The Learning Registry: Sharing Federal Learning Resources

Paul Jesukiewicz
OSD P&R, Training Readiness & Strategy
Alexandria, Virginia
paul.jesukiewicz@adlnet.gov

Daniel R. Rehak, Ph.D.
Advanced Distributed Learning Initiative
Alexandria, Virginia
daniel.rehak.ctr@adlnet.gov

ABSTRACT

The Learning Registry (learningregistry.org), a joint project of the US Department of Defense and US Department of Education, provides an infrastructure that enables instructors, teachers, trainees and students to discover and use the learning resources held by various federal agencies and international partners. There are many good learning resources (both primary source materials and content explicitly created to support learning) from government, institutions and the commercial sector that can be used in many different ways. But these resources are hard to find. It's difficult to tell what resources are available, how they have been used, and, most importantly, if they are effective in training and education. The Learning Registry enables better access to learning resources and the building of interconnected and personalized learning solutions.

The Learning Registry is not another repository, search engine or portal. It is a resource distribution network with open APIs that anyone can use to expose or consume learning resources and information about how they are used. It enables building a business-to-business infrastructure where users can find, share, use and augment learning resources. Organizations build third-party applications and communities on top of the distribution network to facilitate learning resource discovery, access and sharing. These applications let communities of users publish information about learning resources or discover information about learning resources.

The Learning Registry network hosts and shares both metadata and paradata (content about where a learning resource was used, comments, rankings, ratings, etc.), i.e., the Learning Registry provides “social networking for learning resources”. It combines cataloging information, usage, assertions, data exhaust and analytical data into a single, sharable timeline for learning resources. Organizations and users can access the network to share learning resources and to provide information and feedback about the use of learning resources, thereby amplifying knowledge and adding value to the learning resources.

ABOUT THE AUTHORS

Paul Jesukiewicz serves as the Senior Advisor of Advanced Learning Technologies for the Training Readiness and Strategy Directorate, Office of the Deputy Assistant Secretary of Defense (Readiness). Mr. Jesukiewicz advises the Training Readiness & Strategy Directorate for all matters related to science and technology issues associated with Defense training, distributed learning, educational technologies, and performance aiding. Mr. Jesukiewicz was designated the DoD lead advisor to the White House Office of Science and Technology Policy (OSTP) for development of the next generation open learning architecture and the federal learning registry. Mr. Jesukiewicz also was designated as Chairman of the NATO Training Group (NTG) on Individual Training & Educational Developments (IT&ED).

Mr. Jesukiewicz has 25 years experience in training and learning technologies working in government, industry, and academia. Mr. Jesukiewicz received his B.S. in Electrical Engineering from the Pennsylvania State University, his M.S. in Engineering from Catholic University, and is currently working on his Ph.D. in Instructional Technology at George Mason University.

Daniel R. Rehak, Ph.D., is Senior Technical Advisor, Advanced Distributed Learning Initiative (ADL). He provides technical expertise to the DoD OSD P&R (Readiness) Training Readiness and Strategy Directorate, in the areas of systems design, information management and architecture, with emphasis on learning and training technologies, content and knowledge management and related technology interoperability standards. He has held prior positions as the Co-Director of the Workforce Advanced Distributed Learning Co-Laboratory and as the

Technical Director of the Learning Systems Architecture Laboratory at Carnegie Mellon University, which he co-founded in 2000.

Dr. Rehak has over 15 years of experience in the e-Learning industry as a technical strategist, information architect and planner. His expertise is in software design and architecture; data and knowledge management; Internet and Web-based systems; and educational technologies. He has worked with government, academic and commercial partners, focusing on future directions and approaches to learning technologies, content, learning management, and digital libraries. He has participated in the development of many of the current e-learning interoperability standards and specifications (IEEE LTSC, IMS, CEN) and has been an invited speaker addressing the future trends in learning technology standards and systems at numerous international events.

The Learning Registry: Sharing Federal Learning Resources

Paul Jesukiewicz
OSD P&R, Training Readiness & Strategy
Alexandria, Virginia
paul.jesukiewicz@adlnet.gov

Daniel R. Rehak, Ph.D.
Advanced Distributed Learning Initiative
Alexandria, Virginia
daniel.rehak.ctr@adlnet.gov

THE LEARNING REGISTRY

A timeline of second-party usage data and analytical data prioritized over static, curated first-party descriptive metadata; no mandated data standards; publish and add to the timeline from anywhere; access by anyone; learning resource descriptions replicated worldwide; open; cloud and app ready; an enabling infrastructure for building communities; an alternative approach to learning resource discovery, data sharing and knowledge amplification.

FINDING AND SHARING LEARNING RESOURCES

The Problem

Let's imaging that you're a Department of Defense Education Activity middle school physics teacher and you want to build a lesson on orbital mechanics and combine elements of physics, math, history of the space program and a writing assignment. Where would you go to find the learning resources you need, either lessons to reuse or individual pieces?

A search engine *might* help in finding the individual pieces, but formulating a query for the entire lesson is difficult. Even if you can formulate the query, you will probably get back hundreds of results. If you want images and primary historic source material, you'll probably have to search individual collections: NASA, National Archives, Smithsonian, Library of Congress, and probably multiple repositories for each since all of these resources are not cataloged by search engines.

Let's assume you found several video animations on orbital mechanics. Can you tell which of these are right for your students (without having to preview each)? How will you know which ones are of high quality, come from trusted sources, or are aligned to your curriculum? Is there any information about who else has used them and how effective they were? How can you provide your feedback about the resources you used and found effective, both to other teachers and to the organizations that published or curated them?

