MAKING A SPLASH: A HETEREGENEOUS PEER-TO-PEER LEARNING OBJECT REPOSITORY

Marek Hatala
Simon Fraser University Surrey
2400 Central City
Surrey, BC, Canada, V3T 2W1
+1 (604) 586-6053
mhatala@sfu.ca

Griff Richards
Simon Fraser University Surrey
2400 Central City
Surrey, BC, Canada, V3T 2W1
+1 (604) 586-6062
griff@sfu.ca

ABSTRACT
In this paper, we describe the heterogeneous peer-to-peer infrastructure for sharing learning object. The POOL project builds on the three types of nodes: SPLASH is a freely downloadable application which allows individuals to create metadata and maintain their collection of learning objects, PONDs are bigger repositories of learning objects connected to the peer-to-peer network and POOL Centrals increase the speed and breadth of the searches in the peer-to-peer network. The POOL project uses CanCore – a subset of the IMS metadata protocol – to describe learning objects. In the second part of the paper we discuss the future direction of this initiative based on the maturing learning objects community and lessons learned in the deployment of POOL network. We argue that the standardization effort, although very important, currently provides solutions that are too complex. We see the communities where the knowledge is shared to be the main force in the creation of the metadata standards which would support the growth of semantic web. Key to our approach is the development of overlapping ontologies which address global issues for discovery, and local issues for precision classification and usage. We propose a general mechanism which would allow non-expert community members to create their own metadata description in the form of application profiles, automatically updating their tools when changes to the ontological schema occur and still allowing for the exchange of metadata with the users outside of the community using a part of the basic IMS/SCORM/Dublin Core based metadata.

Categories and Subject Descriptors
H.3.4 [**Information Systems**]: Systems and Software – *distributed systems, information networks* H.3.3 [**Information Systems**]: Information Search and Retrieval – *clustering, search process,* K.3.1 [**Computers and Education**]: Computer Uses in Education

General Terms
Management, Design, Reliability, Standardization, Languages

Keywords
Learning object repositories, Peer-to-peer systems, Learning objects metadata, Community portals, Ontologies, Semantic web

WWW 2003, May 20-24, 2003, Budapest, Hungary.
ACM xxx.

1. INTRODUCTION
With a growing number of organizations moving their training and education programs into the web environment, there is an increasing demand for high-quality, reusable components – learning objects [19]. This demand comes from the realization that the development of learning objects is resource intensive and time consuming. The learning object is a definable, reusable chunk of digital content and process elements used for learning and instruction [28].

The huge uptake of web technology in education and training has generated a flurry of un-coordinated activity developing digital learning objects – images, animations, computer applets or textual content which could be used in the processes of education and training. Centralized digital learning object repositories evolved as a means of collecting and cataloguing these assets with hopes of reducing the redundancy of development and enabling others to build on the aggregated ideas and designs, and in many cases to preserve the elements, and protect the rights of ownership and usage.

1.1 Learning Object Metadata
It was immediately recognized that standards are important for interoperability between learning and business systems. Several standards for describing metadata have been developed through collaboration between the private and public sectors. The first result has been a standard for learning object metadata (LOM) developed by the IEEE [8] and standardization of other aspects of e-learning, including packaging of learning resources using IMS Content Packaging [9], or messaging between content and systems using the AICC CMI (at the time of writing this was being formalized as an IEEE standard) and a proposal for the sequencing of learning resources known as IMS Simple Sequencing [11]. The Shareable Content Object Reference Model (SCORM) from the Advanced Distributed Learning Network (ADL) [23] pulls together these standards and specifications and defines how they work together. Standardized metadata descriptions enable content providers to describe different aspects of the learning resources and promote both interoperability and sharing of resources.

However, the business and educational communities have been slow to adopt the full IEEE LOM standard and its IMS implementation mainly due to the high number of the fields (over 80) and vagueness with which the values for these fields have been defined. Too much information results in too much time spent cataloging that no one will bother. On the other hand, too little information in the tagging will result in too many false positive results. The alternate standard, the Dublin Core [4] protocol identifies only 15 fields.

The IEEE LTSC LOM is intended to be comprehensive and to cover most situations in which metadata are applied to learning resources. In actual implementations it is often useful to both constrain and enrich the standard. The standard is constrained by specifying which of the metadata elements in the standard are required and by limiting the possible values of these elements. It can be enriched by applying taxonomies or ontologies to each of these elements. We refer to such implementations as metadata profiles [4] (for example, Cancore is a metadata profile for sharing learning resources based on IMS implementation of IEEE LOM).

The Canadian Core Learning Resource Metadata Protocol (CanCore) [5] has been defined to specifically address these problems. CanCore was developed in Phase I of the POOL project (described below) by the collaboration of Canadian researchers searching for a level of sufficient specificity to enable the efficient search of learning objects. CanCore is based on IMS implementation of IEEE LOM and identifies a sufficient number of fields (36) to be useful for educators, without overburdening the indexing process. CanCore has sufficient flexibility in its protocol that not all fields need be completed, thus developers can ignore many fields that may be inappropriate for their purposes. CanCore is fully compliant with the IMS metatagging specification. As IMS matures, additional development will be required of CanCore. CanCore elements are organized into 9 groups describing different characteristics of the learning object. Detail information can be found in publications available from the CanCore website (http://www.cancore.org).

1.2 Centralized vs Decentralized Learning Object Repositories

Centralized digital repositories evolved as a means of sharing resources for collecting, cataloguing and storing objects of a defined community. In addition to centralized control, a centralized repository offers advantages in rapid indexing and object retrieval. Unfortunately, a single centralized repository is unlikely to be of sufficient size to accommodate all of the web-based learning objects that have or will be created.

