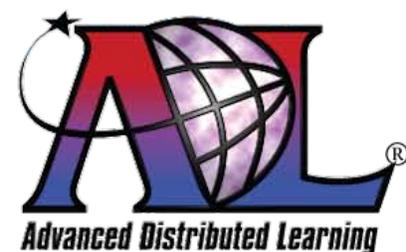


The ADL Registry and CORDRA

Volume 1: General Overview

August, 2008



Dedication

The Institute for Defense Analyses and the Advanced Distributed Learning Initiative dedicate this document set to the memory of Philip Dodds. He provided leadership, vision, and guidance that made the ADL Registry possible. His many contributions to distributed learning and content sharing are of immense significance and abiding value.



The ADL Registry and CORDRA

Volume 1: General Overview

Managing editor:
Philip Dodds
Institute for Defense Analyses

Technical editor:
Scott Lewis
Institute for Defense Analyses

Contributors

| | | | |
|--|---|--|--|
| Institute for Defense Analyses | Corporation for National Research Initiatives | Concurrent Technologies Corporation | Stanley Associates |
| Bill Blackmon Dexter Fletcher G.A. Redding Eric Roberts | Henry Jerez Larry Lannom Giridhar Manepalli | Joanne Mekis Angelo Panar Eric Sakmar Schawn Thropp | Nina Deibler Jacob Marks Damon Regan |

The Advanced Distributed Learning Initiative, Alexandria, VA

Copyright

This work was conducted under contract DASW01-04-C-003, Task BE-2-1624, for the Office of the Undersecretary of Defense (Personnel and Readiness/Readiness and Training Directorate). The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be considered as reflecting the official position of that agency.

© 2008 Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, VA 22311-1882 (703) 845-2000.

Reproduction

This material may be reproduced by or for the U. S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (Nov 95).

The appropriate citation of the source document is used as follows:

The ADL Registry and CORDRA, Volume 1: General Overview, The Advanced Distributed Learning Initiative, Alexandria, VA, May 2008.

Distribution

Permission to distribute this document is granted under the following stipulations:

1. The use of this document, its images and examples is for non-commercial, educational, or informational purposes only.
2. The document, including its images and examples are intact, complete and unmodified. The complete cover page, as well as the COPYRIGHT, REPRODUCTION, and DISTRIBUTION sections, are consequently included.

For additional information or questions regarding copyright, distribution and reproduction, contact:

ADL Co-Laboratory Hub
1901 North Beauregard Street, Suite 600
Alexandria, Virginia 22311
USA
703-575-2000

Trademark information

CORDRA, Corporation for National Research Initiatives, and Handle System are trademarks or registered trademarks of the Corporation for National Research Initiatives.

SCORM is a registered trademark of the U.S. Department of Defense.

About this document set

The *ADL Registry and CORDRA* document set consists of five volumes. Each covers different aspects of the Advanced Distributed Learning (ADL) Initiative Registry Project.

Volume 1: General Overview

Volume 1, this document, is an introduction to the ADL Registry and the Content Object Repository Discovery and Registration/Resolution Architecture (CORDRA™) including background, history, high-level requirements, design assumptions, and rationale. It is intended for those seeking a basic introduction and orientation and provides context for the volumes that follow.

Volume 2: Registry Overview and User's Guide

Volume 2 describes the ADL Registry. It includes an ADL Web-site user's guide and an introduction to producing digital-object and transaction metadata. This is the "how-to" volume for those getting started with the ADL Registry.

Volume 3: Technical Reference

Volume 3 provides detailed information for technical personnel and tool developers on using the ADL Registry.

Volume 4: CORDRA

Volume 4 provides more detail on CORDRA. CORDRA can be used by any community interested in establishing its own set of registry services independently of ADL. This volume is intended for those who are investigating solutions to object registration and discovery in general and for those who wish to understand how the ADL Registry fits within the CORDRA architecture.

Volume 5: Services and Tools

Volume 5 describes value-added services and tools that are currently available or in development and how new services can be created to provide extended capabilities to registry systems. Some services and tools are ADL specific; others have more general applicability and may be of value to those creating their own registry system.

Contents

| | | |
|-------|---|----|
| 1 | Introduction..... | 1 |
| 1.1 | CNRI, ADL, and CORDRA | 2 |
| 1.2 | Context, discovery, and resolution in CORDRA..... | 2 |
| 2 | The problem | 3 |
| 2.1 | Why is it a difficult problem? | 5 |
| 2.2 | Limitations of Web searches..... | 6 |
| 2.3 | Federating object repositories | 7 |
| 2.4 | More than a simple database problem | 9 |
| 2.5 | Summary | 9 |
| 3 | CORDRA and the ADL Registry | 9 |
| 3.1 | High-level assumptions and requirements | 10 |
| 3.2 | Relating context, discovery, and resolution | 13 |
| 3.2.1 | Context..... | 13 |
| 3.2.2 | Discovery | 13 |
| 3.2.3 | Resolution | 14 |
| 3.3 | A unifying architecture | 16 |
| | Appendix A. References | 17 |

The ADL Registry and CORDRA

Volume 1: General Overview

1 Introduction

Great strides have been made in creating and using distributed digital objects that benefit training, education, and performance and decision aiding (henceforth referred to as “learning”). Organizing learning materials into relatively small and concise digital objects promises new possibilities for developers and learners. These possibilities include widespread access to a rapidly growing library of sharable objects that can be reused for multiple purposes. However, the promise of reuse has been largely unfulfilled because of architectural and organizational reasons. This document describes the challenges and opportunities facing the digital-object world and introduces a new infrastructure for registering, identifying (discovering), locating (resolving), and accessing objects.

