responsibilities and the intellectual property rights of the project results etc. The creation of such a structure is the best way to enable research results to be made commercially available.

It became clear that many similar projects, particularly when there are university partners, do not spend any time at all on these questions before starting projects and were not able to rectify any mistakes once the projects were completed. Therefore it was decided to develop basic planning schemes for the project. The schemes on general aspects of planning for multimedia projects are almost complete. Additions will be made regularly. The schemes can be found in the CORI-folder on the PADLR-Server[20] and are called "Schemata zur Planung von Multimedia-Projekten".

The next issue that arose was that not only in the case of planning but also in all the other stages of multimedia projects, the project organisers are not usually legal specialists. Particularly in the case of public partners like universities professional advice is not asked for, not only for financial reasons but also because problems are simply overlooked. Therefore we decided to design some basic schemes on the systematics in legislation of intellectual property rights. Of course, for many questions, these schemes cannot replace professional advice, but they can help to

[19] http://www.cells.de

[20] http://www.learninglab.de/workspace/padlr/index.html

avoid obvious mistakes and furthermore, they can help to formulate the right questions in order to get concrete answers from professional legal advisers. This, at least, will save a certain amount of money. The schemes on the systematics in German legislation on intellectual property rights are also quite advanced. The relevant folder is called: “Schemata zum deutschen System des (geistigen) Eigentums”, again at the PADLR-Server.

The first idea was to simply design these schemes as separate papers. But while working on the schemes we became aware of the benefits of connecting the papers with a software solution. The CORI-PADLR folder “Skizzierung der softwarebasierten Schematalösung” gives a good overview of the system that could be created for the schemes using this software solution.

After working on these basic elementary issues for the planning and realization of multimedia projects, we then started to work on the particular copyright problems that occur when media is exchanged. As this was a major theme during the PADLR workshops in Hannover and Uppsala, we realized that metadata is one of the most important ways of making the exchange via the Internet possible or at least effective. However, we soon noticed that there were a lot of different standards for metadata, but that no standard really offers a satisfactory system for legal metadata. So we decided to investigate the problems and the possibilities of legal metadata. The folder “Skizzierung der Problematiken rechtlicher Metadaten” gives a short description of the possibilities and the difficulties of legal metadata. A summary report will be available soon.

References

[ADK+02] Heidrun Allert, Hadhami Dhraief, Tobias Kunze, Wolfgang Nejdl, and Christoph Richter. Instructional models and scenarios for an open learning repository, instructional design and metadata. In *AACE E-Learn World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education*, Montreal, Canada, October 2002.

[ADN02a] Heidrun Allert, Hadhami Dhraief, and Wolfgang Nejdl. How are learning objects used in learning processes? instructional roles of learning objects in lom. In *ED-MEDIA 2002, World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Denver, Colorado, June 2002.

[ADN02b] Heidrun Allert, Hadhami Dhraief, and Wolfgang Nejdl. Meta-level category role in metadata standards for elearning. instructional roles and instructional qualities of learning objects. In *Proceedings of the 2nd International Conference on Computational Semiotics For Games And New Media*, Augsburg, Germany, September 2002.

[ANS02] Benjamin Ahlborn, Wolfgang Nejdl, and Wolf Siberski. Oai-p2p: A peer-to-peer network for open archives. In *Proceedings of the Workshop on Distributed Computing Architectures for Digital Libraries, 31st International Conference on Parallel Processing (ICPP 2002)*, Vancouver, Canada, 2002.

[Axe00] Margareta Westergren Axelsson. USE – The Uppsala Student English Corpus: an instrument for needs analysis. *ICAME Journal*, 24:155–157, 2000.

[BD99] Lars Borin and Mats Dahllöf. A corpus-based grammar tutor for education in language and speech technology. In *EACL'99, Computer and Internet Supported Education in Language and Speech Technology. Proceedings of a Workshop Sponsored by ELSNET and The Association for Computational Linguistics*, pages 36–43. ACL, 1999.

[BEG+01] Stefan Brantner, Thomas Enzi, Susanne Guth, Gustaf Neumann, and Bernd Simon. UNIVERSAL - design and implementation of a highly flexible e-market-place for learning resources. In *Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT) 2001*, 2001.

[BEK+00] Don Box, David Ehnebuske, Gopal Kakivaya, Andrew Layman, Noah Mendelsohn, Henrik Frystyk Nielsen, Satish Thatte, and Dave Winer. Simple object access protocol (soap) 1.1. Technical report, W3C Note, May 2000. http://www.w3.org/TR/2000/NOTE-SOAP-20000508/.

