The NCore Platform: An Open-Source Suite of Tools and Services for Implementing Digital Libraries

Dean B. Krafft, Cornell University

Introduction

NCore is an open-source suite of tools and services for creating production digital libraries and related repository systems. It provides a general platform for building digital libraries united by a common data model and interoperable applications. NCore currently powers the National Science Digital Library (NSDL)[1], which catalogs over 2.5 million resources from over 130 digital collection providers and thousands of web sites. Developed by the NSDL Core Integration team and its partners, the NCore platform consists of a central repository built on top of Fedora[2]; a data model and web services API that implements the model; a programmers toolkit in PHP, Java, and Javascript; a number of fundamental services, such as full-text search and OAI-PMH aggregation; and a variety of end-user tools for the collaborative creation of context and content. The repository and API code components of NCore are currently available on the NCore SourceForge site[3] under the Educational Community License (ECL 1.0). Over the first half of 2008, the NSDL team will release the code for the full NCore suite of tools and services on the SourceForge project site.

NCore and NSDL

The NCore suite of technologies, software and standards were originally created to power the core technical infrastructure of the NSDL. It was designed to replace NSDL's original metadata-record based digital library system[2] with a far more flexible architecture designed to support collaborative community development of content and context within the framework of a carefully selected and cataloged set of primary library resources. NCore is based on the Fedora repository system, and the fundamental data model and motivation have been previously described in [3].

In the first phase of NCore development, completed in January 2007, NSDL successfully implemented existing services on top of NCore and released that implementation into production. Since the initial production release, NSDL has been both reworking its existing tools and services to take full advantage of the capabilities of the NCore platform, and developing and releasing new library tools and services that greatly expand the capabilities of the library. Some of those tools were described in a presentation at Open Repositories 2007[1].

Most recently, the NSDL Core Integration team has been revising and refactoring the NCore tools and services to remove NSDL-specific dependencies, fully document the system and capabilities, and release the full NCore platform through SourceForge, licensed under the Educational Community License (ECL 1.0).

---

[1] http://nsdl.org
[2] http://www.fedora-commons.org/
[3] http://sourceforge.net/projects/nsdl-core/

NCore Services and Management Tools

NCore provides a number of significant library services. These include:

- Search service: Using a combination of a REST API, Lucene, and the Nutch web crawler, the search service indexes all the metadata and accessible resource text content referenced in the NCore repository. Full information on the REST API and the service is available online[4].

- OAI-PMH ingest service: This harvester automatically aggregates resource metadata records for appropriate collections as specified in the NCore repository. Harvests can be scheduled to happen automatically, and full diagnostics are available to deal with problems.

- NSDL Collection Service (NCS): Derived from the Digital Collection System[5], developed by Digital Learning Sciences[6], this is a highly flexible XML-driven metadata editing system that can be used by catalogers and editors to manage collection and item metadata in the NDR.

- OAI-PMH server: NCore repository metadata is served out using the Fedora Proai system. The system flexibly presents NCore collections as OAI-PMH sets, and arbitrary metadata collections can be optionally blocked from public OAI-PMH distribution.

- Shibboleth Identity Management: The NCore platform uses Shibboleth[7] as the identity management framework for its tools and services. The client code for the tools can be configured to work with any federated set of Shibboleth identity providers.

- Strand Map Service: This is a highly flexible visualization service that can support developer-created knowledge maps in arbitrary subject domains. Within NSDL, it is used to implement AAAS Project 2061[8]'s Science Literacy Maps as a knowledge map tool to allow interactive visualization of the knowledge space for K-12 science[9].

End User Tools

Integrated with NCore are three key collaborative tools, allowing community users to create context around NCore resources, create new resources, and add metadata to new or existing resources. References among resources in the NCore repository are captured as explicit RDF relationships among the Fedora digital objects, allowing these relationships to be used for discovery, browsing, or visualization. The tools are:

- Expert Voices – A blogosphere, built as extensions to WordPress Multi-User, Expert Voices supports annotation of NCore resources with blog entries,

---

[4] http://wiki.nsdl.org/index.php/Community:Search
[5] http://www.dlsciences.org/tools-dls/#dcs
[6] http://www.dlsciences.org/
[7] http://shibboleth.internet2.edu/
[8] http://www.project2061.org/
[9] http://strandmaps.nsdl.org

publication of blog posts as new NCore resources, and creation of metadata and categorization of NCore resources within blog posts. Full documentation[10] and the NSDL production implementation[11] are both online.