Over the past 10 years, the Advanced Distributed Learning (ADL) Initiative’s Sharable Content Object Reference Model (SCORM[®]) [1]¹ evolved to provide a modular, object-based design approach for digital objects that solved key interoperability and reusability issues across many learning systems in industry and government. SCORM has enjoyed widespread international adoption, has become a *de facto* standard in many learning communities, and is supported by Department of Defense (DoD) policy [2]. While SCORM advances the state of the art in designing and creating interoperable and reusable objects, it does not address finding and reusing objects after they have been created.

In 2003, ADL began work to solve this problem. As its sponsor observed, “It doesn’t do much good to have interoperable and reusable digital objects if people can’t find and use them.” ADL launched an investigation into the difficulties and realities of object creation, storage, and management and uncovered the limitations and problems encountered by others in related fields, such as library science, computer and network systems design, and publishing. As ADL investigated these fields and formulated high-level requirements for the learning community, it quickly found that many problems had not been solved by others and that the problem space was much more complex than it first appeared.

¹ The numbers in brackets correspond to those in the bibliography.

To address these issues, ADL set out to

- Define high-level requirements, policies, and business rules for object repositories that would be practical to implement.
- Identify and apply the most relevant technologies and specifications that could be used to define the architecture.
- Define an architecture on which necessary services could be built.
- Define an architecture that would be scalable.

NOTE – This document uses the term “digital object” to refer to resources that may be registered in the ADL Registry. In addition to objects that contain content for display to the learner (content objects) and objects with specific educational and training goals (learning objects), a digital object, as used here, may be anything of value to the learning community, such as a simulation, mathematical model, teaching-strategy algorithm, glossary, technical manual, or style guide. A digital object may also be a collection of resources encapsulated in a package, such as a ZIP file.

1.1 CNRI, ADL, and CORDRA

Early in the process, ADL partnered with the Corporation for National Research Initiatives[®] (CNRI) to survey the state of the art in object collection, management, and use and to develop an architecture that could address the unique requirements of the ADL communities. Through this partnership, ADL has been able to take advantage of CNRI’s considerable knowledge and experience in designing and deploying complex, Internet-based, infrastructure systems and technologies. CNRI’s many years of involvement with the evolution of the Internet and the World Wide Web and its role in providing guidance and support to key industry members combined with ADL’s experience in learning technologies has resulted in a new approach to the discovery and use of distributed digital objects. The result is the Content Object Discovery and Registration/Resolution Architecture (CORDRA), which is described in detail in Volume 4. The ADL Registry is ADL’s implementation of CORDRA to serve the DoD learning community and is described in more detail in Volume 2.

The ADL Registry is the first publicly available CORDRA implementation. The project went live in December 2005. The ADL Registry provides a mechanism to search for digital objects within DoD and enables their discovery and reuse by the DoD learning community and by other communities concerned with learning technology.

1.2 Context, discovery, and resolution in CORDRA

Obtaining digital objects that precisely meet the needs of learners and developers involves three processes: *contextualization*, *discovery*, and *resolution*. Until now, these processes have not been integrated into a single coherent and comprehensive solution. As shown in Figure 1, CORDRA addresses this integration challenge. With CORDRA, *context* defines *discovery* criteria, which then *identify* relevant objects that can be

resolved to a location and then *retrieved* and *delivered* to the user. These processes are discussed further in Section 3.2.