[BKD+01] Jeen Broekstra, Michel Klein, Stefan Decker, Dieter Fensel, and Ian Horrocks. Adding formal semantics to the web: building on top of RDF schema. In *Proceedings of the 10th World Wide Web Conference*, Hongkong, May 2001.

[BkSBa] Lars Borin, Kariné Åkerman Sarkisian, and Camilla Bengtsson. A stitch in time: Enhancing university language education with web-based diagnostic testing. In *Proceedings of the 20th ICDE Conference, Düsseldorf 1–5 April 2001*. ICDE, t a .

[BkSBZa] Lars Borin, Kariné Åkerman Sarkisian, Camilla Bengtsson, and Monica Langerth Zetterman. Developing and evaluating web-based diagnostic language testing in university language education. In *Accepted for the ALTE European Year of Languages conference, Barcelona, 5–7 July, 2001*. ALTE, t a .

[Bod99] Susanne Bodker. Scenarios in user-centred design - setting the stage for reflection and action. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS 32)*, Hawaii, USA, January 1999.

[Bor00] Lars Borin. A corpus of written Finnish Romani texts. In *LREC 2000. Second International Conference on Language Resources and Evaluation. Workshop Proceedings. Developing Language Resources for Minority Languages: Reusability and Strategic Priorities*, pages 75–82. ELRA, 2000.

[Bor02] Lars Borin. What have you done for me lately? the fickle alignment of nlp and call. In *EuroCALL 2002 Pre-Conference Workshop on NLP in CALL*, Jyväskylä, Finland, August 2002.

[Bor a] Lars Borin, editor. *Parallel corpora, parallel worlds. Papers presented at a symposium on parallel and comparable corpora, Uppsala, 22–23 April, 1999*. Rodopi, Amsterdam, t a.

[Bro01] Donald Broady. Digitala arkiv och portföljer (digital archives and portfolios) - mirakelmedicin eller sockerpiller. In *IT i skolan*, number 45/2001. Stockholm: IT-kommissionen, 2001.

[Bru96] P. Brusilovsky. Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction*, 6(2-3):87–129, 1996.

[Bru99] Peter Brusilovsky. Adaptive and intelligent technologies for web-based education. *Special Issue on Intelligent Systems and Teleteaching, (Künstliche Intelligenz),*, 4:19–25, 1999.

[Car00] John M. Carroll. *Making Use: Scenario-Based Design of Human-Computer Interactions*. MIT Press, Cambridge, 2000.

[CES02] Richard Cole, Peter Eklund, and Gerd Stumme. Document retrieval for email search and discovery using formal concept analysis. In *Journal of Applied Artificial Intelligence (AAI).*, 2002. (in press).

[CGM02] Arturo Crespo and Hector Garcia-Molina. Routing indices for peer-to-peer systems. In *Proceedings of the 22nd International Conference on Distributed Computing Systems*, Vienna, Austria, July 2002.

[CS00] Richard Cole and Gerd Stumme. CEM - a conceptual email manager. In *International Conference on Conceptual Structures*, pages 438–452, 2000.

[DAM01] DAML Joint Committee. DAML+OIL, 2001. URL: `http://www.daml.org/2001/03/daml+oil-index.html`.

[DB01] Rael Dornfest and Dan Brickley. The power of metadata. In Andy Oram, editor, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. O'Reilly, 2001. http://www.openp2p.com/pub/a/p2p/2001/01/18/metadata.html.

[DEFS99] Stefan Decker, Michael Erdmann, Dieter Fensel, and Rudi Studer. Ontobroker: Ontology based access to distributed and semi-structured information. In R. Meersman, Z. Tari, and S. Stevens, editors, *Semantic Issues in Multimedia Systems*. Kluwer Academic Publisher, Boston, 1999.

[DNWW01] Hadhami Dhraief, Wolfgang Nejdl, Boris Wolf, and Martin Wolpers. Open learning repositories and metadata modeling. In *Proceedings of the International Semantic Web Working Symposium (SWWS)*, Stanford, CA, July 2001.

[DvHB+00] Stefan Decker, Frank van Harmelen, Jeen Broekstra, Michael Erdmann, Dieter Fensel, Ian Horrocks, Michel Klein, and Sergey Melnik. The semantic web - on the roles of XML and RDF. *IEEE Internet Computing*, September 2000.

[GBC+01] Carole Goble, Sean Bechhofer, Leslie Carr, David De Roure, and Wendy Hall. Conceptual open hypermedia = the semantic web? In *Proceedings 2nd Intl. Workshop on the Semantic Web (SemWeb'2001), Hong Kong, China, May 1, 2001*, pages 44–50, 2001.

