Detailed Notes from DC Fedora User Group Mtg 10.14.11

Twitter feed at #DCFEDORA//.

Hydra

Matt Zumwalt, Mediashelf and Robin McGovern, UVA

-brief overview slide of architecture

-new-ish website: http://hydraproject.org/

Islandora

Mark Leggot, Discovery Garden/Univ of Prince Edward Island

-goal to support full life cycle of research (from emails, raw data, analysis, etc.)

-Islandora is middleware written in PHP, Phyton

-Fedora is storage layer - all data, display, search configurations are stored

-Drupal is the UI

-working on a possible Word Press viewer

-Mark showcased current projects and highlighted future projects

-Mark talked about a future project - Hydra-Islandora - cross framework interoperability

-LIB20 - Dell appliance pre-installed w/Islandora - customer would get support from both Dell and DGI

-Virtual machine for each release of Islandora, have a how to manual to help figure out if Islandora is for you

Goddard Space Flight Center

Mitzi Cole

-Drupal over Fedora (not Islandora)

-live in Feb 2011 internal only

-will be live for the public in June

-3 different projects:

--1) NASA case studies

---for training situations w/in Goddard

---each case study has at least 1 pdf and repo is full text searchable

---some case studies have supplemental material, like audio files

-- 2) Colloquia material - grey material

---audio/video of presos at Goddard

---serve as both catalog and storage

---restricted access rights for some material

---object can have slides too

--3) Authors & publications

---material published by employees of Goddard

---approved for public release - harvested from other publication site sources

---want authors to submit their materials

---repo points to published links outside Goddard

---has relationships for authors - can create author records and RSS feed, can view relationships between authors and their co-authors

---can browse by internal organization code based on NASA authority control

---generating publication related stats

---~~1st approach - which journals were Goddard employees being published in the most? - extract database on code manually~~ >not sustainable

----2nd approach - automated stats - query holding tables, SQL queries to retrieve results from pub/author holding tables, calculate metrics, display data

National Agricultural Library (NALDC)

Don Gourley

-now have repo up: NALDC.NAL.USDA.GOV

-silos of resources - trying to pull them together

-40k docs in DSpace repo, 30k docs in Zylab repo, various other collections in ad hoc software/locations

-so far have replaced the Zylab repo

-want to pull in the other materials

-small staff - found a way to move quickly:

--1) re-use content models done by others

---don't worry about doing it perfectly ourselves

---use Hydra content model - take advantage of their work

---like additive nature of Fedora - more content models provides future flexibility

--2) loosely tied to Fedora

--- doesn't interact w/Fedora, interacts w/SOLR instead

---real work was getting content into SOLR

---need to revisit front end and how we index it on our next generation of projects: need to figure out embargos, XMACA policies, how to use GSearch, JAVA msging svc

National LIbrary of Medicine (NLM)

John Doyle

-historical books and films

-ramp up book ingest/scanning

-2k resources in, 1400 more in the queue

-need figure out how to tackle multi volume books - how to do the content model

-limited by Muradora - trying to figure out the next step - Hydra/Islandora?

-nominated for an award

-Nancy Falgran from NLM wants to share ideas on how to approach multivolume monographs

-lots of data streams - millions - that represent the books

-books will be copied over to Internet Archive, might also end up at Hathitrust - don't know how to get usage stats from either place though

National Techincal Information Service

Donald Hagen, Gail Hodge, Daniel Redman (lead developer)

-Fedora Science Repository Service (FSRS)

-been working on for 6 months

-joint venture w/Gail Hodge from Information International Association (IIA)

-NTIS gets no funds, all $ is from charging for technical reports and microfiche

-want to help multiple agencies develop repos-->true collaboration among groups

-work is incremental-->more effective over time

-Fedora is storage layer

-SOLR is search engine, robust interface, multimedia display, faceted browse, full text

-model written in PHP

-collections that represent agency

-need to preserve - data streams important - now worried more about access - NOAH is focused on preservation

-UI components

--layer of repos (our repos and others)

--can be hosted at agency or hosted at NTIS and take advantage of other svcs

-MODS metadata schema, might also use Geospatial

-Object modeling for NTIS tech reports - lots of numbers associated w/reports to identify - set up object model to reflect that

-Flexibility - modularity of Fedoras architecture is important - different objects/data models, flexible metadata relationships

-Projects:

--1) The National Tech Reports Library

--~~April 2009, 65 institutions, IP subscriptions, 630k records/docs~~ >moving to Fedora in early Q1 2012