- NSDL Wiki – An NDR-integrated wiki system, built on MediaWiki, supporting authenticated/authorized edit/view access based on namespace group permissions. Wiki pages can be published as NCore resources, and existing NCore resources can be referenced, augmented and annotated. Full documentation[12] and the NSDL production implementation[13] are both available online.

- On Ramp – A content management system, based on Fedora-Fez[14], supporting full creation/editing/publication workflow for reusable content components. Full documentation[15] is available online.

Implementing DLESE in NCore

Recently, the NSDL Core Integration team, working together with Digital Learning Sciences, had an opportunity to test the proposition that NCore is truly a general platform for building semantic and virtual digital libraries united by a common data model and interoperable applications. DLESE, the Digital Library for Earth Science Education[16] ported their existing infrastructure and test content to NSDL's NCore repository. By overlaying their data model on top of the NCore model and embedding their existing specialized metadata within the flexible digital objects of the NCore architecture, DLS developers enabled specialized DLESE services, such as the DCS metadata editing tool, to continue without loss of functionality, and to co-exist alongside of, but potentially independent from, the NSDL library.

To support the DLESE Search service, the NCore API was used to build the DLESE index by reading DLESE metadata from the NCore repository, using the NCore API, and then building the index. The key technical capability was to the ability to store and extract the required DLESE-specific metadata from digital objects in the NCore repository.

The DLESE Search Service is run from the Index, and is used to power the DLESE library search and browse functionality. Use of the NSDL Data Repository as the primary metadata repository required no modifications to DLESE Search Service. The indexes and search service remain the same, which means the DLESE Library interface and all other interfaces that use the NCore services continue to operate without modification.

Since the DLESE resources and metadata are true library resource and metadata objects within the NCore model, all the standard NCore tools and services (search, Expert Voices, NSDL Wiki) also work with the DLESE library entries.

---

[10] http://wiki.nsdl.org/index.php/Community:ExpertVoices
[11] http://expertvoices.nsdl.org
[12] http://wiki.nsdl.org/index.php/Community:NSDLWiki
[13] http://wiki.nsdl.org
[14] http://dev-repo.library.uq.edu.au/wiki/index.php/Main_Page
[15] http://wiki.nsdl.org/index.php/Community:OnRamp
[16] http://dlese.org

Conclusion

The NCore platform is a major open-source contribution to the field of open repositories and digital libraries. By providing a set of standard models, tools and capabilities, it can result in major time savings for any project looking to implement a Fedora-based repository system. Moreover, it is an active and developing effort with a growing community of users. In the near future, the NSDL Core Integration team expects to implement NCore extensions and integration for a range of additional tools, including course management systems, online learning activity creation systems, and bookmarking/folksonomic tagging systems. The NSDL Core Integration team is very interested in speaking to any project that wishes to enable a particular tool or service to integrate with the NCore platform.

The proven flexibility, scalability, and extensibility of NCore make it a strong candidate for any group contemplating the implementation of a digital library or cataloged repository system, particularly one where collaborative community contributions to the repository are of significant importance.

Acknowledgements

References

1. Krafft, D.B. NSDL 2.0: Building a Collaborative Digital Library *Open Repositories 2007*, OpenRepositories.org, San Antonio, Texas, 2007. Available at http://openrepositories.org/2007/program/user_groups#session5

2. Lagoze, C., Arms, W., Gan, S., Hillmann, D., Ingram, C., Krafft, D., Marisa, R., Phipps, J., Saylor, J., Terrizzi, C., Hoehn, W., Millman, D., Allan, J., Guzman-Lara, S. and Kalt, T., Core Services in the Architecture of the National Digital Library for Science Education (NSDL). in *Joint Conference on Digital Libraries*, (Portland, Oregon, 2002), ACM/IEEE. Available at http://arxiv.org/abs/cs.DL/0201025

3. Lagoze, C., Krafft, D., Cornwell, T., Eckstrom, D., Jesuroga, S. and Wilper, C., Representing Contextualized Information in the NSDL. in *ECDL2006*, (Alicante, Spain, 2006), Springer. Available at http://arxiv.org/abs/cs/0603024