

Associating Research Data Records using RAIDmap

Alex Ball¹ Uday Thangarajah² Mansur Darlington²
Chris McMahon² Liz Lyon¹

¹UKOLN, University of Bath ²IdMRC, University of Bath

I Introduction

Uday and I will be demonstrating an application called RAIDmap, a tool for recording the associations between research data records, but first I want to explain why we think such a tool is necessary. ¶

The precursor project to REDm-MED was ERIM, and in that project we looked at the data produced by engineering researchers across a range of projects, looking for common features. What we actually found was incredible diversity: it seemed every project was working with a different set of formats and using a different workflow. It's not like in Crystallography, say, where it's pretty obvious how the PRP, CIF and MOL files relate to each other. It's unlikely that anyone coming to directory of data like this (Figure 1) would know what it all means and how it fits together.

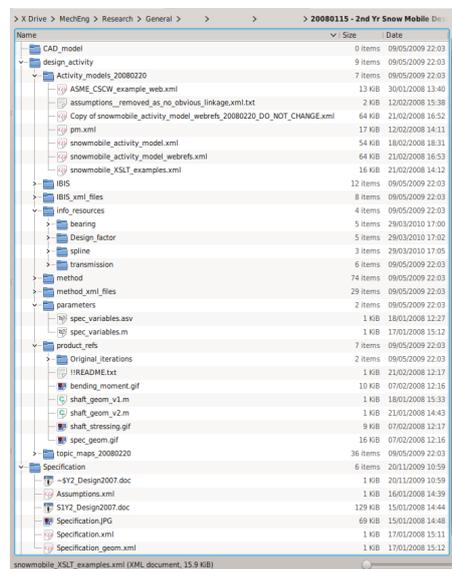


Figure 1: Files and directories relating to the 2nd Year Snowmobile Design Task data case from the KIM Project

When I say anyone, I'm including the researchers themselves in that. When we spoke to them, several had difficulty remembering what was in them and how they interrelated, and in some cases they didn't know where all the files were. How much worse would it be for a curator coming to it fresh and having to decide what needs to be kept, or what might need to be produced in case of an FoI request? ¶

So, in ERIM and REDm-MED we decided the way to solve this would be to create a Project Record Manifest (Figure 2).

Project Data Record Manifest Template for IdMRC Projects

The Project Data Record Manifest (PDRM) constitutes the principal record through which the records relating to a research project may be identified and accessed. It must be located in a publicly accessible and searchable place. The default location is an anonymous top-page of the research project site.

The PDRM should be readily accessible, editing rights being limited to members of the originating research project team and by other nominated individuals such as the data manager. A vetoing system must be in force.

While the PDRM will be generally available, there will be some records associated with the research project which are confidential or sensitive. Access to records of this nature must be limited by appropriate password-protected facilities. This could be done by either: (a) access to which the research project who or other user requires. If in doubt, the advice of the data manager to identify that the project PI should be sought.

Summary of Research Activity

Project name
e.g. Long And Technical Tensile Evaluation (LATTE)
Period of Project
e.g. October 2009 – March 2011
Lead and partner organisations
e.g. University of Bath (lead), University of Cambridge, University of Leeds
Principal Investigator (name and contact details)
Name:
Contact details:
Data access summary
Data access refers to the physical means by which access to records is controlled. The overarching data access provisions for the research project are recorded in the DMP associated with the PDRM for details of confidentiality cases of individual records see the Project Data Record List below. As a guide, data access should be either consistent with or more restrictive than the confidentiality status.
Retrieving repository
e.g. The data from this Research Activity will be deposited according to the IdMRC DMP (see below) or The data from this research activity will be deposited in
Related documentation
<ul style="list-style-type: none"> • RCUK Policy and Code of Conduct on the Governance of Good Research Conduct • The University Research Data Protection Guide for Research • Engineering Research Data Management Plan Specification • IdMRC Project Data Management Plan

Project Management Documentation

Note that some of these records may need to be placed in a password-protected storage area.

- Project Data Record Manifest (public link)
- Project Proposal (public link)
- Project Plan (public link)
- Confidentiality agreement with (partner) (public link; note if this agreement is itself confidential it should be placed in an appropriate password-protected area)
- Participants consent forms (public link; physical location/contract number/contact details)
- IPR statement (public link; physical location/contract number/contact details)
- UK Data Archive deposit agreement (public link)

Project Data Management Documentation

- Project Data Management Plan (public link) (this will be a reciprocal association, since the PDRM will identify the Project Data Record Manifest)
- DMP associated (public link) or
- Other data record associated documents (public link)

Project Data Record List

Every project data record should be listed in the table below in the form: Title, file name, record type, location, owner and contact details, confidentiality status.

Record Type (for both electronic and physical records)

Every data record will be one of the following: research data record, control data record, associative data record, research object data record, experimental apparatus, data record.

Location

If all the files are archived in a single, central location, the location need be identified for the set of records (the Data Case), only. For electronic records it is expected that a hyperlink or request to the location is recorded. For physical records the location should be described.