Alternatively, you might find the resources through one of the K-12 focused educational portals. NASA could "announce" a new animation on its web site. Someone from the PBS Teachers (PBS, n.d.) portal could be checking the NASA site periodically for new animations, and add the resource to their middle school "Science and Tech" stream. The National Science Digital Library (NSDL, n.d.) could also provide it to their community via one of their "Pathways". These portals typically curate resources such that they are aligned with curricula, and provide mechanisms for feedback and commentary.

There are hundreds of such portals and education-focused search engines, some manually curated, some that monitor data feeds and harvest other repositories to build their collections. Some are small and isolated, some have significant holdings, some federate learning resources into larger collections, many have overlapping coverage, and most are idiosyncratic in their approach and interfaces. Simply finding the right place to start the discovery process can be a challenge.

The communities and sites that do gather feedback and usage data typically do not share their data with others. Often significant data trails and *data exhaust* are lost, e.g., an interactive white board knows information such as when a teacher drags a resource onto the white board, how long it is displayed, and where the class is in the curriculum. Similarly a learning management system or an intelligent tutor knows what content is delivered in what sequence and how that content is related to learning outcomes and progression. There is currently no systematic way to gather and share this data. Being able to aggregate and analyze this data more broadly would be beneficial in understanding what learning resources are effective in what contexts.

This current situation is illustrated in Figure 1. While the examples above are from the K-12 schools sector, the problems and situations are the same for all educational and training sectors, both nationally and internationally.

In summary the problem is:

- an abundance of learning resources distributed in many locations, often hidden,
- a lack of data sharing,
- lost valuable data trails and data exhaust,
- limited feedback loops,
- a legacy environment and technical approach.

The net result is that we don't know what resources have a positive effect on learning and we don't have an easy way to find out. Finding, sharing and using learning resources are just too difficult.

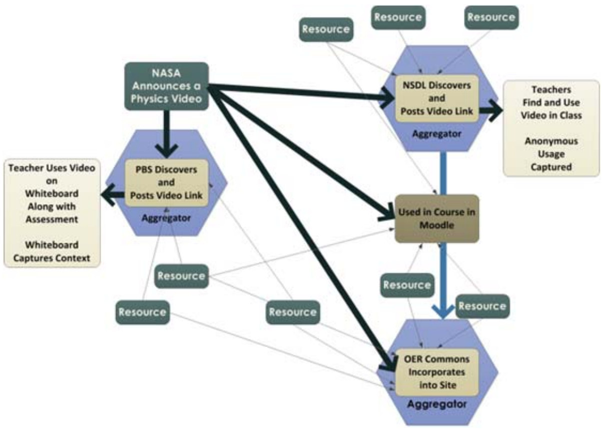

Resource
Resource
Resource
NASA Announces a Physics Video
NSDL Discovers and Posts Video Link
Aggregator
Teachers Find and Use Video in Class
Anonymous Usage Captured
Teacher Uses Video on Whiteboard Along with Assessment
Whiteboard Captures Context
PBS Discovers and Posts Video Link
Aggregator
Used in Course in Moodle
Resource
Resource
Resource
Resource
Resource
OER Commons Incorporates into Site
Aggregator

Figure 1. Sharing Learning Resources

The Social Metadata Timeline

The Learning Registry solution to resource discovery and sharing centers on building and sharing an event timeline around the interactions that occur when a learning resource is found, shared and used.

Revisiting the example, NASA, other resource providers and resource curators publish their descriptions of learning resources to a shared common timeline. Various communities, e.g., NSDL, PBS Teachers, monitor the timeline and detect when new resources and information are available. This monitoring process is automated, and communities can listen to only part of the timeline, e.g., listen only to NASA, only to physics resources, only to middle school resources.

Using the shared timeline eliminates the need to build every connection between each provider and each community. All communities and applications can use a simple, consistent set of interfaces and APIs to put information into the timeline, to monitor it for changes, and to extract data from it for local processing and use.

Any of the communities or organizations that consume the learning resources can capture information about how they are used in the community, usage context, user feedback, user ranking, rating, annotations, etc. User communities such as NSDL can provide this contextualized usage data, denoted as *paradata* (VanGundy, 2010) or attention metadata (Ochoa, 2006), and add it back into the timeline. Tools such as interactive whiteboards and learning management systems (e.g., Moodle, SAKAI, Blackboard) can anonymize and share their usage data back into the common timeline.

This growing collection of data can be consumed by anyone and can be fed to analysis tools such as recommender systems. Just like Amazon and Netflix recommenders (Netflix, 2009), usage data gathered from a diverse educational community combined with the context of an individual user can provide improved discovery results. Likewise sophisticated data mining tools can begin to discern what content is effective in what situations.

Analytical data, analysis and recommendations are also entered into the timeline, further amplifying the available knowledge base. Other tools and systems can then use this data. A resource federation portal like OER Commons (OER Commons, n.d.) can use the timeline as the source of its resources, rather than having to harvest from individual repositories and other federations – integration is simplified. The OER Commons discovery portal can also include the paradata and recommendations when it presents search results. Similarly, a generative intelligent tutoring system can incorporate this data about use, content and sequencing into its decision-making process, dynamically picking the best next resource to deliver to the student based on a wealth of information.

The social metadata timeline illustrating this example is shown in Figure 2.

