Global vs. Community Metadata Standards: Empowering Users for Knowledge Exchange

Marek Hatala and Griff Richards

Technical University of British Columbia
2400 Surrey Place
10153 King George Highway
Surrey, BC V3T 2W1, Canada
{hatala, richards}@techbc.ca

Abstract. The idea of knowledge sharing has strong roots in the education process. With the current development of the technology and moving learning material into the web environment it acquired a new dimension. Learning objects are the chunks of knowledge shared by e-learning community. Organizations and individuals are building repositories of learning objects and annotate them with metadata to describe their educational values and standardization efforts are on the way to provide a franca lingua for the educators. In this paper we describe the peer-to-peer infrastructure for sharing learning object we are building in Canada. The POOL projects builds on the three types of nodes: SPLASH is an 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. The implications of moving the responsibility for schemas and metadata creation on communities poses new requirements on interoperability and tools. We describe those requirements and we outline approach we are developing to address them.

1 e-Learning, Learning Objects, and Metadata Standards

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 (LOs). 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 [1].

I. Horrocks and J. Hendler (Eds.): ISWC 2002, LNCS 2342, pp. 292–306, 2002.

With 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 IEEE Learning Object Metadata [2] defines a set of metadata elements that can be used to describe learning resources. The IMS Global Learning Consortium has identified a minimum set of IEEE metadata elements called IMS Core [3]. The IMS metadata consist of over 80 elements describing different aspects of learning objects.

However, the business and educational communities have been slow to adopt the full IMS standard mainly due to the high number of the fields 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 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 a concerted effort to identify 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. Table 1 provides a description of these groups, more detail information can be found on the CanCore website [5] and concrete examples of the metadata CanCore records are on [16].

Having a standard metadata protocol is of little use unless it is maintained and widely used. Only if a critical mass of educational users standardizes their cataloging with CanCore can the POOL repository protocols search and locate learning objects with efficiency. CanCore must also be embedded in repository systems – this has been promoted by a general agreement in Phase I among Canarie Learning Program projects [6] to use the CanCore protocol. To support CanCore adoption, the documentation, indexing guidelines, training and support is provided to educators by the CanCore team.

Table 1 CanCore elements

General	Groups information describing learning object as a whole. *Active elements:* Identifier, Title, Catalogentry.Catalog, Catalogentry.Entry, Learning Object Language, Description, Coverage
Lifecycle	History and current state of resource. *Active elements:* Version, Contribute.Role, Contribute.Entity, Contribute.Date
Metametadata	Features of the description rather than the resource. *Active elements:* Identifier , Catalogentry.Catalog, Catalogentry.Entry, Contribute.Role, Contribute.Entity, Contribute.Date, Metadata Scheme, Metadata Language
Technical	Technical features of the learning object. *Active elements:* Format, Size, Location, Other Platform Requirements, Duration
Educational	Educational or pedagogic features of the learning object. *Active elements:* Learning Resource Type, Intended End User Role, Context, Typical Age Range, User's Language
Rights	Conditions of use of the resource. *Active elements:* Cost, Copyright and Other Restrictions, Description
Relation	Features of the resource in relationship to other learning objects. *Active elements:* Kind, Resource.Identifier, Resource.Catalogentry
Classification	Description of a characteristic of the resource by entries in classifications. *Active elements:* Purpose, Taxonpath.Source, Taxonpath.Taxon.Entry, Keyword

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.

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.

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: Distributed Infrastructure Based on Standard Metadata

The Portal for Online Objects in Learning (POOL) Project [7] 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 currently 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 [6]. 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 ends 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.

To support the evolutionary nature of learning objects we designed POOL as a network of individual peers communicating together using the POOL protocol (Fig. 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 high speed 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.

