DSpace Futures Metadata Improvement Meeting Notes

The following are notes from the initial community meeting to discuss the proposal to update the QDC registry and add DCTERMS registry from March 20,2013.

Attendees

Sarah Potvin, Texas A&M

Maureen Walsh, Ohio State University

Sarah Miles and Geneva McDonnald, Georgetown University

Bram Luyten, @mire

Ann Devonish, Woods Hole

Danielle Buselle University of British Columbia

Bob Sandusky, Univ of Illinois Chicago

Melissa, University of Missouri

Mark Wood, IUPUI

Jim Holiday, Indiana University

Jennifer, Indiana University

Richard Rodgers, MIT

Tim Donohue, DuraSpace

Valorie Hollister, DuraSpace

Ivan Masár, DSpace commiters

Peter Dietz, The Ohio State University Libraries

Felicity Dykas, University of Missouri

Notes

Overview of DSpace Futures

Valorie Hollister

DSpace Future discussions were inspired by some momentum in the Fedora community.

- A group of Fedora users determined that they were really at a crossroads and needed to make some significant improvements in Fedora -- far more complex and involved than the normal development cycle with volunteer committers could accommodate.
- They formed a stakeholder group and established the Fedora Futures project.
- DuraSpace held a number of calls with the community to explore what the current perceptions the DSpace community had about DSpace and whether or not there could be a similar project/s for DSpace.
- Out of those discussion we identified 3 projects that to address some ideas / needs that came up the most frequently. The first 2 project ideas have already had calls:
    - 1) developing a REST API for DSpace
    - 2) creating DSpace and Hydra integration of some type
- The third issue that came up on the initial DSpace Future calls was on the need for metadata enhancements /improvements
- This project is a bit different -- in that the DCAT group has been working on gathering community feedback and putting together a detailed proposal for the feedback and involvement by the community.
- Because DCAT proposal really lays the groundwork for other metadata improvements we decided to fold the metadata update project into DSpace Futures.

Summery of Sponsor Summit

- Since this meeting is supplanting the normal DSpace committer mtg and ordinarily they would have chatted about the sponsor summit last week, we'll provide a quick summary here

- Great summary that Tim Donohue sent out yesterday along with some of the presentations slides to both DCAT and the committer list.
- Will publish detailed mtg notes along with the all presentation slides for those who are interested in the next day or so.
- Other sponsor meetings held in the last few years, the goal of the meeting was to provide DuraSpace sponsors with an update of the organization and some of the key issues we face. This is the first time that we've held the mtg in person -- usually it has been a virtual meeting -- conference call/webinar syle.
- DuraSpace and the DuraSpace board wanted to clarify:
    - DuraSpace's role relative to DSpace and Fedora -- including debunking some very commonly believed myths as well as how DuraSpace fits into the larger digital preservation ecosystem (see Tim's notes)
    - Have a discussion what the community needed to do to rapidly move forward on an effort to modernize both DSpace and Fedora - including a creating roadmap, governance/steering cmte, finding funding and developer/project resources for the effort
- Good breadth of the type of attendees at the mtg -- obviously many from sponsoring institutions -- but there were university libraries, repository managers, technologist, technology leaders
- Overall the meeting was very energizing - everyone seemed interested in helping to push for some progress and continuing to support the DuraSpace organization
- Specific actions going on for the DSpace community:
    - 1) developing a non-technial roadmap of where the community needs the software to be in the next 3-5 years
    - 2) developing a governance strawman -- so that the community can start a modernization effort, seek funding, etc.
- How does this all relate to DSpace Futures projects?
    - Sponsors driven initiative may take on much bigger modernization project
    - None of the 3 current DSpace Futures projects would conflict with that -- and in fact, would likely end of being an integral part of it.
    - We would encourage the community to press on with current DSpace Future projects and hope that a some point there may even be some resources or funding to help push the work along.