[GMPQ+97] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y.Sagiv, J. Ullman, V. Vassalos, and J. Widom. The tsimmis approach to mediation: Data models and languages. *Intelligent Information Systems (JIIS), Kluwer*, 8(2):117–132, 1997.

[Hen00] Nicola Henze. *Adaptive Hyperbooks: Adaptation for Project-Based Learning Resources*. PhD thesis, University of Hannover, 2000.

[Hen01] Nicola Henze. Towards open adaptive hypermedia. In *ABIS-Workshop 2001, Workshopwoche Lernen - Lehren - Wissen - Adaptivität (LLWA01)*, Dortmund, Germany, 2001.

[HH00] Jeff Heflin and James Hendler. Semantic interoperability on the web. In *Proceedings Extreme Markup Languages*, Montreal, August 2000.

[HMS01a] Andreas Hotho, Alexander Maedche, and Steffen Staab. Ontology-based text clustering. In *Proc. of the Workshop "Text Learning: Beyond Supervision" at IJCAI 2001. Seattle, WA, USA, August 6, 2001*, 2001.

[HMS01b] Andreas Hotho, Alexander Maedche, and Steffen Staab. Ontology-based text clustering. In *Workshop Text Learning: Beyond Supervision, IJCAI*, 2001.

[HMSS02] Andreas Hotho, Alexander Maedche, Steffen Staab, and Rudi Studer. A topic broker for human resource management. *Journal of Universal Computer Science*, 2002. to appear.

[HMSV] Siegfried Handschuh, Alexander Maedche, Ljiljana Stojanovic, and Raphael Volz. KAON – the Karlsruhe ontology and semantic web infrastructure. "http://kaon.aifb.uni-karlsruhe.de/white-paper".

[HN99] Nicola Henze and Wolfgang Nejdl. Adaptivity in the KBS Hyperbook System. In *2nd Workshop on Adaptive Systems and User Modeling on the WWW*, Toronto, Canada, May 1999.

[HN00] Nicola Henze and Wolfgang Nejdl. Extendible adaptive hypermedia courseware: Integrating different courses and web material. In *Proccedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2000)*, Trento, Italy, 2000.

[HN01] Nicola Henze and Wolfgang Nejdl. Adaptation in open corpus hypermedia. *IJAIED Special Issue on Adaptive and Intelligent Web-Based Systems*, 12, 2001.

[HN02] Nicola Henze and Wolfgang Nejdl. Knowledge modeling for open adaptive hypermedia. In *Proccedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2002)*, Malaga, Spain, 2002.

[HNW99] Nicola Henze, Wolfgang Nejdl, and Martin Wolpers. Modeling Constructivist Teaching Functionality and Structure in the KBS Hyperbook System. In *Proc. Computer Supported Collaborative Learning Conference*, Stanford, December 1999. a previous version appeard in AIED99 Workshop on Ontologies for Intelligent Educational Systems.

[Hor01] Ian Horrocks. The FaCT System, 2001. URL: `http://www.cs.man.ac.uk/~horrocks/FaCT/`.

[KBN02] Tobias Kunze, Jan Brase, and Wolfgang Nejdl. Editing learning object metadata: Schema driven input of rdf metadata with the olr3-editor. In *Proceedings of the Semantic Authoring, Annotation and Knowledge Markup Workshop, 15th European Conference on Artificial Intelligence*, Lyon, France, July 2002.

[KKM02] Dan Klein, Sepandar D. Kamvar, and Christopher D. Manning. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *International Conference on Machine Learning*, 2002.

[KM02] Dan Klein and Christopher D. Manning. A generative constituent-context model for improved grammar induction. In *Proceedings of the Association for Computational Linguistics 40*, 2002.

[KSV+02] Barbara Kieslinger, Bernd Simon, Gorazd Vrabic, Gustaf Neumann, Juan Quemada, Nicola Henze, Sigrun Gunnersdottir, Stefan Brantner, Tilman Kuechler, Wolf Siberski, and Wolfgang Nejdl. Elena - creating a smart space for learning. In *Late-Breaking topics session at International Semantic Web Conference (ISWC 2002)*, Sardinia, Italy, 2002.

[KTI+02] Dan Klein, Kristina Toutanova, H. Tolga Ilhan, Sepandar D. Kamvar, and Christopher D. Manning. Combining heterogeneous classifiers for word-sense disambiguation. In *WSD Workshop at ACL 40*, 2002.

[Las98] Ora Lassila. Web metadata: A matter of semantics. *IEEE Internet Computing*, 2(4):30–37, 1998.

[LC97] L.Liu and C.Pu. An adaptive object-oriented approach to integration and access of heterogeneous information sources. *Distributed and Parallel Databases*, 5(2):167–205, 1997.

