

2006


FeDCOR: An Institutional CORDRA Registry

Giridhar Manepalli

Henry Jerez

Michael L. Nelson
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_fac_pubs

 Part of the [Computer Sciences Commons](#), and the [Digital Communications and Networking Commons](#)

Repository Citation

Manepalli, Giridhar; Jerez, Henry; and Nelson, Michael L., "FeDCOR: An Institutional CORDRA Registry" (2006). *Computer Science Faculty Publications*. 33.

https://digitalcommons.odu.edu/computerscience_fac_pubs/33

Original Publication Citation

Manepalli, G., Jerez, H., & Nelson, M.L. (2006). FeDCOR: An institutional CORDRA registry. *D-Lib Magazine*, 12(2), 1-11.
doi:10.1045/february2006-manepalli

ARTICLES

D-Lib Magazine
February 2006

Volume 12 Number 2

ISSN 1082-9873

FeDCOR: An Institutional CORDRA Registry[Giridhar Manepalli](#), [Henry Jerez](#)Corporation for National Research Initiatives
Reston, VA 20191
{gmanepalli, hjerez}@cnri.reston.va.us[Michael L. Nelson](#)Old Dominion University
Department of Computer Science
Norfolk, VA 23529
mln@cs.odu.edu

Abstract

FeDCOR (Federation of DSpace using CORDRA) is a registry-based federation system for DSpace instances. It is based on the CORDRA model. The first article in this issue of *D-Lib Magazine* describes the Advanced Distributed Learning-Registry (ADL-R) [1], which is the first operational CORDRA registry, and also includes an introduction to CORDRA. That introduction, or other prior knowledge of the CORDRA effort, is recommended for the best understanding of this article, which builds on that base to describe in detail the FeDCOR approach.

Introduction

The first instance of a CORDRA registry has been built by CNRI for the ADL-R project. ADL-R is designed to serve the U.S. DoD e-learning community and will be hosted and maintained by the Defense Technical Information Center (DTIC) [2]. Any single CORDRA registry, including ADL-R, provides various value added services, the primary service being the federation of a collection of repositories, but the higher level functionality of CORDRA only comes into play when there are multiple federations, which can themselves be federated. Among the motivations for creating FeDCOR was to begin testing the use of CORDRA to federate heterogeneous communities in a single architecture [3]. DSpace [4] was selected as the institutional repository base to use in the development of a CORDRA registry for the institutional repository community, hence the name FeDCOR (Federation of DSpace using CORDRA).

DSpace is a repository system designed to capture, store, index, preserve and redistribute content in various digital formats. Many research institutions and other organizations fitting the library model have found DSpace useful for meeting a variety of digital archiving needs. Although DSpace has succeeded in meeting most of the needs of these organizations, the interoperability of different services across DSpace repositories is a missing feature. The effective federation of DSpace repositories remains a challenge in the digital library community.

Building a federation of DSpace repositories using a CORDRA compliant registry would serve two purposes:

1. A *federation* for DSpace repositories.
2. A *CORDRA registry* from the library community.

The design and implementation of such a registry should be useful both for the DSpace community and the evolving CORDRA community.

FeDCOR

The logical view of FeDCOR is depicted in Figure 1 below.