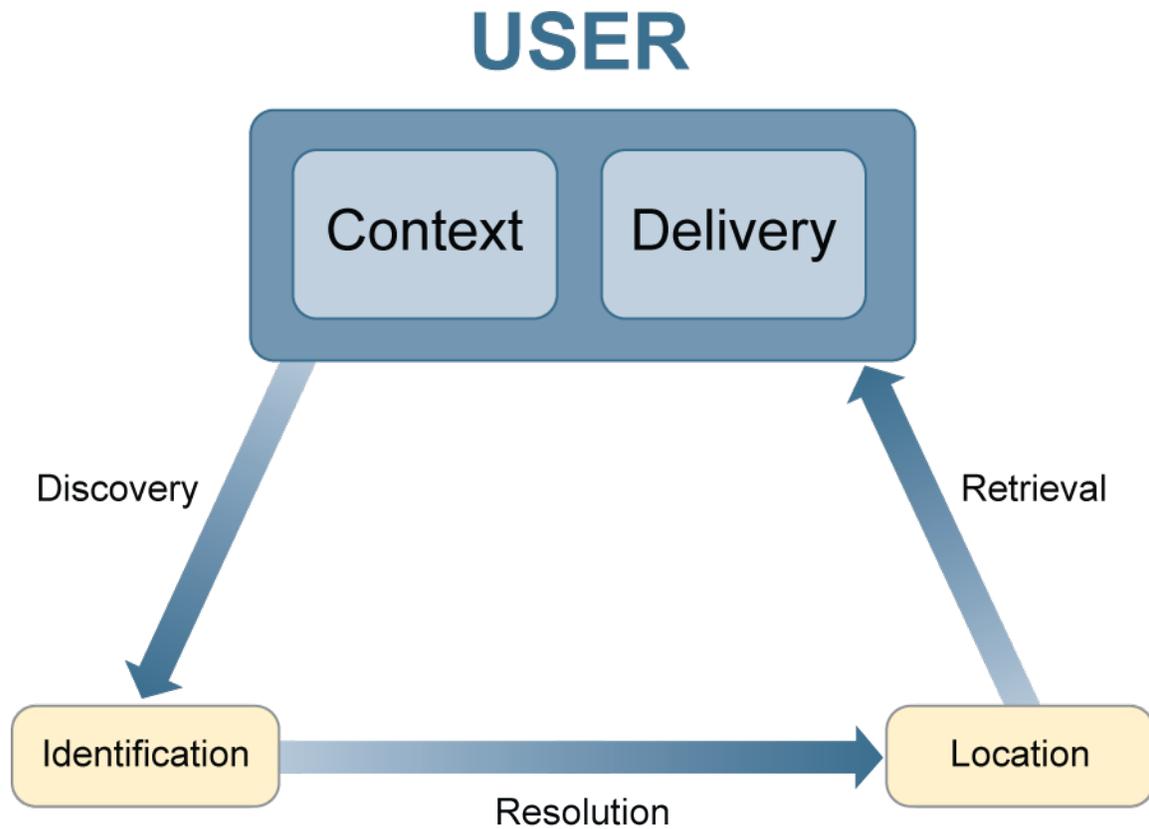


Figure 1 – The CORDRA triangle

2 The problem

Throughout ADL and other learning communities, new digital objects are being created daily by hundreds of developers spread over multiple programs, organizations, and missions. Many of these objects could be applied in many other learning contexts. Every day, the volume of these valuable resources increases. Yet, their existence is virtually unknown outside of local and relatively small groups of users and developers. The resources are created in and exist in isolated islands, limiting their visibility, access, and reuse.

Compared to publishing and library science, the learning community has been described as “very messy” and, therefore, difficult to organize on any large scale because of the lack of common practices for creating, storing, and describing digital objects. As shown in Figure 2, multiple communities of object creators and users exist with multiple repositories but have severely limited means of bridging and reusing learning material between or even within communities.

Multiple Communities of Practice

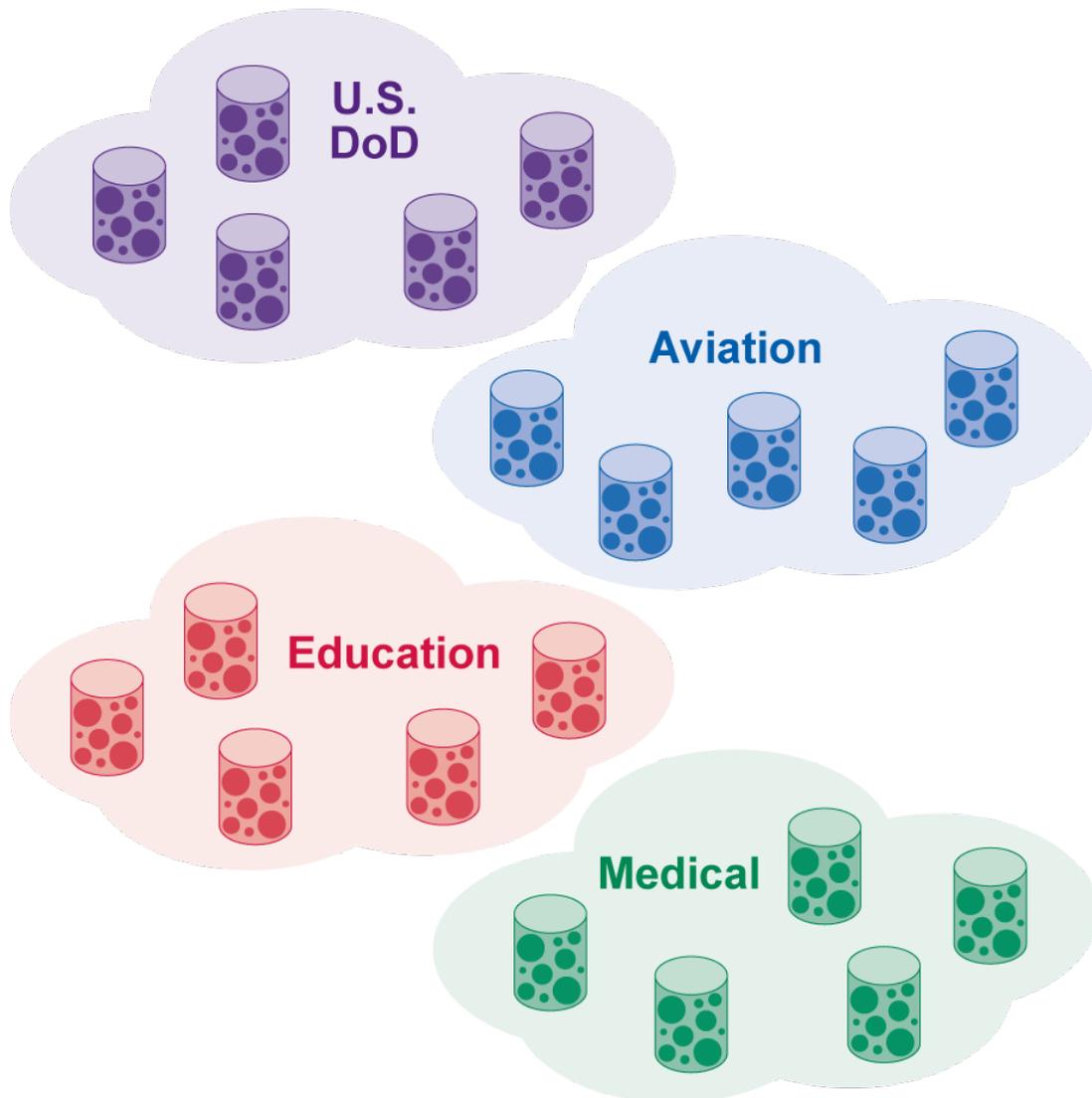


Figure 2 – Isolated communities of digital-object creators and repositories

Creating, managing, describing, and making digital objects accessible is expensive. However, making objects widely visible and accessible should increase the return on investment in them. Because no infrastructure has existed that reinforces such activities, this has not happened.

To increase return on investment, a simple means is needed to make digital objects visible and accessible to primary users as well as to new audiences. Increasing this return also requires a capability to allow delivery of the right objects to the right user at the right time, financial incentives to invest in the production of quality materials, and improved life-cycle management.

