



Demetrius RQF Preparedness

Peter Raftos

This publication is made available by the Australian Partnership for Sustainable Repositories as one of its “Helpful Documents” series issued in association with its work on RQF and Repositories.

www.apsr.edu.au/rqf/http://www.apsr.edu.au/rqf/demetrius.pdf

July 17, 2007

Preliminary activities

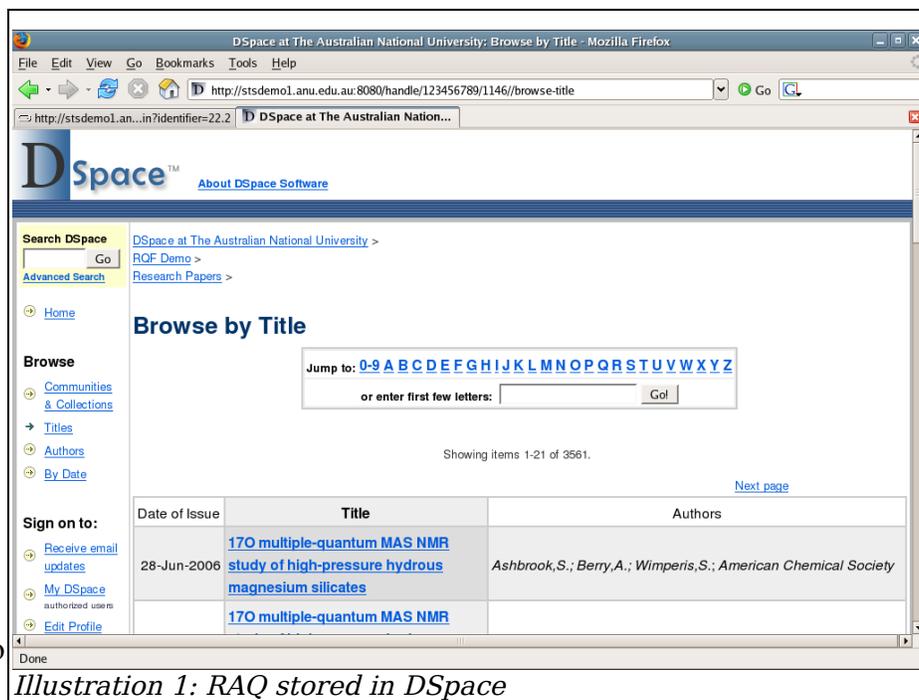
Review of ANU Quality (RAQ)

In 2004, Digital Resource Services (DRS) was involved in the RAQ, an external review of ANU's performance designed to inform the University and the Australian public about the standing of the ANU in the international academic community and about the best ways of developing the ANU for Australia's longer term national interest. Research academics from every discipline represented at the ANU were required to make available to assessment panels a selection of their research output.

There were more than 7,500 items (including 3,500 papers), submitted by 1,500 academics.

As this was a one-off exercise and public access was not a consideration, access to the documents was limited to the assessors. Document metadata were entered by the depositors, by hand.

The documents were stored, mostly as PDF files, in an instance of the DSpace software managed by DRS. The software's internal access control mechanisms were adequate to ensure that depositing academics had access to the Collections within DSpace they needed, and assessors were also able to log in and access the appropriate Collections. In both cases, the lists had to be drawn up manually.



The screenshot shows a web browser window displaying the DSpace interface. The page title is "DSpace at The Australian National University: Browse by Title - Mozilla Firefox". The address bar shows the URL "http://stsdemo1.anu.edu.au:8080/handle/123456789/1146/browse-title". The page content includes a search bar, navigation links, and a list of items. The list has columns for Date of Issue, Title, and Authors. The first item is dated 28-Jun-2006 and has a title starting with "17O multiple-quantum MAS NMR study of high-pressure hydrous magnesium silicates". The authors listed are Ashbrook, S.; Berry, A.; Wimperis, S.; American Chemical Society.