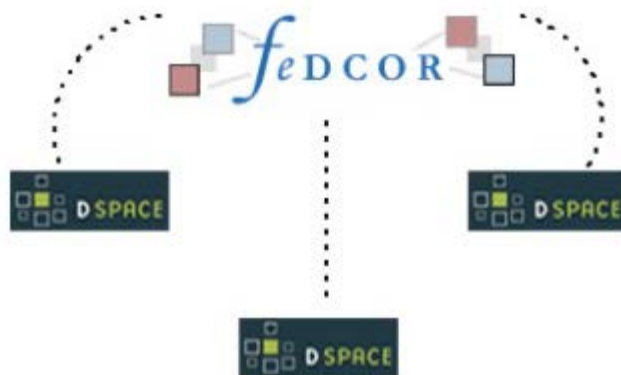


Figure 1: Logical View of FeDCOR

DSpace repositories act as content repositories within FeDCOR. The metadata provided by DSpace for each content object is treated as the metadata instance for that object. DSpace associates handles [5] with the complex object abstraction that incorporates metadata instances and byte streams. At the time of publication of this article, the official distribution of DSpace keeps a handle server per DSpace instance, but work is underway to allow DSpace instances to store and manage their handles through a conventional external handle server [Note 1]. Therefore, FeDCOR assumes the announced future functionality, and assumes each handle for a particular object introduced into a particular DSpace repository to be consistent, even across multiple DSpace implementations.

The design of FeDCOR requires assimilating the data access mechanisms of DSpace and thoroughly defining the CORDRA conformant data structures, business rules and taxonomies.

The following sections address the process of design, implementation and development of FeDCOR, and provide insight into the process of adapting and deploying a CORDRA Registry for a new community.

(1) Design of FeDCOR

A registry contains, by reference, the registered content objects and also holds the metadata instances pertaining to these registered content objects. The community Registry should therefore reflect the community's agreed upon metadata instance format. At the same time, the registry should be able to accommodate an upper layer of metadata in order to integrate it with other registries in a registry of registries federation. Fortunately the original CORDRA registry design provides metadata independence at multiple layers/levels. The three possible levels are (as shown in Figure 2):

1. Content Object specific metadata, also known as Community Level Metadata
2. Registry specific metadata
3. CORDRA specific metadata

In FeDCOR the three different levels must be strictly defined. Leveraging our experience in the development of ADL-R, it is assumed that CORDRA specific metadata is a subset of the union of metadata from the Registry level and the Content Object level.

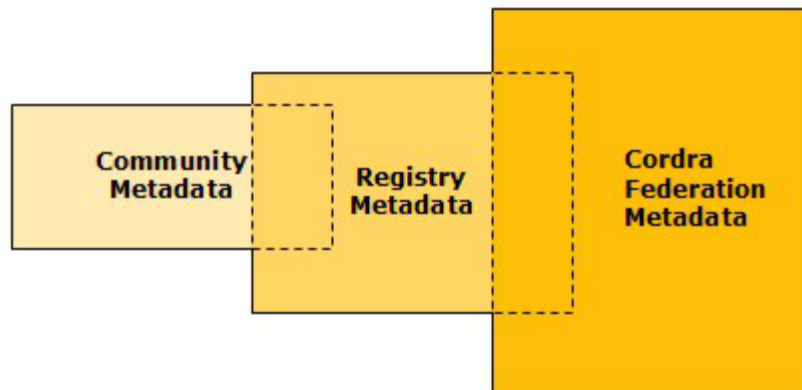


Figure 2: Metadata Instance Layers

Also, CORDRA specific metadata will evolve over a period of time depending on its usage, and the exact usage is unpredictable at this stage. For those reasons, metadata design is derived from registry and content object metadata.

Content Object Metadata Instance Definition

Content Object Metadata Instance (COMI) is the metadata related to each content object retrieved from DSpace. DSpace implements an extended Dublin Core metadata set and most deployments are compliant OAI-PMH [6] data providers. Since OAI-PMH is a standard protocol used in the digital library community, FeDCOR adopted it as a de-facto data access method for DSpace. The Dublin Core metadata format [7] used by the traditional DSpace communities is compatible to both DSpace and OAI-PMH. Hence the content object metadata is conformant with the oai_dc schema of the OAI-PMH metadata record in Dublin Core. The different elements in the metadata include title, identifier, keywords, timestamp, etc. The metadata records are related to the content object handle associated with both the metadata and the byte streams (the complex object mentioned previously). Since we consider the more general case in which DSpace entities are implementing the DSpace remote handle patch, it is feasible to import the same handle into multiple DSpace repositories. Thereby, the persistence of this handle, and its relationship with the Content Object Representation Entity (CORE) [1] identifier of any particular CORDRA repository, is achieved.

