The Evolution of Learning Object Repository Technologies:
Portals for On-line Objects for Learning

Griff Richards, Rory McGreal, Marek Hatala,
and Norm Friesen

Abstract
Learning objects are the digital files that are used to construct e-learning experiences, and repositories provide mechanisms to encourage their discovery, exchange, and reuse. Portals for On-line Objects in Learning (POOL) is a consortium project of the TeleLearning NCE to build a learning object repository scalable to the national level. Funded in part by the CANARIE Learning Program, POOL efforts have resulted in the development of two focal technologies: POOL, POND and SPLASH, a distributed architecture for a peer-to-peer network of learning object repositories; and CanCore, a practical metadata protocol for cataloguing learning objects. The authors conjecture that the technology of learning objects and repositories is in an early phase of development and that significant evolution can be expected as user communities form, protocols emerge for the functional linking of these structures, and the underlying technology becomes less visible.

Résumé
Les objets d’apprentissage sont les fichiers numériques utilisés pour construire des expériences de e-learning et les banques d’objets fournissent des mécanismes pour favoriser la découverte, l’échange et la réutilisation de ces objets. Les portails pour les objets d’apprentissage en ligne – Portals for On-line Objects in Learning (POOL), est un projet de consortium du RCE-téléapprentissage ayant pour but de construire, sur le plan national, une banque d’objets d’apprentissage adaptables autant pour les grands que les petits nombres d’utilisateurs. Financé en partie par le programme Canarie, les efforts de POOL ont donné comme résultat le développement de deux technologies importantes: POOL POND et SPLASH. Il s’agit d’une architecture distribuée pour un réseau de pair à pair de banques d’objets d’apprentissage ainsi que d’un protocole pratique de méta données pour référencer les objets d’apprentissage (CanCore). Les auteurs pensent que la technologie des objets d’apprentissage et des banques d’objets d’apprentissage est dans une toute première phase de développement et que l’on peut s’attendre à une évolution importante au fur et à mesure que des communautés se forment, que des protocoles émergent pour la liaison fonctionnelle de ces structures et que la technologie sous-jacente devient plus transparente.

Overview
During the seven years of the TeleLearning NCE, the use of Web technologies for on-line learning has found widespread adoption. Learners and instructors find the Web a convenient medium for educational and administrative transactions and have spurred an unprecedented investment in time and resources in the creation of materials and network infrastructure for distance and augmented learning. Digital learning objects are the computer files that store graphics, lessons, animations, and other computer-mediated activities that constitute the content and process activities of on-line learning. As knowledge assets in an e-learning economy, they represent an ever-increasing store of intellectual property and educational capability. Although many learning objects could be reused in different instructional contexts, much of this investment is used for highly specific audiences and remains unknown beyond the immediate creators and consumers.

Repositories are built on database technology, but seek to go beyond simple warehousing to provide mechanisms to encourage the discovery, exchange, and reuse of learning objects. This article describes the evolution of Portals for On-line Objects in Learning (POOL), a consortium project of the TeleLearning NCE, and chronicles the lessons learned in our efforts to build a learning object repository scalable to the national level. Funded in part by the CANARIE Learning Program, POOL has contributed to the development of two enabling technologies: “POOL, POND and SPLASH,” a distributed architecture for a peer-to-peer network of variously sized learning object repositories; and CanCore, a practical metadata protocol for cataloguing learning objects.

**Learning Objects: the Building Blocks of E-Learning**

Although their definition varies among authors and organizations, learning objects are essentially the digital files that are used to generate e-learning activities. They include audiovisual media files, Java applets, and interactive exercises that make up the learner’s experiences. Whether aggregated in generic categories or split into groups of information objects, instructional objects, or reusable learning objects, digital learning objects are the basic building blocks of the e-learning experience. There are a number of detailed introductions to learning objects. Downes (2001) compares them to the reusable programming elements of object-oriented computer programming. In their Cisco Systems White Paper, Barritt and Lewis (2000) provide an example of a reusable learning object strategy constrained to a specific instructional model based on Merrill’s component display theory, and Wylie (2001) compiles a number of interesting articles that examine their pedagogic nature and the implications for higher education.