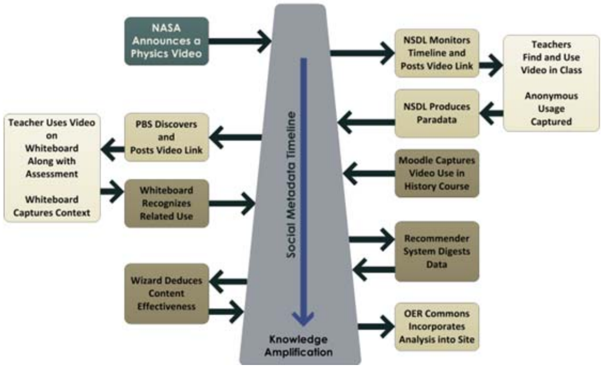

NASA Announces a Physics Video
NSDL Monitors Timeline and Posts Video Link
Teachers Find and Use Video in Class
Anonymous Usage Captured
NSDL Produces Paradata
Teacher Uses Video on Whiteboard Along with Assessment
Whiteboard Captures Context
PBS Discovers and Posts Video Link
Moodle Captures Video Use in History Course
Whiteboard Recognizes Related Use
Social Metadata Timeline
Recommender System Digests Data
Wizard Deduces Content Effectiveness
OER Commons Incorporates Analysis into Site
Knowledge Amplification

Figure 2. Social Metadata Timeline

THE LEARNING REGISTRY APPROACH

Capabilities and Concepts

The Learning Registry is a new approach to solving the problem of resource discovery and understanding what resources are truly effective. It aims to address needs of educators, students, administrators, funders, etc. It eschews the conventional approaches of building portals, search engines and repository federations.

The Learning Registry focuses on four key capabilities:

- **Find**: enabling the discovery of appropriate, effective learning resources for a particular context by using an extensive body of knowledge.
- **Share**: permitting anyone to share any data they think is valuable about any learning resource, without complex technical restrictions of data formats and schema, and letting anyone consume this data for any purpose.
- **Use**: making it easy to use learning resources while gathering usage data of all sorts about how, what, when, where, by whom and why a resource is used.
- **Amplify**: supporting feedback loops, data mining and data analysis to amplify the knowledge base to provide a rich, robust, extensive data collection for everyone.

We believe that in the long term, second-party paradata and usage data, and third-party analytical data will be more valuable and more useful than traditional first-party curated cataloging metadata for discovery and understanding what learning resources are effective. Our conceptual "metadata" model, illustrated in Figure 3, recognizes the importance of the resource in context, its descriptive metadata, the usage paradata, assertions and linkages to other data and analytical data.

The Learning Registry allows any data consumer or producer to talk about any of these, and to share such data in the timeline, either as individual quanta, events, or aggregations. It emphasizes this data and sharing over conventional repository, federation and learning resource management approaches.

Beyond these core concepts and metadata model, the fundamental precept is that it's sufficient to develop an enabling infrastructure. We want to produce an innovative infrastructure supporting learning resources, and not produce interfaces and applications. We want to build and enable a learning and learning resource layer on Web 2.0. We believe that the right infrastructure and data will let others build effective and innovative systems, applications, communities and business models that we cannot anticipate. We leave it to external communities to decide what to do with data in the timeline.

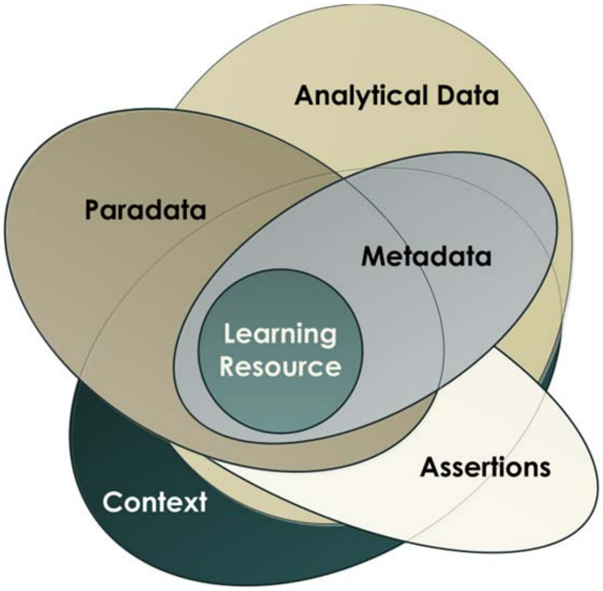


Figure 3. Conceptual Metadata Model

The Learning Registry is designed around a number of key enabling principles:

- *Provide capabilities not solutions:* the Learning Registry provides core enabling capabilities. It is not a complete turn-key solution; organizations and communities need to build their solutions on top of the core or extend the core to meet their needs.
- *Let anyone participate*: by design, there are no inherent restrictions on who can publish or consume data, and who can be a part of the Learning Registry resource distribution network. Individual organizations may place restrictions on their private operations, but these do not extend beyond their local boundaries.
- *No default "winners"*: there are no a priori preferred metadata schemas, vocabularies, tagging strategies, user community features, applications and the like. Any alternative is permitted, and those which are found useful and accepted by the overall community win by gaining mind share and market share.
- *No single point of failure or control*: there is no central point of control, either in the technology or in policies and operations. Fault tolerance and redundancy is part of the model. Data can be widely replicated. Disabling any one part of the resource distribution network does not impact the operations or services provided by other parts of the network. There is no "off switch".