In addition to having a unique identifier for the content, CORDRA requires a unique metadata instance identifier associated with content object metadata. FeDCOR follows the same approach as ADL-R, wherein the registry generates and manages the metadata instance handle for each particular metadata instance found inside a particular DSpace repository. The DSpace handle is used as the content object handle.

Registry Level Metadata

Registry Level Metadata reflects the signature of each registry entry at the level of community specific registry. This signature is important in the realization of CORDRA as a platform for

heterogeneous content. It helps to identify the particular registry to which the given entry belongs. Each content object handle, and its metadata instance handle related to its occurrence in a particular DSpace repository, are stored at this level. We also keep track of their update time, so we include the last updated timestamp as part of the registry level metadata.

CORDRA Specific Metadata

As mentioned above, CORDRA specific metadata is the subset of the union of content object metadata and registry specific metadata. As described in the ADL-R article in this issue of *D-Lib Magazine* [1], the CORDRA level metadata and procedures are still in testing and development. Fortunately, most of this metadata will be produced by seamless additions or agents to the pre-existing CORDRA communities.

Business Rules

The operation of CORDRA with multiple levels of system repositories depends heavily on the identification system. The Handle System is used for this purpose [5]. The presence of persistent identifiers for content objects and metadata instances is mandatory. In addition to the handles, the presence of a timestamp is recognized as important information to support multiple views of the system.

FeDCOR enforces the aforementioned business rules, in addition to a schema validation, before registering the content objects from DSpace.

(2) Implementation of FeDCOR

The designed data structures, taxonomies and business rules have to be implemented in such a way as to produce a scalable and workable implementation model. This involves coordinating a number of technical details. It is important to note at this point that the implementation of FeDCOR is a customization of the implementation of ADL-R, which is scheduled to be released as an open source software in the near future.

The different components required by the implementation model are defined below, and shown in Figure 3.

- Registry Engine -- Coordination module
- CORDRA Web or Applications Interface Module
- Index Engine