The promise of digital learning objects lies in reusability. If constructed appropriately, warehoused wisely, and catalogued accurately, a learning object might find usage beyond its original audience and instructional context. Given the relatively high cost of developing good learning objects, the promise of reusability receives considerable attention from administrators and publishers who are trying to amortize the cost of production and maximize the potential return for each of these digital investments. Reuse and wider use may also bring greater recognition for the author.

For educators the promise of reusability goes beyond the economic argument to encompass notions of quality (Bowden & Marton, 1998) and the reuse of exemplary teaching strategies in other contexts. The Campus Alberta Repository of Educational Objects (CAREO, www.careo.org) and MERLOT (www.merlot.org) are Web portals founded partly on the premise that academic peer review of learning objects can improve the quality of learning objects and enhance the quality of on-line education. Learning objects are posted not just to advertise their availability, but also so that others can observe how they are crafted to suit the needs of the
learners, see how they can be adapted into new instructional settings, or how the instructional strategies might serve as models for other content areas.

Early references to learning objects often oversimplified the notion of their being the building blocks of e-learning: to be combined in many creative ways to suit the needs of the learners. Although this attractive analogy implies that standardization is the key to interoperability, like real building blocks, we can expect learning objects to come in many shapes and sizes and commercial brands that for reasons of functionality, sophistication, and competitive marketing will probably not all be compatible and interlocking. Fortunately, like children, learners will be oblivious of this fact and will integrate them into their learning experiences and use them in ways unimagined by the original designers and creators.

**Learning Object Repositories**

Repositories may be viewed simply as places to put digital objects. A central repository would aggregate a collection of objects for a defined community or organization and store them in a single locality. As objects can vary in number, size, and file type, it is unlikely that a single central repository would be able to collect or even physically hold and effectively serve all of the available learning objects in any given field (Hamilton, 2001). As with libraries, organizations or communities may have a something about everything or everything about something, but having everything about everything is unlikely. Thus a decentralized or “distributed” model that links a variety of learning objects repositories is a likely scenario, where the actual storage of objects is in a number of places that are linked with Internet technology. Using the common Web-browser approach, the physical distribution of the objects need not be apparent to users, as in an ideal setting they would interface with the collection through a single access point or portal. If a number of different organizations are involved, inevitably there will be some technical differences in the computer systems employed; however, a portal serves as a consistent access point to find on-line information. Thus the key to a successful national repository strategy will lie in the ability of repositories to share information and exchange records about learning objects and their ability to grant access to the objects themselves.

Repositories might hold collections of learning objects as a book warehouse might store books, or they could hold collections of information about learning objects as a library catalogue might hold descriptions of books. The catalogue descriptions are referred to as the metadata: data about the data contained in the elements. Continuing the analogy, just as a book’s dust jacket contains metadata about the book, its title, its intended audience, can become part of a learning object’s metadata. Some repositories may specialize in the type of information they carry, for example the Australian AVIRE repository (Shannon, Roberts, & Woodbury, 2001) contains only architectural objects, and the metadata is specialized to describe the needs of the architectural community. MERLOT has a more open approach and welcomes information about learning objects in a wide range of content areas. MERLOT holds descriptions of learning objects; peer reviews of learning objects; lesson plans or assignments that use the learning objects; and in a growing number of cases, marketing information about availability, price, and conditions of sale. Any given repository may offer a wide variety of services based on the service it seeks to give its supporting user community.