[LRK02] H. Lin, T. Risch, and T. Katchanounov. Adaptive data mediation over XML data. *Journal of Applied System Studies (JASS), Cambridge International Science Publishing*, 2002. to be published in special issue on Web Information Systems Applications.

[MEH+02] Alexander Maedche, Marc Ehrig, Siggi Handschuh, Raphael Volz, and Liljana Stojanovic. Ontology-focused crawling of documents and relational metadata. In *Proceedings of the Eleventh International World Wide Web Conference WWW*, Hawaii, 2002. (Poster).

[MS99] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Boston, MA, 1999.

[MS00] Alexander Maedche and Steffen Staab. Discovering conceptual relations from text. In W. Horn, editor, *ECAI 2000 — Proceedings of the 14th European Conference on Artificial Intelligence, Berlin, August 21-25, 2000*, Amsterdam, 2000. IOS Press.

[MS01] Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2), March/April 2001.

[Nae97] A. Naeve. The garden of knowledge as a knowledge manifold - a conceptual framework for computer supported personalized education. Technical Report CID-17, KTH Stockholm, 1997.

[Nae99] A. Naeve. Conceptual navigation and multiple scale narration in a knowledge manifold. Technical Report CID-52, KTH Stockholm, 1999.

[Nae01] A. Naeve. The concept browser - a new form of knowledge management tool. In *Proceedings of the 2nd European Web-based Learning Environments Conference*, Lund, Sweden, October 2001. http://kmr.nada.kth.se/papers/ConceptualBrowsing/ConceptBrowser.pdf.

[NB02] Kristina Nilsson and Lars Borin. Living off the land: the web as a source of practice texts for learners of less prevalent languages. In *EuroCALL 2002 Conference*, Jyväskylä, Finland, August 2002.

[NDW01a] Wolfgang Nejdl, Hadhami Dhraief, and Martin Wolpers. O-telos-rdf: A resource description format with enhanced meta-modeling functionalities based on o-telos. In *Workshop on Knowledge Markup and Semantic Annotation at the First International Conference on Knowledge Capture (K-CAP'2001)*, Victoria, BC, Canada, October 2001.

[NDW01b] Wolfgang Nejdl, Hadhami Dhraief, and Martin Wolpers. O-telos-rdf: An extension of rdf with enhanced mata-modeling and reification functionalities. Technical report, University of Hannover, November 2001.

[Nej02] Wolfgang Nejdl. Semantic web and peer-to-peer technologies for distributed learning repositories. In *IIP-2002*, Montreal, Canada, August 2002.

[Nil00] M. Nilsson. The conzilla design - the definitive reference. http://conzilla.sourceforge.net/doc/conzilla-design/conzilla-design.html, August 2000.

[NNP01] A. Naeve, M. Nilsson, and M. Palmér. The conceptual web - our research vision. In *Proceedings of the First Semantic Web Working Symposium*, Stanford, July 2001. (Position Paper),.

[NP99] M. Nilsson and M. Palmer. Conzilla - towards a concept browser. Technical Report CID-53, KTH Stockholm, 1999.

[NPN02] M. Nilsson, M. Palmer, and A. Naeve. Semantic web meta-data for e-learning - some architectural guidelines. In *Proceedings of the 11th World Wide Web Conference (WWW2002)*, Honolulu, Hawaii, May 2002. (Alternate Track), http://kmr.nada.kth.se/papers/SemanticWeb/p744-nilsson.pdf.

[NSST02] Wolfgang Nejdl, Wolf Siberski, Bernd Simon, and Julien Tane. Towards a modification exchange language for distributed rdf repositories. In *Proceedings Of International Semantic Web Conference*, 2002.

[NW99] Wolfgang Nejdl and Martin Wolpers. KBS Hyperbook – a data-driven information system on the web. In *8th International World Wide Web Conference*, Toronto, Canada, May 1999.

[NWC00] Wolfgang Nejdl, Martin Wolpes, and Christian Capelle. The RDF schema specification revisited. In *Modellierung 2000*, St. Goar, Germany, April 2000. http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/2000/modeling2000/wolpers.pdf.

[NWQ+02] Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjörn Naeve, Mikael Nilsson, Matthias Palmér, and Tore Risch. Edutella: A p2p networking infrastructure based on rdf. In *Proceedings of the 11th International World Wide Web Conference (WWW2002)*, Hawaii, USA, 2002.

[NWST02] Wolfgang Nejdl, Boris Wolf, Steffen Staab, and Julien Tane. Edutella: Searching and annotating resources within an rdf-based p2p network. In *Proceedings of the Semantic Web Workshop*, Honolulu, Hawaii, 2002.