Secondly, there can be workflow disadvantages to the centralized repository as the objects are stored away from their point of origin and away from their point of use. Users have to be connected to the web for even the simplest operation, and off-line creation or modifications of learning objects are not captured until the object and its metadata are re-loaded. We believe the optimal storage sites for learning objects are *close to the creator and close to the user*. Further, as workstation storage increases, it becomes feasible for each learner to amass a personal collection of the learning objects which have influenced their intellectual growth, and to be annotated for future reference and review, much in the manner that study notes enabled classroom learners to keep track of significant content and conceptualizations.

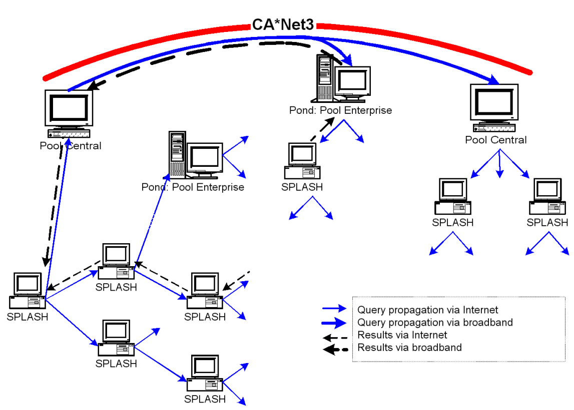

Figure 1. POOL network architecture

Recent developments in peer-to-peer web technology have made it possible for individuals to amass local collections of entertainment content. Although Napster and Gnutella may have been lacking support for rights management, the peer-to-peer model demonstrated that a global community can benefit from decentralized storage of content on the users' own hard drives. For learning objects this means that individual instructors, if provided with the standard metadata and communication protocol, can develop and store their materials so that others may directly search and access their public materials, or become aware of semi-public materials which the individual may wish to negotiate consideration for use. Individuals may also store private materials that are under development, or are not intended for mass consumption.

Several educational projects found the peer-to-peer platform appealing. Two projects exploiting peer-to-peer technology for learning objects are currently under development. Edutella [17] is a prototype peer-to-peer network which builds in a structured query service to help locate learning objects, an annotation service to allow users to comment on learning objects, and a mediation service to join metadata from different sources. In comparison, the POOL project which is the focus of this paper concentrates on the heterogeneous infrastructure and end-user tools utilizing CanCore standard to connect individual and organizational repositories.

2. POOL: HETEROGENEOUS PEER-TO-PEER BASED LEARNING OBJECT REPOSITORY INFRASTRUCTURE

The Portal for Online Objects in Learning (POOL) Project (http://www.edusplash.net) is a consortium of several educational, private and public sector organizations to develop an infrastructure for learning object repositories. It is one of several projects funded in part by Canarie's E-Learning Program – Canadian initiatives to build a national infrastructure for collections of high quality learning objects and related business models [12]. The Phase I of the POOL project ended in June 2001 with two major outcomes: the CanCore protocol and the POOL centralized repository prototype. The lessons learned from the evaluation of the prototype helped us to formulate requirements for the Phase II of the project which ended in September 2002. In this section we describe the redesigned POOL infrastructure – a hierarchical network of nodes communicating via peer-to-peer protocol using CanCore as a core metadata exchange schema.

2.1 POOL Architecture

Learning objects are developed mostly by individuals either for their individual purposes or for their organization's needs. Typically, the learning object evolves during its lifetime as it is getting feedback from its intended usage or is redeployed in new instructional contexts. This evolution is possible through a persistent stewardship that exists throughout the object's lifetime. This stewardship may frequently change as interest in using the object shifts from one person or community to another.

Table 1. POOL Network nodes functionality

	SPLASH	POND	POOL Central
Create/edit metadata record	+++	+	
View metadata records	+++	+	
Search for metadata records locally	++	+++	+
Search for metadata in the POOL network	++	++	+++
Respond to the search request from another peer	++	+++	+
Propagate search query and return collected results	++	+	+++
Robust database support	+	+++	
Management and workflow tools		+++	

+ supported ++ main functionality +++specialized

To support the evolutionary nature of learning objects we designed POOL as a network of individual peers communicating together using the POOL protocol (Figure 1). Three types of peers participate in the network: SPLASH, POND, and POOL Central. The names evolved from the original POOL acronym from Phase I for the centralized repository model. It is our hope that the names of nodes represent their relative size, purpose and persistence level, all linked together by the 'water' analogy.

SPLASH is a desktop client communicating with other peers via the peer-to-peer POOL protocol. It provides the metadata creation tools, a limited storage capacity for metadata records and searching capability for the POOL network. SPLASH is developed using open source code, and distributed freely in the belief that thousands of small repositories held by learners and instructors will create a wide acceptance and use of both learning object technology and the CanCore protocol.

The wide distribution of SPLASH will not obviate the existence of community repositories. There is a role to be played for established collections of mature, accepted learning objects with common themes or purpose that can be stored in a selective gallery of learning objects. Indeed, an advantage of the SPLASH is that such galleries or PONDS can be set up with ease. Within the project we have tested the concept by incorporating several community repositories to create PONDs – repositories that are accessible using the POOL protocol and searchable using the CanCore metadata standard. A POND may be simply a larger, community implementation of SPLASH, or it may involve building an interface to a third party repository system. The ability to include such proprietary systems is expected to be an advantage over a single centralized pool, and will hopefully enable organizations already committed to a particular repository technology to contribute their content to the larger POOL movement. POND typically comes with a robust database support and a suite of tools for managing the learning objects workflow. These features are essential for organizations with intensive production of learning objects.