Overview of Metadata Improvement Proposal

Detailed proposal: Updating the Qualified Dublin Core registry in DSpace to the latest standards of the DCMI

Maureen Walsh and Sarah Potvin (other proposal team members include: Amy Lana and Bram Luyten)

- project arose from community survey
- DC default schema currently
- Local instances modify dc default schema, having a more standard schema would be better
- Laid out possible phases of update - ultimate goals to move to DCTERMS, done in phases
- Start with update to current QDC ('dc' would stay), add DCTERMS, would also be a local schema and an administrative schema open to modifications, but we are urging that we lockdown standard schemas

Questions/Feedback

- Richard, MIT: What is the scope and vision? Agree metadata models are in need of refurbishment. What is the thinking about actions to take on current installs or new systems moving forward?
- Maureen
    - we do want to help current instances upgrade, not just new instances
    - lots of technical issues - what is the graceful path forward? need to provide tools and update crosswalks and review everything this would affect - add-ons, features, interfaces, etc.
    - we are talking about helping current instances
    - we would provide tools to do an update - these are the registries and here is what you have to do to overlay - have to help people address
    - always going to have a schema named dc (not repurpose the name)
- Bram: Want to mention a bit about the benefits/inconveniences - further we are away from standards, the more overhead for importing and exporting - goal is to lower overhead costs on import/export and to be in compliance with standards

- Mark: Think this is basically a sound proposal, like the overall shape, something we need to do
- Maureen: Lot of areas we need to flesh out / some areas to make decisions yet on - put comments in wiki page
- Mark: Concerned about possibility moving DSpace-specific terms into institution-local - local should be local
- Maureen: Yes, local should be just what local needs are, could move DSpace admin metadata into a schema for admin metadata, maybe 'dspace'
- ?: Does the proposal suggest a flat DCTERMS vs. or does it imply we will need a new data model?
- Richard: If we stick with flat we have a chance of reaching that with existing sites - expanding registry. If we change the underlying data model we can't, because older versions of DSpace didn't allow editing the schema (approx. sometime before DSpace 1.5)
- Maureen:
    - Phase 1 and 2 is backward compatible (with versions allowing multiple schemas), phase 3 may not be /not sure of implications - maybe that full implementation of DCTERMS means that there would be a break
    - Try to update old version w/tools - but tools would really be designed for upgrade process (like SWORD upgrade)
    - Write off older versions that don't have ability to add schema -- but 1.5 and above we may want to consider
    - Some of the tools may be usable with earlier versions w/schemas - backport tools
- Mark - Phase 1 seems cheap
- ?: Confusion when we have multiple schemas that have semantically the same fields - how do we allow for almost identical fields?
- Maureen: Both DC and DCTERMS would be there - so we'd have to set a default schema
- Tim: Phase 1 - only adding a parallel schema - cheap, starting to create migration tools, parallel schema to look at it and think about and start to migrate before jumping to DCTERMS
- Maureen: People would need help with local schema - migrate them
- Sarah: Bring us to compliance - have local terms become compliant, give people who want to start to experiment
- Felicity: (Univ of Missouri): We've added a lot of qualifiers for more granularity - are we locking down at the qualifier or element level? I imagine there is not a DSpace instance out there that hasn't added qualifier.
- Sarah: Elements are made up and added - non-compliant, added qualifiers is still an open question - should they be moved into the local schema, probably outcry if we don't allow for qualifier - but to think about moving them to local at some point during the migration
- Tim: No new elements and then figure out how to manage custom qualifiers later
- Maureen: could lock down on UI, would allow people to change code (Richard: customizing the code will be always an option)
- Mark: Maybe the question is do you want 1 or 2 revolutions - qualifier and element lock down all at once or in stages

Next Steps:

- Val to post notes to wiki and to mailing lists, including a call for more participation – developers and non-developers to work on implementing phase one and refining plans for subsequent phases.