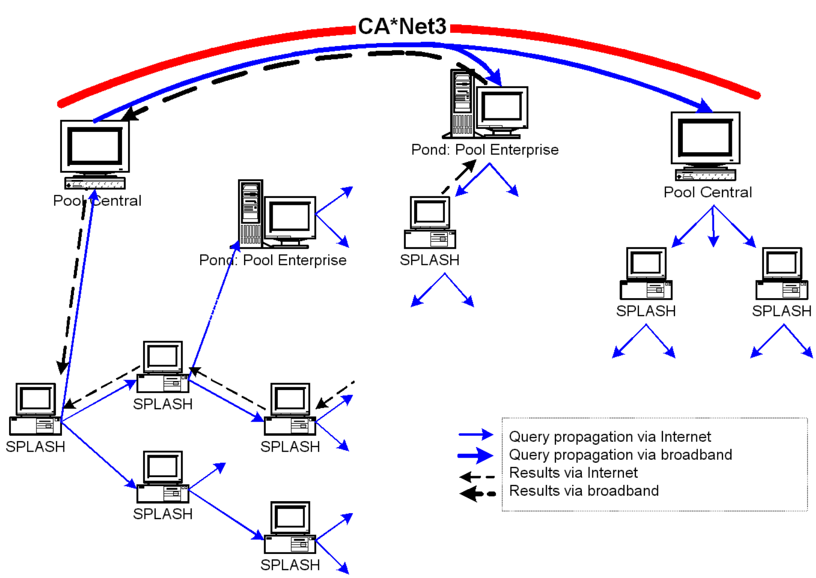

Fig. 1. POOL network architecture

Table 2. 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

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 currently being incorporated into the POOL network [8]. 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 2 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 [9], 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 [10] application 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.

Fig. 2 shows a template for the query request. The query parameters specify both format of the query and response as well 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. [17]. 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. The reasons for having this ability are outlined in the Section 3.2.

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. Table 3 shows examples of two metadata elements. The meaning of most of the attributes is self-explanatory; the cardinality attributes specify how many copies of the same element appear in the form, HasOtherOption specifies whether an extra text field for new values should be displayed, Expandable specifies whether a newly defined values should be automatically added to the existing vocabulary. Full details are available in [11]. 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 users' 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. A set of default profiles is shipped with the SPLASH reflecting various roles the user can take in the

```
<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>
```

Fig. 2. Query specification in POOL protocol

Table 3. Representation of *Title* and *Intended User Role* elements

Attributes	***Title***	***Intended User Role***
ElementKey	general.title	educational.intendedenduserrole
PrintName	Title	Intended User
Tooltip	Learning objects name.	Normal user of the learning object, most dominant first.
SubmitStyle	TEXT	COMBO
QueryStyle	TEXT	COMBO
MaxSubmitCardinality	1	4
MaxQueryCardinality	1	2
HasOtherOption	False	False
Expandable	False	False
ElementValues	[]	[teacher, author, learner, manager]
NewValues	[]	[]
Mandatory	True	False
SubElements	[]	[]
ParentKey	General	Educational
AllowTextSearch	True	False
AllowVocabSearch	False	True

metadata creation process, e.g. educator, learner, editor, media developer, license specialist, etc. Two types of profiles 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. For example, CanCore groups element into 9 groups as shown in Table 1 above. 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 Fig. 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 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 *general.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 it 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 most of the technical development 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. The version available implements all the functionality listed in the Table 2, 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. We anticipate a number of revision cycles during the spring of 2002 to incorporate feedback from our beta testers.

Technically, SPLASH is a Java application running on the user's desktop computer. It uses mySQL as a database engine (bundled with the installation), which is replaceable by another SQL database. Fig. 3 shows a snapshot of the SPLASH search page.

PONDS. We have implemented and deployed so far three 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 BC 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.

The same mechanism has been used for the Canadian Learning Objects Metadata Repository (CanLOM). CanLOM has been built on top of the existing TeleCampus [12] 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 enabling the user to register the learning object in the CanLOM.

In the third case, we have connected the CAREO [13] repository that uses Web services. In this case, we have implemented a simple component translating between the POOL protocol and Web services.

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 Lessons Learned

The deployment of the infrastructure of such scale as POOL has a potential for a big impact on the education communities and the business active in this area. The benefit of a publicity accompanying big projects is that the potential users are more serious about considering new opportunities and willing to invest more effort in participation in such initiatives. For developers this brings an enormous opportunity to discuss the users' real needs and identifying their challenges in adopting new technology.

3.1 Communities Mature

The technology and its public acceptance are evolving so rapidly that even during a project as short as POOL the requirements of the communities are growing. On the one side the developers are getting satisfaction from the acceptance of their technology but at the same time the new requirements are arising even before the