POOL Central is a specialized peer connected to the network and a broadband Internet. The purpose of the POOL Central is to replicate the queries through the other POOL Central peers over the broadband connection and enhancing the reach of the network. POOL Central does not necessarily have a storage capacity, although caching of records might be possible.

Also, within the proposed hybrid architecture we can see a role to be played by specialized nodes. An example of such a node is the LORI (Learning Object Review Instrument), which can be embedded in SPLASH, and link reviewers to specialized nodes for learning object collaborative assessment (LOCA). A prototype of this is incorporated into the POOL network [26]. The appeal of specialized nodes is that they enable any user or interest group to add intrinsic value to the network without the need for centralized planning, resources or control.

Table 1 illustrates how the network functionality is spread over the network nodes.

2.2 POOL Protocol and Metadata Exchange

The POOL networking component builds on JXTA (http://www.jxta.org), the publicly available peer-to-peer platform from Sun Microsystems Inc. JXTA provides basic protocols for peer discovery, sending messages, obtaining information, routing and group membership. These protocols are low-level protocols and it is up to the developers to implement the content part of the messages being exchanged. Although POOL does not build on the JXTA Search application (http://search.jxta.org) it builds on its Query Routing Protocol that specifies message types, message formats, and message routing rules that must be supported. The POOL protocol expands the JXTA Search protocol by building in more control for distributed searches (Section 2.5) and provides for flexibility in metadata schemas used for queries and responses.

Figure 2 shows a template for the query request. The query parameters specify both format of the query and response as well

```
<request xmlns=http://www.edusplash.net
   query-space=[(required)unique URL id for query space]
   query-format={XML|RDF|SPLASH|EDUTELLA}"
   response-space=[unique URL ids for response space]
   response-format={XML|RDF|SPLASH|EDUTELLA}
   query-uuid=[globally unique id of this query]
   query-lifetime=[number of milliseconds this query is valid]
   max-depth=[maximum number of peers to hop]
   max-fanout=[maximum number of peers to forward the query to]
   max-hits-per-provider=[return only n results from each peer]
   max-results=[maximum number of collected results peer sends back in one flush]
   flush-after-providers=[flush the output stream to the client after receiving
responses from n peers]
   flush-after-ms=[flush the output stream to the client after this time]
>
   [query specification (arbitrary valid XML)]

</request>
```

Figure 2. Query specification in POOL protocol

as parameters of the distributed search. The query space defines the metadata schema which is used to specify the query (with current possible values representing CanCore and IMS). The query-format attribute specifies the binding for the schema specified in the query space. Currently we have implemented formats are native XML-based SPLASH format, we plan to provide generic binding for XML, RDF, and specific binding to the Edutella set of query languages within the eduSource project described in Section 3.1. The same applies to the response space and response format fields. Being able to specify formats and spaces for queries and responses separately gives us an ability to work with the metadata in different formats.

Returning results in the pre-specified format requires translation of the records from the “native” format in which they are stored in the repository to the specified format for transmission. We are using XSLT technology to transform between different schemas. Of course, such transformations may cause some loss of information when a direct mapping between schema elements does not exist.

2.3 Modeling Metadata

As might be deduced from the previous section in the design of the SPLASH, we are considering an option of handling of more than one format of the metadata represented by different schemas and exchanged using different formats. In SPLASH we model metadata at the element levels with the full information about the type of data the metadata element can hold (e.g. free text value, defined vocabulary, etc.) and how it is rendered on the entry form and search form screens. In SPLASH we provide a tool for definition of new elements.

The metadata schema is defined as a collection of the elements. Even standard schemas such as CanCore or IMS are defined in this way. This enables us to treat each schema as a real core and create a community tailored schemas around this core.

The third component of our metadata creation tool is the user’s ability to control how many elements from the particular schema are displayed in the forms and views. The user has an ability to define profiles using the profile editor (Figure 3). A set of default profiles is shipped with the SPLASH reflecting various roles the user can take in the metadata creation process, e.g. educator, learner, editor, media developer, license specialist, etc. Several types of profiles for metadata creation form are supported. First type preferred mainly by the professionals familiar with the metadata reflects semantic groupings of element as it is defined in the standards. Second type of profiles is preferred mainly by the naïve users and organizes element into levels of relative importance or detail. The elements in the first level are those considered to be mandatory for the minimal valid record and those are the only ones displayed in the new form. The second level contains the more specific but still common elements; third and subsequent levels contain more specific but less used elements. The second and subsequent levels are hidden to reduce the mental load of the user and the user has to explicitly choose to display them.

Finally, the fourth feature of the metadata creation support in SPLASH is directed toward users creating high volumes of metadata. In such cases, the values in many fields as creator, educational level, etc., are the same and records differ only in some content specific fields. To speed up the metadata creation process the user can store any (partial) record as a template and later on can use this template as a starting point for metadata creation process. The template editor is provided to support the template management.

2.4 Discovering Peers

POOL is designed as a network of independent peers that are both the providers and consumers of learning objects. Discovery and communication with the other peers is the key to the sharing of learning objects.

The JXTA platform provides the basic discovery functionality including a mechanism for crossing firewalls. An interesting feature we take full benefit from is JXTA’s concept of rendezvous nodes. The rendezvous node is a specialized node that collects the list of other peers and provides this list to peers to speed up the discovery process. We have developed our POOL Central node as a rendezvous node positioned on the high speed Internet (40GB Ca*NET3).

