

DevMtg 2012-07-09 - OR12 Meeting

- [Developers Meeting before OR12 on Mon, July 9, 2012](#)
 - [Who is invited?](#)
 - [Logistics](#)
 - [General Agenda](#)
 - [Possible Open Discussion Topics](#)
 - [Sign Up to Attend!](#)
 - [Meeting Notes](#)
 - [Jonathan Markow on Duraspace Managed Projects](#)
 - [Valorie Hollister on DCAT](#)
 - [Sarah Shreeves open discussion](#)
 - [Robin Taylor DSpace 3.0 Release discussion](#)
 - [Mark Diggory on Github and new JIRA Workflow](#)
 - [Richard Rodgers Summary and "Blue Sky" Discussion](#)

Developers Meeting before OR12 on Mon, July 9, 2012

Face-to-face developer's meeting on topic of DSpace RoadMap before [OR2012 in Edinburgh, Scotland](#).

Who is invited?

- All Committers,
- All [DSpace Community Advisory Team \(DCAT\)](#) members,
- Any other interested DSpace developers or technology-savvy individuals

If you don't fall into one of the above categories, you are still welcome to attend. However, be warned that discussion will likely get very technical at times (which is why we recommend you be a developer or have a technology background).


Space is limited. Please sign up on the [Sign Up sheet](#) below. If we begin to achieve capacity, preference will be given to Committers or DCAT Members. However, we welcome any interested DSpace developers to join us and take part in the discussion!

Logistics

- Day: Monday, July 9th, 2012
- Time: 9:00 AM - 4:00 PM
- Where: **Room 1.11 in Main Library, University of Edinburgh**
 - Google Map: <http://www.ed.ac.uk/maps/buildings/main-university-library>
 - Printable PDF campus map: http://www.ed.ac.uk/polopoly_fs/1.20490!fileManager/campus_maps.pdfhttp://www.ed.ac.uk/polopoly_fs/1.20490!fileManager/campus_maps.pdf
 - Someone will be at the reception desk to direct you to the room. Instructions/Directions will be left at reception for any latecomers.
- Refreshments
 - Coffee / Tea / Caffeine: The Library has a Café (and others are nearby). Some coffee/tea will also be in the room.
 - Lunch will be provided from 12:30pm-1:30pm
- Attendance Limit: Likely 30 people maximum
- Wifi: Accounts are provided at the conference registration desk. But, as it's not strictly necessary to have registered before attending this meeting, accounts will also be available at the meeting.

General Agenda

Format of Meeting

 This meeting will be organized as a group discussion. Although at times one or more of us may lead discussion sections, it is meant to be a group discussion and **not** a series of presentations.

Meeting Notes will likely be taken on PiratePad: <http://piratepad.net/>

AGENDA:

"Master of Ceremonies (MC)": [Richard Rodgers, MIT](#)

9:00am: Introductions, Overview of Day (Discussion Director: Richard Rodgers)

- Introductions

9:15am: DuraSpace / Fedora / "DSpace with Fedora Inside" Discussion & Updates (Featuring Jonathan Markow)

- *First 1/2 hour* : Jonathan Markow, DuraSpace's Chief Strategy Officer, will discuss and lead a Q&A regarding the latest DuraSpace strategies, Fedora RoadMap/Plans and provide updates about the "[DSpace with Fedora Inside](#)" concept.
- *Second 1/2 hour* : Extra time for Q&A (as needed) or open discussion on DuraSpace / DSpace with Fedora Inside / Fedora plans and how they relate to DSpace plans/roadmap.

10:15ish : Break

10:30am: DCAT Discussions (Discussion Director: Val Hollister)

- Developer / [DCAT](#) collaboration - how has it been, what improvements can we make

11:00am: Open Discussion Period (Discussion Director: Sarah Shreeves)

- We will discuss "hot topics" decided by all of us (See [#Possible Open Discussion Topics](#) listing)
- This time can also be used to touch back on topics from earlier in the days' discussion.

12:30 - 1:30pm: Lunch Break (LUNCH WILL BE PROVIDED)

(Please note that the after-lunch sessions are meant to be a bit more "technology heavy". Non-developers are still welcome to sit in, if you wish. Though you are also welcome to join up again later at the chosen local pub. It's entirely up to you.)

1:30pm: 3.0 Planning / Updates / Digging Deeper (Discussion Director(s): Robin Taylor)

- More detailed planning or updates around [DSpace 3.0](#)
- Are there any areas of high interest that need volunteers/support (perhaps based on earlier discussion)? I.e. can we identify areas of "high priority" or potential "low hanging fruit" that needs a person or two to just help out?
- "Finalize" our 3.0 Release Schedule / Feature Freeze dates?

