REVIEW OF DCC TOOLS AND GUIDANCE

ALEX BALL AND MANSUR DARLINGTON

redm6rep120202ab10.pdf

ISSUE DATE: 31 May 2012

Catalogue Entry

Title	Review of DCC Tools and Guidance
Creator	Alex Ball
Contributor	Mansur Darlington
Subject	research data management
Description	In the course of its work, the REDm-MED Project has used various tools and guidance produced by the DCC, most notably CARDIO and DMP Online, the latter in both its checklist and software forms. The Project team found CARDIO to be promising but in need of further development before being used widely. The process of setting up a DMP Online template was relatively straightforward, but unfortunately there was no opportunity to solicit feedback from researchers on using it in the context of the tool.
Publisher	University of Bath
Date	2 February 2012 (creation)
Version	1.0
Type	Text
Format	Portable Document Format version 1.4
Resource Identifier	redm6rep120202ab10.pdf
Language	English
Rights	© 2012 University of Bath

Citation Guidelines

Alex Ball and Mansur Darlington. (2012). *Review of DCC Tools and Guidance* (version 1.0). REDm-MED Project Document redm6rep120202ab10. Bath, UK: University of Bath.

1. INTRODUCTION

As part of its work in preparing a data management planning regime for the Department of Mechanical Engineering at the University of Bath, the REDm-MED Project has had occasion to use tools and guidance produced by the UK Digital Curation Centre.[1] As part of its requirements elicitation process, it performed a 'health check' regarding the department's readiness for research data management using the DCC's CARDIO tool.[2] It also used the checklist of data management issues underlying the DCC's DMP Online tool to ensure a broad range of challenges were considered in assembling the requirements specification.[3] Furthermore, the REDm-MED Project chose to express its data management plan template for new research projects within the department as a DMP Online template.

This document records the experiences of the REDm-MED team in using these tools, in the hope that this will aid the developers in strengthening and improving them for the benefit of the wider HE community.

2. CARDIO

These comments refer to CARDIO as we experienced it between November 2011 and January 2012, and do not necessarily reflect the current state of the tool.

2.1 Initial impressions

The site as a whole has a friendly design and, apart from some issues noted below, we found it intuitive and simple to use. One initial problem we had was that one of the team only had access to version 2 (specifically 2.0.0.2.0) of the Firefox browser, which is not supported by CARDIO; fortunately the version of Internet Explorer to which the team member had access was supported. We also found that the site was rather slow and unresponsive at times.

We (as RSOs) had no issues with registering, logging in and beginning a new process.

2.2 Stage 1

2.2.1 Contextual information

Other than encouraging participants to perform a Data Asset Framework audit, the benefit to the user of filling out many of the fields here remains unclear. Some fields seem inappropriate for certain levels of granularity; for example, when spanning an entire institution, the answers to 'Storage required', 'From what date', and 'Data to be managed for [duration]' will vary enormously. Indeed, even within a single project one could imagine several triples of answers to these three questions.

---

[1] Digital Curation Centre website, URL: http://www.dcc.ac.uk/

[2] CARDIO website, URL: http://cardio.dcc.ac.uk/

[3] DMP Online website, URL: http://dmponline.dcc.ac.uk/

2.2.2 *Three Legged Stool*

Some of the guidance notes for the statements are directed personally – e.g. 'Do you understand the need to document data?' – which implies a collection methodology involving a statistically significant sample of the members of the institution, department, research group, etc. being assessed. Experience of later parts of CARDIO, though, suggests that it is not suited to data collection on that scale. We felt the more typical scenario would be that users of the system would be answering on behalf of themselves and their colleagues; we in the 'Research Support Officer' (RSO) role were certainly attempting to answer for the whole department. This being so, we felt the questions should be directed more generally, e.g. 'Do you and your colleagues understand the need to document data?' (c.f. 'Is there an awareness of legislation that affects research data management?' under Legal Compliance).