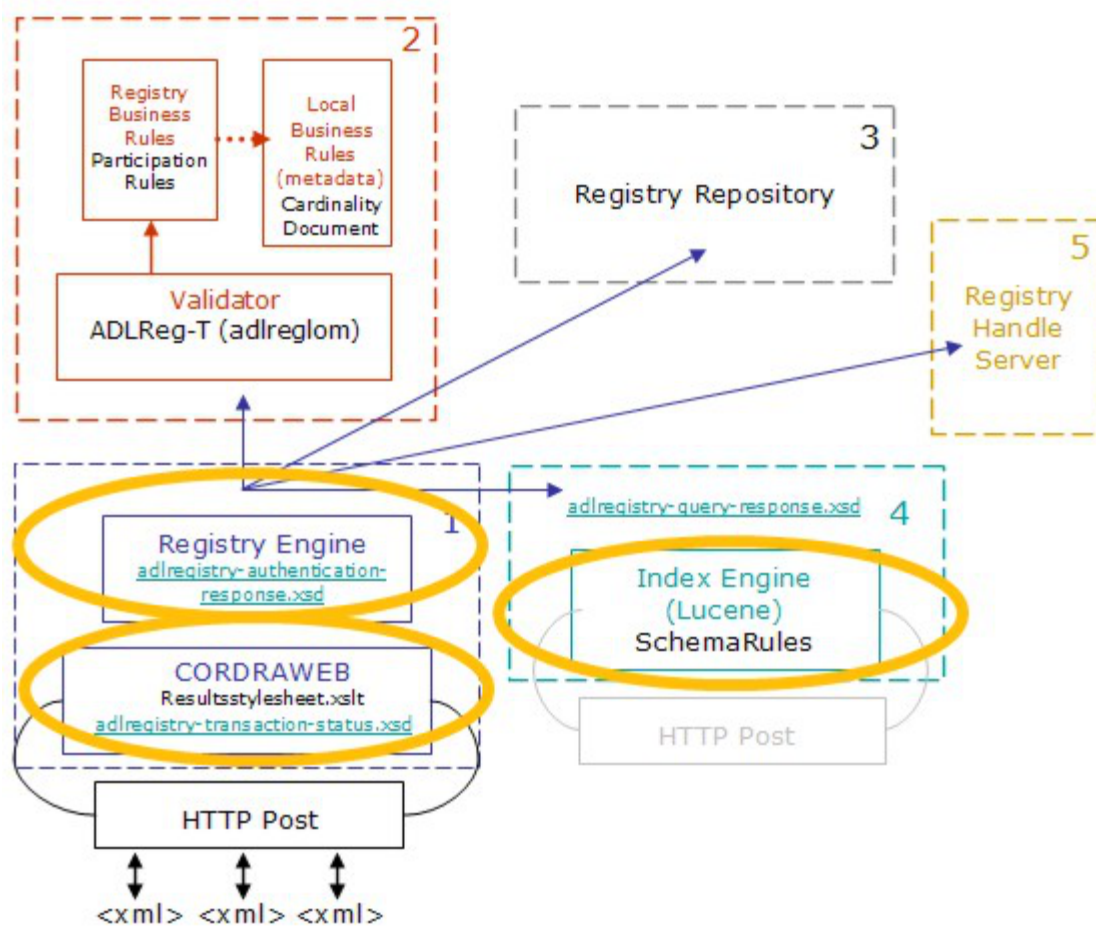


Figure 3: Components from ADL-R Requiring Modification

Registry Engine

The Registry Engine, through its main programming component, Registry Lib, is the core library that coordinates the enforcement of business rules, executes operations, and defines structural components in FeDCOR. The metadata accessed from institutional repositories is stored inside digital objects held by the registry. Registry Lib coordinates the enforcement of various business rules for validating the metadata before registering it. The various operations coordinated by Registry Lib are insert, update, withdraw and delete. In the case of an insert, the metadata accessed from an institutional repository undergoes business rules validation. Upon validation, the metadata is indexed and stored. This applies to any DSpace repository. It is interesting to note that each metadata item stored is added as a datastream to a digital object inside the storage repository, called Content Object Representation Entity (CORE). Each CORE has an identifier determined from the first metadata instance registered for any given content object. If the same content object is represented in two DSpace repositories, each metadata instance accessed from DSpace is added as a separate datastream for the same CORE. Thus, multiple instances or copies, of the same content object can be contained by reference within a single FeDCOR registry. Note that this does not by itself solve the problem of knowing when two records in two DSpace repositories are copies of each other, since it is only based on the actual handle used to identify them.

In ADL-R, the validation module validates the entire submission to check for XML compliance first, and then for adherence to the registry and community business rules. In FeDCOR, the community business rules are enforced by the DSpace repositories, so FeDCOR only needs to validate the registry business rules and does not use the community business rules validator.

Additionally, since every DSpace repository is related to a locally managed handle server, FeDCOR does not administer these handles. This is in contrast to the ADL-R approach that optionally performs content object handle administration for several of its repositories.

CORDRAWeb (Applications Interface)

The CORDRAWeb, or applications interface, provides an accessible interface for the various services made available by the registry. The federation of DSpace content is achieved with the help of a harvester agent, which interacts with the OAI-PMH data provider service of DSpace. (See the 'Populating FeDCOR' section below.) The content and metadata registered in FeDCOR is made available through the search interface. The results contain the metadata record and the corresponding handles (metadata instance and content object). The metadata instance handle may be resolved to reach the corresponding ingested record in DSpace, thus providing the traversal path from FeDCOR to the DSpace record.

Unlike ADL-R, which has a more passive approach, FeDCOR automatically interacts with the DSpace repositories to retrieve their information. Extensive modifications to this component were therefore needed to provide the registry with a harvester interface. These changes came in the form of a harvester agent that monitors DSpace repositories and automatically registers their changes, communicating with CORDRAWeb and registering or updating the content object metadata. The resultant behavior changes are explained in detail in the 'Populating FeDCOR' section below.

