Developing a Repository at the Library of Congress

LIBRARY OF CONGRESS

Leslie Johnston

CENDI/NFAIS Workshop
November 30, 2011

How do we know what our most basic needs are to put a repository into place at the Library of Congress?

We start by collecting requirements

Collecting requirements for an organization of this scale required an extended effort.

A team of contract consultants held a series of more than 50 interviews with representatives of every division and every major initiative in the Library, with questions covering 6 topics:

- Collection (ingest) – description of content to be stored in repository
- Storage – quantities, size of media and metadata, growth rate etc.
- Cataloging and Indexing – aggregation, creation on ingest, text search, etc.
- Preservation planning – document integrity, authenticity, how long, etc.
- Read access - rules, availability, rendering
- Scholarly access (expert annotation and tagging) – subject matter experts who need to add information to an object

This took almost a year, and produced a 200+ page document that described desired Repository Services.

But this document didn’t really have requirements; it produced a list of desired functionality, from the mundane to the truly blue sky.

How would we get at real requirements?

The Library already had in place a Digital Content Lifecycle model, with the following categories:

- Plan
- Produce
- Get
- Select
- Describe
- Prepare/Assemble
- Sustain
- Make Available

Each is broken down into the necessary Technical Infrastructure, Policies and Best Practices, and Activities

The collected requirements were mapped to the model’s categories and activities.

Mapping the requirements to the Lifecycle model showed the gaps in both the requirements, and in the model.

We have now identified new requirements where there were none mapped to the Model, and places where the Model must be extended, primarily around interaction by users with content.

But these still are not requirements that a software developer can truly develop against.

The Library works in an agile development model.

Requirements are chunked out into quick development iterations, each of which has its own requirements gathering, design, implementation, and testing phase. More progress is made more quickly.

So where is the Library now?

Existing systems are in place that will provide the core of repository functionality.

Additional development will be needed, which will require going out for bids for various aspects of the work.

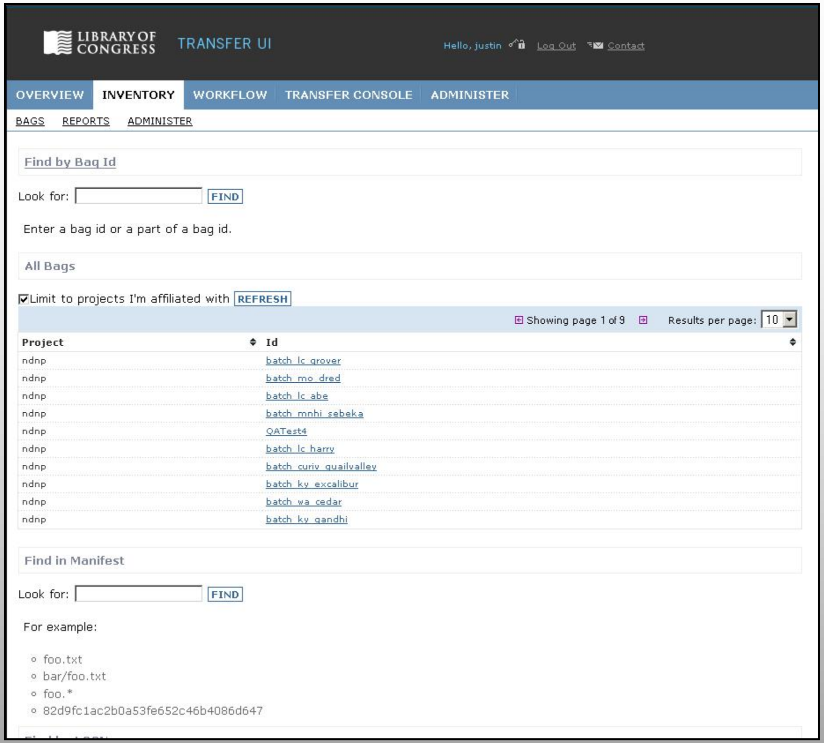
What are our most basic needs in putting a repository into place?

How do we know what we have, where it is, and who it belongs to?

How do we know what events have occurred in a object’s life cycle?

How can the object be accessed and used?

Content Transfer Services and Inventory Services



Find by Bag Id

Look for:

Enter a bag id or a part of a bag id.