In addition to the peer discovery we provide users with the utility to store information about the favorite peers (e.g. based on the previously obtained results) and organize them into the groups. The user can direct queries directly to selected peers or groups of peers.

2.5 Distributed Search

The search in the POOL network is a combination of the distributed peer-to-peer search and a deep search similar to the one in the JXTA Search application. The peer-to-peer search provides for the breath of the search by broadcasting the query to the neighbor peers in the network. There are several parameters controlling the scope of the search which are shown in Figure 2. The deep part of the search occurs when the query reaches the POND built on top of an existing repository. In such case the local specific search algorithm is invoked and the

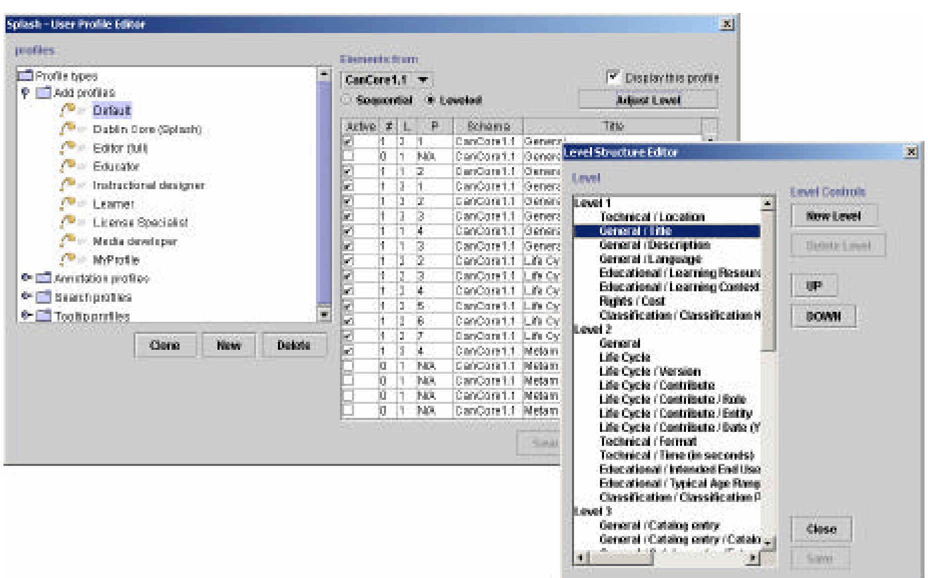

Figure 3. Profile editor for creation of metadata entry forms with the tool for organizing elements into different levels.

results are passed back to the POOL network.

When the search request is received by each individual SPLASH peer a local search is invoked. Local search results are combined with the results received from the peers the query was forwarded to. The growing chain of results is eventually passed back to the originating node. The local search is a combination of four different approaches:

Text search for the text fields. The metadata record contains several fields which enable the creator of the record to enter a free text. Examples of this type of the field might be *general.title* or *general.description* elements. To search for the records with the specific values stored in these fields is possible using a full-text search.

Text search in the vocabularies. Some fields in the schema can be filled only with values selected from the predefined vocabularies. For example, the *educational.intendedenduser* element can only have values from the vocabulary *[Teacher, Author, Learner, Manager]*. In the most cases, these values are proper natural language words and therefore they should be searchable by the full text search.

Value-matching search in the vocabularies. This type of search applies to the fields with values selected from vocabularies. In this case the user specifies exact values for specific elements the record should satisfy.

Taxonomy-based search in the hierarchical vocabularies. This is a third type of the search applicable that applies to the fields with values selected from vocabularies. Some vocabularies can have values organized in the taxonomy. For example, the *pool.mediatype* element has vocabulary consisting of values *[Text, Text.Correspondence, Text.Correspondence.Discussion, Text.Correspondence.email, etc.]*. The taxonomy-based search uses this information to find relevant records. For example, when searching for the record with the value *Text.Correspondence*, the record marked with *Text.Correspondence.Discussion* should be retrieved as it represents a more specific value of *Text.Correspondence*. Currently we use simple string parsing algorithm relying on the 'dot-notation' which we plan to replace with the full ontology search.

The results from each type of search are ranked and multiplied with the coefficient representing the relative importance of the search type. Results from four types of searches are combined and a cumulative rank value is computed for each record. Because the number of results returned from each peer is limited, only the specified number of the best results is returned from each peer. The cumulative rank value is used again when merging local results with those coming from the network.

2.6 Implementation and Deployment

Although we have finished the technical development activities within POOL project we are still working on the improving the technical solutions. Our main focus now is on deploying the solution by supporting a creation of the community repositories both by connecting existing repositories into the POOL network as well as working with the repository solution providers to make their products pluggable into the network.

SPLASH. The beta version of the SPLASH desktop application is available for the public download free of charge as of February 2002[1]. The version available implements all the functionality listed in the Table 1, it includes a tagging engine enabling the user to create metadata records using CanCore and search engine searching through POOL network. SPLASH is designed with nearly all components designed to be customizable by the user which makes it easy to tailor it for the individual user needs or for other metadata schemas.

Technically, SPLASH is a Java application running on the user's desktop computer. The current version comes with two database options: hsql a small pure java-based database or mySQL. The database engine is replaceable by another SQL database. Figure 4 shows a snapshot of the SPLASH search interface.

PONDS. We have implemented and deployed several PONDS to test different ways how to build or incorporate large repositories going beyond desktop level. The repository at the Center for Curriculum, Transfer and technology in British Columbia has evolved from the SPLASH by dedicating one SPLASH application to play a role of the common repository for the community. Individual community members run they own SPLASH but have a choice of storing the metadata records either at the centralized repository or locally.