[Pai02] Mark Painter. Padlr submodule modular content archives, first technical report. Technical report, TU Braunschweig, January 2002.

[Pet00] D. Pettersson. Aspect filtering as a tool to support conceptual exploration and presentation. Technical Report TRITA-NA-E0079, KTH Stockholm, December 2000.

[PNN01] M. Palmér, A. Naeve, and M. Nilsson. E-learning in the semantic age. In *Proceedings of the 2nd European Web-based Learning Environments Conference*, Lund, Sweden, October 2001. http://kmr.nada.kth.se/papers/SemanticWeb/e-Learning-in-The-SA.pdf.

[QN02a] Changtao Qu and Wolfgang Nejdl. Towards interoperability and reusability of learning resource: a scorm-conformant courseware for computer science education. In *Proc. of 2nd IEEE International Conference on Advanced Learning Technologies (IEEE ICALT 2002)*, Kaza, Russia, September 2002. IEEE Computer Society Press.

[QN02b] Changtao Qu and Wolfgang Nejdl. Towards open standards: the evolution of an XML/JSP/WebDAV based collaborative courseware generating system. In *Proc. of the 1st International Conference on Web-based Learning (as part of the 28th International Conference on Very Large Databases (VLDB'2002))*, Hongkong, August 2002. Springer LNCS.

[RA02] Christoph Richter and Heidrun Allert. Bericht über die interviews zum olr-2 und ki-skript - kurzfassung. Technical report, University of Hannover, April 2002.

[RJ01] T. Risch and V. Josifovski. Distributed data integration by object-oriented mediator servers. In *Concurrency - Practice and Experience J.* John Wiley & Sons, 2001. http://www.dis.uu.se/ udbl/publ/concur00.pdf, to be published.

[RTK99] F. Ruth, J. Tholander, and K. Karlgren. Trusting the tutor - design aspects and trust issues in a prototypical pedagogical assistant. In *7th International Conference on Computers in Education*, Chiba, Japan, November 1999.

[SD02a] Michael Sintek and Stefan Decker. TRIPLE—A Query, Inference, and Transformation Language for the Semantic Web. In *1st International Semantic Web Conference (ISWC)*, Sardinia, Italia, June 2002.

[SD02b] Michael Sintek and Stefan Decker. Using TRIPLE for Business Agents on the Semantic Web. In *AI-2002 Workshop on Business Agents and the Semantic Web (BASeWEB)*, Calgary, Alberta, Canada, May 2002.

[SS00] Anne Morgan Spalter and Rosemary Michelle Simpson. Reusable hypertext structures for distance and JIT learning. In *Hypertext 2000*, pages 29–38, 2000.

[SS02] L. Stojanovic and N. Stojanovic. Evolution in ontology-based knowledge management systems. In *ECIS 2002 - The European Conference on Information Systems*, Gdansk, Poland, 2002.

[SSDN02a] Mario Schlosser, Michael Sintek, Stefan Decker, and Wolfgang Nejdl. HyperCuP—Hypercubes, Ontologies and Efficient Search on P2P Networks. In *International Workshop on Agents and Peer-to-Peer Computing (AP2PC 2002)*, Bologna, Italy, July 2002.

[SSDN02b] Mario Schlosser, Michael Sintek, Stefan Decker, and Wolfgang Nejdl. HyperCuP—Shaping Up Peer-to-Peer Networks. In *Submitted to: 16th International Symposium on Distributed Computing (DISC 2002)*, Toulouse, France, October 2002.

[SSDN02c] Mario Schlosser, Michael Sintek, Stefan Decker, and Wolfgang Nejdl. Ontology-Based Search and Broadcast in HyperCuP. Poster at 1st International Semantic Web Conference (ISWC), June 2002.

[SW00] Gerd Stumme and Rudolf Wille, editors. *Begriffliche Wissensverarbeitung — Methoden und Anwendungen*. Springer, Heidelberg, 2000.

[TKRR99] J. Tholander, K. Karlgren, F. Rutz, and R. Ramberg. Design and evaluation of an apprenticeship setting for learning object-oriented modeling. In *7th International Conference on Computers in Education*, Chiba, Japan, November 1999.

[TM02] Kristina Toutanova and Christopher D. Manning. Feature selection for a rich HPSG grammar using decision trees. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, 2002.

[TRJ+99] J. Tholander, F. Rutz, P. Johannesson, K. Karlgren, and R. Ramberg. A pedagogical assistant for learning object-oriented design: Nagging students into self-reflection. In *Proceedings of the 10th International PEG Conference*, Exeter, England, July 1999.