Date of Issue	Title	Authors
28-Jun-2006	17O multiple-quantum MAS NMR study of high-pressure hydrous magnesium silicates	Ashbrook, S.; Berry, A.; Wimperis, S.; American Chemical Society
	17O multiple-quantum MAS NMR	

Illustration 1: RAQ stored in DSpace

Report generation

In 2005, DRS ran an experiment to determine if metadata stored with the RAQ collections could be extracted and used to generate reports.

For this exercise, the Cocoon XML publishing framework was used to query DSpace, extracting the needed metadata as XML. In this form, it could be converted to any of the following:

- Web tables
- Excel spreadsheets
- Comma- or tab-delimited tables
- Graphs (SVG, JPEG or PNG image formats)

It could also be exported as raw XML, such as via a web service.

This experiment was deemed a success. However, it needs to be emphasised that metadata can be extracted and summarised for reports only if it is already stored with the Items in DSpace in the first place. For instance, since the original RAQ did not require academics to record RFCID codes when they deposited papers in DSpace, this information was not available in the experimental reports.

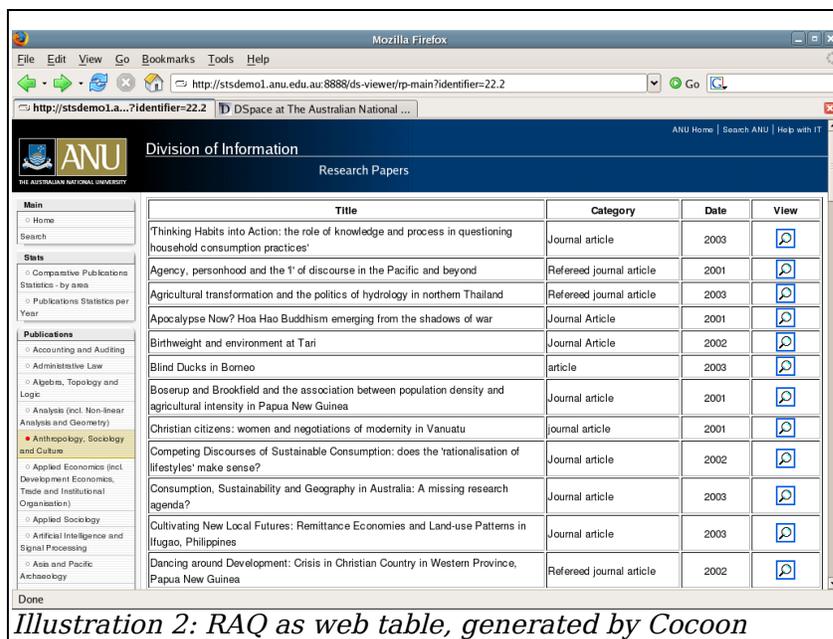


Illustration 2: RAQ as web table, generated by Cocoon

It is possible to add metadata to Items stored in DSpace after the event, but this may be difficult and time-consuming, and where metadata are not directly describing the content it is worth considering whether these should reside in the content repository.

Research Office

RQF reporting compliance will be the responsibility of the ANU's Research Office (RO). However, it is anticipated that the ANU's digital repository staff (DRS) will be called upon to provide advice and technical assistance to the RO.

In the first instance, I expect that DRS will be called on to:

1. Deploy DSpace as the storage mechanism (or the primary one of a number of mechanisms)
2. Provide access controls to the holdings
3. Provide reports and summaries