- *Anyone can provide information on anything*: anyone can add metadata, add paradata, provide assertions, or submit analyses. It is up to the consumers to evaluate the value of the data and the reputation of the provider. But just as anyone can be a provider, anyone is permitted to use just the parts of the data they want to use, not all of it.
- *Identity and trust exist*: To understand "who is speaking", a simple model of identity is incorporated in the overall model. All data is associated with an identity. Data analysis and provided assertions can be used to determine the reputation of an identity, and thus the data consumer can decide who to trust.
- *Re-aggregation, amplification and sharing are natural*: All data is open and shared globally, and automatically. Anyone can use the data, augment it, extend it, recombine it and re-aggregate it any way. All this leads to more data and amplification of the value of the data.
- *Web 2.0 technology*: The design is based on current approaches and best practices for web 2.0, including RESTful interfaces (REST, 2000), JSON data structures (JSON, 2006), NoSQL databases (NoSQL, n.d.), map-reduce computation (Dean 2004) and open APIs. Technology is not, however, selected just to be new, but is chosen because it simplifies the approach and solution.
- *Separation of design from implementation and deployment*: there is a formal specification for the model and the core features of the Learning Registry, but it only specifies *what* needs to be done, how interfaces need to behave to provide interoperability, and minimal requirements for operations. Implementation and deployment are not specified.
- *As simple as possible*: the approach is to provide only essential features, and those that solve 80% of the problem. When choosing alternatives, we favor those that can be generalized and provide extensibility over those that hardwire solutions or limit choices. We leave it to others to extend the system and provide more features.

Technical Approach

The Learning Registry concepts and principles are reflected in the four key elements of the technical solutions that enables resource discovery, sharing, use and data amplification:

- *Network distribution model*: At its core, the Learning Registry is centered on a learning resource data network distribution model, used to distribute and replicate metadata and paradata throughout the network. Data published anywhere flows throughout the network; consumers anywhere can listen for and grab the data they need.
- *Data model*: The data model is used to describe the metadata, paradata, assertions, analytical data, identities, and reputations that flow through the distribution network.
- *Base service APIs*: The base APIs provide essential features that organizations can use to build applications and communities.
- *Layered stack approach*: The Learning Registry stack provides a layered model that enables organizations to develop communities of users on top of the APIs that provide access to the resource descriptions that flow through the resource distribution network.

Network Model

The Learning Registry enables the creation of a resource distribution network. A distribution network consists of an arbitrary collection of *nodes*, connected into a directed graph of arbitrary topology. Each node provides a data store for information about learning resources, i.e., the store of metadata and paradata.

A node is connected to one or more other nodes through a distribution link. Periodically, a node will replicate its data to the nodes it is connected to. The distribution process understands data uniqueness, conflicts and versioning such that if a node receives conflicting versions of the same data from different sources, it will store only the most recent version.

If there is at least one acyclic path from each node to all other nodes, the entire network will reach "eventual consistency", i.e., each node will contain an identical copy of all data. Thus data is redundantly stored; data may be published into the network from anywhere, and accessed from anywhere, meeting the requirements for "publish anywhere" and "no single point of failure".

Our network model adds some additional features to support discovery and sharing that may eliminate full consistency of all nodes in particular implementations, but only if done intentionally.

Any node may *filter* its incoming data, and store only data that meets node-specific criteria and policies, e.g., it might accept data that only comes from certain trusted or reputable providers or only data about certain types of learning resources (grade level, subject, etc.). Only the concept of filtering is part of the model; each node can design and implement its own filters, and can filter on any attribute or data value.

In the model, a *network* is a collection of connected nodes with shared operational policies. To support security requirements, by default, a node may only be connected to other nodes within the same network. Different networks with different policies may be created. Special *gateway* nodes, with restricted capabilities, provide the mechanism for controlled transfer of data between networks.

Networks may be assembled into communities, where communities have additional shared policies. A network may only be a part of one community and, in general, communities cannot be inter-connected (a "social" community can be connected to other social communities). For example, an organization can establish a private instance of the Learning Registry for internal resource sharing, and by design and implementation, this instance cannot be connected to another Learning Registry community and cannot share the private data outside of the private community, e.g., data in a private DoD community cannot be shared outside of DoD, but external data could be brought into the DoD community.

For simplicity, the model is restricted to these three levels; an arbitrary hierarchical nesting structure is not permitted. Nodes store data, networks provide connections and policy boundaries, and communities provide isolation and security boundaries. This overall network structure is illustrated in Figure 4.

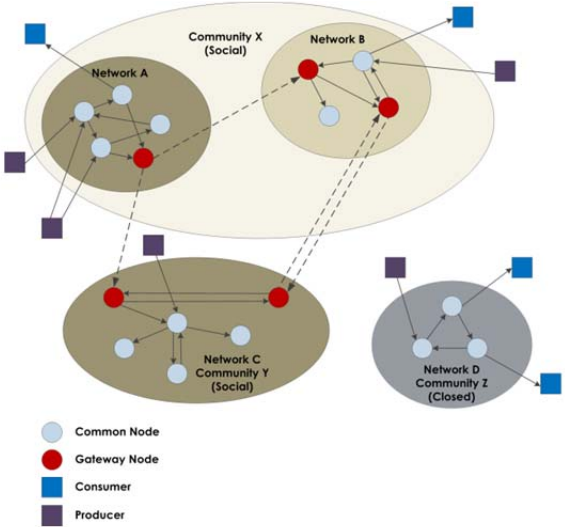


Figure 4. Learning Registry Network Model

Data Model

The data stored at the nodes that describe a learning resource is held in a two-part *resource description* data model:

- *Envelope*: general information common to all data objects.
- *Payload*: optional detailed object type and schema-specific data about a learning resource.

The envelope contains key attributes:

- *Resource Locator*: the resource locator (a URL) provides a unique identifier for the learning resource. It provides a way to reference the learning resource, to access it, and to assemble collations of data that describe it.
- *Data Provider*: the separate identity for each person or organization that:
  - owns the resource being described,
  - owns and curates the metadata or paradata about the resource, and
  - is submitting the resource description.