Because not all repositories store the actual object files, a key function of repositories is to identify the storage location of the objects and provide an indexing system that enables the efficient search and discovery of the objects. How repositories accomplish the first is a function of their architecture; the latter is a function of their catalogue information or metadata.
The Architecture of POOL, POND and SPLASH

POOL, POND and SPLASH evolved as a catchphrase to explain a distributed architecture that could flexibly meet the needs of many groups. Designed to support the individual instructor or learner, SPLASH is conceived of as a small, single-user repository that would be made freely available for download from the Internet. SPLASH combines a database program and a peer-to-peer search engine with a CanCore meta-tagging interface. Built on Sun Microsystems JXTA platform (www.jxta.org), each SPLASH site holds those objects of immediate importance to the owner and has the ability to search other SPLASH peers and to exchange learning objects or learning object metadata with other members of the network.

SPLASH development is partly driven by the notion that the most important place to hold a learning object is close to the developer and close to the user. SPLASH enables instructors, developers, and learners to become consumers of, and contributors to, a network of learning object repositories. SPLASH enables individuals to collect and manage learning objects, perhaps creating portfolios of their personal learning experiences to reduce the transience of the e-learning experience. At the community level the main success of SPLASH may simply be its proliferation of desktop tools that encourage and assist the learning community to metatag their objects and create a large virtual pool of otherwise undiscoverable learning objects.

Communities and organizations are a reality of the world of education and training. Ministries of education, universities, colleges, school boards, schools, and employers are typical of organizations that will have an interest in providing their constituents with access to specific collections of learning objects (they may also have interest in denying access to other “unauthorized” learning objects). These organizations will also have special needs to govern the access, workflow, and life-cycle management of their learning objects. They may have access to financial and technical resources that will support them in building specialized and robust databases such as the initial POOL prototype developed by IBM Canada or the CAREO repository developed in the Province of Alberta. Although serving the defined needs of their communities, these sites also have the potential to become PONDs in our network architecture: community sites that primarily serve the interests of their clients, but through their inter-operability with the POOL network provide those outside the community with access to POND resources.

Finally, a third level of aggregation, POOL Central, was devised to replicate search requests in topological regions of the repository network and overcome the horizon effects that arise in decentralized peer-to-peer networks such as Gnutella. The designation of a number of “super nodes” could facilitate a faster and more exhaustive search of all the member repositories via a high-speed and high-bandwidth connection to the CA*Net 3 optical highway (Figure 1).

The significance of POOL, POND and SPLASH is that it defines not so much a repository structure as a method of linking repositories. Any repository of any size can be cross-searched simply by adding on the SPLASH search mechanism, adhering to the IMS, CanCore, or SCORM metadata protocols, and being willing to be included in the aggregate repository initiative.
The Internet has become one of the best research tools for students and their teachers. Many, however, are finding that there is just too much information online. Powerful search engines such as Google (www.google.com) return a plethora of data for almost any search term. This lack of precision makes it difficult to unearth and identify quality learning resources. Learning objects include nontextual elements such as images, sound tracks, and executable Java applets or dynamically generated Web content, which are usually poorly indexed by conventional Web-crawlers. The POOL project, working in conjunction with the CAREO project has created a national metadata protocol known as Can-Core (Canadian Core Learning Resource Metadata Protocol) to help address this problem and facilitate the sharing and management of online educational materials or learning objects.

As noted above, a learning object is a resource with an explicit educational application. It can be digital, for example, a simple Microsoft Word, PDF, or WordPerfect text, an e-book, or a Flash animation. Or it can be physical like a textbook or CD-ROM. For the purposes of interoperability, POOL is concerned only with those learning objects that can be accessed from the Internet. Metadata facilitates access to learning objects by providing a controlled and systematic way of describing each object. Metadata describes and links learning objects (Innes, McGreal, & Roberts, 2002).

Metadata, quite simply, is data about data, like a label that is placed on a package. Common examples of metadata are the card catalogue at a library or the dust jacket of a book. Metadata is the term used to describe a package of information about an electronic resource. It acts much like a catalogue record in a library, describing the resource it refers to by providing information such as the title, author, subject, and the like.
as author, title, subject matter, copyright information, and location (McGreal & Roberts, 2001). Just as the metadata for a book can appear in the on-line card catalogue, in Books In Print, or printed inside the title page, the metadata for a learning object can appear separately or contiguously with the object it describes. The process of creating metadata for a learning object is often referred to as meta-tagging, the resultant meta-tag being the explicit metadata that describes a given learning object.

