



Project Document Cover Sheet

Project Information			
Project Acronym	DCC SCARP		
Project Title	Disciplinary Approaches to Sharing, Curation, Reuse and Preservation		
Start Date	01 January 2007	End Date	31 March 2009 extended to 31 July 2009
Lead Institution	The University of Edinburgh		
Project Directors	Chris Rusbridge; Liz Lyon		
Project Manager & contact details	Colin Neilson, UKOLN, University of Bath c.neilson@ukoln.ac.uk		
Partner Institutions	The University of Bath (UKOLN), The University of Edinburgh (DCC), Science & Technologies Facilities Council (STFC)		
Project Web URL	http://dcc.ac.uk/projects/scarp/		
Programme Name	<i>e-infrastructure</i>		
Programme Manager	Neil Grindley		
Document Name			
Document Title	<i>Final Report</i>		
Reporting Period	1 Jan 2007 – 31 July 2009		
Author(s) & project role	Liz Lyon, Chris Rusbridge (project co-directors) Colin Neilson (project manager) Angus Whyte (research officer)		
Date		Filename	Scarp-final-project-report.pdf
URL	http://dcc.ac.uk/scarp/scarp-final-project-report.pdf		
Access	<input type="checkbox"/> Project and JISC internal	<input checked="" type="checkbox"/> General dissemination	
Document History			
Version	Date	Comments	
1.0	17/09/2009	Draft for amendment / comment	
1.1	09/10/2009	Re-draft for amendment / comment	
1.2	16/10/2009	Draft for amendment / comment	
1.3	22/10/2009	Re-draft (LL email 19/10/09) for amendment / comment	
1.4	28/10/2009	Consolidation of AW input for comment	
1.5	5/11/2009	Final draft for comment	
1.6	23/02/2010	Consolidation of recommendations	
1.7	4/03/2010	Final, including workshop highlights	



JISC Final Report

DCC SCARP

Disciplinary Approaches to Sharing, Curation, Reuse and Preservation

FINAL REPORT

Author(s):

Liz Lyon

Chris Rusbridge

Colin Neilson

Angus Whyte

Contact Person

Liz Lyon, E.Lyon@ukoln.ac.uk

Date

4/03/2010

Table of Contents

Executive Summary.....	4
Background	4
Aims and Objectives.....	5
Methodology	6
Implementation	9
Outputs and Results	10
Outcomes	12
Conclusions	17
Implications.....	18
References	21

Acknowledgements

This project was funded by JISC under the e-infrastructure programme.

The Project Partners were:

DCC at the University of Edinburgh
UKOLN at the University of Bath
STFC (Science & Technology Facilities Council) at Rutherford Appleton Labs.

The synthesis was provided by Key Perspectives Ltd
48 Old Coach Road, Playing Place, TRURO, TR3 6ET
+44 (0)1392 879702
www.keyperspectives.co.uk

The authors are very grateful for the generous support of participating research groups and curators.

Executive Summary

The DCC SCARP project investigated attitudes and approaches to data deposit, sharing and reuse, curation and preservation over a range of research fields in differing disciplines. The aim was to investigate research practitioners' perspectives and practices in caring for their research data, and the methods and tools they use to that end. Objectives included identification and promotion of 'good practice' in the selected research domains, as expressed in DCC tools and resources such as the Curation Lifecycle Model.

The approach combined case study methods with a survey of the literature relevant to digital curation in the selected fields. A range of methods was applied to fit the differing research contexts. The aim was for SCARP researchers to be 'immersed' in groups of practitioners. The success of this was limited by a number of factors, the most general being that full' immersion (i.e. mostly or wholly embedded in a research group) was neither feasible nor desirable in the circumstances. Alternative strategies were used and the varying degrees of success shed additional light on differences in data practices between research fields.

The resulting 7 case studies, each with recommendations and findings for the research groups and for the range of stakeholders in digital curation, demonstrated that the curation tools, such as the Digital Curation Lifecycle Model, were useful provided effort was applied to adapt them to the context of 'live' data creation and curation. Participating research groups generally did not have formalised curation or preservation processes, so case studies aimed to understand expectations, risks and constraints, and find appropriate ways to build on current capabilities.

The SCARP synthesis report, carried out by Key Perspectives for the project, identifies factors which help understand how curation practices in research groups differ in disciplinary terms. This provides a backdrop to different digital curation approaches. However the case studies illustrate that the discipline is too broad a level to understand data curation practices or requirements. The diversity of data types, working methods, curation practices and content skills found even within specialised domains means that requirements should be defined at this or even a finer-grained level, such as the research group.