In other cases, we felt the guidance notes were asking slightly the wrong questions. Under 'Data Ownership and Management', for example, CARDIO asks, 'Who owns data and associated documentation?' and, 'Who has responsibility for data management?' Users should not be answering these questions in their reasoning, though, but rather the question of whether they could, if asked, *discover* who owns the data and who has management responsibility. By way of contrast, the questions under Data Policies and Procedures (does the organisation have written policies, are they implemented) are rather more directly helpful in assigning a rating.

We also had some comments about specific statements.
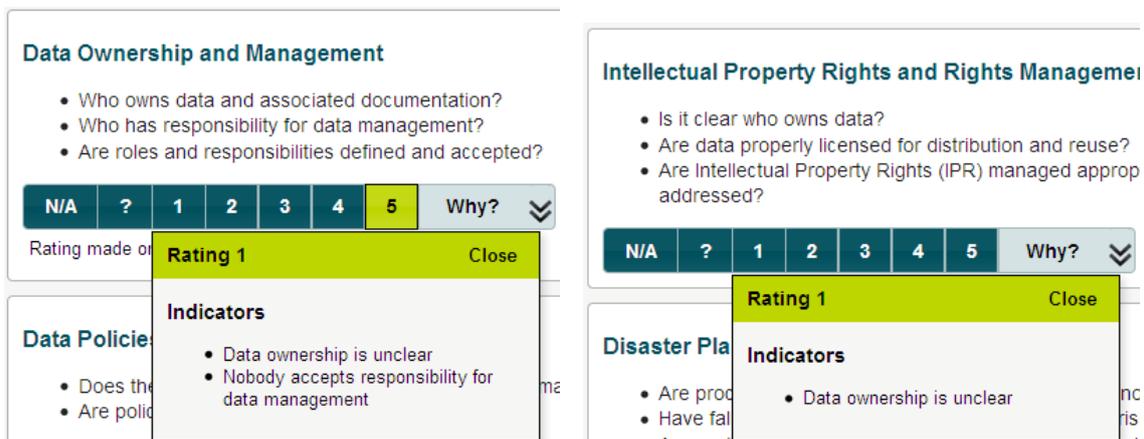
Figure 1: Repetition of data ownership

- There seems to be a major overlap between Data Ownership and Intellectual Property Rights and Rights Management; in particular, we do not think users should be asked twice about identifying the owner of data (see Figure 1). We suggest that the IPR statement should ask whether there is clarity over what actions can legally be performed on data with respect to IPR, how permissions might be obtained, and so on.

- The Business Planning statement is hard to interpret at the departmental or project level.

- The typographical inverted comma in the description for Staff Development Opportunities suffers an encoding issue in some contexts, displaying as 'â€™'.

After a rating has been assigned, to the bottom of the rationale box is added the characteristics of that rating, each with an adjacent tick-box. The purpose of these tick-boxes was not apparent to us from the interface. It appears from the manual that they are to allow users to express which characteristics led them to choose that particular rating (previous iterations of tool provided more characteristics and suggested possible actions). We suggest that the free-text rationale box is sufficient for allowing users to clarify their decision, but that if these tick-boxes are to be retained, the logic should be reversed. Users choose the rating for which the set of characteristics most closely matches their perceptions; users should therefore be invited to *deselect* characteristics they feel do not fit the situation, *despite* them having chosen that rating.

2.2.3 Completing stage 1

There was a save button for the whole page; it was unclear to us why this was needed in addition to the save buttons for each statement. (This button has now been removed.)

2.3 Stage 2

There is a mismatch between the actual workflow the user needs to adopt and the way CARDIO moves the user through the process. RSOs proceeding from Stage 1 to Stage 2 are presented with a summary of their decisions and a button that says 'Activate Stage 2 for all users and send notification emails'; at first glance it appears that pressing the button will provide a screen allowing the RSO to do those things, while in fact the button does exactly what it says. The RSO must instead know to navigate to a page that appears after Stage 5 to invite participants. It would be rather more intuitive if, at the end of Stage 1, the button said 'Invite participants', and the next screen presented was one where the RSO could invite/manage participants. (Indeed, the 'invite participants' page could safely be a sub-page or pop-up page reached via the 'manage participants' page.) The RSO should then proceed from that page to the Stage 2 page.