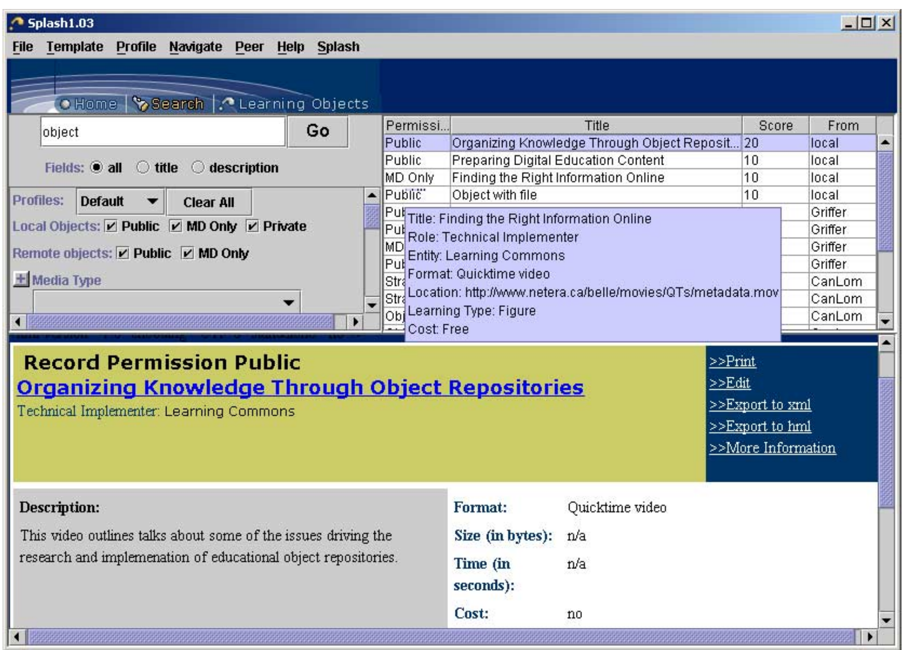

Fig. 3. SPLASH Search page (will provide new snapshot in final version)

system is completely deployed. The idea of metatagged resources brings new ways in which communities share knowledge and it is being cordially accepted. Individual communities are starting to articulate their requirements to support their needs.

Working with the real users and communities always helps to sharpen the research ideas and keep them focused. In this section we would like to generalize our findings from the deployment of the POOL system and formulate a new set of requirements for the metadata sharing system supporting diverse communities.

3.2 Too Much Metadata

The role of metadata was to help solve the problems of search and discovery of learning objects. Ideally, each record could be tagged with catalogue information about its title, content, the intended audience, and other information about cost, availability, authorship, rights management, language of availability, revision history, relationship with other objects, etcetera. These metatag elements then become the targets for a sophisticated search as described in Section 2.5 The more precise the metadata, the more precise the search results. The quest for the management of learning objects became a two-pronged infrastructure-building program (of which POOL project became a part). Part one being the creation of repositories that facilitate common searching, and part two, the creation of international standards for the metadata that described the objects housed within. Corporate and national players are invited by the IMS Global Learning Consortium Inc. to identify their interests by paying $100,000 USD for a seat at the discussion table at a on-going series of globe-trotting workshops to discuss the precisely on what kind of information needs to be kept in each field of the IMS record. Right now the number of fields available to index an IMS learning object sits above 80, and we are graced with the prospects that the cost of providing this metadata can in some cases be greater than the original cost of creating the object itself. Although SCORM [14] based its metadata scheme on IMS, the marketplace pragmatically regards SCORM as a *de facto* standard because the US Department of Defense has instituted a clear and pragmatic policy of only buying instructional materials that are SCORM compliant.

The problem lies with the extent to which metadata should be conducted, and who should do the metadata tagging. It is not a question of if we should do metatagging, but a simple application of the law of diminishing returns – to what extent is metatagging really required. How many field are really necessary to describe a learning object to a useful level, and how much of the effort becomes counterproductive – a logistical barrier that has the unfortunate result in actually reducing the ability and interest of the educational community to participate.

Adding to this complexity is Doctorow's [18] observations about "metacrap", the foibles of human-generated metadata. As if deceit, stupidity, ignorance and sloth in metatagging efforts were not problematic enough, Doctorow points out that the schemas themselves are open to subjective interpretation and can be skewed by false metrics. However, he does not totally throw out metadata, but instead refines his quest for "observational metadata", the kind more likely to be automatically generated by a community of users than the originators of the object.

Fortunately, it is in the service of community interests that metadata is created. Moreover educational communities are starting to look at the technology and how they could benefit from it. It is interesting to look at the kind of metadata already in use in some collaborative learning object collections. For example, The Slice of Life (http://medlib.med.utah.edu/kw/sol/sss) is a consortium of health educators who have been pooling their pathology slides and other medical images since the early 1990's. As the medium of preference moved to digital images, the Slice converted many of their videodisc images to digital media. What has not changed however is the nature of the Slice of Life database, which continues to follow a rather idiosyncratic mix of community and contributors' information, and much of it adhering to Medline, a standardized labeling common to the medical field. A similar situation can be found in the Australian AVIRE collection of architectural objects – here an example of prime consideration is the fenestration of the buildings, hardly an IMS field. Truly, these examples show that there is a need for two levels of meta-data – some generic perhaps following the global standard of IMS, while the other specific to the needs of the specific community of practice.