2.1 Why is it a difficult problem?

Successful digital object discovery, location, and access requires a variety of technical, architectural, and organizational solutions. Some requirements can be met partially or wholly through known technologies. Others have not been addressed, in part because of the fragmented and specialized nature of the learning community and because most research and development addresses major business areas of the Internet, such as general searching and social networking. Special requirements of smaller communities, such as the learning community, have been lost in the shadows of major Internet markets. Enabling successful discovery, location, and access of digital objects is difficult because many of the requirements of learning communities are not addressed by mainstream Internet developers and because many of the requirements are unique to learning.

Requirements for the ADL communities and learning communities in general to effectively discover, locate, and reuse digital objects include system capabilities and tools that

- Register and expose objects with minimal impact on developers and managers.
- Provide immediate value to those who register objects within their own organization, as well as to others outside their organization.
- Accommodate variations in local administrative policies.
- Provide a common means for describing objects, which can vary according to subject and organizational domain.
- Provide a means of federating information about objects from multiple sources.
- Include a framework to make information about objects visible across multiple communities.
- Enable objects to be precisely located with references that remain valid over time despite changes in location and ownership.

- Support and encourage the development of value-added services that meet local policy, business rules, and life-cycle management and that increase return on investment in object creation.

None of these requirements are particularly challenging when considered individually. Taken together, however, the requirements are challenging because of their number and complexity. CORDRA provides a relatively simple, robust solution.

2.2 Limitations of Web searches

Few would question the value of Web-searching technologies such as Google and Yahoo. However, these technologies fail to address important, learning-community requirements. Two key limitations are as follows:

- General Web searches use indexed data that are created by “crawling” through the *entire* World Wide Web to see what is accessible and machine readable. Many digital objects are inaccessible to Web crawlers because of policy or lack of infrastructure, and if these objects are accessible, they are either in formats Web crawlers cannot decode or do not include information that is easily interpreted and indexed.
- General Web searches find everything that *might* be relevant and offer no reliable means to filter digital objects for authenticity, validity, currency, and other criteria to limit the results more precisely to relevant learning material.

Web searching using Google and similar services provides great value on many levels and will no doubt improve over time as search algorithms become more sophisticated. However, mission-critical use cases cannot afford the hit-or-miss nature of today’s index-everything search strategies. The solution to finding material that meets specified learning objectives, has well-crafted instructional and informational value, and has been vetted, authorized, and made available for use by those who need it remains a serious unsolved technical and organizational problem.

Key to solving this problem is the way in which indexes are built for digital objects. The ADL Registry uses descriptive information called metadata to provide the needed consistency. Objects are “tagged” with a common set of metadata that describes them. Specific processes for registering metadata are provided, and business rules specific to the ADL DoD community are defined and enforced. As shown in Figure 3, unlike currently available search engines, the registry approach creates a master metadata index to allow the discovery of specific objects described by associated metadata – an approach that is both more precise in discovering relevant objects and faster than the “text crawling” in more common use today.

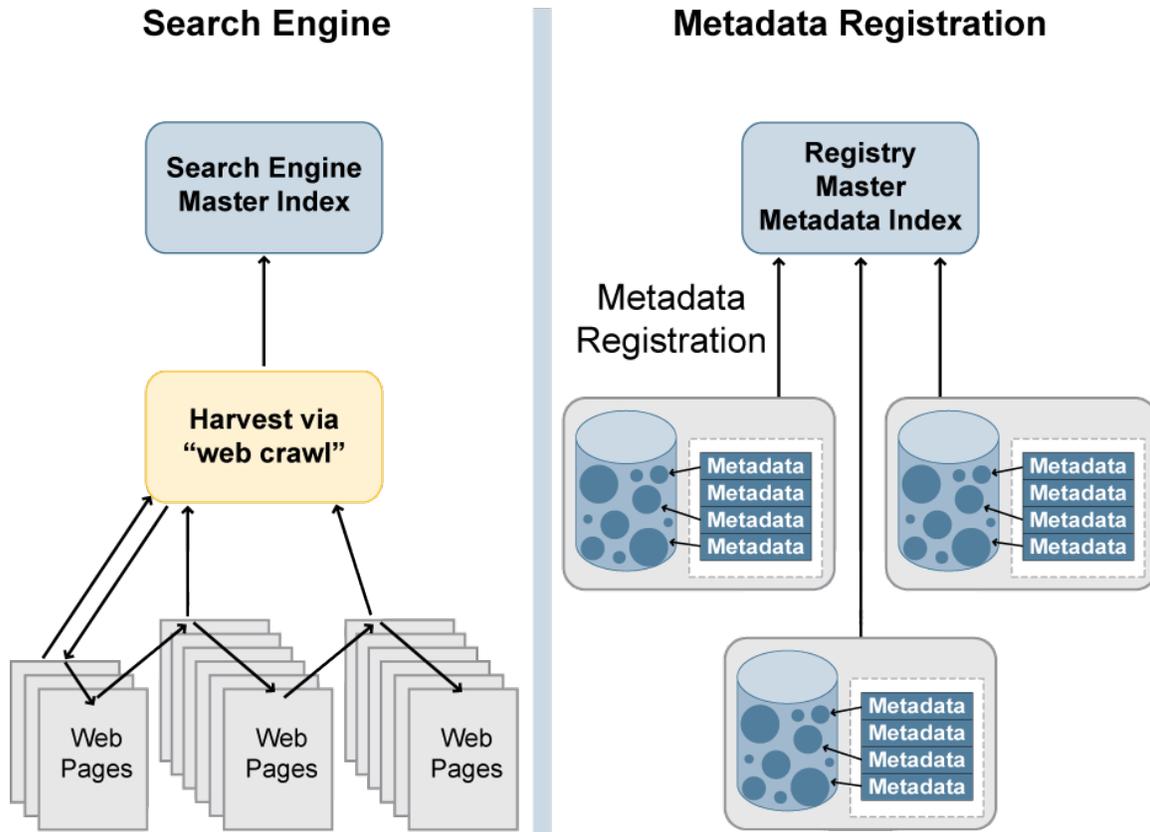


Figure 3 – Search engines vs. metadata registration

2.3 Federating object repositories

The primary goal of the ADL Registry is to provide the means to register, search for, and discover digital objects developed by many independent developers. The registry assumes that objects created locally will be stored in digital repositories. It makes no assumptions about how repositories are implemented or administered. This approach has minimal impact on repository business practices by accommodating a wide variety of implementations, business rules, and workflows.