All Bags

☑ Limit to projects I'm affiliated with

Showing page 1 of 9   Results per page: 10

Project	Id
ndnp	batch_lc_grover
ndnp	batch_mo_dred
ndnp	batch_lc_abe
ndnp	batch_mnhi_sebeka
ndnp	QATest4
ndnp	batch_lc_harry
ndnp	batch_curiv_quailvalley
ndnp	batch_ky_excalibur
ndnp	batch_wa_cedar
ndnp	batch_ky_gandhi

Find in Manifest

Look for:

For example:

- foo.txt
- bar/foo.txt
- foo.*
- 82d9fc1ac2b0a53fe652c46b4086d647

Keeps track of and enables the querying of important events in the preservation lifecycle of packages and files, beginning with transfers to the Library.

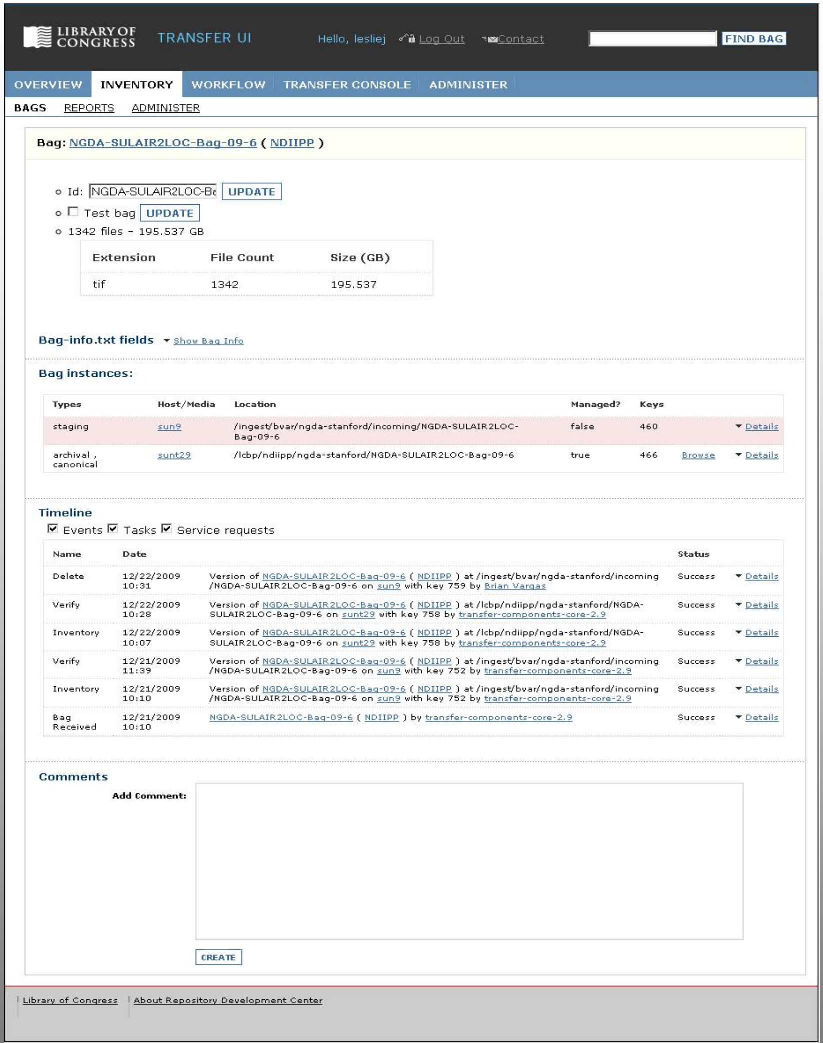
Satisfy needs identified through the process of doing transfers.

These needs include keeping track of package transfers for a project, tracking life cycle events associated with packages, and maintain an inventory of the files that make up each package and their locations.

Files are associated with a program or project, a custodial unit, a content type (textual, still image, audio, etc.), a content process (partner transfer, digital conversion, web archiving, etc.), and an access category.

It records location events and locations (with datetimestamps) at a package level and on a file level

Multiple copies of content can be recorded as related instances, each with their own event history.