--2) The NOAA Deep Water Horizon Repository

---different audience, different content

---tech reports, but not that complex

--~~diversity of content/resources - full text, media, etc~~ >different metadata (MARC XML mapped to FSRS fields) makes for a more complex schema

---Phase 1: prototype by the end of Oct, Phase 2: add'l media formats & features by late spring

--Next steps

---implemented GSearch in NOAA environment

---interface improvements & admin capabilities

-FSRS supports collaboration - share info resources and costs in order to create cross agency collections

Rutgers University

Ron Jantz

-RUCore - instutional repo

-plans for a science data repo - need to be able to archive large files for research data

-UI & svcs

--lots of collections coming into 3 different portals: scholarship, cultural heritage and research data

--MODS descriptor metadata

--want to make sure everything in scholarship portal is indexed by Google Scholar - used a site map and that wasn't sufficient, had to contact Google, seemed to flip some kind of switch and then everything was indexed

-flexible architecture - some faculty are creating there own portal

-data - a new library svc

--collect data from lots of different disciplines

--lots of different process and tools

-~~15-20 GB in size for some of the data sets~~ >trying to come up w/solutions

--new role for the library - joining researchers on grants

--developing a model for metadata - to track life cycle of data - from grant, to research protocol, data gathering, analysis, etc.

--compound data object model - data sets, surveys, research protocol, instruments, maybe some custom software to collect data

--when searched you bring up data object

--RUCore's large data set issues - interium archival mgmt external directory, different ingest process for large and small data sets

--proposed solution >move everything in Fedora now to a backend archival system - object is in Fedora, point to it w/eformat

Smithsonian Institute

Thorny Staples

-Smithsonian NOT founded as a library, founded to advance and diffuse knowledge

-our department was created to support researchers at SI: earth science, biodiversity, endangered species

-9 science centers around the world

-current project is the "repository enabled virtual research environment"

--goal is to preserve research and make it durable

---start with a conceptual model of how scientists work

---content creation (new data)

---analyze data

---disseminate data

--->need a place to do this -> why not a repo

--architecture to support research projects

---objects focal point - objects relate to one another

---understand how to look at data - how data relates to each other

--Projects

---1) SI Wild Project: camera trapping project

----set up heat sensitive camera, take photos

----gather digital photos and preserve

----create simple metadata about each project - location, time/date, etc.

----different territories for each project

----using DarwinCore metadata

---2) Botany field project

----collect plants - reseachers go on collection expeditions, write in field notebooks

---3) Archaelogical Excavation data

----feature, lots, context (chunk of dirt - photos, describe, dig/brush away and destory), finds (images)

---put context into lots (a building) >researchers create drawing based on pieces and what they imagine the building was

----currently developing research project content model

University of Virgina

Julie Meloni, Adam

-Virgo: digital items integrated w/the rest of the library catalog - patrons don't know the difference between digital items or actual books

-Libra: using a Hydra head, looks just like Virgo (the library catalog), search results look the same

-freedom for the depositor to put content and metadata in - not unmediated by library staff, but only a small amount of metadata is req'd >goal is to sustain, not provide descriptors

-"two-store"

--technical solution to organizational problem

-repo was very successful, library kept asking for more server space >created all sorts of problems for the server guys

-3 part solution to help w/storage issue: maps areas in bitstreams, Akubra mounts, multiplexing combining into repo tree>make storage req'mt look like other reqmts on server - it is cheap, but when it looks strange, it is expensive -> this solution makes the storage reqmts look like other reqmts on the server

United States Geological Survey

Helen Tong, Richard Huffine

-repo project started 2 years ago

-currently a pilot project, not a live system yet - hoping to launch by the end of the year

-goal: repo to manage and curate the library's digital collections and link relevant data to facilitate integration and data re-use

-2011: chose Fedora because of flexible object model, journaling used for replication, using USGS lower level storage

-2012: implement Islandora for end user interface and for staff, need to configure new storage space - have 6 TB but hardware wasn't installed

-collections

-~~USGS publications warehouse - all pubs for USGS, all pdfs and docs, some scanned, some born digital~~ >need to ingest

-~~USGS photographic library - photos and sketches done by scientists dating from 1870s - 400k photos, only 38k are digital~~ >need to digitize quickly

--USGS library online catalog - inventory of library's print material

-object model - seperate metadata and asset record

-object to object relationships in order to represent more complex objects

-future challenges

--want to reference individual parts of books as a seperate object (like a map in a book)

--some material not for public consumption - just USGS employees only

--authority control for author and subject

DuraCloud

Andrew Woods, DuraSpace

-talked in more detail about tools for integration of Fedora and DuraCloud - SyncTool and CloudSync