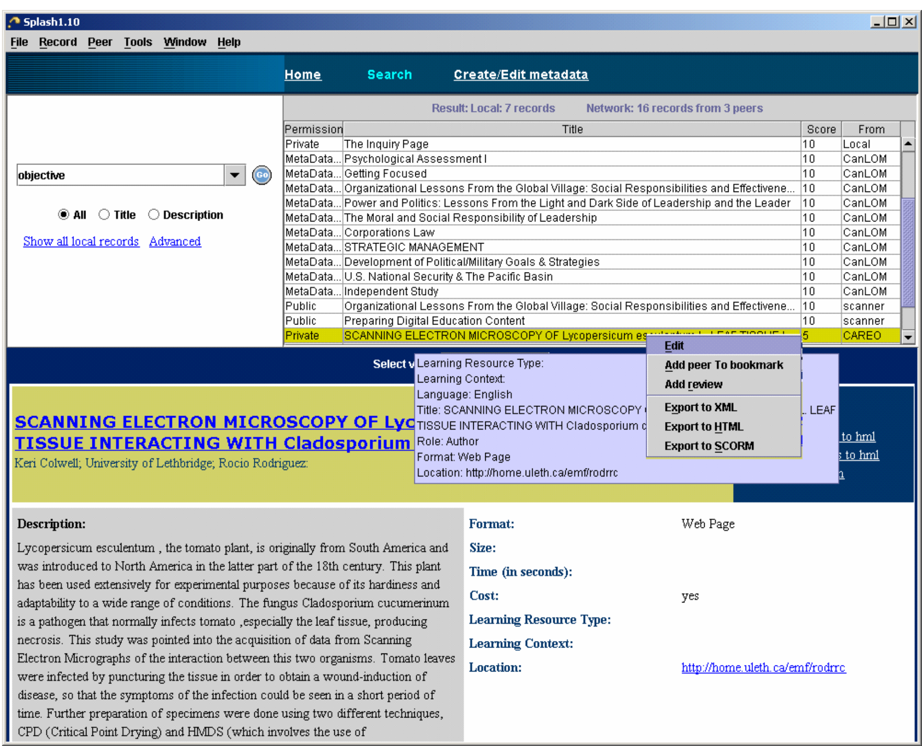

Figure 4. SPLASH search interface

---

1 At the time of writing we have registered over 500 downloads of the SPLASH.

The same mechanism has been used for the Canadian Learning Objects Metadata Repository (CanLOM). CanLOM has been built on top of the existing TeleCampus (http://www.telecampus.com) database of over 50,000 learning objects which uses slightly different metadata schema. In the CanLOM case, the repository was a pre-existing system where we built a wrapper to connect it to the POOL network. The wrapper is a networking module from SPLASH connected to the TeleCampus system that is built on Oracle database and ColdFusion. CanLOM functions as a specialized node to expedite search of "published" learning objects. It also implements a registry protocol for the learning objects. The SPLASH user can decide to register and upload metadata to CanLOM. As copies of the metadata record spread over the peer-to-peer network the SPLASH application keeping a copy of the record can check with the CanLOM whether a new version of the record has been uploaded. In this way we provide for the notification of the updates within the peer-to-peer network.

In the third case, we have connected the CAREO [14] repository that uses Web services. In this case, we have implemented a simple component translating between the POOL protocol and Web services.

Especially interesting is the POND adaptation of the Australian AVIRE repository [22] so it can be searched from POOL. This work will not only give SPLASH users access to an extensive collection of architectural learning objects, but also provide international exposure for POOL and CanCore. As AVIRE uses different metadata schema, a communication layer between SPLASH and AVIRE maps CanCore metadata fields to AVIRE metadata fields, and return search requests coming from SPLASH.

POOL CENTRALs. The purpose of the POOL Central is to provide the network with the specialized nodes which function as reflectors broadcasting queries into the 'distant' areas on the network. Their efficiency depends on the speed of the connection between them. In Canada, most of the inter-university traffic is routed via Canarie's Ca*Net3 broadband connection, so once a SPLASH node communicates with a peer located on a university network, POOL Central is automatically invoked. So far we have placed POOL Central nodes into four provinces: Calgary in Alberta, Montreal in Quebec, Fredericton in New Brunswick and Vancouver in British Columbia.

3. CURRENT PROJECTS

In this section we briefly describe two follow on projects to the POOL project. The eduSource project described in the next section is a collaborative project funded by Canadian government organization Canarie Inc. under its e-learning initiative [12]. The project aims at the integration of most significant previous initiatives in e-learning in Canada. The eduSource will integrate POOL infrastructure as one of its components. The second project called Talking Objects is an implementation of the POOL approach for the community preserving music and sound collection and using it in the music education in the province of Quebec in Canada.

3.1 eduSource: A pan-Canadian Learning Object Repository

The eduSource project [15] is a pan-Canadian collaborative project to create a testbed of linked and interoperable learning object repositories. The 18 month project which started in October 2002 is providing leadership in the ongoing development of the associated tools, systems, protocols and practices that will support such an infrastructure. The primary delivery mechanism for this testbed will be the broadband Internet, and in particular CA*Net 3/4. This project is based on national and international standards; it is fully bilingual; it will be accessible to all Canadian including those with disabilities; and it will share and disseminate its findings across Canada and internationally.