Bag: NGDA-SULAIR2LOC-Bag-09-6 (NDIIPP)

- Id: NGDA-SULAIR2LOC-Ba UPDATE
- □ Test bag UPDATE
- 1342 files - 195.537 GB

Extension	File Count	Size (GB)
tif	1342	195.537

Bag-info.txt fields ▾ Show Bag Info

Bag instances:

Types	Host/Media	Location	Managed?	Keys		
staging	sun9	/ingest/bvar/ngda-stanford/incoming/NGDA-SULAIR2LOC-Bag-09-6	false	460		▾ Details
archival , canonical	sunt29	/lcbp/ndiipp/ngda-stanford/NGDA-SULAIR2LOC-Bag-09-6	true	466	Browse	▾ Details

Timeline

☑ Events ☑ Tasks ☑ Service requests

Name	Date		Status	
Delete	12/22/2009 10:31	Version of NGDA-SULAIR2LOC-Bag-09-6 (NDIIPP) at /ingest/bvar/ngda-stanford/incoming /NGDA-SULAIR2LOC-Bag-09-6 on sun9 with key 759 by Brian Vargas	Success	▾ Details
Verify	12/22/2009 10:28	Version of NGDA-SULAIR2LOC-Bag-09-6 (NDIIPP) at /lcbp/ndiipp/ngda-stanford/NGDA-SULAIR2LOC-Bag-09-6 on sunt29 with key 758 by transfer-components-core-2.9	Success	▾ Details
Inventory	12/22/2009 10:07	Version of NGDA-SULAIR2LOC-Bag-09-6 (NDIIPP) at /lcbp/ndiipp/ngda-stanford/NGDA-SULAIR2LOC-Bag-09-6 on sunt29 with key 758 by transfer-components-core-2.9	Success	▾ Details
Verify	12/21/2009 11:39	Version of NGDA-SULAIR2LOC-Bag-09-6 (NDIIPP) at /ingest/bvar/ngda-stanford/incoming /NGDA-SULAIR2LOC-Bag-09-6 on sun9 with key 752 by transfer-components-core-2.9	Success	▾ Details
Inventory	12/21/2009 10:10	Version of NGDA-SULAIR2LOC-Bag-09-6 (NDIIPP) at /ingest/bvar/ngda-stanford/incoming /NGDA-SULAIR2LOC-Bag-09-6 on sun9 with key 752 by transfer-components-core-2.9	Success	▾ Details
Bag Received	12/21/2009 10:10	NGDA-SULAIR2LOC-Bag-09-6 (NDIIPP) by transfer-components-core-2.9	Success	▾ Details

Comments

Add Comment:

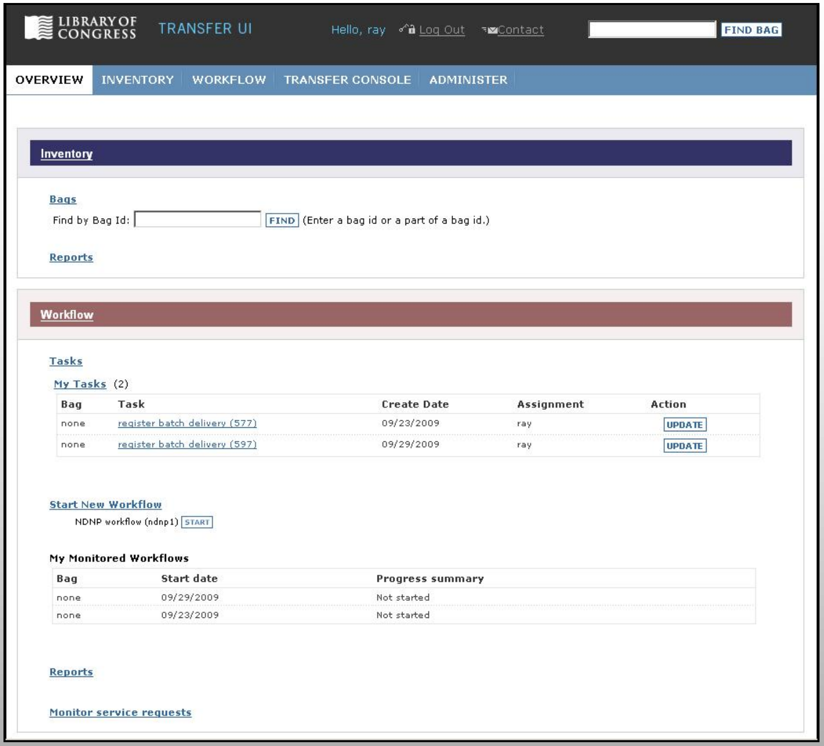
Workflow Services

The Transfer and Inventory services are tied together by workflows.