As shown in Figure 4, the ADL Registry addresses the needs of DoD. Other CORDRA registries can be implemented for other communities. The ADL Registry provides a means to gather, or federate, information about digital objects from multiple sources and then index that information so that developers and users can discover, locate, and access the objects they need.

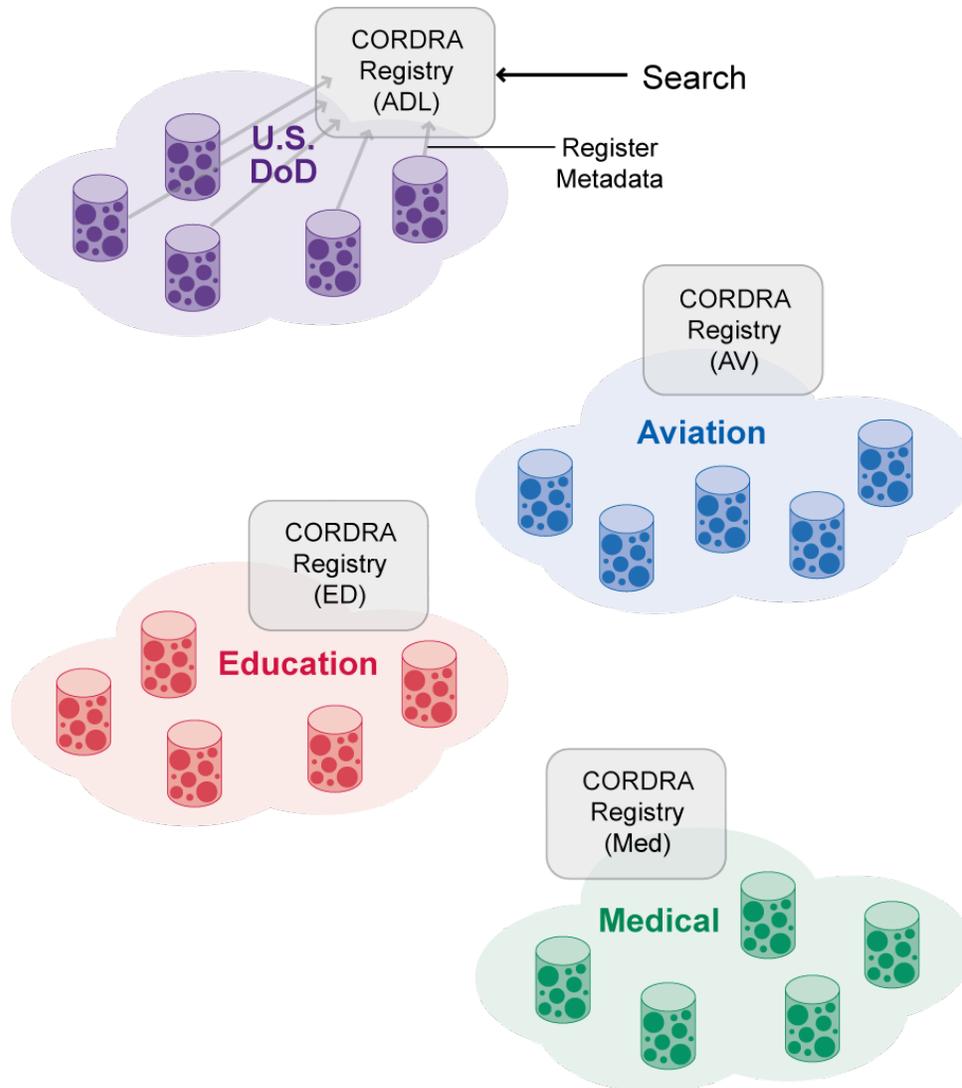


Figure 4 – Searching within a domain: examples of different communities, each with a searchable CORDRA registry

To successfully federate many separate repositories, common metadata approaches must be developed, local business practices reflected, and individual access rights and policies administered. These requirements present a set of ADL-specific design and technical challenges that must be addressed. However, ADL’s goal is a system that not only meets the requirements of the ADL communities, but can also be modified and adapted by other communities so that the underlying systems and interfaces are common, even if some aspects are localized. This means that the problem of federating repositories is two tiered. One tier serves a general, customizable model (CORDRA), and the other supports a specific community, such as the ADL DoD community. The design of an architecture that can accommodate and federate multiple communities is non-trivial.

2.4 More than a simple database problem

At first glance, accumulating metadata from multiple sources that can be centrally indexed and searched appears to be a simple database problem. Once the metadata have been assigned and collected, this is basically true. The problems lie in designing an infrastructure for which services and tools can be built to collect relevant metadata from multiple sources, populate the database, and properly administer, process, and use the metadata in it.

For example, the ADL Registry and CORDRA include a set of robust services for creating persistent, unique identifiers that can be used to determine information about digital objects. This system is as critical to the registry as the Domain Name System (DNS) is to the Web. It is called the Handle System[®] and was developed by CNRI [3]. The same system is used by the publishing industry, among others, and has been proposed as a more valuable and useful approach than DNS.