To simplify its activities, eduSource has six designated primary partners across Canada. The overall project budget is $9.8 million. EduSource takes advantage of, and enhances, the capabilities of each of its members in building a comprehensive LOR implementation infrastructure. NBDEN will work with the NewMIC Foundation to explore the development of Repository Content. NBDEN will also work with TeleEducation NB, the National Research Council, and private sector partners such as RightsMarket Inc. to investigate Digital Rights Management for LOR's. Télé-université with le Centre de recherche LICEF and the Advanced Technologies for Learning Unit of the University of Alberta will provide a testing and evaluation framework for the project. Athabasca University will lead in the ongoing development of the CanCore metadata application profile. Télé-université will coordinate the integration and development of software packages such as Explor@ - II, CAREO (Campus Alberta Repository of Educational Objects), ALOHA (from the Learning Commons at the University of Calgary) and SPLASH (from the POOL project) and lead the development of new software packages. The University of Waterloo will contribute its knowledge of and international collaborations in community building, as well as its expertise developed in the CANARIE Learning Program's PLIANT (Partnerships for Learning, Innovation and Technology) project.

3.2 Talking Objects: Repository of Sound Learning Objects

Our second project is aiming to build a knowledge repository to pool a large collection of specialized items – sound and music recordings. The recordings will become basic components blocks for Talking Objects – a format enabling to build a 30-45 minute audio learning object that can be downloaded and cached for the in-class use at the primary education.

The coordinating organization is concerned that key elements of the distributed collection should be centralized to allow for better maintenance and control. However, much of the collection is held in small private collections or in collections in small community organizations that are reluctant to release control of their collection. The third confounding issue is that although the key elements should be held centrally, there is recognition of need and desire to build follow-on consumer traffic – users who would want to add-value to the key collection through annotations, stories, examples and instances. The concern is that a large exponentially growing collection and the swelling user base would strain technical resources. These factors influenced our decision to blend a small number of community repositories with foundational content built along the pattern of the community portal, with a secondary network of peer-to-peer nodes for the value-added content. However, this decentralization into self-evolving communities poses again the problem of potentially incompatible local metadata schemas and ontologies – somewhat analogous to the devolution of Latin into the romance languages.

4. FUTURE WORK: SUPPORTING LOCAL AND GLOBAL METADATA

Both projects provide a framework for our research in the area of interoperability and support for self-evolving communities. The lessons learned from the application of the SPLASH within the communities let us believe that rather than specifying and encouraging more work on standardized global metadata systems, it may be more appropriate to take stock of a simple schema as CanCore, deeming mandatory a handful of the fields as a global standard, and then encouraging the adoption of the remaining fields from the standard and definition of community based metadata for local operations [6].

Although the solution looks simple, it has unprecedented consequences on the infrastructure, protocols and tools. We will address consequently different aspects of proposed scenario and imply the requirement for the tools supporting them.

Locally defined metadata schemas. The benefit of metadata is obvious when there is a community striving for getting a formalized understanding of different aspects of the (learning) objects. To enable the communities to define their own metadata the tools should support seamless process of creation, advertisement, and adoption of new schemas. In the process of creation of new schema not only the collection of metadata elements has to be compiled but also a mapping of new schema to the desired other standards has to be created. All this has to be supported by the tools in the way enabling the local developers without metadata expertise to achieve this task. In our experience in the most communities of practice some people volunteer to play the role of local ‘gurus’ and adopt and learn the tools for the benefit of community [16]. Once the new schema is defined the infrastructure has to support automatic sharing of this schema with the tools used by the community members. CanCore may be of use in this regard if it is used as a basis to define application profiles for specific collections of learning objects.

Global exchange with community schemas. The sharing culture within communities rarely limits its scope to the community itself. The solutions allowing for the global exchange are preferred over the closed ones. There are two ways of how to support the sharing beyond the ones community boundaries. First, the globally agreed minimum set of the elements provides a necessary nucleus. Secondly, if there is a greater need for the exchange between several communities and the mapping exists between local parts of the community schemas then the search and exchange mechanism should support the automatic translation between schemas using this mapping.

Supportive infrastructure. The infrastructure should provide the services to enable the communities to define themselves (including the definition and sharing of the community schemas), to define their relation with other communities and support the global exchange between the communities. For example, a community would define its schema, incorporating a handful of mandatory fields, and customizing those fields required to support the community functions. The resultant schema would then be advertised on the community repository. Upon discovery of the object, the search tool could read in the community schema and enable a broader search within that community’s repository. Although this simplistic solution might work well at the schema level, the solution at the ontological level are more complex and require more work. This work is in the very center of semantic web research.

We are implementing these principles within the scope of two above mentioned project. The infrastructure we currently developing is described in the next section.

4.1 Generic Architecture for Self-evolving Communities

Two main premises shape the architecture and tools of our system:

1. the communities have to be in charge of the metadata schemas and ontologies they are using, and
2. in most cases the communities want to share resources with other communities, i.e. they need a common ontology to understand each other.

As a result, the local ontology combines concepts selected from the global ontology plus the community specific ontology. Moreover, this combination is defined by the community itself.

Figure 5 demonstrates the main components supporting the knowledge sharing within the community. There are three distinguished groups of community members: local developers, web users and peers.

1. Local developers use specialized tools which allow them to modify local ontology,
2. Web users are users who don’t use the web tools on the server to create and share knowledge and objects.
3. Peers are users using specialized tools which allow them to create their own collection of objects annotated with the ontology. They can either upload objects to the community server or they can share their objects with other peers via peer-to-peer protocol.

The local ontology is created by a group of local developers who may be a community members or specialist contracted for this purpose. The local ontology builds on existing ontologies and adds concepts which are in the focus of the community. The local developers use tools which allow them to modify and upload the ontology to the community server. Once the local ontology is defined, it is available for the download to other community members. The members’ tools automatically detect new versions of the local ontology and update themselves to the new version as required. The web portal tools are updated to the new versions as well.