Begin forwarded message:

From: Tim Donohue <tdonohue@duraspace.org>

Subject: Report from the DuraSpace Sponsor Summit

Date: March 18, 2013 5:53:00 PM EDT

To: "dspacecommunityadvisoryteam@googlegroups.com" <dspacecommunityadvisoryteam@googlegroups.com>

Reply-To: DSpaceCommunityAdvisoryTeam@googlegroups.com

Hi DCAT members,

As you may know, DuraSpace hosted a Sponsor Summit last week in

Baltimore. Val & I attended (along with some of you) and wanted to report back to all of you on some of the discussions that took place there. (I've sent this same below summary to Committers as well.)

The discussions at the Summit will be summarized/advertised publicly

(likely in the next week). But we wanted to get some early notes to all

of you ASAP, just to bring you into the loop.

The attendees to this summit were mostly Repository Managers and

University Librarians / Heads of Departments (with a few tech folks

sprinkled in). The discussions took part over the span of two "1/2 day"

sessions.

What follows are some rough notes (from me), which summarize the main

concepts & discussions from the meeting. I'd encourage others who attended to share their own notes (especially if I missed something or misrepresented something).

We're all glad to answer any questions that my come up! Just ask.

- Tim

========================
Day 1 - Overview for All
========================

The first day's discussion was essentially an overview of the issues at
hand and some very initial brainstorming.

James Hilton, CIO at U of Virginia (and President of DuraSpace Board of
Directors) : Where DSpace / Fedora fit into larger ecosystem

(Slides attached: "James Hilton - duraspace summit.pdf")

* Detailed what he sees as the "Emerging Digital Stack" to support
Access & Preservation. DSpace and Fedora are KEY components in that
stack. We need to invest in them as such
* Not a crisis meeting. Rather this is an opportunity to "move the
needle" towards a future (for DSpace & Fedora) that better meets all our
needs.
  - DSpace and Fedora were well funded initially (ten years ago) -
Millions of dollars were spent. But, that spigot is shut off, and we are
working towards making these open source projects sustainable.

Michele Kimpton & Jonathan Markow of DuraSpace : Detail the purposes of
the Summit & Debunk some DuraSpace Myths

(Slides attached: "Summitmeeting2013.pdf")

* Debunked some "myths" about DuraSpace. NONE of these are true:
  - DuraSpace has teams of developers working on DSpace and Fedora
(False - we just have a tech lead for each project)
  - DuraSpace is funded by the Moore Foundations and others (False -

DuraSpace does not receive grants to support DSpace or Fedora or itself)

  - DuraSpace is DSpace on Fedora (False - DuraSpace is not software,

it's an organization to support software)

  - Sponsorship dollars are funding DuraCloud (False - DuraCloud

development was funded by Library of Congress)

  - Bringing on more projects reduces effort on existing projects

(False - there's some shared infrastructure with various OSS projects,

but they all require support staff and funding)

* DuraSpace is a small team. We separate our work into "OSS Projects"

and "Services"

 * OSS Projects (DSpace, Fedora) are funded directly by Sponsorship and

Registered Service Providers.

  * Services (DuraCloud, DSpaceDirect, Educational Webinars) are funded

by service revenues and occasional grants.

* Unfortunately, the reality is that the funding for DuraSpace OSS

Projects is lacking. DuraSpace has had to pay "out-of-pocket" to help

fund/sustain these projects. The amount of DuraSpace support provided to

these projects is currently no covered by the sponsorship $$. (see

slides for charts/graphs)

* This isn't a crisis. DSpace & Fedora & DuraSpace are all fine. But, we

all need to realize that this is NOT SUSTAINABLE over the long term.

DuraSpace cannot be constantly losing money on these OSS Projects --

otherwise DuraSpace may have to eventually cut back staffing of these

OSS projects (or find other ways to fund them).

* DSpace & Fedora are really COMMUNITY OWNED projects. DuraSpace helps

steward them, but they are owned by everyone.

* The purpose of the meeting is to brainstorm ways for us to increase

levels of community contribution and engagement to ensure long term

success of DSpace and Fedora.