IndexEngine

The IndexEngine provides indexing and searching features to the registry. The present version of FeDCOR accesses the Index Engine module using the HTTP protocol. It provides extensible searching based on different elements inside the metadata.

The actual mapping between these elements and their indexed form is controlled by the schema rules configuration inside the IndexEngine. The configuration of the schema rules defines the xpaths to the elements in the metadata record from which they are indexed. Since FeDCOR uses a different XML schema to ADL-R, corresponding modifications were made to the schema rules.

(3) Populating FeDCOR

The population of FeDCOR with metadata records is quite dependent on the different institutional repositories (DSpace repositories) participating in the registry. In order to manage the different participants and also to conform to the CORDRA requirements of having a Repository Registry [8] to register the participating content repositories, the IRR (Institutional Repository Registry) was designed. The IRR is a smaller image of a typical registry. IRR entries are related to the authentication and accessing details of the participating institutional repositories, and are managed specifically by two models that build FeDCOR: the PUSH model and the PULL model. The IRR entries may be registered, unregistered, and searched by accessing the IRR interface.

PULL

In a PULL model, the participating institutional repositories are pre-registered with the IRR. The IRR holds the authentication details and the URL for the corresponding DSpace data provider. The DSpace repositories for the corresponding entries in the IRR are monitored on a timely basis, and queried for content modification using OAI-PMH run by a software agent, as part of the building process. Any changes that occur within the DSpace repository are captured and registered with FeDCOR.

The PULLED (harvested) metadata record from DSpace is processed to validate the cardinality of registry level metadata. The validated records are registered with FeDCOR. The process of registration includes storing, indexing and creating/updating handles.

PUSH

In the PUSH model, a DSpace plug-in is made available to the DSpace community. The plug-in may be installed at the DSpace repositories. Once installed, the plug-in registers the DSpace instance with IRR. The purpose of the plug-in is to trigger IRR when it finds new or updated information in the corresponding DSpace repository. Because the plug-in resides with the corresponding DSpace repository administrators, considerable rights are reserved for them.

The plug-in is more suited to DSpace repositories that hibernate. In other words, if a given DSpace repository has only limited uptime, it can install the plug-in to trigger the agent running at the CORDRA registry to harvest the records when it is available.

The workflow logic of the harvesting agent is shown in Figure 4.

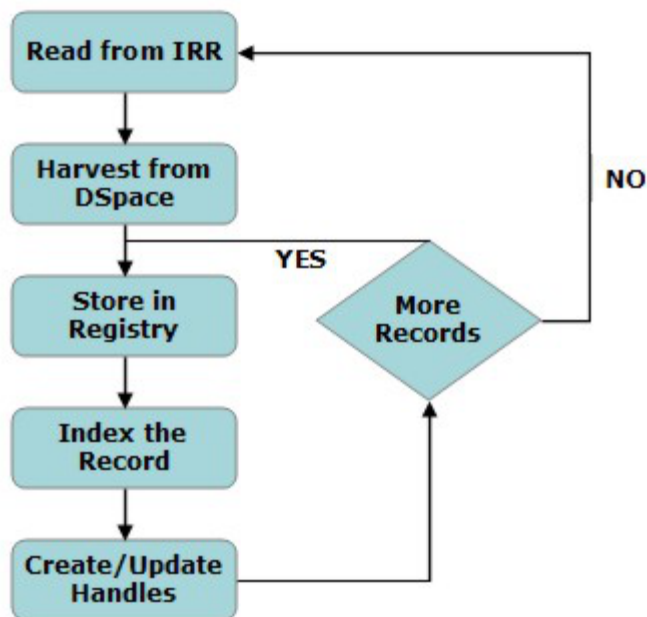


Figure 4: Workflow of FeDCOR Agents

FeDCOR Implementation of CORDRA Architecture

As shown in Figure 5 below, the architecture preserves most of the original ADL-R components, but adds the first implementation of an Institutional Repository Registry, and an intelligent population agent and plug-in that make the registration process automatic. The result is a useful and relatively seamless CORDRA federation of DSpace repositories that reuses most of the original ADL-R code and provides a different community registry with a set of basic CORDRA Services.