The dilemma of this two-tiered requirement poses a whole set of new requirements on the tools supporting knowledge sharing at the community level and at the same time enabling global exchange.

Thus, the question of metadata is not in dispute, the question is which combination of global and local fields is required to generally search for relevant objects, allow for observational metadata creation in addition to that produced by object originator, and which should be retained at the community level to select specific items once the general collection has been discovered.

4 Potential Solution and Future Goals

To summarize the previous section, the amount and type of meta-data is best decided by the local user community in consideration of the global requirements. 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.

Although the solution looks simple, it has an 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 [15]. 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.

Observational metadata. The simplest example in POOL will be the construction of a review mechanism to enable consumers to flesh out metadata with new context and evaluative annotations. Nesbit [8] is working closely with POOL to integrate learning object review instrument with SPLASH.

Flexible tools. To summarize the requirements implied by the above functionality the tools should support defining new metadata elements and schemas, mapping between schemas, and they should update themselves automatically reflecting newly defined and upgraded schemas. Of course, tools by themselves will not solve the ontological problems associated with the creation and agreement of metadata requirements for any given community as the quest for collaborative meaning and shared understanding challenges most human conversations.

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, and the name and location of this schema would be incorporated into the metadata. 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. These issues are at the heart of the semantic web construction

Alternative interfaces. Tim Bray (www.antarcti.ca) has suggested non-textual methods for simplifying the task of metadata, is to create visual sorting bins through which simple drag and drop classification might be possible. Either explicit or implicit relationships derived from the sorting process would automatically fill in the metadata – this strategy may succeed by reducing the choices available, and also enable metatagging by association.

In POOL we were aware of the several of these requirements and we have incorporated some of these features into the SPLASH. Although SPLASH already supports a creation of new schema we are still working on the mapping issues. Another major task we are working on is the infrastructure support for the schema exchange and exchange between communities.

Acknowledgements

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 Technical University of British Columbia, TelesTraining Inc, BC's Centre for Curriculum, Transfer and Technology, and the Open Learning Agency. The development team at TechBC also includes Dr. Tom Calvert, Gordon Yip, April Ng, Timmy Eap, Nolan Boulanger and Jason Toal. A special acknowledgement is given to the nurturing community of the Technical University of British Columbia which has been recently subsumed by Simon Fraser University.

References

1. Porter, D.: Object Lessons for the Web: Implications for Instructional Development. In: Farrell G. (ed.): The Commonwealth of Learning (COL): The Changing Faces of Virtual Education (2001) 47{70.
2. IEEE Information Technology - Learning Technology - Learning Objects Metadata (P1484.12), Learning Object Metadata Working Draft (WD5), http://ltsc.ieee.org/doc/wg12/LOM WD6 3a.pdf
3. IMS Learning Resource Meta-data Information Model Version 1.2 - Final Speci_cation, http://www.imsproject.org/metadata/imsmdv1p2p1/imsmd infov1p2p1.html
4. Dublin Core Metadata Initiative, http://purl.org/dc
5. CanCore, http://www.cancore.org
6. Learning Program, Canarie Inc., http://www.canarie.ca/funding/learning/backgrounders(99).html
7. POOL Project, http://www.newmic.com/pool/
8. 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)
9. Project JXTA, http://www.jxta.org
10. JXTA Search, http://search.jxta.org
11. EduSplash, http://www.edusplash.net
12. TeleCampus, http://www.telecampus.com
13. CAREO, http://www.careo.org
14. Sharable Content Object Reference Model (SCORM), http://www.adlnet.org/
15. 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
16. CanCore Examples, http://www.imsproject.org/metadata/mdv1p2p2/samples/cancore/cancore_ex1.xml, cancore_ex2.xml
17. Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M. Ambjorn Naeve, A., Nilsson, M., Palmer, M., Risch, T.: EDUTELLA: A P2P Networking Infrastructure Based on RDF, Edutella White Paper, http://edutella.jxta.org/reports/edutella-whitepaper.pdf
18. Doctorow, C.: Metacrap: Putting the torch to seven straw-men of the meta-utopia, Version 1.3, August 2001, http://www.well.com/~doctorow/metacrap.htm