  These values are used to allow consumers to access and filter resource descriptions that come from specific providers.
- *Terms of Service*: An optional URL identifying the terms of service (ToS) under which the resource description for the learning resource is being provided, e.g., is the resource description public, private, copyrighted, licensed. ToS apply only to the resource description, not to the identified learning resource. Nodes advertise the ToS that they support and may reject resource descriptions that do not contain an acceptable ToS. ToS provide a mechanism to support approved reuse and sharing of resource descriptions.
- *Digital Signature*: Submissions are digitally signed using OpenPGP (OpenPGP, 2007). The signature supports message integrity and tamper resistance, and provides a provable assertion that the submitter is in control of their identity, thus enabling a mechanism to build trust and reputation for submitters.
- *Hashtags*: A collection of unstructured tags that describe the learning resource. Hashtags provide an alternative to formal metadata, and a way to associate arbitrary information with a learning resource.
- *Weight*: A scaled value (-100:100) used by the submitter to associate a weighted trust or reputation assertion with the data.
- *Workflow Data*: information such as message IDs, versions, time stamps, transit nodes, etc., used to manage the resource description as it flows through the distribution network.

The optional payload contains additional detailed data about a learning resource. For metadata and paradata, the payload consists of:

- *Payload Schema*: the designation of the schema and storage model for the payload.

- *Resource Data*: the payload data that describes the learning resource, e.g., the typical metadata that describes a learning resource. The data can be inline or referenced via an external link.

Any schema and type of metadata or paradata is permitted, e.g., IEEE LOM (LOM, 2002), METS (METS, n.d.), Dublin Core (DC, 2010). The resource data may be in any form: JSON, CVS, XML, binary, etc. The data consumer may process only the schema and data types that it understands.

One of the alternative payloads is an assertion. An assertion states an n-way relationship of a specific type between learning resources, identities, resource descriptions, nodes, etc. For example, you can assert that two different resource locators refer to the same learning resource, that a learning resource is aligned to a specific curricular standard or learning objective, that a submitter is injecting spam into the network, that a node is no longer operational, etc. The vocabulary of assertions is open and a recommended vocabulary of assertions is defined.

The resource description data model is open; additional attributes and additional structured types and payload may be added. Most values are optional, and data types are flexible.

In addition, each node has a collection of data that describes the node, the current status of the node, the network and community that the node is a part of, the policies of the node, network and communities, and the available APIs and descriptions of services at a node. All of this information is stored in data rather than in code, enabling discovery of node and network capabilities and self-documentation. Each node contains a full copy of all data that describes the network and community that it is a part of, eliminating centralized storage of this data.

Except for payloads, all data is represented in JSON key-value pairs. All data is stored in a NoSQL, document-oriented, schema-free database (non JSON values are string encoded). Resource descriptions are segregated from node and network configuration data since resource descriptions flow through the nodes in the network and node and network descriptive data is local to a node and is not distributed.

Services and APIs

Each node in the network may offer one or more services, exposed through RESTful APIs. Services are grouped by functionality. Base services include:

- *Distribution Service*: used to distribute resource descriptions from one node to another node, providing the essential capability to flow data through the network. The service can filter incoming resource descriptions on any criteria specific to node policies. Distribution from a node to its outbound partners is triggered periodically, based on node policies and data volumes.
- *Publish Services*: used to *push* learning resource descriptions (metadata, paradata), assertions, etc., from external data providers into a node for distribution through the network.
    - *Publish*: The publish API supports publishing one or more resource description directly to a specific node. The node checks the data, and if acceptable, adds it to the node's data store.
    - *SWORD*: The SWORD API (SWORD, 2009) lets a SWORD client submit resource descriptions directly to a node.
    - *OAI-PMH Intermediary:* Nodes do not harvest data sources. A data provider may configure an OAI-PMH intermediary that harvests data from the source, transforms it into a resource description and uses the publish API to push it to a node. Other similar *data pump* utilities can be created as proxy publishers for other data sources.
- *Access Services*: used by consumers to *pull* resource descriptions and other data from a node for external processing.
    - *Obtain*: The Obtain API gets the list of all resource descriptions or all resource identifiers held at a node, gets the data for a list of resource description identifiers, or gets the collation of all data for a given resource identifier.
    - *Harvest*: The Harvest APIs support OAI-PMH-like (OAI-PMH, 2008) harvest of data from a node. The APIs may be used to return the complete raw JSON resource descriptions (by document ID or collated by resource identifier), or the payload data, in XML, for a specified payload schema.
    - *Slice*: The Slice API returns a subset of the resource description data to a consumer. The API provides limited views into the data, e.g., subset by identity, by schema, by keyword value.
- *Administrative Services*: used to discover the state of a node and the network, e.g. its description, status, policies, services. Used for discovery and reporting.
- *Broker Services*: used to manipulate the resource descriptions as they flow through the network, e.g., make assertions, perform deduplication. Broker services augment data.

You may notice the lack of a search or query API. This is intentional. Given the variety of query languages, search models (faceted, keyword, free text), ranking algorithms, etc., devising a simple, effective search API is difficult. We defer to consumers and third parties to define and implement appropriate community-specific search APIs.

Each service and API is documented in a machine- and human-readable service description stored at the node. The description contains sufficient information for service discovery.

A node can decide which services are deployed on the node. Except for the distribution service, all services are optional (for security, certain services may not be deployed on gateway nodes). The set of services is not closed. Anyone is free to deploy any other services at any node.