A case study approach to understanding researchers practice in producing, caring for and exploiting research data can support uptake of digital curation techniques, as part of a broader engagement strategy based on dialogue with researchers to ensure the appropriateness of more formal digital curation models and tools.

Background

The Digital Curation Centre SCARP project through its seven published case studies has explored data curation practice across a variety of clinical, life, social, humanities, physical and engineering research communities. The case for in-depth case studies emerged from the Warwick Workshop "Digital Curation and Preservation: Defining the research agenda for the next decade". Held in November 2005, the workshop found that "there is currently little understanding in any detail of the real heterogeneity of data management and preservation practices across different disciplines".

A better understanding of the diversity of researchers' data practices, and of their needs and expectations of support for data management and preservation remains vital for at least four reasons. Firstly, the data policy environment for data management has changed quickly;

funding bodies place increasingly explicit responsibilities on researchers and institutions to manage and preserve data (e.g. the RCUK *Code of Conduct and Policy on the Governance of Good Research Conduct*). Secondly, institutions are responding to the challenge by embarking on data repositories and support services but the requirements for these are complex and difficult to establish. Reporting on the problem of 'empty data archives', recent articles in *Nature* point to major issues in the "staggering diversity" of the data institutions seek to curate, as well as "wide variations in the culture of data sharing" and attendant legal issues across research fields, Nelson, B (2009). Thirdly, the need for tools and technical infrastructure to deal with the quantity and quality of data are still an active area of research. Fourthly, there is a widely-seen need to support creators to document their data 'upstream' of any ingest process, partly to control the costs of this, for example see Beagrie, N, Chruszcz, J and Lavoie, B. (2008). This requires an understanding of the practicalities of documenting the data creation context, the process followed, and the background knowledge that research communities need to reuse curated data productively.

The JISC report "Dealing with Data: roles, rights, responsibilities and relationships", Lyon, L. (2007), sets out recommendations for an institutional framework supporting digital working with research data (in the context of an e-infrastructure for the UK). The SCARP project aimed to complement this approach by looking at practitioners views and practices in relation to data sharing, curation, access and preservation.

The results complement other recent studies, such as Fry, J. et. al. (2008), on a cost-benefit analysis of data curation and preservation. SCARP provides a detailed qualitative account of curation and preservation in context, identifying common practices, motivators for change, and barriers to accomplishing it. The project highlights the necessity for institutional data service providers to understand the diverse forms of data researchers already curate, and to provide data exchange and support mechanisms that fit with their requirements.

Aims and Objectives

The DCC SCARP Project aimed to investigate the disciplinary attitudes and approaches to data sharing, curation, access, reuse, and preservation through a range of methods, combining traditional survey/literature review approaches with up to 10 immersive case studies in selected discipline areas. The studies were to apply known good practice, where possible identify new lessons from practice in the selected disciplinary areas, and potentially develop new good practice. Seven case studies were successfully completed.

The case studies followed the main themes below:

- Policy drivers and barriers: Organisational and institutional factors including different skill levels, deposit and preservation policies and arrangements, willingness to use these and the relationships to incentives and reward structures within disciplines.
- Stewardship practices: The research process and methods; how these relate to the digital objects created (e.g. data, drawings and models, software tools and methods, repositories and databases). How the digital assets produced (outputs) are used (including reuse and sharing) and linked to publication. Attitudes to doing this and to the usefulness of prior data, and the sustainability of collected digital information.
- Tools and infrastructure: the tools and facilities used to collect, deposit, find, cite, discuss and annotate the digital objects and to ensure persistence and preservation over long time periods.

- Preserving context: how communities of practice and their knowledge bases can be characterised, and how understanding of the lineage and provenance of digital objects may be documented.

The themes did not change during the project but the emphasis varied with the research group and disciplinary setting. Participating research groups varied widely in the extent to which they had explicit curation or preservation processes. The studies therefore aimed to understand expectations, risks and constraints, and so find appropriate ways to build on current capabilities.

One of the aims of the SCARP project was to integrate the work of the project into existing DCC structures and channels. This entailed a double learning curve for those working on SCARP, firstly to learn enough about the host groups to contribute to their practice, and secondly to keep pace with DCC's ongoing development of digital curation tools and methods.

Methodology

The methodology combined a case study approach with a landscape review of relevant prior studies in each of the research fields examined, building on prior surveys of attitudes to sharing and deposit carried out in, for example, the JISC-funded project StORe¹. The methodology envisaged that SCARP researchers would then be 'immersed' in their host research communities through in-depth case studies of 4 to 6 months. Embedded in research groups or established curation centres, they would work on one or more specific curation problems relevant to their respective communities, with the aim of helping move these problems forward towards a practical solution.