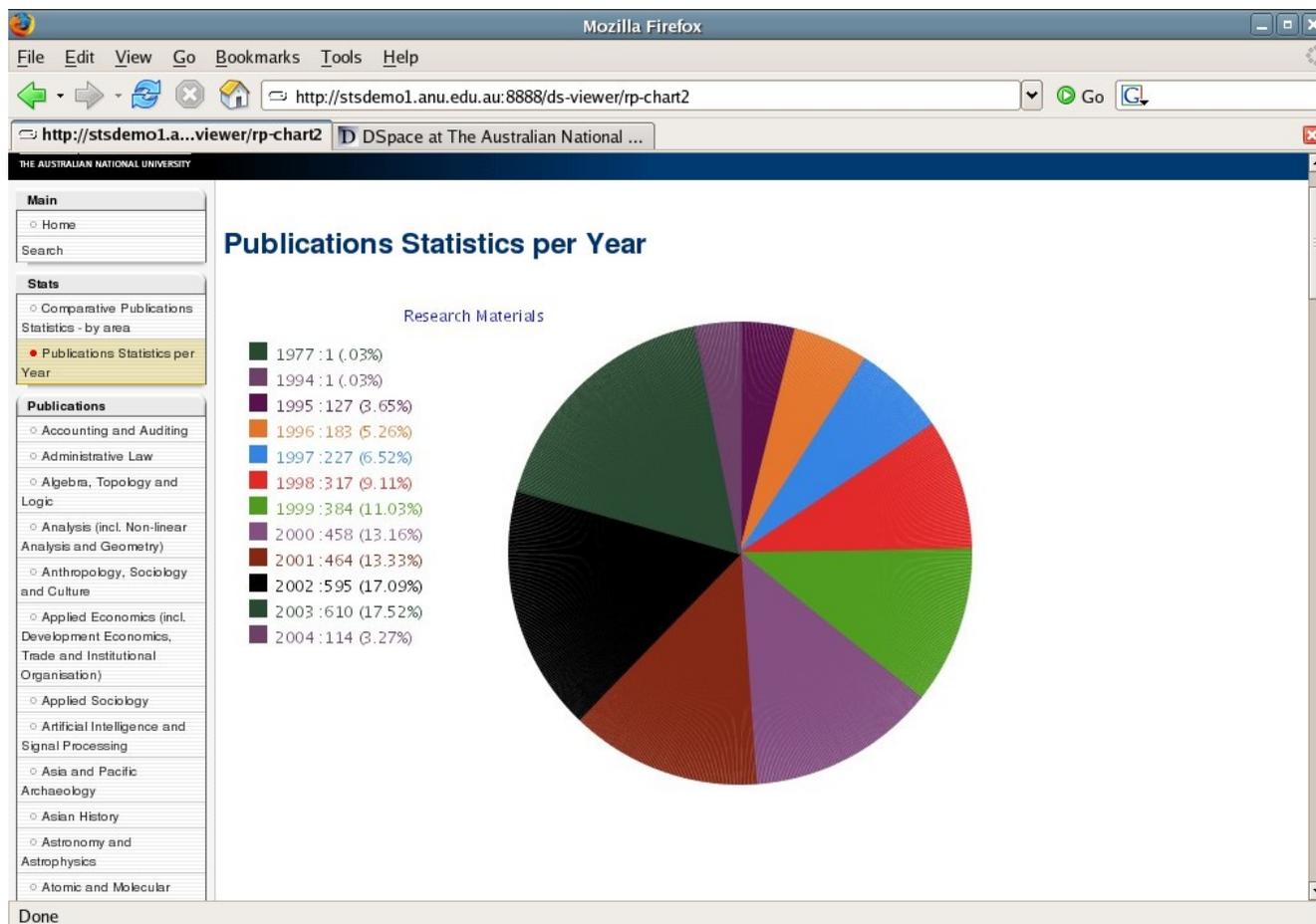


Illustration 3: RAQ Output summary generated by Cocoon

Content to be stored

There are a number of unknowns around the RQF. It's expected that the majority of papers deposited in DSpace will be PDF (Portable Document Format) files. This will simplify storage, retrieval and rendering. However, a number of disciplines may require storage of other formats. If the information/paper/report or other item can be adequately stored in a single file, then storage will not be an issue; however, external players may be needed for film and sound.

Formats might include:

- Tex/Latex
- Word
- Still images: TIFF or other
- Sound: WAV, BWF or other
- Moving image: variety of formats here
- Flash animation

More involved are true complex objects (such as web pages or multimedia installations) which consist

of a number of separate files, and the relationships between them. DSpace does not currently handle these objects well: further thought and experimentation may be needed in order to avoid dealing with each object individually.