It was not immediately apparent from the 'Invite participants' screen what the email notifications would say. As the 'message' is optional, presumably there is some standard text sent out, but the RSO is unaware of what it is and therefore cannot judge what would be appropriate to include in the message (e.g. whether to introduce the CARDIO tool). The system does not appear to record what was sent out to the participants, nor when the invitations were sent, and there is no facility for including an attachment (such as a help sheet).

It was found not to be clear whether it was possible to stagger the invitation of participants or whether the panel had to be invited in its entirety at a single shot. We took the first approach: in the first instance five panellists were invited and invitation emails sent out. A further two participants were then successfully invited and registered. However, we could find no way of sending members of the second or later tranches of participants instructional emails through the CARDIO system independently of those already invited and processed.

Much confusion occurred and time was lost as a result of participants delaying their registration. It was not clear that (*a*) there was a registration time limit, (*b*) what that time limit was, and (*c*) how to rectify matters once the registration period had expired. Attempts to reset accounts at the system end were not always successful at the first

attempt. 'Uninviting' a participant did not clear the earlier email details from the CARDIO record.

The system will accept as input the common email form: 'Joe Bloggs <jbloggs@lincoln.ac.uk>' but then fails to process the input without flagging an error. As a result the RSO thinks the invitation has been extended but the recipient does not receive it.

We would have liked the ability to have more than one RSO involved in the process.
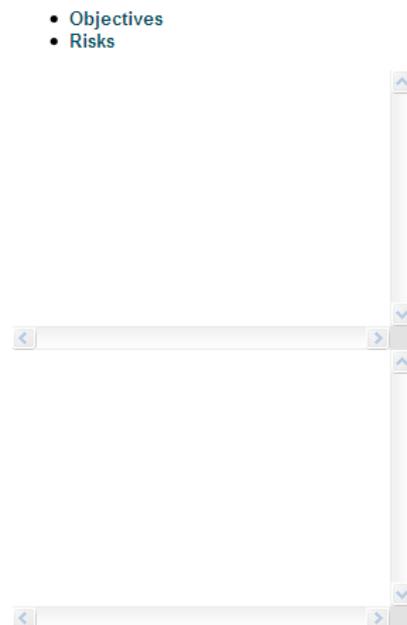
2.4 Stage 3

This stage was more self-explanatory, but there were a few niggles in the interface.

Figure 2: Pop-up box obscuring information icons

- When trying to decide the overall rating, the RSO scans across the row and looks at the ratings and rationale given by the other participants. The box that pops up to reveal this rationale, though, comes up to the side of the information icon, thus obscuring the neighbouring ratings and icons. The RSO must therefore click on the 'Close' link each time to remove the popup and reveal the other ratings (and gain access to their rationale). It would be far more convenient were the other ratings and information icons left exposed.

- In the discussion pages, on the left hand side, there are links to objectives and risks (see Figure 3). These links do nothing and go nowhere. Beneath them are two areas with scrollbars which serve no apparent purpose.

- With a large group of participants, it might be useful to have the standard deviation of ratings included in the statistics; it is a more appropriate measure of the agreement of participants than the minimum and maximum.

- In the statistical tables, where the modal average is 'Unknown', this displays as 'k'.

- When we found no cause to change the automatically calculated overall rating, we did not re-select that score, so the system never registered that as our overall rating. It would be good if the system would register the


Figure 3: Mysterious text areas

automatic rating as the RSO's overall rating if no change has been made by the time the RSO finalizes Stage 3 and moves on to Stage 4.

