

KRDS/I2S2 Digital Preservation Benefits Analysis Tool

Workshop: July 2011

1. Introduction

On 12th July 2011 a workshop was held to disseminate the results of the KRDS/I2S2 Digital Preservation Benefits Analysis project to the research community. The aims of the project were to test, review and promote combined use of the Keeping Research Data Safe (KRDS) Benefits Taxonomy and the I2S2 Value Chain Analysis tools for assessing the benefits of digital preservation of research data. The project partners are [UKOLN](#) and the Digital Curation Centre at the University of Bath, [Centre for Health Informatics and Multi-professional Education](#) (CHIME) at University College London, [UK Data Archive](#) (University of Essex), [Archaeology Data Service](#) (University of York), [OCLC Research](#), and [Charles Beagrie Limited](#).

Sharon McMeekin attended on behalf of the DPC and members in attendance included Neil Grindley (JISC), Diana Murray (RCAHMS), Catherine Hardman (ADS), Adrian Brown (Parliamentary Archives), Matthew Woollard (UK Data Archive), Keith May (EH), Max Wilkinson (BL), Liz Lyon (DCC), Neil Beagrie. Thanks go to Neil Grindley for providing the notes included here for the first two presentations.

(These **informal notes** are intended to give DPC members an informal briefing about what was discussed. They are partial, are not intended as an official record and should not be understood to represent the views of the DPC.)

2. Neil Beagrie – The Toolset

The guide for the framework will be released in July. Have looked at the common benefits across 16 or so projects, generating lists of generic benefits. People can add to, look out or delete examples allowing them to work from these and stimulate their own work from common generic benefits. The tool includes a pick list of popular generic benefits which can be used to populate the form on the right. Toolset builds on the KRDS Activity Model and a number of lifecycle models.

It supports services and indirect costs like estates and training. It provides a framework for activity modelling. Tool itself takes the phases and the activities and then maps against points and benefits within the lifecycle. It is possible to use stuff from tool 1 into tool 2. Can map across from one worksheet into the other.

<http://beagrie.com/krds-i2s2.php>

3. Disciplinary Test Sites and Applications

Dipak Kaira (Centre for Health Informatics and Multi-professional Education [CHIME] at UCL) MRC longitudinal health studies case study

An investigation to try and understand how data sharing takes place and can be improved across MRC health studies and to contribute back to MRC data sharing and policy reviews. Did this work across 6 studies to understand on how data is handled. How and where good quality data management strategies can be invoked and used. This is a virtual study that takes all the studies into account.

The way that the studies operate is different to other longitudinal studies that are held in archives. They undertake their own collection of data and have their own personal relationship with the subjects. They are responsible and have designed the study. They do their own cleaning and derivation and they share voluntarily as a complement to its core mission. (They may get commercial exploitation at some point). The wide range of study types and their maturity means the opportunities and benefits will vary between studies. But the principal thing is that there is no curation WITHOUT sharing.

Stimulating new networks and collaborations. If we think ambition and money, and we do richer x-disciplinary research then there are rewards to be had in larger pots of money. KRDS outcome type - direct benefit in 2-4 years. The metric over time will be INCOME.

But also in disciplinary circles, where they suddenly see my research and remember me, then kudos is a real benefit. Benefit - measure = 3 How can I benefit from re-purposing and re-use of data? Localised expression - lower costs of data collection and curation, faster route to usable data. There is often a saving to be made by using someone else's data. It's a cost avoided but it's score = 4 because it's cheaper to use the data and quicker to publish.

Not every part of Neil's lifecycle applies to MRC because they don't tend to go to service. Working through the toolkit may be of benefit to the studies because they might work out what is important and what isn't in terms of workflow and may help them to design new and additional pieces of curation-based work. Whilst initial population might be done by one person completing the spreadsheet and working out the weighting might be nicely undertaken in a team workshop.

Catherine Hardman (ADS)

Archaeology Data Service Case Study

ADS have undertaken a lot of costing analysis, which the toolset has helped with, and this has then had an impact on decision making in relation to the Digital Archive. Has also helped articulate the benefits gained by undertaking digital curation work. In particular worksheets explain benefits in a very structured fashion which is useful with a range of audiences including projects, funding agencies and partner organisations. With the latter it is particularly useful to help identify partners' relative strengths and weaknesses. The toolset can also be used to produce different levels of information for different audiences.