[TRV98] A. Tomasic, L. Raschid, and P. Valduriez. Scaling access to heterogeneous data sources with DISCO. *IEEE Transactions on Knowledge and Data Engineering*, 10(5):808–823, 1998.

[VT99a] V.Josifovski and T.Risch. Functional query optimization over object-oriented views for data integration. *Intelligent Information Systems (JIIS)*, 12(2–3):165–190, 1999.

[VT99b] V.Josifovski and T.Risch. Integrating heterogeneous overlapping databases through object-oriented transformations. In *Proc. 25th Conf. on Very Large Databases (VLDB'99)*, pages 435–446, Edinburg, September 1999.

[WBN02] Martin Wolpers, Ingo Brunkhorst, and Wolfgang Nejdl. Using an O-Telos peer to provide reasoning capabilities in an rdf-based P2P-environment. In *Proceedings of the International Workshop on Agents and Peer-to-Peer Computing*, Bologna, Italy, July 2002.

A Biographies

Dr. Lars Borin. Lars Borin has his academic background in Languages (Slavic and Finno-Ugric), Eastern European Studies, and Computational Linguistics. He has a Ph.D. (1992) in Computational Linguistics from Uppsala University. Currently, he divides his time between Uppsala University and Stockholm University, teaching Computational Linguistics and doing research and publishing on the use of Natural Language Processing techniques in IT-supported learning in the Humanities, in particular Computer-Assisted Language Learning, and Corpus Linguistics (especially learner corpora).

Prof. Donald Broady. Donald Broady is Professor at ILU, Uppsala University. He directs the research unit Sociology of Education and Culture and the research program Digital Literature. Director of Uppsala Learning Lab. He is research fellow at IPLab (Interaction and Presentation Laboratory) and CID (Centre for User Oriented IT-Design), Nada (Department of Numerical Analysis and Computing Science), KTH (Royal Institute of Technology), Stockholm.

Prof. Stefan Fischer. Stefan Fischer is a Professor for Computer Science, especially Distributed Systems, at the Technical University of Braunschweig, Germany. Before he joined TU BS in 2001, he was with the University Mannheim, where he received his PhD in 1996, with the University of Montreal as a postdoc, and with the International University, where he was assistant professor for information technology. His research interests currently focus on eLearning and on applications in mobile environments. He is a member of ACM, IEEE Computer Society and Gesellschaft fur Informatik. Together with Wolfgang Nejdl, he heads the ELAN-Pilot Hannover/Braunschweig in the context of the ELearning Academic Network Lower Saxony.

Prof. Christian Floto. 1956 Born in Lübeck, Germany. Studied medicine at the University of Kiel and the Medical University Lübeck; 1991 Approbation (full medical license) granted; 1992 Doctorate: Dr. med. 1977-81 Teacher at a vocational school and a high school. 1982-84 Research fellow at the Institut für medizinische und pharmazeutische Prüfungsfragen, Mainz; main focus on testing theory and computerized exams in medical training. 1984-86 Medical business manager for the Hessian State Medical Board of Registration. 1986-92 Akademischer Rat in Theoretical Medicine at the University of Osnabrück in the interdisciplinary course of studies "Health Sciences." Areas of concentration: medical education, preventive medicine / epidemiology, medical/science journalism. Author, reporter, and program moderator for numerous programs for the ARD national radio station. 1992 Editor for the ZDF national television station; 1993-2001 Director of the ZDF department "Gesundheit und Natur" (health and nature); Other activities included: Directing and moderating the world's oldest television magazine program, "PRAXIS- das Gesundheitsmagazin" (Practice - the magazine for health); Planning and developing of multi-media materials to accompany the program; ZDF representative on various national and international boards (e.g., the European Union's "Health Broadcasting"). March 2001-current Professor for Media Application in Science, Department of Media Sciences at the Technical University (TU) of Brunschweig and Director of the IWF Wissen und Medien gGmbH (IWF Knowledge and Media, gGmbH)

Prof. Carl Gustaf Jansson. Carl Gustaf Jansson is Professor of Artificial Intelligence at KTH. He received his PhD in Technical Computer Science at KTH in 1986 and has been a permanent faculty member at KTH since 1987. He is currently director for the Knowledge and Communication laboratory (at the department for Computer and Systems Sciences) which he founded in 1996. Current administrative tasks at KTH, include member of the KTH board, member of the board for the school of Information Technology and Deputy Head of dept. He is also academic co-founder of the center Wireless@KTH. Jansson has worked extensively with curricula development in artificial intelligence, human machine interaction and programming methodology and is currently program chairman for the KTH Information Technology Master of Science program. In the early ninetees he cofounded the center for IT and Cognitive science in cooperation between the three Stockholm Universities: SU, KTH and KI. He was a senior scientist in the ESF program on Learning in Humans and Machines. He has published papers and been responsible for both national and EU projects in areas such as knowledge representation, machine learning, learning environments, intelligent interfaces and ubiquitous computing. He has supervised more than 20 PhD theses, and has in the last five years been active as a senior supervisor in the scientific commitees in two national swedish graduate school, one in human-machine

interaction (HMI) and one in telecommunications (PCC). Current research interests include ubiquitous computing and applications in learning. Has been PI in initial WGLN projects: PADLR (APE) and iSpace (DILS).