Designing and building a custom system to manage these processes would be difficult but not overly so. Such systems are built and deployed frequently using a combination of off-the-shelf products, such as databases and servers, and custom middleware. However, these systems typically work for one client in one specific environment. Building an extensible system based on a scalable architecture that can be adapted across multiple communities based on state-of-the-art Internet technologies is more difficult – but more useful in the long run.

2.5 Summary

Rather than build an isolated system with brittle features and capabilities, CNRI and ADL proposed an extensible architecture, CORDRA, that would accommodate the needs of different communities and include new services and capabilities. CORDRA is modular, customizable, and scalable, and accommodates problems that will inevitably arise in the future.

3 CORDRA and the ADL Registry

As Figure 1 suggests, CORDRA is an architecture on which a family of services and tools can be built to support discovering and resolving digital objects for both specific and general contexts. An application of CORDRA in a particular domain involves establishing CORDRA registry services that implement a community's policies and practices. Distributed repositories can then register objects and related metadata so that they can be discovered through search services. A CORDRA application, such as the ADL Registry, is called an "instance" of a CORDRA registry. As shown in Figure 5, instances of CORDRA registries can, in turn, be federated in a "registry of registries."

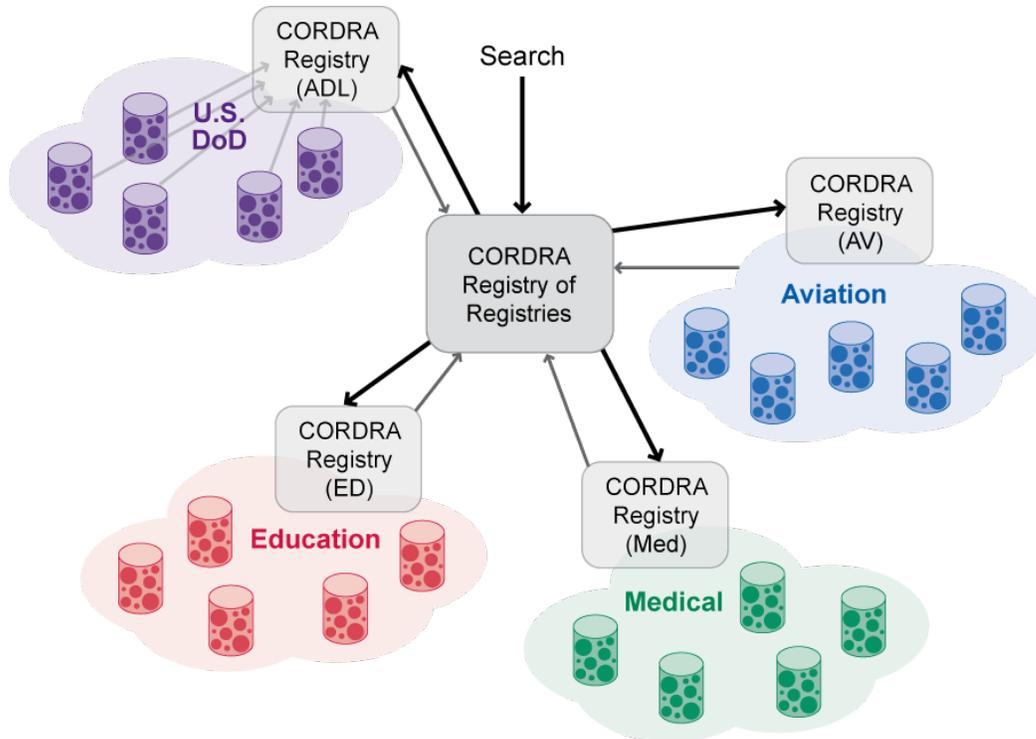


Figure 5 – Different communities federated by a CORDRA registry of registries

ADL and CNRI built and deployed the first operational instance of this architecture, the ADL Registry. CNRI was responsible for much of its development and implementation. The registry went live in December 2005. ADL continues to manage registry development and implementation for DoD.

3.1 High-level assumptions and requirements

Early in the development process, ADL defined a set of assumptions that led to high-level requirements for digital-object registration and discovery [4]. These assumptions are the foundation of the initiative and serve as both guiding and constraining criteria. They also distinguish the goals and objectives of this effort from seemingly related solutions developed within other domains with different criteria. The requirements are summarized in Figure 6 and discussed below.

