D2.4: User acceptance

We selected five "power users" from those who were interviewed or answered the online questionnaire and that have some datasets ready to be published, and asked them to submit their data. Of the five users, three (A, B, and C) gave us detailed feedback until now.

The email sent to the test users:

Hi everyone,

You received this email because you answered the online survey about data management, and we think you are a good candidate to test the first version of the C4DM data repository. Please bear in mind that this is a TEST version, so any feedback will be very much appreciated.

The repository, which is only accessible from within QMUL, is at:

(address of the repository)

The repository is based on DSpace (www.dspace.org). A DSpace instance is comprised of Communities (E.g. departments, labs, research centres). Each community contains Collections (e.g. Theses, a project, …). Collections can belong to several Communities. Communities and collections have a limited, fixed set of metadata (Title, Description, News, Copyrights).

Collections contain Items, for examples a dataset or a publication. Items have customisable metadata schemas. Each item may contain several bitstreams (i.e. files). Each bitstream has a fixed set of metadata (i.e. name, filetype).

An item (i.e. a dataset) is created via the web interface through the submission process. The submission process is divided into several steps:
- select a collection
- add the metadata
- upload the files
- select a licences
- accept the terms and conditions

After completing these step, the data set is available on the repository, is indexed and searchable.

In this first test version of the repository, we want to focus on the submission process through the web interface. To make it easier for us to follow the tests, we created a "sandbox" collection for each user in which to try the submission process. To submit any data, you will need to register first. We will then create the sandbox collection for you. Please try to create some datasets and give us some feedback, for

example regarding the metadata fields (are there any fields missing?), the licenses (would you like to have other licenses?), and in general on the submission process.

Notes:
- If you have many files to upload, you might want to create a zip file and just upload that instead
- The URIs currently provided by the system are not registered, so don't count on them!
- If you want people to cite the dataset using a publication, add it to the metadata. Otherwise, the system will suggest a citation based on the DataCite standard.
- Because of the way the submission process is designed, although we can have as many metadata fields as we like, these cannot be chosen during the process, but must be defined beforehand by creating a "submission template". For this reason, to make the template as generic as possible, we tried to limit the metadata to the essential. If you have specific metadata that describes your dataset, attach it as a bitstream for example as an XML file. Although this is not a standard procedure, it is possible to index the text in the bitstreams and use it in the free text search.

We are working to find a way to upload files via a command line or desktop client, but for the time being only the web interface is available.

Also, we are experimenting with access rights in order to create a private and a public collection, but consider this test as if you were publishing on the public repository (although for now the test repository can only be accessed from within QMUL).

Again, if you have any questions don't hesitate to get in touch with us!

This is the text of the original feedback we received. For a summary, see the blog: http://rdm.c4dm.eecs.qmul.ac.uk/repo-user-feedback

User A:

I managed to create a collection and submit an item - it worked quite well. Some random thoughts on the process:
- *Licence options: I am not very familiar with licences, perhaps you could add some brief explanations of each licence option?*
- *When adding a citation for an item, perhaps there should be a specific input format, e.g. bibtex.*
- *I couldn't do anything with the 'Related resources' option when creating an item.*
- *When uploading an item (especially with a large file size), it would be nice to have a progress bar or some indication that the upload is actually happening.*
- *When I tried to download my submitted item, the .zip extension disappeared from the downloaded file (it might have to do with the fact that the original file contained spaces).*
- *When entering keywords for an item, is there a standard character (e.g. comma, semicolon) for separating different keywords?*

I uploaded another item and have some additional feedback:

- *When editing the metadata for an existing item, instead of having that user-friendly form when first creating the item, a page appears that is considerably less user-friendly, where the user has to edit some fields like 'dc.date.created' etc.*
- *I tried to upload a large file (~120MB) and received an error when trying to upload it for the first time (worked for the 2nd time though).*

User B:

It would be useful to have a short explanation on the structure/organisation of the repository, maybe even on the homepage? I think a simple tree diagram would do that shows how communities, (subcommunities,) collections and items relate to each other. And also an explanation on what a 'collection' and an 'item' actually is. I think I got that wrong when I did my submission. I assumed an 'item' is a file (which seems to me quite intuitive), but apparently you consider an 'item' a dataset. I only found that out when I did a submission. Maybe it would be worth considering changing those names? Also I thought a 'collection' is a collection of files and thus a dataset. I still don't see what the use of a 'collection' in your understanding is? Is it often the case that you want to combine different datasets to a collection?

On the 'create collection' process:
- *Some of the fields in the description step (in the leftmost pane) were not clear to me. It would be useful to have a small link or a pop-up to some more information about the fields.*

On the submission process:
- *It was not clear to me from the start, if a submission can only contain a single file, or multiple files. Maybe you can explain that somewhere?*
- *I was not sure, what the preferred data format is. Should I rather submit single zip-file, containing my own folder structure? Or should I submit multiple zip files for my audio and midi? I understand that you don't want to restrict that, but maybe you can give some suggestions somewhere about the usual practice.*
- *In the upload step, it would be very useful to provide a progress bar. I assume that many files in the collections will be large files and it will take a while to upload these. It can be a bit annoying to wait for the page to reload after you pressed 'upload file', without knowing if it is still actually doing something. (... after waiting for about 15 Minutes, I got a 'bad gateway' error... :/)*

(These other comments were based on my false assumptions about 'items' and 'collections':)
- *Why is there a 'standard citation' filed for an item and none for a collection? I would assume that people would rather cite the collection instead of a single item (e.g. an audio file).*
- *The same applies to the 'keywords' field and the 'related resources'. Should the keywords not be related to the collection rather that the items?*

Generally I was wondering if you really need to have 2 different steps for creating a collection and submitting files. Is there a case where you just want to create a collection but you don't want to submit files to it?

User C:

- *When creating an "Item" it's very odd to have to manually add myself as admin etc, by browsing through a list of users. This should be the default, or there should be an "add myself" button.*
- *I have a dataset with 46 files. The web interface is slow one-by-one upload. I have uploaded a zip rather than individual files, which is suboptimal.*
- *The CC License / License steps are confusing.*
    - *Why two license pages, does one override another? I don't know, and I guess others won't.*
    - *I choose "No CC license" and it automatically flips the menu back to "Please select", as if I have not made a choice, when in fact I have.*
    - *It is evil to lead people through a many-page submission process and have a new, unexpected license appear just near the end. I assume that this extra license is compatible with the cc licences (is it compatible with "no-derivs"?) - but since this second license is compulsory, this page should be at the START not the end.*
- *The completion message is confusing: "Your submission will now go through the review process for this collection" - since I am the approver, in reality it is immediately approved. The message should reflect this.*
- *In my abstract I put line breaks; these are not rendered faithfully on the webpage.*
- *I can't see any way in the web form to change the licence. I need to change the licence, since I set it as "no CC" but that was just as a precaution.*
- *In the top breadcrumbs, I could see "C4DM-RDR Home→Items→Curator". If I click on "Items" it takes me to the URL <http://c4dm.eecs.qmul.ac.uk/smdmrd-test/xmlui/admin/item> and says "Insufficient privileges".*

All the problems and feature requests were added to the bug tracking system in the project's repository on code.soundsoftware.ac.uk, and will be address in order of importance. Individual answers to the emails were also sent out.