We have extended SPLASH to provide a means to modify the

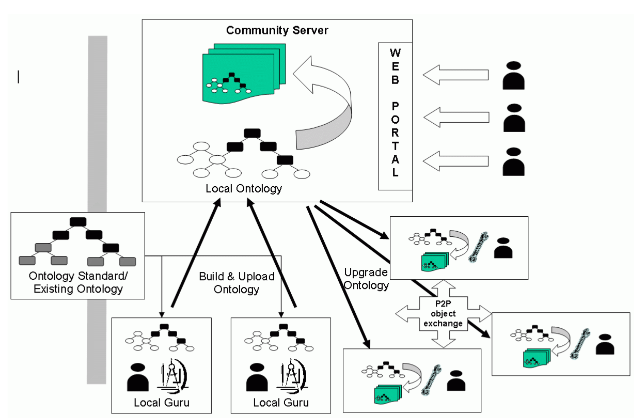

Figure 5. Architecture for self-evolving knowledge community

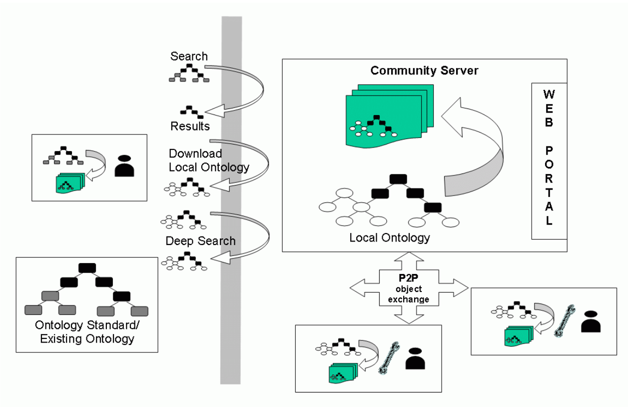

Figure 6. Sharing knowledge with the agents outside of the community

local ontology and also to keep track of the current version of the ontology. SPLASH will update itself automatically to the latest ontology version available on the community server.

The community server has two main functionalities: ontology support and primary collection building. The server provides the community with means to organize a collection building via distributed authoring, community support tools, discussion spaces, management support (if required), and quality control. The community server also serves as a repository of objects created either by web users via web portal tools or uploaded by the peer users. The ontology support provided by the community server is both internal and external with regard to the community. Internally, the server registers all peers and local developers and keeps them updated with the latest ontology changes. The server's external role is to make the community resources available to the larger community.

In Figure 6we explain how the community server communicates with the users/agents outside of the community. The user depicted in Figure 6 (on the left of the gray bar) uses some ontology standard or existing ontology to search for the objects. If this ontology overlaps with the local community ontology, the results described by those shared concepts are returned. If the user finds those results interesting and worth further exploring, s/he can download the community ontology referenced by those results. Ideally, the tool then adapts its search interface to the newly downloaded local ontology which enables the user to specify more precise search. The deep search is initialized and the fully specified results are returned. It should be emphasized that search can be directed to any peer in the community, not only to the community server. We are modifying SPLASH to support this functionality.

5. DISCUSION

The POOL framework presented in this paper presents a significant alternative to the traditional notion of the learning object repositories. Our work builds on the premise that optimal storage sites for learning objects are *close to the creator and close to the user*. Further, as workstation storage increases, it becomes feasible for each learner to amass a personal collection of the learning objects which have influenced their intellectual growth, and to be annotated for future reference and review.

The work which is closest to our approach is project Edutella [17]. The Edutella aims to build a network of connected repositories and provide a set of well defined languages. The main purpose of the network is to dispatch queries to the appropriate repositories based on their previous registration with the one of the specialized hub nodes. Edutella is based on RDF and provides bridges to the XML based repositories [21]. POOL is using XML and XML-based metadata standards for the search and transfer of the metadata.

We can see a significant overlap of our future work with and the research around the semantic web. The semantic web initiative [1] is working towards a vision of having data on the web defined and linked in a way that can be used by machines for various applications and not just display purposes as is generally the case today [7]. Ontologies are a basic component of the semantic web. There is a direct connection between semantic web and metadata systems. The values for the metadata elements can be successfully represented using formal ontologies supporting computational reasoning. A more specific area of semantic web research related to our work is that of mapping between ontologies [20][18][13]. The idea of semantic portals [2][24][25] correlates with our notion of community portal. However, we do not see the portal as the only place where the community can exchange objects and ideas. We see an additional role of the portal in supporting the community members in reaching their goals by providing a necessary infrastructural support for the users' tools. At the same time this will leave the users complete freedom and control over their tools and collections.

6. ACKNOWLEDGMENTS

POOL is partly-funded under the Canarie Inc. Learning Program, and is a consortium involving The NewMedia Innovation Centre (NewMIC), TeleLearning Network of Centres of Excellence, New Bunswick Distance Education Network, The Electronic Text Centre at University of New Brunswick, the Simon Fraser University, TelesTraining Inc, BC's Centre for Curriculum, Transfer and Technology, and the Open Learning Agency. The development team at SFU also includes Dr. Tom Calvert, Gordon Yip, April Ng, Timmy Eap and Jason Toal.

7. REFERENCES

[1] Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. Scientific American, May. 2001.