The methodology was influenced by similarly 'immersive' approaches to case studies of technology development in other fields. Ethnographic approaches – typically involving prolonged periods of interviewing and participant observation in a fieldwork setting (e.g. a laboratory) – are common in Science and Technology Studies (STS) and in Computer Supported Collaborative Work (CSCW). In the latter more especially, ethnography is used to document and convey an understanding of the social organisation of activity as a first step to identifying the requirements for collaborative systems development. It is also typically done in weeks or months rather than years (a timescale more typical of STS). SCARP adopted a relatively cut-down approach involving:-

- Semi-structured interviews with members of the participating groups; typically Principal Investigators, data managers, post-doctoral researchers and research assistants.
- Observation of data management activities and meetings
- Walkthroughs or demonstrations by participants of the methods and tools they normally work with.

SCARP sought to address problems as well as contribute to understanding of curation in the problem domains. It therefore also involved an element of action research, i.e. a process of identifying problems or situations in need of improvement, agreeing relevant courses of action with those affected, intervening to take the appropriate action, and reflecting on the outcomes. This strand of the methodology drew on the tools and methods that DCC was developing in parallel with SCARP. In one case (Atmospheric Sciences) an approach to

¹ <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/store.aspx>

analysing preservation requirements was piloted, in another (Neuroimaging) the risk assessment approach DRAMBORA was used, and in most the DCC Digital Curation Lifecycle Model was applied. The studies generally involved:-

- Identifying the relevant actors and stakeholders in the participating group
- Identifying the policy context, in terms of funder data policies and relevant legal or ethical compliance regimes
- Identifying curation drivers, challenges and risks
- Assessing data types, assets and current activities to manage these
- Identifying feasible and desirable next steps

Alongside the above, the cases variously used questionnaires, meetings, workshops and participant review to feedback and consolidate results.

Case Selection

Research fields were targeted for the case studies aiming to cover wide areas of research, but also on the pragmatic grounds of DCC partners' prior contact and engagement with:

- branches of engineering and architecture based around Bath and Bristol (UKOLN contacts)
- science data and science images based around Oxfordshire (STFC contacts)
- medical and social sciences based in Edinburgh (DCC Edinburgh contacts)

In each broad disciplinary area groups were approached where they met the following criteria:

1. Substantial volumes of research data were being gathered, i.e. substantial enough for the host team to regard data management as a specialist task and project management issue.
2. Project managers or others responsible for data management were instigating changes in practice or policy, or saw a need to review their existing procedures, policies, technologies, or methods of preservation and reuse.
3. Some aspect of the host's research process was accessible; for example allowing the SCARP researcher to observe members of the host team working with their data.

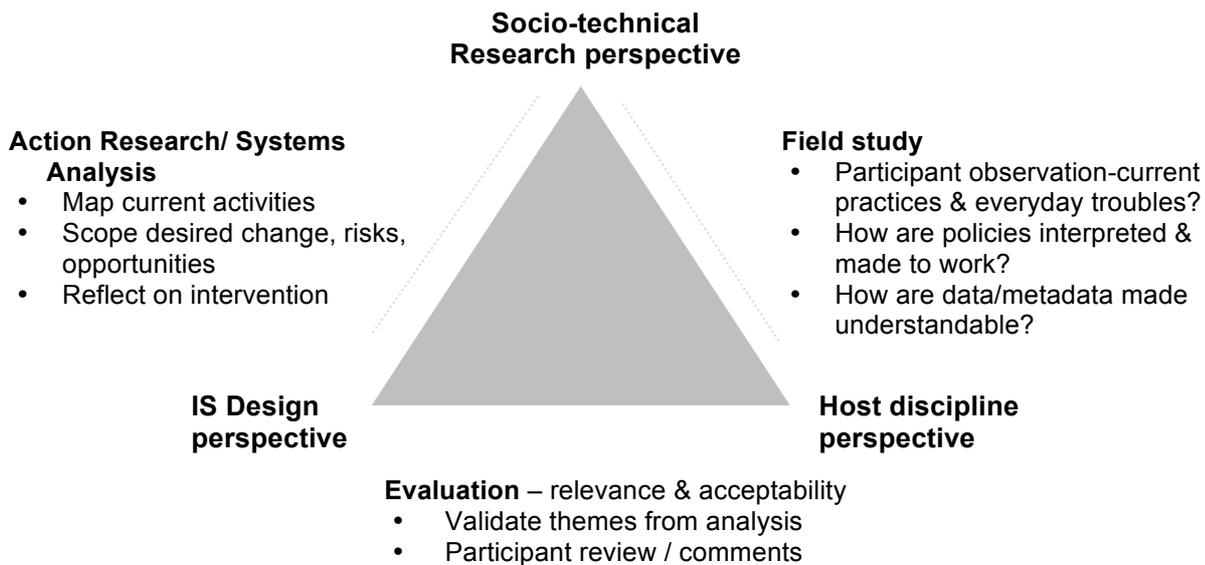
More specific criteria were applied within the disciplinary areas. For example with the focus in Edinburgh on social and medical/health fields, criteria included: -