CanCore has been developed as the metadata application profile for POOL and other CANARIE-sponsored projects. It is already becoming widely known in the Canadian and international e-learning communities (see http://www.cancore.ca). CanCore provides a streamlined version of the complex Instructional Management System (IMS) metadata standard (http://imsproject.org/), which forms the base for the SCORM (Sharable Courseware Object Reference Model) metadata standard (http://www.adlnet.org/scorm/). CanCore is fully compliant with these international standards and with the emerging IEEE Learning Object Management protocol P1484.12 (http://ltsc.ieee.org/index.html).

The CanCore Protocol is a set of elements for the uniform description of modular, digital educational resources. These elements represent a subset of the data elements specified in the IMS Meta-data Information Model. The CanCore specification takes a middle-ground approach between the minimalism of the 15-element Dublin Core Metadata Initiative (DCMI, http://purl.oclc.org/dc/) and the structural approach of IMS.

CanCore has taken only the active core elements from IMS that are considered essential for learning object implementations. The researchers produced a set of formal guidelines explaining the purpose and use of each element and provided standard vocabularies where they were appropriate. This simplification provides developers of learning materials with a prescribed and recognized formula that can speed up the development of standards-based projects. By simplifying the IMS element set and providing guidelines and assistance, the CanCore group hopes to save time for developers and project managers who wish to adopt the metadata standard. The full IMS element set is not suited to direct implementation. Conformity with all 86 elements [of the IMS metadata specification] is often not taken up by vendors because it creates a huge job of classification.

The CanCore Protocol has been developed to provide a common element-set for Canadian and international educational object repository projects. Besides POOL, these projects include other CANARIE-supported projects including the Broadband Enabled Lifelong Learning Environment (BELLE, http://www.netera.ca/belle/), and LearnCanada (http://www.learn canada.ca/). projects, as well as New Brunswick’s TeleCampus (http://telecampus.edu) and CAREO. Funding and support for the development of the CanCore Protocol has been provided through these projects, and by the Netera Alliance, TeleEducation NB, and the Electronic Text Centre at the University of New Brunswick.

**CanCore Rationale**

CanCore has been considered an essential piece of the total POOL project from the beginning. The existing metadata solutions were considered either inadequate or too complex for a real-world implementation. In implementing a distributed learning object repository project, interoperability among various content repositories is indispensable.

Neither the Dublin Core nor IMS specifications present ready-made metadata solutions for the collection and sharing of learning objects. The Dublin Core provides a minimalist set of 15 elements for the description of information resources in general, but does not provide elements for describing educational resources in particular. The IMS Metadata Information Model,
meanwhile, takes a structural approach to metadata and uses 86 elements to cover an extensive set of attributes specifically intended for learning objects. Even the IMS realizes that many implementers have no interest in developing products that require a set of metadata with over 80 elements.

Moreover, the IMS provides only the briefest descriptions of the purpose and character of each of its 86 metadata elements (e.g., element 1.3 general.catalogentry is described only as the “designation given to the resource”). Consequently, the actual implementation of the IMS element set is necessarily a complex, resource-intensive undertaking that requires elements to be chosen, interpreted, used, and tested by those sharing, collecting, or developing educational resources. Also, widely varying interpretations of the utility, purpose, and scope of individual elements threatens to cause considerable interoperability problems. CanCore provides a model or benchmark interpretation of the meaning, purpose, and scope of 36 IMS fields that are considered important for promoting the interoperability (exchange and communication) of learning objects. Is it the learning objects that become interoperable or the repositories?

The 36 elements of the CanCore Protocol already save users the task of interpreting, selecting, and coordinating the use of metadata elements to achieve a basic level of interoperability. As a Canadian and international initiative, CanCore presents the possibility of supporting further economies of scale by discouraging the emergence of duplicate, redundant, or inconsistent implementation efforts. It ensures that educational resources can be shared seamlessly across Canada and internationally.