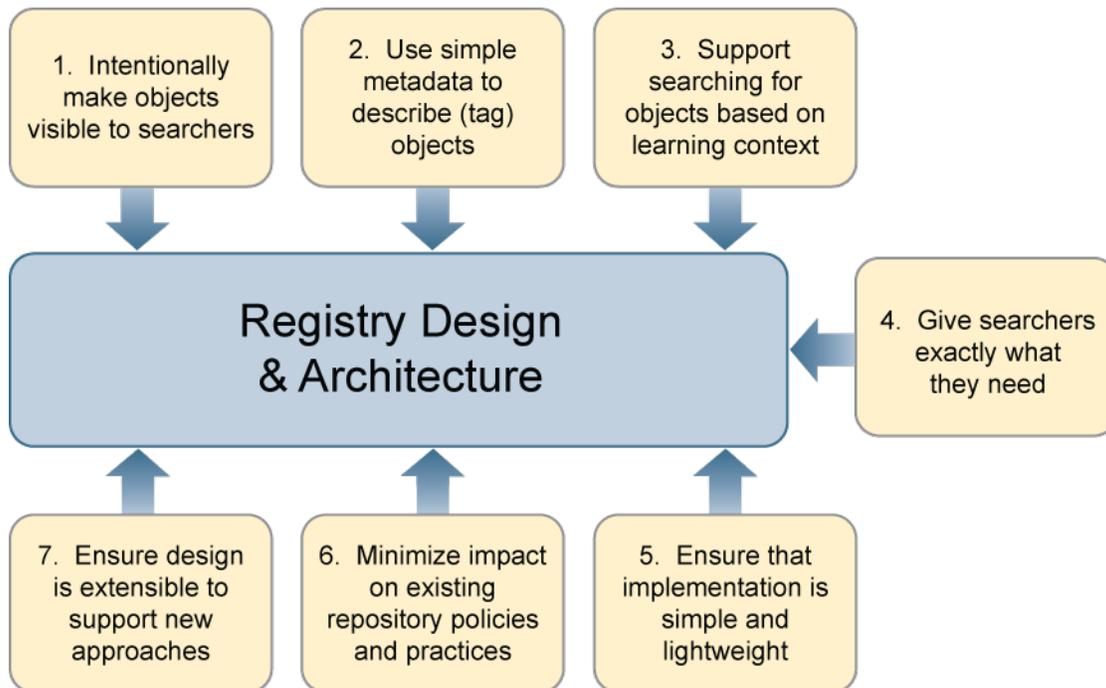


Figure 6 – ADL high-level requirements for CORDRA-based registries

- **Assumption 1:** Publishers want their digital objects to be found.

The act of registering an object in a repository means that its publisher wants to make it discoverable and retrievable. It is probable that the publisher will want to apply rules that regulate who can discover and access the object and under what circumstances, but these are separate issues.

Requirement: A means to discover where contextually appropriate objects are available.

- **Assumption 2:** Most developers are not skilled at tagging digital objects.

Just because an object has been tagged with metadata does not mean that it was tagged with useful search information. Interpretations for when to tag and with what vocabulary vary widely.

Requirement: Guidance and simple interfaces for tagging digital objects.

- **Assumption 3:** Searchers have specific criteria in mind.

Searchers for objects have specific contexts and criteria in mind. They often know what they need based on specific requirements. This might be expressed with a simple description, key words, a specific skill or piece of knowledge, a relationship to other processes, or the state of a learner's profile.

Requirement: A means to relate context-based search criteria to descriptions of specific digital objects, such as mapping a skill definition to an object designed to address that specific skill.

- **Assumption 4:** Searchers want only what they need.

Searchers want discovered objects to match to their requirements as precisely as possible. They do not want every object that *might* pertain. They want the ones that *do* pertain.

Requirement: A means to ensure that discovered digital objects are relevant, accredited, authorized, and deliverable.

- **Assumption 5:** A rigid information service and protocol model will not scale.

Experience and studies have shown that distributed networks that insist on standardizing the details of all protocols, data models, and services to be applied do not scale. A better approach is to standardize as few of these elements as possible while still allowing for the needed level of interoperability.

Requirement: An approach that is low cost and easy to implement with minimal alteration of existing approaches.

- **Assumption 6:** Local policies and business rules must be permitted and enabled.

Widely varying requirements and needs must be accommodated. Enforcing strict procedural compliance will not be broadly supported.

Requirement: The means to institute and expose local policies and business rules so that they can be used or mapped to and from other systems and across different communities.

- **Assumption 7:** We cannot foresee all services and capabilities that eventually will be needed.

Many assumptions need to be made about the functions and features that might be needed or wanted for this architecture. Some of the functions and features that are developed may turn out to be wrong or superfluous. Others will be missing.

Requirement: An architecture that enables services and capabilities to be added or modified without changing the underlying structure.

3.2 Relating context, discovery, and resolution

Obtaining digital objects that precisely meet the searcher's needs involves three related processes: *contextualization*, *discovery*, and *resolution*. These processes and their integration by CORDRA and the ADL Registry are discussed below.

3.2.1 Context

Context provides the criteria required for discovery. It may be simple or complex, and its definition may be automated, semi-automated, or manual.

Searchers for digital objects are driven by specific needs arising from specific contexts. In a simple context, a learner might want to find an object that relates to a particular topic. A more complex context will consider factors such as the learner's characteristics, background knowledge, current progress, learning style, subject matter, environment, instructional objectives, and the instructional strategies being applied.

Imagine that an electronics technician is learning to troubleshoot faults in an unfamiliar avionics system but that he or she has been certified to work on similar systems. A search for relevant objects would need to be informed by (a) the make, model, and version of the avionics subsystem; (b) the skills needed to repair it; (c) the skills the technician has already mastered; and (d) the specific procedures and skills needed for the subsystem at hand. This scenario assumes that

- A database exists with the exact configuration of each avionics system.
- A skills taxonomy exists for the system under consideration.
- A profile exists of the technician's proficiency.
- Digital objects exist for the system.
- Instructional and performance aiding strategies exist to prepare the technician to troubleshoot the system.

Assuming this information exists and is accessible, a service or agent can be developed that can derive the criteria needed to identify objects that are contextually appropriate.

3.2.2 Discovery

Ideally, discovering digital objects might involve a process such as

- Develop search criteria from the local context.
- Go to a master index of relevant repositories.
- Go to one or more likely repositories.
- Discover relevant objects that are available within the repositories.

The problem is that until now no common practice existed for defining search criteria for objects, a master index of repositories did not exist, and the ability to access and search individual repositories was limited at best. To support this discovery approach, a new architecture was needed.

A registry of repositories provides a single place to find digital objects that reside in many locations. Its metadata index can be searched directly or mined by discovery services. Publishers who want their objects to be found can register their objects and provide information to allow their discovery. This approach is relatively simple and scalable because only information about the object and its content is exposed and centralized – not the object’s content itself. This approach also increases the precision of searches by enabling sophisticated search services and by narrowing the scope to intentionally published objects.

We are not adding value to distributed collections by introducing any new or untried search algorithms. Instead, we are providing a framework for collecting metadata that resides in distributed repositories and doing so in a highly structured and predictable fashion.