The services are all designed to be functionally independent from each other. The services deployed at a node may share a common code base, or they may be developed on different software stacks (e.g., Java, .net, Ruby, Perl/PhP/Python) and deployed independently. The node is a logical entity, not necessarily a single physical hardware/software platform.

A node's services may be deployed across multiple real or virtual platforms, all sharing a single node data store. This data store may also be sharded, distributed or replicated across different physical stores. The service model provides a data abstraction layer that separates the data storage from the producer- and consumer-facing APIs.

Stack

The network, data models and services are assembled into a stack model that provides for resource sharing and the development of applications and user communities as illustrated in Figure 5.

At the bottom level are the learning resources, held external to the Learning Registry Network. Resources may come from any source: federal, individual, private, commercial. They may be open (OERs) or commercial.

The second level—and core of the Learning Registry—is the resource distribution network, with its nodes, each offering a set of services and each holding a collection of resource descriptions. Descriptions of resources are held in the bottom layer, and other data and assertions are pushed into the network through the publishing APIs. The resource distribution network flows the data to other nodes.

At the third level, organizations and businesses build the systems and applications on top of the resource distribution network and its APIs. These applications can be used for publishing, to manage and curate data, , providing feedback, for access and discovery, etc. The Learning Registry places no constraints on how the data is used, the types of applications built, or the business models supported.

At the top level are the communities of learners and educators that use these applications and tools. These communities only see the applications and tools; they do not know that these are supported by the Learning Registry.

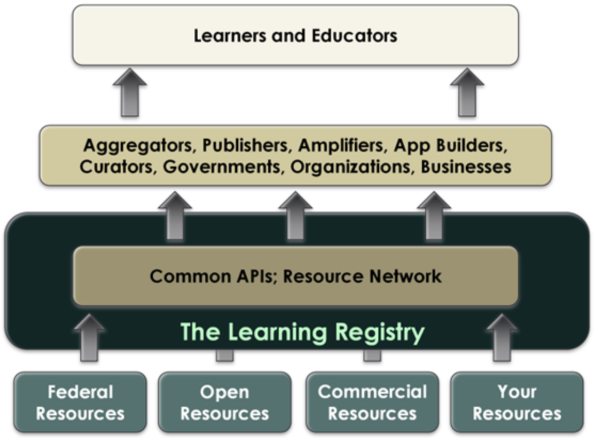

Learners and Educators
Aggregators, Publishers, Amplifiers, App Builders, Curators, Governments, Organizations, Businesses
Common APIs; Resource Network
The Learning Registry
Federal Resources
Open Resources
Commercial Resources
Your Resources

Figure 5. Learning Registry Stack

THE LEARNING REGISTRY PROJECT

Project Team

The Learning Registry project (learningregistry.org) is an informal collaboration among several federal agencies that share the same goal: making federal learning resources and primary source materials easier to find, access and integrate into educational environments. The lead members of the collaboration are the Office of the Under Secretary of Defense for Personnel and Readiness, Training Readiness & Strategy (OUSD P&R, TR&S), and the Office of Educational Technology at the US Department of Education.

In addition we have been working with: the National Science Foundation (NSF), the White House Office of Science and Technology Policy (OSTP), the Department of Energy, the Federal Communications Commission (FCC), the National Institute of Standards and Technology (NIST), the National Archives and Records Administration (NARA), the National Aeronautics and Space Administration (NASA), the

Smithsonian Institute, the Institute for Museum and Library Science (IMLS), the data.gov team, the Federal CIO and CTO, and the Department of Defense Education Activity (DoDEA).

While we focus on the availability of federal resources, our approach and goals are shared with others worldwide. We are talking with and working to leverage the activities of many others, including: The National Science Digital Library (NSDL), American Institutes for Research, ISKME / OER Commons, PBS, BBC, Connexions, Creative Commons DiscoverEd. Mozilla, European SchoolNet, Globe, Ariadne, UK Joint Information Systems Committee (JISC), Education Services Australia, Agilix, Capstone Digital, Cambridge Publishing / Global Grid for Learning, Navigation North, Team Carney, University of Nottingham, UK / Open Nottingham, Oxford, SAKAI, University of Michigan / Open Michigan, Tennessee STEM Innovation Center, University of Catalonia, and New York University.

Technical support is provided by SRI International under contract to the US Dept. of Education and Lockheed Martin Global Training and Logistics under contract to the US Dept. of Defense.

Project Timeline

The project was officially kicked off in July 2010 in an announcement by US Secretary of Education, Arnie Duncan. Over several months we explored over 300 different existing approaches, projects, resource collections, alternative technologies and potential tools. We settled on our base approach in late 2010.

We have a draft technical specification (Learning Registry, 2011), a code base that provides the base functionality outlined above and an initial collection of user, developer and deployment documentation. As of June 2011 we have deployed an initial public distribution network based on our 0.2x code release, and plan to have a 0.5 version code release later in 2011. We work in 6-week development sprints and are focusing on essential enabling features. Our planned focus for 4Q 2011 and beyond is partner integration and supporting the development of user communities, value added services and tools.

Project Operations

All of our activities are open and transparent. We believe in open processes, open products and open data. Anyone who wants to make a difference is welcome to join with us. We have open mailing lists for information sharing, and open teleconferences for technical and design discussions. We disseminate our results in near real time and engage broadly with the community. We build on and use open source tools and components. All of our documents and materials (http://goo.gl/amOYF) are released under open licenses, e.g., CC-BY-3.0. Our code is available on GITHub (http://git.learningregistry.org/), and is released under the Apache 2.0 license. These liberal licenses allow others to build commercial businesses on top of our work.