[2] Domingue, J., Motta, E., Buckingham-Shum, S., Vargas-Vera, M., Kalfoglou, Y., and Farnes, N. Supporting ontology driven document enrichment with communities of practice. Proceedings of the international conference on Knowledge capture, 2001, Victoria, British Columbia, Canada, 155-16230-37

[3] Dublin Core, www.dublincore.org

[4] Friesen, N., Mason, J., and Wand, N. Building Educational Metadata Application Profiles. DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence, Florence, October, 13-17, 2002. 63-69

[5] Friesen, N., Roberts, A., and Fisher, S. CanCore: Learning Object Metadata, Canadian Journal of Learning and technology, Special issue on Learning Objects, Volume 28, Number 3, Fall 2002, 43-53

[6] Hatala, M.., and Richards, G. Global vs. Community Metadata Standards: Empowering Users for Knowledge Exchange, in: I. Horrocks and J. Hendler (Eds.): The

Semantic Web – ISWC 2002, Springer, LNCS 2342, pp. 292-306, 2002.

[7] Horrocks, I., and Hendler, J. (eds.) The Semantic web – ISWC 2002 Springer, LNCS 2342, 2002

[8] IEEE P1484.12.2/D1, 2002-09-13 Draft Standard for Learning Technology - Learning Object Metadata - ISO/IEC 11404, http://ltsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf

[9] IMS Content Packaging Information Model. Version 1.1.2 Final Specification. http://www.imsproject.org/content/packaging/cpv1p1p2/imscp_infov1p1p2.html

[10] IMS Learning Design Information Model. www.imsglobal.org/learningdesign

[11] IMS Simple Sequencing Information and Behavior Model, Version 1.0 Public Draft Specification http://www.imsproject.org/simplesequencing/v1p0pd/imsss_infov1p0pd.html

[12] Learning Program, Canarie Inc., http://www.canarie.ca/funding/learning/backgrounders(99).html

[13] Maedche, A., Motik, B., Silva, N., and Volz, R.: MAFRA - An Ontology MApping FRAmework in the context of the Semantic Web, Workshop on Knowledge Transformation for the Semantic Web (KTSW) Lyon, France, July 23, 2002

[14] Magee, M., Norman, D., Wood, J., Purdy, R, Iwing G. Building Digital Books with Dublin Core and IMS Content Packaging. DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence, Florence, October, 13-17, 2002. pp.91-96

[15] McGreal, R., Anderson, T., Friesen, N., Sosteric, M., Hewitt, K., Ring, J. MacLeod, D., Richards, G. Hatala, M. Calvert, T., Chiasson, M., Roberts, T., Carey, T., Harrigan, K., Paquette, G. and Downes, S. (2002) "eduSource: A pan–Canadian learning object repository" In Proceedings of the E-Learn 2002 Conference. Montreal: Association for the Advancement of Computing in Education.

[16] Mulholland P., Zdrahal Z., Domingue J., Hatala M., Bernardi, M.: A Methodological Approach to Supporting Organizational Learning. International Journal of Human-Computer Studies 55(3), September 2001, 337-367

[17] Nejdl, W., B. Wolf, C. Qu, S. Decker, M. Stintek, A., Naeve, M. Nilsson, M. Palmer, and T. Risch, Edutella: A P2P Networking Infrastructure Based on RDF, in Proc. Of the 11 th International World Wide Web Conference (WWW 2002), Hawaii, USA, May 2002.

[18] Noy, N.F., Musen, M.A.. Anchor-PROMPT: Using non-local context for semantic matching. In Workshop on Ontologies and Information Sharing at the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-2001), Seattle, WA, 2001.

[19] Porter, D.: Object Lessons for the Web: Implications for Instructional Development. In G. Farrell (ed.): The Commonwealth of Learning (COL): The Changing Faces of Virtual Education, July 2001, pp.47-70.

[20] Prasad, S., Peng, Y., and Finin, T. Using Explicit Information To Map Between Two Ontologies. AAMAS 2002 Workshop on Ontologies and Agent Systems, Italy, July 2002.

[21] Qu, C., Nejdl, W., Schinzel, H. Integrating Schema-specific Native XML Repositories into a RDF-based E-Learning P2P Network DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence, Florence, October, 13-17, 2002. 81-89

[22] Shannon, S.J., Roberts, I.W. and Woodbury, R.F.: vGallery: Scaffolding Reflection-in-Action for Students and Teachers, Proceedings of the 18th Conf. of the Australasian Society for Computers in Learning in Tertiary Education, ASCILITE 2001, Dec 9-12, 2001, pp 141-144

[23] Shareable Content Object Reference Model (SCORM), www.adlnet.org

[24] Staab, S., Angele, J., Decker, S., Hotho, A., Maedche, A., Schnurr, H.-P., Studer, R. and Sure, Y. AI for the Web - ontologybased community Web portals. In AAAI 2000/IAAI 2000, pages 1034-1039, AAAI Press/MIT Press, 2000

[25] Stojanovic, N., Maedche, A., Staab, S., Studer. R., and Sure, Y. SEAL: a framework for developing SEmantic PortALs. Proceedings of the international conference on Knowledge capture, 2001, Victoria, British Columbia, Canada, 155-162

[26] Vargo, J., Nesbit, J. C., Belfer, K., Archambault, A.: Learning object evaluation: Computer mediated collaboration and inter-rater reliability. Int. Conf. on Computers and Advanced Technology in Education (CATE), Cancun, Mexico (2002)

[27] Wenger E 1998 Communities of Practice: Learning, Meaning, and Identifying. Cambridge University Press.

[28] Wiley, D. (ed.) The Instructional Use of Learning Objects, Association for Educational Communications and Technology, 2001