The ADL Registry provides a search service for users on its Web site and an interface for tool and Web-site developers. Similar services could readily be developed for other communities.

3.2.3 Resolution

The ADL Registry uses the Handle System for resolving a digital object’s location. The Handle System was developed by Robert Kahn and his team at CNRI in the mid 1990s. It defines globally unique identifiers that can be associated with a variety of information about an object. In the Handle System, an object’s unique name or handle is stored on a handle server along with a pointer to the object’s location. The process of obtaining an object’s location is executed by a resolution service that asks the handle server for the location information, among other things. As shown in Figure 7, the key difference between the Handle System and the use of uniform resource locators (URLs) is that when an object’s location changes, its handle does not, allowing the object to be located despite its migration.

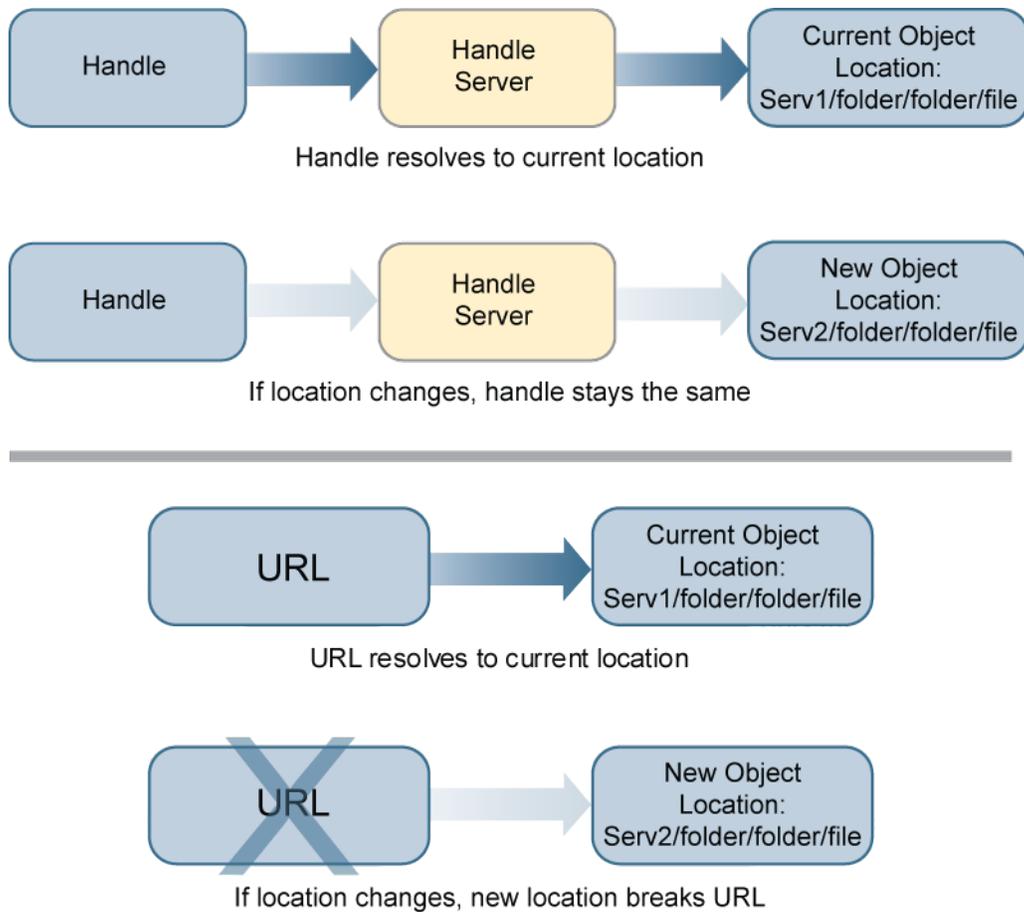


Figure 7 – Resolution of handles vs. URLs

CNRI hosts a global handle server that is a registry for other second-tier handle servers. These second-tier servers have been established by a variety of organizations and offer packaged services, such as object-location resolution, authentication, application of business rules, and metadata storage and use. The packaging of services with the basic Handle System capability creates a community-specific capability that ensures persistent storage and retrieval of digital objects.

The stability and increasing use of handle-based registries make the Handle System a logical foundation for repository search-and-retrieval services. ADL chose it for implementation because of these features.

In addition to a digital object’s location, the ADL Registry provides a variety of information, services, and functions for managing the object’s use. For example, services can be described that determine who can obtain access to specific objects or repositories, enforce business rules to protect intellectual property rights, and allow life-cycle management policies and services to be specified.

3.3 A unifying architecture

Approaches to repository and digital-object management typically emphasize one of the three elements – context, discovery, or resolution – over the others. ADL needs all three to be considered equally. Also, efforts to federate repositories often over define access protocols, metadata sets, and complex services. ADL’s plan is to remain on a higher level and concentrate on means to expose services, data, and capabilities so that use can be negotiated rather than predefined. Communities can then develop independently of one another as required but still create bridges of interoperability.

Appendix A. References

[1] *SCORM 2004 3rd Edition*, Advanced Distributed Learning Initiative, 2006.

Available at: <http://www.adlnet.gov/index.aspx>

[2] *Development, Management, and Delivery of Distributed Learning*. Department of Defense Instruction 1322.26, dtd 16 June 2006. Washington, DC: Department of Defense.

[3] Corporation for National Research Initiatives, “The Handle System,” general information.

See: <http://www.handle.net>.

[4] Dodds, P. V. W., & Fletcher, J. D. (2004). *Opportunities for New “Smart” Learning Environments Enabled by Next Generation Web Capabilities* (IDA Document D 2952). Alexandria, VA: Institute for Defense Analyses.