CanCore is continuing to develop, and as it matures the CanCore team is creating support documents and services for a general audience of developers, designers, and educators. These activities are promoting the expansion of the acceptability of CanCore and thus supporting enhanced interoperability. CanCore, like Dublin Core and IMS/SCORM, is being implemented using RDF (Resource Description Framework) and XML (eXtensible Markup Language).

The real-world trials of the POOL applications and CanCore are being conducted using the TeleCampus and other on-line repositories. The Tele-Campus houses metadata with links to more than 55,000 on-line programs, courses, modules, or lessons (McGreal, 2002). It is being adapted as a specialized node of POOL called the Canadian Learning Object Metadata Repository (CanLOM) to gather learning object metadata from SPLASH sites, and as such it is presently implementing CanCore into its IMS/SCORM-compatible database structure.

The Evolution of Learning Repositories
As we approach the close of POOL’s two-year project life, we contemplate more the coming challenges than our accomplishments. Although conceptually elegant, POOL, POND and SPLASH are just entering their test phase. To be declared a success, POOL must use this infrastructure to unite a community of users with useful content. Much like the fire triangle of fuel, heat, and oxygen, the absence of any one element will extinguish the flame. We know there are users and they have content, but only through a program of iterative evaluation and simplification of the tool set can we lower the learning curve to the point where everyone can join the pool.

A significant hurdle lies in the current complexity of metatagging. Indeed, explicit metatagging has recently been described as an overly complex activity (Doctorow, 2001), and a number of methods for simplifying the schemas and automating the task with visual interfaces are under examination (Bray, 2001). Developers of metadata application profiles—the specification of metadata fields and proscribed vocabulary for a particular collection of learning objects—are caught on the horns of a dilemma as they try to determine the degree of effort that
should be allocated to global issues of discovery versus local issues of detailed descriptions such as AVIRE’s “fenestration.” This ongoing battle will continue as each community of practice defines itself through its shared vocabulary, ethics, and practices. SPLASH and CanCore are designed with flexibility to accommodate these needs; however, they are conceived of as generic tools. It may be necessary to develop multilevel strategies for managing learning objects: perhaps a global search that returns not only probable hits, but identifies the specialized node and the native metadata schema necessary for a thorough localized search. It may be necessary for SPLASH to accommodate a number of interchangeable metadata schemas flexibly while keeping this complexity hidden from the user. Natural language is also a barrier, although we are now working closely with another CANARIE project, SavoirNet (Paquette, 2002), to develop French-language versions of the interface and the CanCore application profile.

Security is an issue that will require much attention in a network of learning object repositories. Security affects the willingness of content creators to see their wares exchanged on the network. POOL does not have an inherent rights-management system, but it could work with a third-party system where encrypted files can be located and distributed in POOL with distribution of decryption keys handled by a secure broker using whatever business model it feels is appropriate. Persistence and integrity are two more issues of great interest to the repository community. Persistence implies that the object will actually be retrievable on demand, and although this is often used by proponents of centralized servers to criticize peer-to-peer systems, the reality is that every instructional delivery system must grapple with this issue, and short of local caching the file, there is no fail-safe way to guarantee persistence on any networked system. Indeed, centralized services are prone to failure when demands for service exceed capacity, whereas peer-to-peer systems actually increase their capacity to serve data each time a new copy is made of a file.

A final concern is that although learning object repositories may provide a better means of discovering and distributing learning objects, repositories do not in themselves address issues of pedagogy. To those educators who ask about quality assurance for learning object repositories, our response is to look to the communities of use for self-regulation of these issues. For example, SPLASH will have user-defined specifications to restrict searches to particular lists of nodes. Thus a self-defining community need only maintain its own list of adherents to its community norms and advertise those norms just as it might post its metadata schema and defined vocabulary. Nesbit, Belfer, and Vargo (2002) have proposed a Learning Object Review Instrument (LORI) to enable the creation of consumer reviews, and a test implementation of this strategy is now being developed for SPLASH. User reviews not only add value for all members of the community, but as they also encourage the metatagging of content by the consumer, they increase the number of objects that are tagged in the network. Nesbit et al. also see the process of collaborative community review of learning objects as a mechanism for professional development for educators and a way of establishing and sharing design heuristics and benchmarks that will enhance the overall quality of the learning objects.