While any node can establish any policies over its operations, we impose open data policies on the nodes in the public Learning Registry network, i.e., the network that the Federal partners operate. We require, via the ToS, that resource descriptions submitted to the public network are licensed under liberal terms, e.g., CC-BY-3.0 (CCBY3.0, n.d.) or a public domain declaration such as CC0 (CC0, n.d.) or PDDL (PDDL, n.d.). This ensures that the metadata, paradata, assertions, etc., are open and can be freely shared, reused, remixed and aggregated. Note, these ToS apply only to the resource data and assertions submitted to the nodes in the public Learning Registry network. The described learning resources are held in collections outside of the network, and may be subject to other license terms and conditions. We permit descriptions of both open educational resources and restricted commercial resources.

Learning Registry 0.2x

Our prototype implementation is based on CouchDB (CouchDB, n.d.). Couch (cluster of unreliable commodity hardware), an Apache open source project, is a document-oriented NoSQL database for JSON documents, accessed and managed through RESTful APIs. It provides map-reduce and view functionality and the ability to execute JavaScript code stored in the JSON store. It is built on the Erlang OTP platform. Individual Couch instances can be connected into a high-latency, loosely connected network of master-master synchronizing database instances.

While the Learning Registry design is not tied to Couch, it is inspired by its approach, e.g., JSON documents, use of map-reduce, JavaScript code, and creating a network of distributed, synchronizing data stores for metadata and paradata. Learning Registry nodes map to CouchDB instances and data distribution is performed through Couch replication. The Learning Registry APIs abstract the underlying Couch functionality, provide a data abstraction layer and add business logic.

While Couch can provide some business logic, not all functionality can be embedded in JSON and JavaScript. The prototype service and API code uses Python and the Pylons web application framework for business logic. The code can be installed on Linux, MacOS or Windows.

We have deployed three Learning Registry Network communities:

- *Test*: a community with a single network used by project collaborators to test their data and processes before publishing to the public network.
- *Development*: a community with two networks used to develop and test code.
- *Production*: the public production network that the Federal partners operate. We currently operate 3 nodes in the public network; more will be added as dictated by demand.

Several collaborators have deployed private nodes or are planning to deploy nodes on the public network.

A Learning Registry node can be deployed on dedicated hardware or virtual machines. The nodes in the test, development and public networks are currently deployed in the cloud on Amazon EC2. An AMI (Amazon Machine Instance) is available for anyone who wants to establish their own node in EC2.

We are currently testing with a small collection of approximately 0.5M metadata and paradata resource description documents. Collaborators are beginning to publish data into the production network.

Community Applications

In addition to publishing metadata and paradata into the network, our collaborators have started to build a collection of tools and applications on top of the Learning Registry and integrate it into their communities.

A search and discovery interface has been developed using Elastic Search (thus providing a query API). Elastic Search interfaces directly with Couch, and indexes JSON documents via Lucene (by ADL).

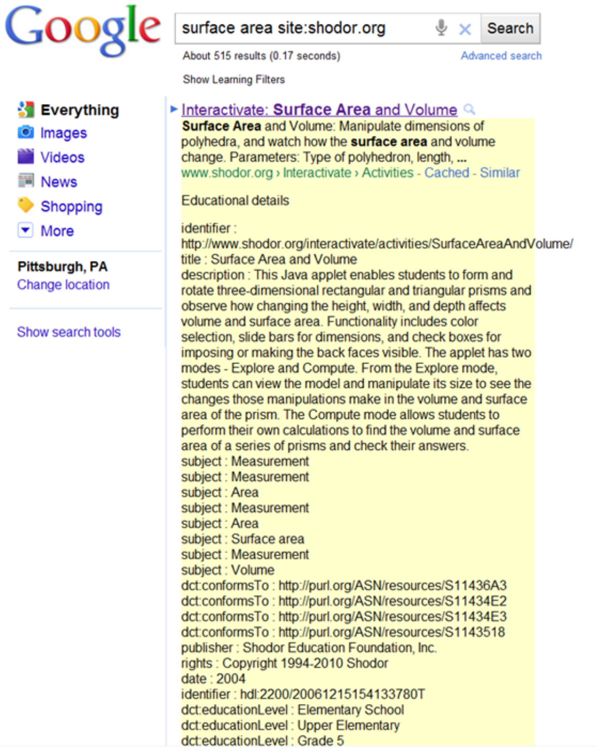
A Drupal publishing module has been created that can be installed in any Drupal instance. When a learning resource is published to a Drupal site, the corresponding resource description is automatically added to the Learning Registry (by Open Michigan).

Bookmark data mining takes resource identifiers from the Learning Registry and looks up these resources in social bookmarking services, e.g., Delicious, Diigo. The bookmarking service APIs are used to retrieve additional data about the resource, e.g., folksonomy keywords used to classify it, and this data is added to the Learning Registry, amplifying the description of the resource (by OER Commons).

Various organizations have aligned learning resources with curricular maps, e.g., assigning the ASN identifier to a learning resource. This enables discovery of resources aligned to curriculum (by Agilix).

While the Learning Registry is designed to store data about learning resources, it can also be used to store data about learning tools. A service description model for tools that have a BLTI (Basic Learning Tools Interoperability) API has been developed and service descriptions of tools have been published into the registry. When a tool is discovered from the Learning Registry, an application can automatically build the tool API invocation string and launch the tool for the learner (by University of Catalonia).

Learning Registry data about a learning resource can be surfaced in search engine output. A browser plugin has been developed such that when a user searches Google, the results are looked up in the Learning Registry and the entire collation of metadata and paradata for the resource is added to the results page. This is illustrated in Figure 6 (by Oxford).