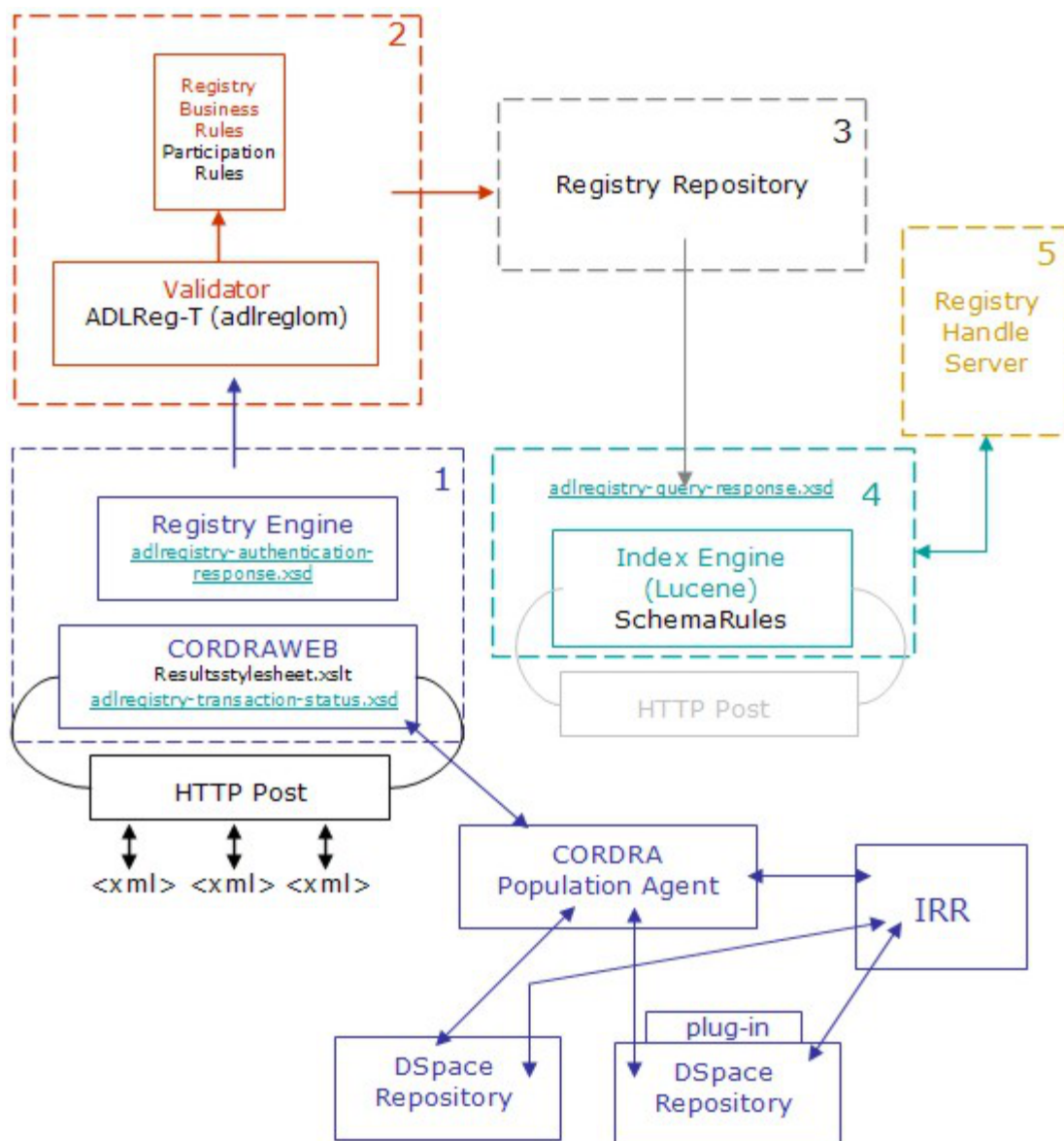


Figure 5: FeDCOR Components

Future Research

Integration of FeDCOR into CORDRA

FeDCOR must be integrated into CORDRA to leverage the advantages of content discovery from multiple levels of a CORDRA hierarchy. Figure 6 depicts the overall model of CORDRA with the integration of FeDCOR into the system. Users would then have heterogeneous communities from which to draw content, thereby increasing the audience for, and the benefits of, the combined systems. The model would follow the original concept of searching/discovering content at any level of registries. Users with specific community knowledge may search at the lowest level and those with only generic knowledge may search at a higher level.