Tyler Walters, Dean of Libraries at Virginia Tech : A University

Librarian Perspective

(His slides are included in "Summitmeeting2013.pdf")

* He gives to DuraSpace (and many other programs/projects in the same
ecosystem) because he feels the Return on Investment (ROI) is excellent.
* At Virginia Tech, they could fund 1.4 Full time staff for the money
they invest in various programs/projects (DuraSpace/Fedora/DSpace just
being one of them). But they would never get as much from that one staff
member as they get from this investment.

General Brainstorming

A general brainstorming session took place. I unfortunately didn't
capture all the general discussion. But, the main concepts were that
people really do feel that DSpace & Fedora are essential, highly
important platforms. Everyone wants to find ways to move both platforms
forward and help them modernize (and possibly even merge -- some
mentions of "DSpace on Fedora" / "DSpace on Hydra"). Also want to find
ways to improve the available $$/resources for the projects. It may be
that new governance models will be needed for these projects (e.g. a
possible "Steering Committee" for each project -- more on this below).

==
Day 2 - Project Specific Breakout sessions (DSpace specific notes below)
==

The second day, we brainstormed ways to move both projects (DSpace and
Fedora) forward more rapidly and modernize the software. The reality is
that both projects are currently underfunded (and arguably understaffed,
except by a very avid volunteer base). Is there a way to find
funds/resources to help modernize both platforms?

DSpace "Challenges"

----

* Committers are doing a great job maintaining and helping add new
features. But it's difficult to always find the bandwidth to do a
long-term roadmap. So, Committers concentrate more year-by-year with
some overarching concepts/brainstorms in mind.
* Even if the long-term roadmap was there, we also may have a
resource/funding problem. We'd need to find the $$ or developers to make
larger initiatives happen.

* Some vague ideas still floating around as to whether DSpace should change to use Fedora or Hydra. There's several directions DSpace could go, or it could remain it's own separate system.

* Currently there's a bifurcated user community. We have a small group of very active, advanced users of the software (e.g. most of the Committers and their institutions). There's also the "long tail" -- many many small institutions who use DSpace for its ease of getting started quickly. The latter is actually the majority of our user base.

* Any longer term roadmap would need to take into account both parts of our user community -- the advanced users and the "long tail"

* We have a bit of a "chicken & egg" problem. We need a roadmap/vision to go get extra funding/resources. But, we also may need resources to develop a compelling / sufficiently detailed technical roadmap.

DSpace "Next Steps"

Out of the discussions summarized briefly above, we decided upon the following next steps:

1) Develop a compelling high-level "vision" for DSpace for the next 3-5 years.

    - This will be a non-technical, but sufficiently detailed vision statement which we could take to Library Directors (or similar) to try an obtain some extra funding/resources to make it happen

    - This vision will also attempt to place DSpace within the larger "ecosystem" of Scholarly Communication / Preservation. Describe its niche.

    - A small group (10 or less - TBD) of folks will meet in Chicago in coming months to draft up a "vision statement". It will then be distributed for broader feedback from Committers, DCAT, community, etc.

2) Once we have a vision statement, we will likely need to create a "Governing Board" / "Steering Committee" of institutions. The goal of this group would be to help make the DSpace vision a reality (and to help find the funding/resources to make it happen)

    - This "Steering Committee" would likely consist of a mixture of University Librarian types, Repository Manager types, and Tech types.

    - Members of this Steering Committee would likely be those that are willing to provide extra $$ or resources to put towards the vision. Members may also include some individuals who are "voted in" by the Community.

- What this governance structure would look like still needs to be
decided. But it'd obviously have to work in conjunction with our DSpace
Committers and DCAT.

    - David Lewis (Dean of Libraries at IUPUI), Ann Wolpert (Director
of Libraries at MIT) and Tyler Walters (Dean of Libraries at Virginia
Tech) will develop an initial draft / concept and distribute it for
feedback.

--

Tim Donohue

Technical Lead for DSpace

DuraSpace.org

--

Tim Donohue

Technical Lead for DSpace

DuraSpace.org