2:30ish: Break

2:45pm: Discussion of DSpace Developer Practices (Discussion Director: Mark Diggory)

- GitHub usage: Any questions/comments from GitHub experience so far? Tips to share?
- Are there ways we can better leverage our developer tools (JIRA, GitHub, Continuous Integration, others) to make things easier on us all?
 - Capture various brainstorm / ideas and perhaps even find volunteers to investigate them (if plausible)
- Any suggestions for how we can improve our development workflow in general? Or, even just improve our patch/pull request workflow(s)?
 - One suggestion of a possible JIRA workflow improvement: [JIRA Workflow Improvements](#)
- What do you need as volunteer developers to make your "lives" easier?
- It would be nice to come out of this session with a ranking of "high priority" infrastructure or workflow changes & a list of volunteers who are willing to help.

3:45pm: Summarizing & Bringing it all Together (Discussion Director: Richard Rodgers)

- Ensure any "To-Do's" or Task Assignments are documented
- Any topics that need further discussion & may warrant a "Special Topic Meeting" or even another "Developers Summit"?
- Meeting Wrap-Up

4:00ish: End Meeting & Depart for the [Pear Tree House Pub](#) for Food/Drink

- The [Pear Tree House](#) is in walking distance from the conference site, and has plenty of seating and a large beer garden.
 - [Google Map](#)
- Even if you were not able to attend the meeting, you are still more than welcome to join everyone at the pub for food, drinks or just interesting discussion!
- NOTE: Attendees of the parallel [OR12 Fedora Developers Meeting](#) will also be joining us at the Pear Tree House. So, this is an opportunity for DSpace & Fedora developers to mingle!

Possible Open Discussion Topics

*This is a list of topics which we may wish to touch upon either during the "Open Discussion" period, or during another part of the day. This list is **unordered** and **unranked**.*

Please feel free add any topics to this list that you feel would be worth discussing!

- Various Metadata enhancement work/proposals
 - [Updating the Qualified Dublin Core registry in DSpace to the latest standards of the DCMI](#)
 - [Standardizing the default namespaces](#)
 - [Adding metadata authority controls/vocabularies to the data model](#)
 - [Moving metadata related configurations from dspace.cfg to the database](#)
 - [Develop support for additional metadata standards \(use the "Other" field to specify which standards\)](#)
 - [Improved \(or more transparent\) flexibility](#)
 - [Enhancing the metadata available for Communities, Collections and Files \(bitstreams\)](#)
- Re-examination of Authentication / Authorization (RELEVANT LINKS NEEDED)
- [XMLUI Theme Repository](#)
- [i18n Improvements Proposal](#)
- [Configurable submission](#)

Sign Up to Attend!

If you're planning to attend this meeting, please add your name to the sign up sheet. This will allow us to determine a proper headcount. If you only plan to attend for part of the day, you can also let us know (e.g. "AM only", "PM only", etc.).

Sign Up Sheet - Will Be Attending

1. Kevin Van de Velde

2. [Bram Luyten \(Atmire\)](#)
3. Valorie Hollister
4. Stuart Lewis
5. Amy Lana
6. Sarah Shreeves
7. Bill Ingram
8. Ryan Scherle (until 1:30)
9. Richard Jones
10. Andrea Schweer (from lunch onward)
11. Elin Stangeland
12. Robin Taylor
13. Richard Rodgers
14. Mark Diggory
15. Graham Triggs
16. [Hardy Pottinger](#)
17. Pratyusha Doddapaneni (EDINA)
18. Gareth Waller (PM only)
19. Jonathan Markow (AM only)
20. Tom Johnson

Want To Attend / May Be Attending

- Andrea Bollini (CILEA - I will arrive at 13.30 to the EDI airport)
- Davide Vitale (CILEA - I will arrive at 13.30 to the EDI airport)
- (add your name here)

Will Not Be Attending

- [Tim Donohue](#) - family conflict

Meeting Notes

The following is a brief summary of the meeting. No official note taker was appointed. These notes are cobbled together, based on the shared recollection of those in attendance, and the notes they've made available. If you have a revision or an addition, please share. Thanks!

Jonathan Markow on Duraspace Managed Projects

Jonathan Markow began the day with an explanation of a new initiative from Duraspace called 'Managed Projects', this was further expanded on in the Duraspace plenary. In response to promptings from the Fedora committers Duraspace will attempt to organise funding and/or resources to facilitate more sizeable pieces of work. (More on "Managed Projects" can be found at: [Fedora Committer Mtg Notes \(2012-05-10\)](#) - This was the first mtg where Managed Projects were discussed with the Fedora Committers) He continued with a brief discussion on 'Fedora inside DSpace'. It was noted that no institution has been willing or able to take on this work and as such it has not progressed. This is not to say that the idea is dead, just a statement of reality.