Prof. Klaus Jobmann. Klaus Jobmann is Professor at the University of Hanover. He is a governmental advisor for Niedersachsen in telecommunications, advisor of the Industrie- und Handelskammer Offenbach for charging in public telecommunication networks as well as advisor of AOL Germany regarding telecommunications. His teaches courses in telecommunication networks (ISDN and Broadband-ISDN, Mobile Networks, Public Data Communication Networks, Traffic Theory and Performance issues). His main fields of research are resource allocations and management of mobile networks, where he heads various third party funded projects as principal investigator. His experience in telecommunication networks is based on a long period of service in the telecommunication industry in the fields of development, network design, and manufacturing.

Prof. Christopher Manning. Christopher Manning is an Assistant Professor of Computer Science and Linguistics at Stanford University. He received his Ph.D. from Stanford University in 1995, and served on the faculty of the Computational Linguistics Program at Carnegie Mellon University and the Linguistics Department at the University of Sydney before returning to Stanford. His research interests include probabilistic models of language, statistical natural language processing, constraint-based linguistic theories, syntactic typology, information extraction, text mining, and computational lexicography. He is the author of three books, including *Foundations of Statistical Natural Language Processing* (MIT Press, 1999, with Hinrich Schütze, [MS99]). His recent work includes unsupervised grammar induction [KM02], parse selection models [TM02], clustering with constraints [KKM02], and word sense disambiguation [KTI+02].

Dr. Ambjörn Naeve. Ambjörn Naeve received a Ph.D. in Computer Science in 1993 from the Royal Institute of Technology (KTH) in Stockholm. He has a strong background in mathematics and has been an early advocate of using projective geometry in computational vision. Ambjörn Naeve has designed a learner-centric knowledge-pulling educational architecture called a Knowledge Manifold[21], and he is the inventor of the concept browser Conzilla[22], which is a navigation and presentation tool for such manifolds. Since 1996 Ambjörn Naeve works as a senior researcher at the Centre for user-oriented Information Technology Design (CID) at KTH, where he is heading the Knowledge Management Research group[23]. The KMR-group is basically concerned with developing principles and tools for the efficient design and implementation of interactive learning environments based on semantic web techniques. The group also specializes in developing new paradigms and tools for mathematics education with a focus on rehabilitation of learners' confidence in their own thinking abilities. Another focus area of the KMR-group is how to support trust building in a shared electronically mediated environment.

Prof. Wolfgang Nejdl. Wolfgang Nejdl has been full professor of computer science at the University of Hannover since 1995, after being associate professor of computer science at the RWTH Aachen (1992-1995). In 1988, he has been visiting researcher at Stanford University, in 1992 to Xerox PARC, 2001 and 2002 again at Stanford University (where he currently spends a half year sabbatical at the CS Department). The Institute for Information Systems / Knowledge Based Systems does research in the areas of artificial intelligence, adaptive hypermedia systems, as well as metadata based learning repositories and peer-to-peer systems. He is member of two ESPRIT Networks of Excellence in the area of artificial intelligence. Prof. Nejdl is academic and administrative director and founding member of the Learning Lab Lower Saxony (L3S)[24]. Together with Stefan Fischer, he heads the ELAN-Pilot Hannover/Braunschweig in the context of the ELearning Academic Network Lower Saxony.

Prof. Helmut Pralle. Helmut Pralle, University Professor, Dr.-Ing., studied engineering sciences at the Technische Hochschule Karlsruhe and Hannover. Afterwards he started work as assistent at the Institute fuer Praktische Mathematik /Lehrstuhl fuer Elektronische Rechenanlagen and at the computer centre/center of the Technical University of

[21] http://kmr.nada.kth.se/km/index.html
[22] http://www.conzilla.org
[23] http://kmr.nada.kth.se/
[24] http://www.learninglab.de/

Hannover. 1967 gaining of doctorate of Regelungstheorie, 1968 head of the university computer centre/center, 1971 director of the Regionales Rechenzentrum fuer Niedersachsen (RRZN)/Regional Computer Centre of Lower Saxony. Since 1986 Chair of Rechnernetze und Verteilte Systeme (RVS) at the faculty Elektrotechnik/Technische Informatik at the University of Hannover. Till 1996 member of the administrative council and the executive committee of the DFN (Deutsches Forschungsnetz). Expert reports and vocational guidance for Deutsche Forschungsgemeinschaft (DFG), ministries, industry. Area of responsibilities besides others are distributed information systems and high bandwidth networks.