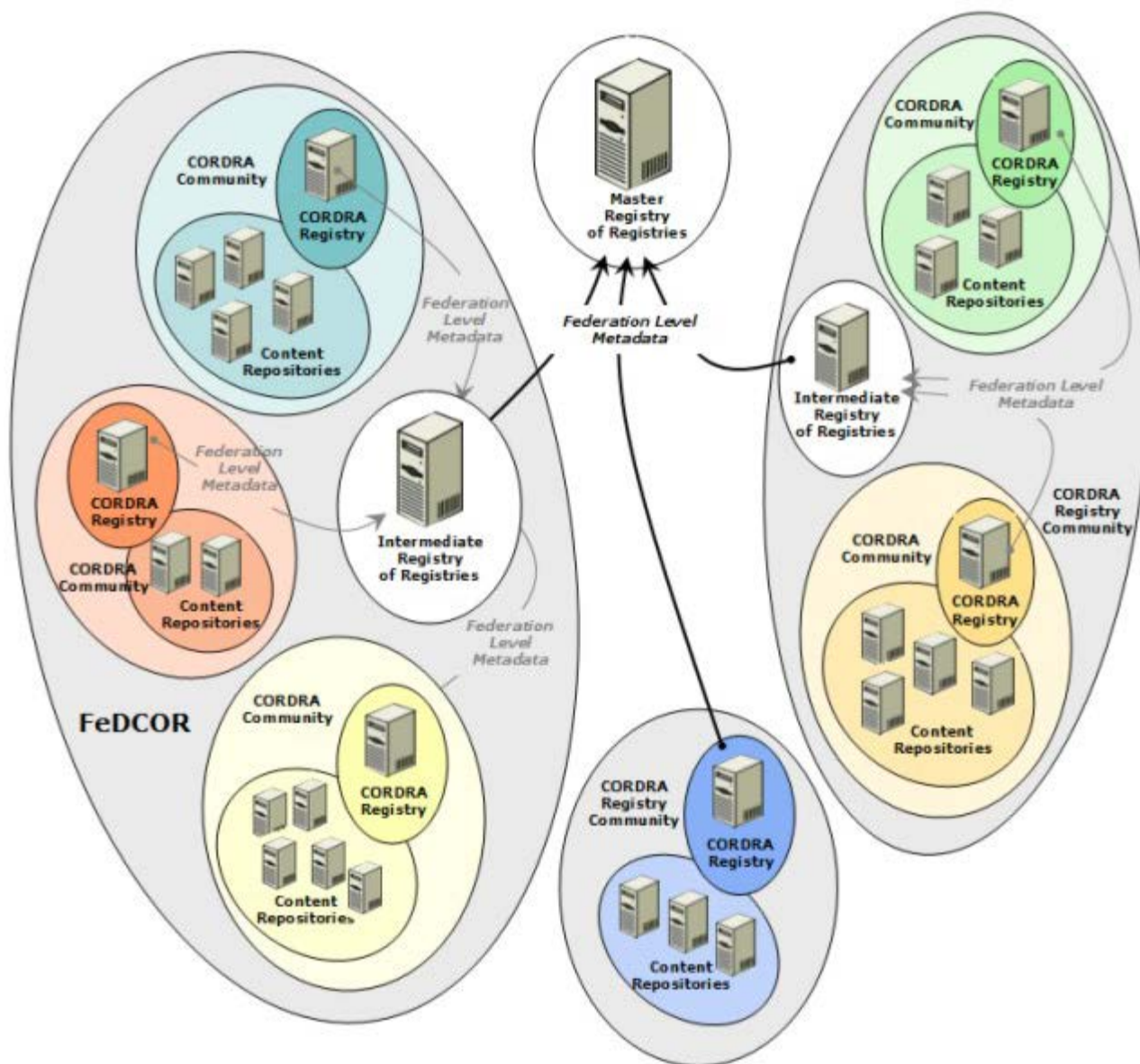


Figure 6: FeDCOR integrated into CORDRA

In order to verify the functionality of this system, a heterogeneous CORDRA Registry of Registries (RofR) needs to be built that integrates metadata from multiple communities. Once the heterogeneous registry is built, the extensibility of CORDRA across diverse communities can be tested.

The experience related to the design of FeDCOR, integration of FeDCOR into CORDRA, and all associated metrics will be made publicly available over time, perhaps in future *D-Lib* articles. Current tests are taking place through a partnership with the [Ibero American Science and Technology Consortium \(ISTEC\)](#), which will run a test version of FeDCOR that is scheduled to be released to the public on April 15, 2006.

Conclusions

FeDCOR enables the federation of DSpace communities by following the CORDRA infrastructure. FeDCOR is not only the first CORDRA registry from the library community, but also the first operational DSpace content federation registry available of which we are aware. Aside from standing as a proof of concept for the merging of library communities with learning communities in

CORDRA, FeDCOR also benefits the DSpace community as a generic federator.

Acknowledgements

The authors would like to thank the Advanced Distributed Learning Initiative Co-Labs who funded most of the ADL-R work on which FeDCOR has been based and the DSpace Users Development Community for their knowledge sharing through their forums. The authors would also like to acknowledge the early works of the LANL prototyping team on OAI-PMH harvesting from DSpace repositories. Their work inside the OAI-PMH federator on their ADORE architecture [9] was an inspiration for this work.

Notes