Some cross discussion about best ways to foster big development projects/large architectural changes in community-supported open source projects. A discussion of a need for better communication between committers and the community, possibly an information exchange about active development projects. Valorie Hollister pointed out that there is a page for this information on the wiki, the [Development Proposals page](#), though it was later pointed out that the wiki page doesn't really provide a good way to **track project status**. Bram Luyten suggested the possibility to automatically track an "outside" project status by linking to other issue trackers and wikis, perhaps via RSS feeds.

Valorie Hollister on DCAT

Valorie Hollister spoke about DCAT:- a description of what it is, who participates, etc. Valorie observed that the "most successful meetings have been when committers have been present," which leads to a discussion of ways to encourage this synergy in the future. The idea of a regular joint meeting of DCAT and the committers was brought forward. Everyone in the room seemed to agree that making a specific agenda available prior to a joint meeting would be key to the success of the meeting.

Some ideas were proposed to create better awareness of DCAT:

- Invite new repositories (through their managers after they sign up for a listing on [dspace.org](#))
- invite duraspace sponsors
- find ways to approach the non-english speaking community.
 - Translated wiki pages? [Confluence plugin](#)

Currently, DCAT meetings are one hour phone conferences (where you can also use Skype to dial in). The developers present in the room shared benefits of using IRC for text based chat meetings (could be in parallel with phone conversation):

- Automated transcript of everything being typed
- Subject overlaps, easy to pick up on something someone else has said before in parallel to another topic going on. This takes some time getting used to
- It tends to be short & very focused. Barrier is higher to keep typing as opposed to keep talking on the phone.

The developers commented that if DCAT will create new issues in JIRA, it would be best if:

- it's thoroughly verified if the bug or feature request isn't already logged in another JIRA ticket. If variants or related problems are logged, use the functionality to link JIRA tickets together.
- the issue category is carefully selected: New feature, improvement or bugfix is carefully considered.

Commenting on existing issues is no problem at all and should be encouraged at all times.

Sarah Shreeves open discussion

This discussion segued into the subsequent session hosted by Sarah Shreeves on 'hot topics'.

The following topics were offered for further discussion: Metadata, Auth (both N and Z), XMLUI Repository, i18n, Configurable Submission. Taking the topics in order, the discussion around "metadata" (generally understood to mean the concept of "metadata for all objects") was lively, and seemed to hover closer to the need for a better definition of what "metadata for all" really means. Richard Rodgers discussed his MDS work, where it's clear that the existing metadata system can easily be extended to apply to all DSpace objects. Richard acknowledged that there's some question and need for further discussion whether this is "good enough" to serve the use cases envisioned for "metadata for all?" Some stumbling blocks for acceptance of this approach would be the current state of date handling in the code base, and the "simple" or "flat" metadata approach may not be as expressive as necessary. At least one committer declared there is value in simply extending the existing design to all objects, citing the usefulness of a consistent interface. Richard Rodgers then asked what the status of the DCAT recommendation for out of the box metadata schemas might be. Bram Luyten offered that DC was the clear winner, just from a skimming of the results. Amy Lana pointed out that DCAT is still evaluating the survey results. Mark Diggory urged caution before proceeding, since a recommendation from DCAT is still in the works. Richard Rodgers asked whether DCAT should also consider recommending a migration path from the existing model to their recommendation.

There was also a discussion of the need for better integration of controlled vocabularies and identifier systems, such as ORCID. There appeared to be consensus that integration with ORCID would be a clear win for the vast majority of DSpace repositories given that most are focused on people.

Configurable submission was the next topic tackled, Elin Stangeland is keenly interested, but says that things appear to be a bit messy: DCAT has a wiki page up for [discussing/prioritizing](#), there was a Google Summer of Code project on the topic, Robin Taylor adapted some of the code from GSoC and put it into a patch, but only added the back-end code, not the interface work (nodding heads from many committers in the room), MIT is working on Context-guided ingest. Elin stated, and there seemed to be agreement, that what is needed is a developer to champion and make sense of the state of the work. Robin offered to take the topic up with the committers. Bram cautioned that "use cases are important for this feature," especially how one actually customizes the workflow (i.e. via a config file, or a web-based interface). For example, a delegated admin (collection or community admin) should have the functionality to modify the submission process for his or her collections, without being able to access or change those of any other collections. Furthermore, the person who can take the design & functionality decisions for the submission process might not necessarily be someone who knows how to modify an XML config file, or have the authority to do a server update & restart to put certain changes to effect.