Owner

The 'owner' is the person currently responsible for the management of the record, and who is in a position to consider matters such as availability and security. Ownership does not imply the right to use or disposal. During the period that the research project is under way it is likely that the owner will be a research officer or an individual in a supervisory role. At project end the ownership should be transferred to an appropriate individual, such as the project PI or the data manager responsible. In many cases it will be appropriate for a research officer to retain ownership.

Confidentiality Status

Confidentiality status indicates whether classes of people and other information (information-gathering systems may have a role) of the data record it does not provide information about how such records are protected. It is likely that the confidentiality status will change during the life-cycle of the data record in which case the public number assigned. Access to whether the record is confidential is made, that the work public domain should be used. If the access is limited, then the entities who are permitted to use this data should be identified either by naming groups or individuals.

Record Title	File Name	Owner	Contact Details	Data Record Type	Confidentiality Status
Example:					
IdMRC Research Project Data Record Manifest	idmrcManifest102170.pdf	IdMRC/ Cambridge	idmrc@bath.ac.uk	associative data record	public domain

History of this PDRM

Figure 2: Project Data Record Manifest Template for IdMRC Projects

The main component of this (let me zoom in ☹) is a table listing all the records associated with a project, showing the record title, file name and location, owner and contact details, record type, and confidentiality status. With this at least we have a chance of working out what's what. But filling out that table is laborious, and very easy to forget to do until the task becomes monumental. Plus, there's a lot a mere table can't do, such as indicate which files derived from which other files. So we came up with the idea of a RAID diagram.

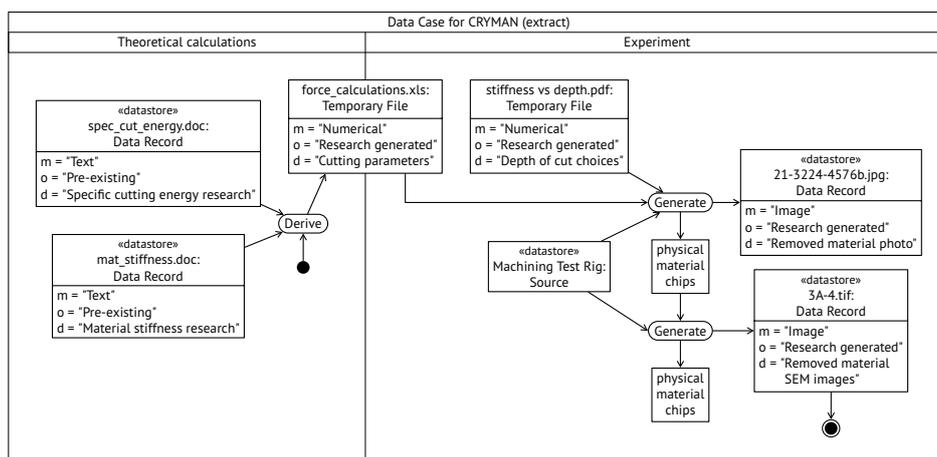


Figure 3: Example RAID diagram

This (Figure 3) is an extract from a RAID diagram we did for an investigation into machining cryogenically frozen materials. You can see fairly clearly where the machining parameters came from and which runs these two images came from. It's better, but still a bit cumbersome to do by hand, so Uday has written a RAIDmap application to take the pain out of it. Here's how it might be used for this data case.

2 Demonstration

- First, Uday is setting up a data directory for our project files and configuring the RAIDmap daemon to monitor it.
- Now, we're going to do things slightly out of order. We're grabbing a photo of the machined material and adding it to our watched directory. Immediately we get a metadata form popping up. It has already been partially completed by the metadata extractor, and all we have to do is fill in the blanks and save it.
- Now, let's go to the RAIDmap application proper. We're starting a new project, and adding a new data case.
- If we now add the photo to our data case, you can see that the application has found the metadata we wrote for it. We can get a summary of it if we mouse-over the asterisk at the top right corner.
- It's not just files we can add to the data case, we can also add physical objects. So let's add the machining rig. Because it's physical, there's no network location to fill in, and there are fewer metadata fields to worry about but they all start off blank.
- So now we have two items in the data case and we need to show how they are related. We just click and drag to add the relationship, and we can select the type of relationship from the context menu. Here, we used the rig to generate an artefact that we photographed, so we choose 'Generate'.
- We don't have time to construct the whole thing for you now, so here's one we made earlier. All these items have metadata we could look at. If anything's not clear we can zoom in and out, drag icons around, or get the computer to lay the icons out for us.
- We can export this map as an XML file, as a JPEG picture, or as HTML.
- The HTML version looks like this.

3 Questions

That's all we wanted to show you. We would be grateful for feedback on this, particularly if you can think of anything we could do or include to make it more useful. But we're also happy to take any questions you might have.



REDm-MED is funded by JISC.

For more information, please visit the project Website at <http://www.ukoln.ac.uk/projects/redm-med/> or the blog at <http://blogs.bath.ac.uk/redm-med/>