

Publishing in the NSDL: Fundamental Concepts for Creating and Reusing Content

Carol Minton Morris
National Science Digital Library
Cornell University
Ithaca, NY 14850
607-255-2702
clt6@cornell.edu

Lynette Rayle
National Science Digital Library
Cornell University
Ithaca, NY 14850
607-255-2702
elrayle@cs.cornell.edu

ABSTRACT

The National Science Digital Library (NSDL) dramatically broadens the information about STEM resources that it can accept and make available to its users with the introduction of the NSDL Data Repository (NDR) architecture. [1]

In this paper we present concepts for On Ramp (ONR), a content and communications system that supports handling of multiple types of content collected together as a single entity; that facilitates reusability and distribution of content to multiple locations; and that leverages existing and emerging technologies to enable far-flung networks of content contributors to create and widely share context and insights about Library resources that can transform information into warranted knowledge. [2]

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Communications Applications.

General Terms

Management, Design

Keywords

Content Management, Content Dissemination, Reuse, Repurpose

1. INTRODUCTION

The NSDL community requires the ability to create, organize, and disseminate content in a variety of formats to a number of distribution targets. Currently, content is created and maintained by multiple systems dependent on differing distribution requirements. Re-use of content is limited within various systems. Manual intervention is required to move between systems. The potential reuse of content is inhibited because content is difficult to find and duplicate. Additionally, synchronization problems arise when the same content exists in multiple systems. On Ramp provides a workflow environment for the creation or uploading of content in a wide range of formats

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '04, Month 1–2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

within a centralized repository capable of disseminating content to multiple distribution targets. A searchable archive facilitates reuse and repurposing of content. A web-based user interface provides access through a distributed user network organized into working teams. Disseminations can range from publications like the *NSDL Annual Report* to educational workshop materials and online displays like the Highlights exhibit on NSDL.org's homepage.

Unlike existing content management systems, various types of content in ONR can be arbitrarily repackaged and repurposed. The system is designed to address issues of large-scale, complex, and distributed content management. Some core ONR concepts are described in the following sections.

2. ON RAMP FUNDAMENTAL CONCEPTS

2.1 Content Collected as a Single Entity

Content is collected from increasingly diverse data streams (e.g., video, images, text, websites, etc.) that require integration of related content. [3] At the lowest level of granularity content often has little meaning unless it is related to other content. Retrieval and re-use of content chunks can be confusing and limited when content is discovered out of context.

ONR uses a package object to collect multiple pieces of content into a single entity. The package provides a higher level of granularity and meaning that allows related content chunks in various formats to be collected together. All relevant content is stored in a package and passed within the system at this level. Aggregating content pieces into a package object provides context that a single piece of content lacks.

2.2 Repackaging and Repurposing Content

Packages of content are created within a flexible workflow environment utilizing the Fedora Workflow Orchestration Service. [1] The underlying Fedora repository provides a means for archiving packages as major changes are made to content and also provides a scalable environment that can be efficiently searched. Content that has been retired from active use can be found by searching the archive and may be reused in several ways.

The package can be reactivated and pulled back into the workflow to allow for content modifications, and re-released as an updated version of the same package. The package can be copied as a whole and modified in the context of the new package. Or individual pieces of content within one package can be copied to a different package.

In addition to repackaging archived content, a fundamental ONR implementation strategy is the dissemination of packages to multiple targets (e.g., NSDL website, print materials, workshop materials, newsletters, etc.) as the package is released from the workflow process.

2.3 Dissemination to Multiple Targets

Emergence of new technologies requires that information systems be prepared to select, adapt, and deliver content to diverse systems. [4] Three primary types of disseminations have initially been identified for content delivery related to NSDL communications.

RSS feeds are used to pass content from ONR to targets that process the feed. Currently, RSS feeds are used to generate webpage content such as the exhibits featured on NSDL.org. However, any target that understands RSS can use this dissemination mechanism.

A Bundle Target collects content that users want to be able to retrieve as a collection of documents. There is no physical target for the delivery of content. On Ramp's interface provides access to content distributed to bundled targets. In this way, targets such as those designed to hold workshop materials (e.g., PowerPoint presentations, workbooks, etc.) can be created, vetted, and retrieved through the On Ramp system.