These issues are only the tip of the iceberg when it comes to learning object repositories. As we move forward and the user community and content base expands, we hope to have created in POOL and CanCore a sufficiently flexible base and an open enough mindset that these and other issues yet to come may be easily addressed. Indeed, one deliverable of the POOL project will be an examination of the community governance structure to ensure the long life and usefulness of CanCore.

Summary
POOL is but one of many international efforts to create learning object repositories. Others such as MERLOT and CAREO have been created to meet specific community efforts, and there is a growing abundance of Learning Content Management Systems (LCMS) in the commercial e-learning market (Washburn, 1999).

CANARIE, which has sponsored POOL and other repository initiatives through its Learning Program, also recognizes the need for convergence of effort and has initiated a series of informal strategy sessions through the Canadian Repository Action Group (CRAG). This would see the regrouping of the various repository projects into a single pan-Canadian effort to create a national strategy for the advancement of learning object repositories (McLeod, 2001). These community consultations will be used by CANARIE to shape future work under the e-Learning Program.

Canada is not alone in these efforts to build repository tools. Australia, Sweden, and the Netherlands are also moving rapidly in repository research and development. Indeed, the POOL team is in ongoing correspondence with a Swedish group building Edutella (http://edutella.jxta.org/), a peer-to-peer model that is being built using the same JXTA platform as SPLASH. We would hope to see convergence of these international efforts so that a universal repository model can emerge.

Learning object repositories are the catalogues of the e-learning era. They will be the fundamental first step in knowledge discovery and object exchange. They will provide the foundation for future learning and commerce in the knowledge market. They will fuel e-learning as the stock exchanges fueled the industrial era. This is why they are of priority interest.

Acknowledgments
The original idea for a Portal for On-line Object in Learning must be credited back to Nicholas Galan and Ian Dodswell, then of TeleLearning, and Chuck Hamilton of IBM Canada who wanted to create an E-Bay for learning objects. As momentum for the project grew, so did a wider recognition of general need for the technical needs of a learning object repository. Thanks to Rory McGreal and Terry Anderson (both now at Athabasca University), the need for a common metatagging protocol was recognized. They facilitated the CanCore project with team members: Alan Burk and Sue Fisher, University of New Brunswick; Norm Friesen, University of Alberta; and Anthony Roberts, TeleEducation NB. Maria Callas at IBM headed the initial development team. As the design team met to scope the second phase of POOL, it needed to resolve conflicting descriptions of the target users for POOL. Were they to be individuals or organizations, publishers or consumers, experts or novices? Griff Richards conceptualized POOL, POND and SPLASH as a way of fostering interoperability between independent repositories developed by individuals or communities regardless of scale, but with a common goal of facilitating the discovery and exchange of learning objects. Also contributing were Steve Petschaluk of IBM, Randy Bruce of C2T2, David Porter, then of Open Learning Agency, and Marek Hatala. Marek went on to lead the TechBC team of Gordon Yip and April Ng that produced the early version of SPLASH. Collaboration between large projects requires special leadership, and we have been fortunate to have the support of Joanne Curry of TL-NCE, Fred Lake of NewMIC, and Jamie Rossiter of CANARIE.

References

Griff Richards is the Research Integration Officer for the TeleLearning NCE. Rory McGreal is the Associate Vice-President of Research at Athabasca University. Marek Hatala is an assistant professor of information technology at Simon Fraser University, Surrey, British Columbia. Norm Friesen is a graduate student in the Faculty of Education and Director of the CanCore initiative at the University of Alberta.