[1] A software patch currently exists for DSpace that allows the use of remote handle servers. See http://sourceforge.net/mailarchive/message.php?msg_id=12760613.

References

[1] Jerez, Henry, Giridhar Manepalli, Christophe Blanchi, and Laurence W. Lannom. "ADL-R: The First CORDRA Registry". *D-Lib Magazine*, Volume 12, Number 2, February 2006.

[doi:10.1045/february2006-jerez](https://doi.org/10.1045/february2006-jerez).

[2] Defense Technical Information Center (DTIC). <http://www.dtic.mil>.

[3] Kraan, Wilber, and Jon Mason. "Issues in Federating Repositories". *D-Lib Magazine*, Volume 11, Number 2, March 2005. [doi:10.1045/march2005-kraan](https://doi.org/10.1045/march2005-kraan).

[4] DSpace Federation. <http://www.dspace.org>.

[5] The Handle System. <http://www.handle.net>.

[6] Open Archives Initiative Protocol for Metadata Harvesting. Document Version 2004/10/12T15:31:00Z. Open Archives Initiative, October 19, 2005.

<http://www.openarchives.org/OAI/openarchivesprotocol.html>.

[7] *Dublin Core Metadata Initiative*. November 7, 2005. OCLC Research. November 13, 2005.

<http://dublincore.org>.

[8] Rehak, Dan; Philip Dodds and Larry Lannom. "A Model and Infrastructure for Federated Learning Content Repositories", *Interoperability of Web-Based Educational Systems Workshop*, Volume 143 or *CEUR Workshop Proceedings*, May 10, 2005.

<http://cordra.net/cordra/information/publications/2005/www2005/cordrawww2005.pdf>.

[9] Jerez, Henry; X. Liu; P. Hochtchenbach; and H. Van de Sompel, "The multi-faceted use of the OAI-PMH in the LANL Repository". *Proceedings of the fourth ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL 2004.

(The URL in Reference [8] was corrected on February 20, 2006.)

Copyright © 2006 Corporation for National Research Initiatives and Michael L. Nelson

[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

doi:10.1045/february2006-manepalli