2.5 Stage 4

We had some immediate issues with the radar graph showing a comparison of ratings. One of the colours chosen to code ratings was white. When the participant in question was selected, we did not have the problem one might expect, that of the data line being invisible. Instead, all the lines on the graph turned orange, and the markers that show where the lines intersect with the axes disappeared.
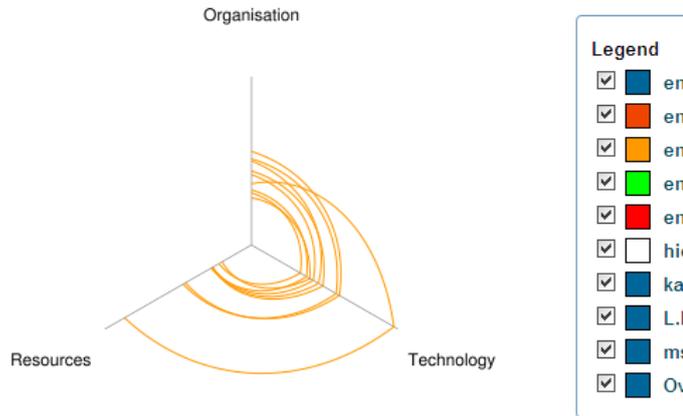
When that participant was deselected, the graph displayed correctly. There remained two other issues, however: CARDIO only provided six colours where we needed ten, and the arrows that marked where the red line intersected with the axes gave that set of ratings prominence over the others.
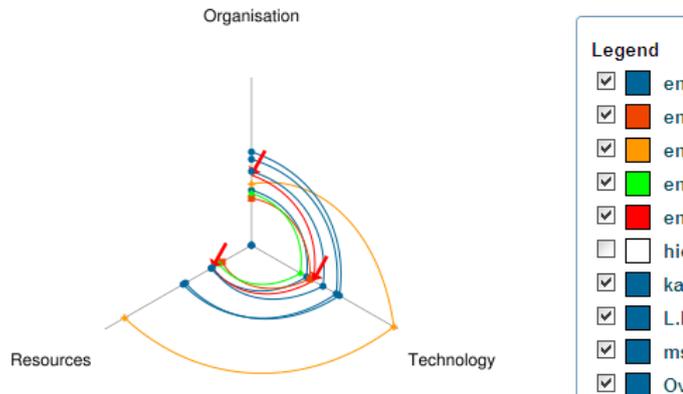
The issue with the colour coding turned out to be worse with the graphs for the individual legs, as all lines were shown in orange regardless of which participants were selected. Also, as ratings of 'N/A' and 'Unknown' are treated as zero, a radar plot may not be the most suitable way of representing the figures: once several lines are plotted it is less apparent to which axis these zero scores belong.
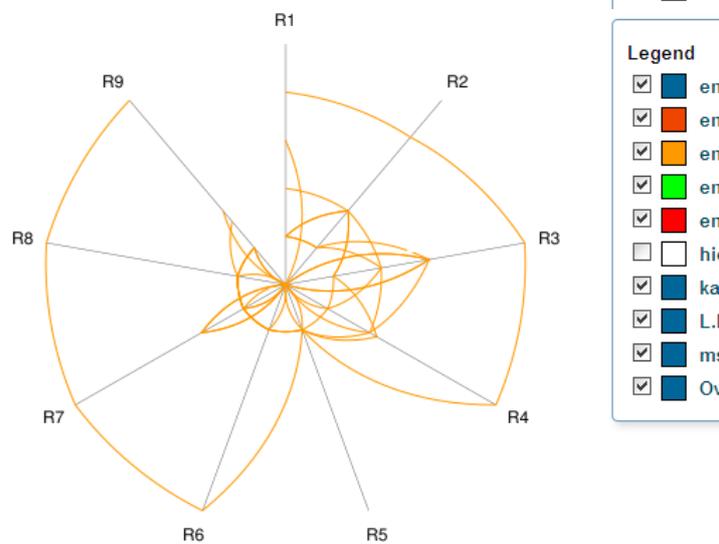
The colour coding problem with the graphs occurred again in the generated PDF report, only worse as the uniform orange colour had spread to the legend as well. One has the option of including objectives and risks in the report, but this just inserted 'No known objectives' and 'No known risks' under every statement. The PDF provided no explanation of the statistics – that the ratings run from 1 to 5; that 'overall' and

'assigned' refers to all answers, and answers from those who were assigned to look specifically at that statement, respectively – making it less useful out of context.