The final type of delivery mechanism identified is the merging of content. The result of the dissemination process is a single document including all content released to the target. The released document can itself be fed back into the workflow process to be further modified. An example use of this type of dissemination mechanism is the creation of an annual report which has outlined content, as well as various articles and images that need to be combined into a single document for final print. (Existing software creates layouts for print documents outside of On Ramp.)

Each dissemination target can have different format requirements for incoming content. In some cases, conversions can be made automatically from one format to another. If this is not possible, the same content can be stored in a package in multiple formats. Additionally, some targets require specific content to be in place prior to release. For example, a typical RSS feed might need content to identify a title, a description, a link to more details, and an image. The package would hold a content item for each of these designations.

By facilitating dissemination of content to multiple targets, any package can have content going to one or all of the above types of dissemination mechanisms. The content is maintained in a single package and used in multiple contexts immediately once released from the workflow process.

3. EXPLORING EXISTING SYSTEMS

Prior to beginning work on On Ramp in the summer of 2005, existing systems and open source initiatives were investigated that could fulfill the fundamental concepts while maintaining a scalable data store environment. Although many systems met some of the requirements, none were found that could fulfill all the needs of the NSDL community. Systems primarily fell into the following categories: 1) focused on efficient storage and retrieval of content regardless of format, 2) storage of content for a specific data type (e.g., digital media such as video, images, and

sound; institutional repositories of journal articles), 3) creation of content, with or without a basic workflow, for a specific distribution domain (e.g., web pages; print documents), 4) a few systems provided a means for content to be released through multiple distribution domains (e.g., web pages and news feeds). The majority of content management systems were commercial products that could not be expanded to meet additional ONR requirements.

4. PHASED DEVELOPMENT PROCESS

Phase one development of On Ramp included creating the underlying architecture to support multiple content in a package, integration with the Fedora Workflow Orchestration Service, and a user and group management system, implementation of the RSS Feed dissemination mechanism, as well as the user interface for interaction with these subsystems. Workflow was limited to a single, static workflow supporting creation of packaged content, review approval, and release to targets.

Phase two will include the development of additional dissemination mechanisms to support Bundle Targets and Merge Targets. A scheduling component will control the timed-release of content to a target to handle various scheduling scenarios including delayed release, release for a specified period of time, and collection of content into regularly released issues of a publication. A facility for batch loading existing content from legacy systems into On Ramp will also be developed.

Phase three will address issues related to allowing users to personalize their ONR usage. Anticipated enhancements include a user homepage with a 'Watch Packages Portlet' in which packages can be identified that are considered a high priority to track. Users will specify that they would like to see all documents in a particular state(s), such as those "Waiting for Review." The user will specify the starting page when entering On Ramp based on the most commonly performed tasks in the system. Phase three enhancements will be finalized based on additional user testing.

5. CONCLUSIONS

An On Ramp Survey was conducted using the prototype system during the Fall 2005 to learn more about users' basic understanding of such a system [see poster submission: "Finding a Metaphor for Collecting and Disseminating Distributed NSDL Content and Communications"]. The alpha release of ONR is scheduled for May 2006 with additional testing and adoption by the NSDL community to follow.

6. REFERENCES

- [1] Lagoze C., Krafft D., Payette S., Jesuroga S. "What is a Digital Library Anymore, Anyway? Beyond Search and Access in the NSDL." *D-Lib Magazine*, Nov. 2005.
- [2] Schneiderman, B. *Leonardo's Laptop*. The MIT Press. Cambridge, MA. 2003. 111-131.
- [3] Sreekumar, M. G., and Sunitha, T. Seamless Aggregation and Integration of Diverse Datastreams: Essential Strategies for building Practical Digital Libraries and Electronic Information Systems. *International Information and Library Review*, 2005.
- [4] Beckers, T., Oorts, N., Hendrickx, F. and Van De Walle, R. Multichannel Publications of Interactive Media Documents in a

News Environment. In *Proceedings of the Twelfth International World Wide Web* (Chiba, Japan, May 10-14, 2005). ACM Press, New York, NY, 2005.