Workflow tasks formalized through the system include transfer, validation by an format validation application, manual quality review inspection, inventorying, and file copying to archival storage and production storage.

A workflow UI allows users to initiate, monitor and administer processes; and notify the workflow engine of the outcome of manual tasks, including task completion.

Status views and a variety of auditing reports are available.



Find by Bag Id: FIND (Enter a bag id or a part of a bag id.)

My Tasks (2)

Bag	Task	Create Date	Assignment	Action
none	register batch delivery (577)	09/23/2009	ray	UPDATE
none	register batch delivery (597)	09/29/2009	ray	UPDATE

NDNP workflow (ndnp1)

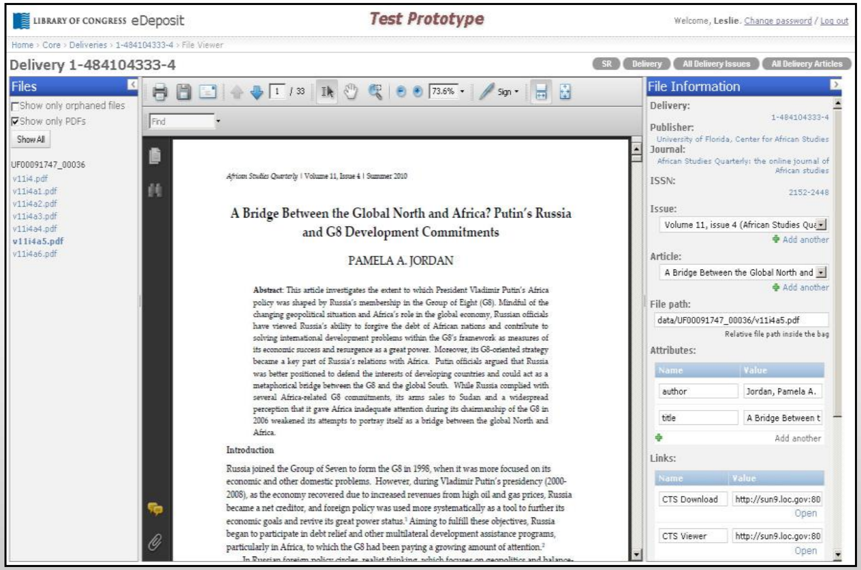
My Monitored Workflows

Bag	Start date	Progress summary
none	09/29/2009	Not started
none	09/23/2009	Not started

Delivery Management Service

The first tool built on top of CTS and the Inventory services to map files to collection objects.

It’s a service that constantly checks the Content Transfer Service for new deliveries, records them, extracts descriptive metadata where possible, and make the files available to staff for examination and processing, which includes the mapping of files to objects and the assignment of metadata.



Delivery 1-484104333-4

UF00091747_00036
v11i4.pdf
v11i4a1.pdf
v11i4a2.pdf
v11i4a3.pdf
v11i4a4.pdf
v11i4a5.pdf
v11i4a6.pdf

African Studies Quarterly | Volume 11, Issue 4 | Summer 2010

A Bridge Between the Global North and Africa? Putin’s Russia and G8 Development Commitments

PAMELA A. JORDAN

Abstract: This article investigates the extent to which President Vladimir Putin’s Africa policy was shaped by Russia’s membership in the Group of Eight (G8). Mindful of the changing geopolitical situation and Africa’s role in the global economy, Russian officials have viewed Russia’s ability to forgive the debt of African nations and contribute to solving international development problems within the G8’s framework as measures of its economic success and resurgence as a great power. Moreover, its G8-oriented strategy became a key part of Russia’s relations with Africa. Putin officials argued that Russia was better positioned to defend the interests of developing countries and could act as a metaphorical bridge between the G8 and the global South. While Russia complied with several Africa-related G8 commitments, its arms sales to Sudan and a widespread perception that it gave Africa inadequate attention during its chairmanship of the G8 in 2006 weakened its attempts to portray itself as a bridge between the global North and Africa.