The issue of value is key in archaeology. Due to the financial crisis things have been difficult in the discipline and so have been trying to articulate the value of archaeology. IFA have produced the Southport report to help with this, showing that that value of archaeology is informational rather than directly financial. Linked to this the tool provides a way of articulating different types of value. In addition the value chain allows you to identify value for different audiences.

'Quantification' of impact can help in a number of ways: in research to justify resources; in archiving can help with selection and retention decisions. Results from use of tool can be used to persuade projects and funders of the importance of data curation. Can also help to plan for this when finite resources are available.

Monica Duke (UKOLN)

SageCite Case Study

SageCite – citation in the domain of disease network modelling. Case study carried out a range of activities, review of data citation – technologies, issues. About understanding the domain. Developed a demonstrator, worked with publishers, did benefits analysis. Main partner Sage Bionetworks, US-based non-profit, trying to provide open tools and resources for community. Identified 7 stages of data management/lifecycle, each then broken down further. Figuring out where data citation fits in. Many different sources of data coming into repositories. Terminology issues, the discipline used 'data curation' to mean a slightly different thing – more about sorting/cleaning of data. Looking more at benefits of data citation rather than curation, benefits articulated:

- Better discovery of network models, better access (direct)
- Increased trust and reproducibility (indirect)
- Machine readability, recognition for contributors, better evidence for use in articles, more attributions (near term)

Who benefits?

- Funders, Policy makers, scientists, Datacite/BL, publishers. (Internal)
- Society, Funders, other scientists, other publishers, increased public trust (external)

Ongoing work: validate analysis with domain experts, need to take into consideration new versions of the tool
What has been learned: Easy to apply and articulate benefits, may need intermediary to facilitate process, digital data management background useful. Terminology important.

In Summary: can be used on just one aspect of curation, can have positive effect on practice and tools, analysis with the input of experts most useful.

Matthew Woollard, (UKDA)

UK Data Archive Case Study

UKDA is an umbrella data service organisation, providing access to social science data generated by government and academics.

Tool helps articulate why what you are doing is good, also helps identify what you should be doing more of (helps prioritise activities), justify costs to stakeholders, also understanding of the service impact. A check on what you are doing internally and externally.

Shows: where there's the need to add value (internal), where value is added (external), who can benefit and when. Hoping that it will help build a business case to help increase income. Help to prioritise in planning activities to maximise benefit and impact. [Long list of generic benefits – see slides, online soon]. Stakeholder may receive multiple benefits, different stakeholders get different levels of (same) benefits.

Tips:

- Start with worksheet rather than value chain, start simple and expand out, then reduce...
- Use a customised spreadsheet to sort and configure for your own needs.
- Quantitative benefits **must** be measurable
- Qualitative benefits can be used to create case studies which people like
- Impact weights are subjective: Decided to put benefits into sentences to help create the value chain.

Early days with tool and it takes a lot of time but has helped already: UKDA implementation of strategic plan, advice to researchers on data management and planning, influenced ESRC's call for the economic impact of the research data infrastructure.

4. Neil Beagrie – Implications for Funders

Work on KRDS has almost always been from the point of view of researchers and projects, but should also look from PoV of the funder. Implications: generic but easily tailored for specific domains and funders; variable levels of expertise needed; consistent, powerful ways of stimulating thinking on benefits and value.

Next steps: will be applying it to specific disciplines and organisations. Uses of toolkit: designed for a wide range of users which could include funders.

JISC Examples: When evaluating responses to a call can be of particular use when looking at questions such as 'benefits to wider JISC community'.

Questions: Can the tool be used to pull out benefits specific to a project, eliminating generic benefits that all projects of the type will bring? No clear answer on this specifically. Suggested that a list of generic benefits provided be provided by funders to rule them out. Comment that it is good to have established metrics to allow the expression of these benefits.

5. Plenary Discussion and Questions

What does this actually mean for funders? More work? Will need to use benefits framework carefully to help shape calls. Might be seen as another administrative burden, but better realisation of benefits should make it worthwhile. Shouldn't take more than a couple of days effort max, and perhaps limits should be placed. Need to help direct applicants to good selection of benefits to include within application. Need to limit time spent. Benefits are assumed based on **proposed** actions. Likely that standard benefits will become agreed within particular disciplines and so the tool might fall out of use. Wonder who will use it in the long term?

6. About this document

Version 1	Document initiated	13/7/2011	SMc (with content from NG)
Version 2	Checked and approved	15/07/2011	WK
Version 3	Distributed to members	15/07/2011	SMc