It's also important to consider the volume of material expected to be deposited in DSpace. This may include:

1. All materials produced within the ANU (received under mandatory deposit). Only those relevant to RQF are flagged
2. All those materials relevant to RQF
3. Most materials relevant to RQF

In all cases, the required metadata will also need to be stored, probably at the time of deposition.

The most likely scenario, in my view, is the final option.

What, then, will be the life expectancy of the material deposited in DSpace for the purposes of RQF? Once deposited, papers will be read and analysed by a review body. At the end of this phase, the material will presumably need to be kept for some period in case of legal challenges or some other need for reviewing content. At the end of the second phase, it is not at this point clear whether the material need be kept (at least, not in its RQF-ready form). Also, materials deposited for the RQF will, presumably, not be eligible for inclusion in later iterations, so should be cleared out.

Although the individual papers may have long-term archival value (indeed, are assumed to have such value), I propose that the mechanisms for RQF compliance be kept separate from the ANU's digital repository. Thus, there would be an RQF instance of DSpace, which is not part of the production or development DSpace instances used in Demetrius (the ANU's Institutional Repository). Papers may be lodged in both Demetrius and RQF DSpace; since access to RQF DSpace is limited and storage there is for a set time, this will not be confusing.

This would provide the following benefits:

- RQF and Repository administrative requirements would not come into conflict
- RQF DSpace can be refreshed (with old instances archived) for each iteration of the RQF
- The DSpace Community and Collection architecture can be used to facilitate RQF, rather than forcing RQF materials to work within the arrangements already in place for Demetrius
- User management and access control would be simplified, thus minimising the likelihood of security breaches
- Complex copyright issues would be obviated: all materials lodged in RQF DSpace are unavailable to the public; for public access they can be stored in Demetrius or, if appropriate, elsewhere

Since DRS already has strong expertise in using DSpace (and using DSpace for similar exercises), DRS would manage RQF DSpace in exactly the same way as it manages the DSpace instances in Demetrius.

Submission control

Currently, the ANU does not have an identity management system that would enable the remote creation of submitter groups for DSpace; however, the ANU LDAP directory service can be used in

this way, and DSpace could use information stored there. For the first RQF exercise, it may still be necessary to create the groups within DSpace, and so lists of academics, with their rights and responsibilities clearly defined, will need to be provided to DRS staff in a timely manner.

Assuming that a separate DSpace instance is used for the RQF exercise, internal Communities and Collections will reflect the subject areas defined for RQF.

DSpace comes with its own deposit workflow tools. These can be customised to suit the RQF.

Access control

Both reviewers and administrators will need access to holdings.

Reviewers will need to be able to find and then download holdings. This can be done within DSpace, but access lists with email addresses will need to be provided in a timely manner. In effect, reviewers will need to log in to DSpace.

RO administrators will need to access metadata, possibly for importation within the Aries system. The simplest method would be a harvest of metadata, which can be automated.

Other Considerations

A further option worth consideration is to develop an application satisfying RQF requirements based around DSpace that could be deployed to other institutions (call it RQFSpace). This application could be managed from ANU as we have a DSpace committer who could continue to ensure the DSpace software versions were kept up to date, as well as make the necessary modifications to support RQF requirements. This solution may be useful in assisting institutions to store their documents in a proper repository.

Conclusions

In short, DSpace is a viable storage mechanism for RQF materials. Effective submission and access controls exist within the system, it is highly configurable and information can be exported for Research Office use.

Should ANU elect to store RQF documents in Demetrius, the RQF function itself should not be tightly tied into the repository software. Repository software is not RQF software and embedding RQF functions within the repository makes it difficult (read costly) to swap-out repository software or research(er) management software in the future. In our opinion the materials and metadata for the RQF should be stored in and extracted from the systems they are suited to.