Introduction

Russia joined the Group of Seven to form the G8 in 1998, when it was more focused on its economic and other domestic problems. However, during Vladimir Putin’s presidency (2000-2008), as the economy recovered due to increased revenues from high oil and gas prices, Russia became a net creditor, and foreign policy was used more systematically as a tool to further its economic goals and revive its great power status.[1] Aiming to fulfill these objectives, Russia began to participate in debt relief and other multilateral development assistance programs, particularly in Africa, to which the G8 had been paying a growing amount of attention.[2]

In Russian foreign policy circles, realist thinking, which focuses on geopolitics and balance-

Delivery:
1-484104333-4
Publisher:
University of Florida, Center for African Studies
Journal:
African Studies Quarterly: the online journal of African studies
ISSN:
2152-2448
Issue:
Volume 11, issue 4 (African Studies Qua
Article:
A Bridge Between the Global North and
File path:
data/UF00091747_00036/v11i4a5.pdf
Relative file path inside the bag

Attributes:

Name	Value
author	Jordan, Pamela A.
title	A Bridge Between t

Links:

Name	Value
CTS Download	http://sun9.loc.gov:80
CTS Viewer	http://sun9.loc.gov:80

Why is Transfer so Important?

While our initial interest in this problem space came from the need to better manage transfers from external partners to the Library, the transfer and transport of files within the organization for the purpose of archiving, transformation, and delivery is an increasingly large part of daily operations.

The digitization of an item can create one or hundreds of files, each of which might have many derivative versions, and which might reside in multiple locations simultaneously to serve different purposes.

Developing tools to manage such transfer tasks reduce the number of tasks performed and tracked by humans, and automatically provides for the validation and verification of files with each transfer event.

Why are we looking at close integration between transfer and inventory functions in a repository?

Inventory services can bring several benefits, including collection risk assessment and storage infrastructure audits.

Realizing any benefits for effective data management relies on knowledge of data holdings.

Knowledge of file-level holdings and recording of life cycle events related to those files from the moment that they enter the collection and in every future action reduces future risk by storing information that can be used in discovery, assessment, and recovery if and when a failure occurs.

Why is Modular So Important?

Identifying needed services as modular rather than monolithic has allowed the Library of Congress to research and implement each of these functions in a more nimble way, all the while planning to fit those services into a larger scheme of repository services.

The integration of modular transfer and inventory services as well as workflows allows for separation of tasks based on project or collection or format needs while supporting backend data integration where required.

Modules can be independently re-implemented in the future when the need arises. This also allows for extensions to services and functionality that we have not yet even considered, let alone planned for.

Is This a Repository?

Not yet. These modular services do not yet equate to everything needed to call a system a repository.

- Most of the service functions outlined in the Requirements document are not yet implemented.
- There are only detached end-user discovery and delivery applications.
- Descriptive metadata is not yet tracked with the media files.
- There are currently no granular rights and access policies nor means to enforce them.
- Preservation monitoring is not yet in place.

But there is a set of services that equate to many aspects of “ingest” and “archiving” – the registry of a deposit activity, the controlled transfer and transport of files, and an inventory system that can be used to track files, record events in those files’ life cycles, and provide basic file-level discovery and auditing.

Through the Inventory tools we expect to be able to provide persistent access at a file level. In other words, it may not be a full-blown repository yet, but is the first stage in the development of a suite of tools to help the Library ensure long-term stewardship of its digital assets.

What are the next steps?

Systematic inventorying of all Library content

Development of additional workflows

Development of the remainder of the Repository modules

Outcomes:

The development initiative has informed the Library’s preservation efforts, building our understanding about what we need to know about files and what events in their life cycle we need to record and track.

The Inventory Tool is supporting the Library's initial efforts in a file-level preservation audit.

The Library’s developing Repository Services provide observability of the state and location(s) of files, enabling querying, auditing and reporting.

Questions?

Leslie Johnston

Chief of Repository Developemnt

Library of Congress

lesliej@loc.gov