For our purposes, the rationale given for each rating was the most important information gathered by the exercise, so we would like to have had the option of including it in the PDF report, or in some other digest. In the cases where we agreed with the automatic overall rating, and therefore did not override it, this rating was not included in the report (our suggestion above, that in such cases the automatic rating should be registered as the RSO's overall rating, would avert this issue).

2.6 Stage 5

We did not make full use of this stage. As we were using CARDIO as part of a requirements gathering exercise, we selected participants whom we thought representative of the needs of the department's academics, researchers and students, rather than those with an influence on the research data management infrastructure of the department or institution. It did not, therefore, seem appropriate to allocate tasks and responsibilities to these participants as an outcome of the process.

3. DMP ONLINE

The DCC's DMP Online tool uses a checklist of just under 120 questions as the basis for creating Data Management Plan templates, each of which is designed to satisfy the requirements of a particular stakeholder. Researchers can then fill in one (or more) of these templates through a simple Web interface.

The process of designing a template for use with the tool is simple. It involves filling out a meta-template in the following form.

Template Title

Explanatory Note

Custom Section Title	
Checklist question number	Custom guidance (if required)
Checklist question number	Custom guidance (if required)
Custom Section Title	
Checklist question number	Custom guidance (if required)

Staff at the DCC enter this into the DMP Online system, and the template designer is given the chance to review the results before the template is made live.

The template that the REDm-MED Project produced for project-level data management plans can be found within *A Research Data Management Plan for the Department of Mechanical Engineering, University of Bath*, under the section 'DMP Contents'. (A

generalized version is also available in *A Research Data Management Plan for Engineering Research*.) The questions and guidance from this template were translated into the terms of the checklist underlying DMP Online, entered into a spreadsheet and sent to the DCC for loading. This version of the template is also presented in more readable form as the *University of Bath Department of Mechanical Engineering DMPOnline Template*.

Filling out the meta-template presented no problems from a process perspective. Also, for the most part it was easy to find questions from the DCC checklist that corresponded to the questions in the REDm-MED template. There were just a few places where we found difficulty.

- There were conflicts of emphasis with a couple of mappings: DCC Question 4.1.2 asks why data many not be shared, while the REDm-MED template asks for a summary of access restrictions; DCC Question 6.1 asks for a full long-term curation strategy, while the REDm-MED template simply asks where the data will be stored (and curated, if applicable) in the long term.

- The REDm-MED sections on data generation and manipulation on the one hand, and data organization on the other, had to be combined as, for example, DCC Question 2.3.2 conflates issues of acquiring data and organizing them into named files.

- There was no DCC question that asked about data packaging; the nearest we could find concerned data transfer.

Version 3 of DMP Online gave the opportunity to provide multiple templates relating to different research lifecycle stages, but we did not take advantage of this. Primarily this was because there are no differences in questions asked at the proposal stage and at the project working stage by the REDm-MED template; the differences lie in the detail researchers are expected to give, and we felt this was adequately addressed in the guidance. Nevertheless, we felt the facility was valuable and we may well take advantage of it in future revisions of the template, depending on the feedback we receive from the researchers who use it.

4. CONCLUSIONS

We found CARDIO to be for the most part easy to use, although some aspects of navigation and data entry could be improved, and the UI was cluttered by defunct features that had not been cleanly removed. We felt that more thought needed to be put into questions asked. We also uncovered some bugs in the reporting template and experienced some issues administering the process. In summary, we felt the tool showed promise but was not ready for widespread roll-out.

We did not get the opportunity to test DMP Online with researchers, but found the checklist on which it is based sufficiently mature, and the process of setting up a template pleasingly simple.