About 515 results (0.17 seconds)

Interactivate: **Surface Area** and Volume

Surface Area and Volume: Manipulate dimensions of polyhedra, and watch how the **surface area** and volume change. Parameters: Type of polyhedron, length, ...

www.shodor.org › Interactivate › Activities - Cached - Similar

Educational details

identifier : http://www.shodor.org/interactivate/activities/SurfaceAreaAndVolume/
title : Surface Area and Volume
description : This Java applet enables students to form and rotate three-dimensional rectangular and triangular prisms and observe how changing the height, width, and depth affects volume and surface area. Functionality includes color selection, slide bars for dimensions, and check boxes for imposing or making the back faces visible. The applet has two modes - Explore and Compute. From the Explore mode, students can view the model and manipulate its size to see the changes those manipulations make in the volume and surface area of the prism. The Compute mode allows students to perform their own calculations to find the volume and surface area of a series of prisms and check their answers.
subject : Measurement
subject : Measurement
subject : Area
subject : Measurement
subject : Area
subject : Surface area
subject : Measurement
subject : Volume
dct:conformsTo : http://purl.org/ASN/resources/S11436A3
dct:conformsTo : http://purl.org/ASN/resources/S11434E2
dct:conformsTo : http://purl.org/ASN/resources/S11434E3
dct:conformsTo : http://purl.org/ASN/resources/S1143518
publisher : Shodor Education Foundation, Inc.
rights : Copyright 1994-2010 Shodor
date : 2004
identifier : hdl:2200/20061215154133780T
dct:educationLevel : Elementary School
dct:educationLevel : Upper Elementary
dct:educationLevel : Grade 5

Figure 6. Google Search Results Amplified by Learning Registry Data

In this example, the learning resource comes from one organization (Shodor.org), the curriculum description from a second (ASN), the resource metadata is curated by a third (NSDL), alignment of the resource to the curriculum was done by a fourth (CTE Online), and a fifth organization submitted the resource description to the Learning Registry (SRI). There are multiple metadata records and paradata records describing the learning resource in the Learning Registry. The original metadata and paradata pre date the Learning Registry. Oxford independently developed the browser extension that leverages the Google search infrastructure, allowing anyone to use Google to expose Learning Registry data. None of these organizations directly collaborated with any of the others; the work was enabled by utilizing the common social metadata timeline of the Learning Registry and the open sharing and amplification of knowledge that it enables.

These exemplify the types of capabilities that we thought could be developed, but none were planned as part of the design of the Learning Registry. We anticipate that the community will develop many other interesting, innovative and unanticipated uses for the Learning Registry.

ACKNOWLEDGEMENTS

Our chief collaborator, Steve Midgley, Deputy Director, Office of Educational Technology, US Department of Education, has filled the fundamental role of shaping the goals and directions of the Learning Registry. The efforts of many participants in the Learning Registry community have contributed to the planning, design, development, implementation and testing of the Learning Registry. While too numerous to individually identify, their individual and collective contributions are graciously acknowledged.

REFERENCES

CC0, (n.d.). Creative Commons CC0 Public Domain Dedication, http://creativecommons.org/publicdomain/zero/1.0/

CCBY3.0, (n.d.). Creative Commons Attribution 3.0 Unported (CC BY 3.0), http://creativecommons.org/licenses/by/3.0/

CouchDB, (n.d.). *The Apache CouchDB Project*, http://couchdb.apache.org/

DC, (2010). *Dublin Core Metadata Element Set*, Version 1.1, http://dublincore.org/documents/dces/

Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters, in *OSDI'04: Sixth Symposium on Operating System Design and Implementation*.

JSON, (2006). *The application/json Media Type for JavaScript Object Notation (JSON)*, RFC 4627, IETF.

Learning Registry, (2011). *Technical Specification*, http://goo.gl/2Cf3L

LOM, (2002). *IEEE Standard for Learning Object Metadata*, IEEE Std 1484.12.1™-2002, IEEE Computer Society, September 2002.

METS, (n.d.). *Metadata Encoding & Transmission Standard*, http://www.loc.gov/standards/mets/mets-schemadocs.html

Netflix, (2009). *Netflix Prize*, http://www.netflixprize.com/

NoSQL, (n.d.). *NoSQL*, http://en.wikipedia.org/wiki/NoSQL

NSDL, (n.d.). *National Science Digital Library*, http://nsdl.org./

Ochoa, X. & Duval, E. (2006). Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects, *Proceedings of the 1st International Workshop on Contextualized Attention Metadata*.

OAI-PMH, (2008). *The Open Archives Initiative Protocol for Metadata Harvesting*, V2.0, http://www.openarchives.org/OAI/openarchivesprotocol.html

OER Commons, (n.d.). *OER Commons, Open Educational Resources*, http://www.oercommons.org/

OpenPGP, (2007). OpenPGP Message Format, RFC 4880, IETF.

PBS, (n.d.). *PBS Teachers*, http://www.pbs.org/teachers/

PDDL, (n.d.). ODC Public Domain Dedication and Licence (PDDL), http://www.opendatacommons.org/licenses/pddl/1.0/

REST, (2000). Fielding, R. T. “Representational State Transfer (REST)”, in *Architectural Styles and the Design of Network-based Software Architectures*, Doctoral dissertation, University of California, Irvine.

SWORD, (2009). *SWORD AtomPub Profile*, V 1.3, http://www.swordapp.org/docs/sword-profile-1.3.html

VanGundy, S. (2010). *What is Paradata*, http://nsdlnetwork.org/stemexchange/paradata