A number of "repository" ideas were discussed next, an XMLUI "theme garden" has long been on the wish list, and now add to that a curation task and SWORD package garden/repository. From the discussion, it seems like GitHub can handle the actual code storage handily, it's just an issue of discoverability. The wiki may work OK for this task. To motivate contributions to the XMLUI theme garden, it could be nice if those themes could easily be surfaced in the demo repository. Richard Rodgers asked, "who takes the lead?" on implementing this idea? The consensus seemed to be that we should push the community to use the wiki at first and that as the number of shared themes, etc grows we should revisit a more formal strategy.

Wrapping up before lunch, a final question: "how do we deprecate code?" (The question was raised in relation to the JSPUI but was relevant to other code). Do we need a formal process, with a comment period? How could a feature or an area of functionality be identified for deprecation? The easiest case is when there is something better replacing it, guaranteeing to cover all of the original functionality while still being backwards compatible. It's a good question, but we're all hungry, something to think on in the weeks to come.

Robin Taylor DSpace 3.0 Release discussion

After lunch Robin Taylor hosted a discussion on the upcoming DSpace 3.0 release. A list of major contributions is still to be identified, more work required here. The provisional schedule was discussed and no objections were raised.

It was observed that there are a number of pull requests listed on the DSpace 3.0 wiki page, almost all of them needing further testing. If you're itching to contribute to the 3.0 process, please do consider testing these. In addition, Richard Jones has a new SWORD2 contribution (general improvements, pull request is "imminent").

Robin pointed out that we have about five weeks until feature freeze on august 17th.

Mark Diggory would like to do some Maven restructuring/refactoring, which makes the most sense to do as part of the RC release process, post-feature-freeze, so as to minimize the impact on active development.

In relation to the testathon, the question came up whether there is still an accurate checklist of things that require testing, [like this one that was created for 1.5](#)

Mark Diggory on Github and new JIRA Workflow

Mark Diggory began a discussion on developer practices with a tutorial on Git and Github and how they relate to DSpace. Mark referred to the [ThinkUpApp's "Contributor Workflow"](#), which closely mirrors the suggested DSpace Git workflow. He used a [prezi](#) version of the diagram, for walk-through purposes.

Tim Donohue has formulated compelling [improvements to the current DSpace JIRA Workflow](#). In short, the current JIRA statuses (Open, Received, In Progress) do not clearly reflect what a JIRA ticket is "*waiting for*" exactly. Therefor, following improvements are proposed:

- Only use "Received" for newly arriving tickets, instead of both Received and Open
- Introduce a "Needs More Details" status for those bugs or feature requests that are too vague for a developer to jump on and resolve
- Introduce a "Needs Volunteer" status for bugs or feature requests that have enough information and just wait to be implemented
- In progress would then be used as the status where both enough information is present and a volunteer is actively working on it.
- Instead of having both Closed and Resolved, just use one status, Closed.

In the discussion that followed it was debated whether there should be a DCAT specific status to indicate that DCAT is taking on specific issues, especially in the "Needs more details" and "Needs volunteer" statuses. I don't think there was a unanimous opinion here so just noting that I personally think the fewer different statuses, the better and DCAT can show its activity in the comment logs within particular tickets. Activity on tickets will put them higher on the list of "last changed" tickets which will get them more attention anyway.

Richard Rodgers Summary and "Blue Sky" Discussion

Quite a few interesting points popped up on the future potential of DSpace in different contexts. Stuart Lewis argued that DSpace is in pole position to offer more extensive support for **digital preservation**. The platform might not be up to par with the digital preservation promise right now, but he assured that all the necessary hooks are present. As an example, he indicated that hooking in JHove // Pronom would only be a matter of a few days of development.

Richard Rodgers mentioned LOCKSS for items & SafeArchive (*BramL: my notes are vague on this :(*)

The group felt like the game is still wide open when it comes to leveraging DSpace as a platform for **managing research data**. No other platforms are arising at the moment with substantial different feature sets than repositories, so again, an opportunity for DSpace.

There was discussion about the relation between **CRIS systems** (like PURE, Avedas, Symplectic, ...) and repositories. Particularly in Europe, CRIS systems are aggressively on the rise as they facilitate compliance with national reporting requirements, such as the RAE in the UK. Apparently in the US this trend is less visible. With CILEA's upcoming contribution of CRIS functionality for DSpace, there was discussion on whether DSpace could and should offer more CRIS functionality. There didn't seem to be a unanimous opinion on this topic.