Prof. Bernd Rebe. Bernd Rebe was born at September 5th, 1939 in Braunlage / Harz. He studied law in Kiel and in Berlin and made his first and second state examination in Berlin also. He did his Ph.D. there in 1969 and habilitated in 1977 in Bielefeld. He got the venia legendi for civil law, trade law and commercial law. From 1975 until 1983 he was professor for civil law and commercial law at the university of Hannover, from 1979 until 1981 he was the vice president there. Afterwards he was the president of the technical university of Braunschweig, where he is professor at the institute for social sciences now.

Prof. Ulrich Reimers. Ulrich Reimers studied communication engineering at Technische Universität Braunschweig, Germany. Following research at the University's Institute for Communications Technology he joined BTS Broadcast Television Systems in Darmstadt. Between 1989 and 1993 he was Technical Director of Norddeutscher Rundfunk (NDR) in Hamburg. Since 1993 he has been a Professor at Technische Universität Braunschweig and Managing Director of the Institute for Communications Technology. Prof. Reimers is chairman of the Technical Module within the DVB Project and president of Fernseh- und Kinotechnische Gesellschaft (FKTG) (the German society of engineers in the field of the electronic media). He is a board member of Deutsche TV-Plattform (the German institution coordinating the interests of all organisations involved in TV). In 1995 he was awarded the Montreux Achievement Gold Medal for his contributions to the development of the DVB technology. In 1998 he received the IBC John Tucker Award and was made an Honorary Fellow of the Royal Television Society of the United Kingdom. In July 1999 he received the J. J. Thomson Medal of the Institution of Electrical Engineers (IEE). In June 2000 he was awarded the 1st class Cross of Merit of the Lower Saxony Order of Merit. In October 2000 he was awarded a "Diploma of Honour" of the National Association of TV and Radio Broadcasters (HAT) of Russia. In June 2001 he was the recipient of the Leibniz Ring and of the IEEE Masaru Ibuka Consumer Electronics Award.

Prof. Tore Risch. Dr. Tore Risch is Professor of Database Technology at Uppsala University (Sweden) where he leads the Uppsala DataBase Laboratory (UDBL) research group[25]. He was previously Professor at Linköping University (Sweden). Before Linköping he was staff member in the Database Technology Department at Hewlett-Packard Laboratories (Palo Alto, California), and Visiting Scholar from HP at Stanford University. Prior to joining HP, he worked for Syntelligence Inc. (Sunnyvale, California) on large scale knowledge bases combining AI and database technologies. He also worked on the Prospector expert system (SRI, Menlo Park, California), on integrating Prolog with relational databases (Uppsala U., Sweden), and at IBM's Almaden Center (San Jose, California) on functional knowledge representation. He made his PhD 1978 at Linköping University (Sweden), on query optimization in a meta-database system.

Prof. Rudi Studer Rudi Studer is Full Professor in Applied Informatics at the University of Karlsruhe, Institute AIFB[26]. His research interests include knowledge management, Semantic Web technologies and applications, ontology engineering, knowledge discovery and eLearning. He obtained a Diploma in Computer Science at the University of Stuttgart in 1975. In 1982 he was awarded a Doctor's degree in Mathematics and Computer Science at the University of Stuttgart, and in 1985 he obtained his Habilitation in Computer Science at the University of Stuttgart. From 1977 to 1985 he worked as a research scientist at the University of Stuttgart. From 1985 to 1989 he was project leader and manager at the Scientific Center of IBM Germany.

[25] http://www.dis.uu.se/~udbl
[26] http://www.aifb.uni-karlsruhe.de/WBS

Rudi Studer is also director of the Knowledge Management Group at the FZI Research Center for Information Technologies[27] at the University of Karlsruhe, a member of the L3S Learning Lab[28] in Hannover as well as co-founder of the spin-off company ontoprise GmbH[29] that develops semantic technologies.

He is engaged in various national and international cooperation projects being funded by e.g. the European Commission, DARPA or industry. He is a member of AAAI, ACM, IEEE, IFIP Working Groups on Databases (WG 2.6) and on Knowledge-oriented Development of Applications (WG 12.5), and German Informatics Society (GI).

[27] http://www.fzi.de/wim/
[28] http://www.learninglab.de/
[29] http://www.ontoprise.de/