Trainer's notes for module 1:
Research data and lifecycles

Good practice in research data management

# Session Details

## Aims and Objectives / Learning Outcomes

By the end of this module participants will have:

* Considered the holistic nature of research data
* An understanding of three lifecycle models

## Session Topics

* Considering research data
* Introducing lifecycle models

## Structure

This module is planned to be delivered via one session lasting 40-60 minutes with a group of 12-20 attendees. The expectation is that these are postgraduate students and/or early career academics.

### Indicative timings

|  |  |
| --- | --- |
| Considering research dataActivity 1: Your research data  | 10 minutes15 minutes |
| Data lifecycle modelsActivity 2: Examining the DCC Curation lifecycle model | 10 minutes20 minutes (optional) |
| Session review  | 5 minutes |
|  | 40-60 minutes total |

## Set-up

Slides are provided as detailed in the "notes to accompany slidedeck". Note the use of hidden slides by default for optional activities and slides offering further detail on a topic. . (Hiding/unhiding slides is best done in 'slide sorter' view; select slides, right-click and toggle 'hide slide')

Activities are indicative rather than prescriptive. The assumption is that you are used to tailoring training outlines to meet your own needs, space and available resources! (Post-it notes, pens, flipcharts etc.)

# Notes to accompany Slidedeck for module 1

## Considering research data (slides 3-6)

### Slide 4: What is research data?

First off, let's consider the nature of research data; challenging the preconception that research data is on**ly** quantitative data organised in a spredsheet or database. In the context of Research Data Management it's important to re-think what counts as data. It includes qualitative material – whether photograph, sculpture, letter, speech or music – and any other material required to re-validate the results of the research being undertaken.

You may introduce the RDM@Ncl site and refer to:
<http://research.ncl.ac.uk/rdm/rdmncl/whatdoyoumeanbydata/>

### Slide 5: Research data: All of these, and more…

To re-confirm the wide interpretation of what is research data, here is a graphical representation of a wide range of things that could be considered research data. Highlight a few as appropriate, but note that this is not an exhaustive list!

A good resource you may wish to demo, providing an overview of research data in the visual arts is:
<http://xerte.ucreative.ac.uk/xertetoolkits/play.php?template_id=35>

Activity 1: Your research data (15 minutes)

Run a high-paced, short activity to energise the early stages of the training session.

Using post-in notes, ask attendees to spend 5 minutes writing down all the different types of research data they work with. (Offer a prize – real or hypothetical! – for the greatest number?!)

Review responses and run a feedback session for 5 minutes.

For the last 5 minutes, mark up an area (flip chart, table top etc.) for qualitative data and another for quantitative data, and ask attendees to add their post-its to the relevant area. Discuss the range of data types, is it diverse or restricted? Digital or analogue? Traditional or contemporary?

This exercise should lead in to the next few slides and the consideration of different models of the research/data lifecycle.

## Introducing lifecycle models (slides 7-14)

### Slide 8: The life of data

Researchers are often understandably focused on the specific project they're working on. However it's important to think about the full life of a particular dataset; that's lifespan should, in most cases, outlive the research project that created it.

Indeed there is a shift happening here; funders are mandating that research data is made available for validation and re-use, and researchers are realising that there are benefits to be had. For example, in the field of linguistics, there is significant growth in the re-use of existing corpora – large collections of language research data – rather than every project undertaking new data collection.

### Slides 9: UK Data Archive data lifecycle

This is the first of three slides showing different lifecycle models.

The UK Data Archive offers a data-focused lifecycle, stating that "Well organised, well documented, preserved and shared data are invaluable to advance scientific inquiry and to increase opportunities for learning and innovation."

The model on the slide is active, linked to: <http://www.data-archive.ac.uk/create-manage/life-cycle> where a clickable interactive version of this lifecycle is available offering detail about the individual stages. You may wish to go through the various stages on the UK Data Archive site.

Note the inclusion of "re-using data" and its position after "Giving access to data". i.e. Once you've given others access to your data, you can still continue to use it yourself! (This challenges the culture of "sharing only once you've finished with your own research".)

### Slides 10: DCC data activity lifecycle

This is an activity-focused lifecycle used by the Digital Curation Centre.

Note the similarity with the UK Data Archive lifecycle, but the addition of the need to 'document'.

### Slides 11: DCC curation lifecycle model

This is the more detailed model of data curation, developed and used by the DCC.

The model shown on the slide is active, linked to: <http://www.dcc.ac.uk/resources/curation-lifecycle-model> where further detail, downloads and FAQs are available.

If time is short, point out the two outer ring of the model and the next inner-ring, "curate and preserve" noting the overlap with the previously-shown simpler models.

Activity 2: Examining the DCC Curation lifecycle model (optional, 20 minutes)

In advance, familiarise yourself with the model! Also unhide the slides (12-13).

#### Running an activity

1. Use slide 13 to describe the model in a little more detail (5 minutes)
	1. Note the difference between full lifecycle, sequential and occasional actions.
		1. **Full lifecycle actions** are shown in concentric rings around the data objects at the centre of the model. These are activities which take place at any time during the digital curation lifecycle and are relevant to many different sequential actions.
		2. **Sequential actions** are the steps which are repeatedly taken to ensure that data is curated according to best practice. This sequence is not simply performed once from start to finish but forms the basis of the curation chain and continues as long as data is being curated. Re-use and transformation of data can lead to the creation of a subset, by selection or query, or create newly derived results which themselves need to be curated.
		3. **Occasional actions** are those which interrupt or reorder the sequential actions as a result of a decision. For example, upon appraisal it may be decided that the data in question does not fit the remit of a digital repository in which case data may be transferred to another archive, repository, data centre or other custodian. In some instances data is destroyed, perhaps for legal reasons. Other occasional actions are the reappraisal of data which fails validation procedures or the migration of data to a different format to protect it against hardware or software obsolescence.
2. Distribute handouts of the model, which can be downloaded here:
<http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>
3. Ask researchers to consider the model for 5 minutes and to assign a 1-10 score of relevance for them and their research.
4. Run a feedback session. (10 minutes)
	1. What were people's scores?
	2. What was their reasoning for that score?

Compare reasoning with Lifecycle Model FAQs:
<http://www.dcc.ac.uk/resources/curation-lifecycle-model/lifecycle-model-faqs>
in particular:

* + 1. I create data, why is the model relevant to me?
		2. I want to reuse other people's data, why is the model relevant to me?
		3. How can I use the model in practice?
		4. What are the benefits of this model?
	1. Ask if people would keep the same score, or change?

## Session review (slides 15-17)

### Slide 16: In summary

This is a round up slide, with one summary point for each of the topics covered – you may wish to edit to align with your particular emphasis.

### Slide 17: Acknowledgements

Cited here are acknowledgements for resources used to create this module.

This module itself is licensed under a Create Commons Attribution-NonCommercial-ShareAlike license, meaning:

* You must attribute the work [[ADD HOW!!]]
* You may not use this work for commercial purposes.
* If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.