- a. Projects with an interdisciplinary focus – combining health and social research questions and methods, and digital research data
- b. Projects with a social (or socio-technical) research focus, and health/medical application (e.g. eHealth)
- c. Data archiving or curation services that address the needs of social or medical research teams (or both)

The SCARP deliverable Whyte, A. et al (2008) "Methods and Tools for SCARP" reviews the approaches considered at the outset of the project, within the broad definition of a 'case study' as an empirical inquiry that investigates a contemporary phenomenon within its real-life context. The SCARP approach had to trade-off the needs for a descriptive account of current practice, and (on the other hand) interventions to 'apply best practice' and evaluate the outcomes. This drew on the few previous attempts in the data curation field;

Karasti (2001) for example attempted “integration of the ‘systemic analysis’ of current practices with ‘appreciative intervention’ to introduce change. This means bridging “the disciplinary discrepancies between ethnographic studies of work and system design, such as descriptive vs. prescriptive, particular vs. general, concrete vs. abstract, present vs. future, and understanding vs. intervention.” (p.239). One influential framework for doing this is given by Suchman and Trigg (1991); Figure 1 below adapts this to SCARP.

Figure 1 A Hybrid Methodology



(Adapted from Suchman & Trigg, 1991)

Figure 1 illustrates that methods from three traditions; action research in information systems, evaluation of information systems requirements, and field studies of science and technology would be used where they helped bring together the three perspectives needed to address the project aims. At the outset it was envisaged that case study outcomes would include high-level system requirements in the form of roles and goals or (re)use cases, although not working systems.

A strong factor in any “immersion” is that the methods used have to be adapted according to the resources and time available from the host research practitioners, teams and organisations. A particular consideration is that in relation to any extra work generated for the researchers what will the research team get from taking part in the study? For the case study methods to be acceptable to the host group care was needed to ensure that the normal working of the group, with their own priorities and deadlines, was not adversely affected. For instance Case Study 4 adopted a combination of methods to achieve a degree of immersion

- informal interviews with staff in key roles
- demonstration of operational processes by key staff
- attendance as observer at research group meetings
- lab-based observation over a series of half days
- questionnaire to staff

The critical factor is that the degree of immersion needs to be sufficient to support useful analysis in relation to digital curation processes but light enough in impact not to interfere with the practitioner's normal working. In terms of scaling up of SCARP like approaches, for instance to cover a wider range of research fields and larger number of case studies, there is a need to moderate the degree of immersion according to the effort required but sufficient to produce useful results and points that can be acted on to produce benefits.

Implementation

Project Startup

The project plan envisaged recruiting staff with knowledge and skills in a disciplinary area that matched a participating DCC partner's local strengths and contacts. The project had a start date of January 2007, although it was not until August 2007 that staff were in place, as shown in Table 1 below.

Partner (part-time / full-time)/ Researchers initials	Start – End Date	Disciplinary Areas
STFC Researcher 1 (PT)/ EC	Apr 2007 – Mar 2008	Scientific Data Centres associated with the Science and Technology Facilities Council (e.g. British Atmospheric Data Centre).
STFC Researcher 2 (PT) / AM	Apr 2007 – Jan 2008	Biology especially with regard to image based working
UKOLN Researcher 1 (FT) / CN	Jul 2007	Architecture, Engineering
DCC Edinburgh (FT) / AW	Aug 2007- Oct 2009	Social Science, Medical Science
STFC Researcher 3 (FT)/ MC	Jul 2008 – Jul 2009	Astronomy

Table 1. Embedding researchers in disciplines

EC was embedded with scientific data centres based at STFC, for instance working on the specific area of preservation analysis techniques for scientific datasets in the British Atmospheric Data Centre (BADDC). AM, working from STFC, was embedded with the Images Bioinformatics Research Group, in the Zoology department of Oxford University.

This initial phase involved surveying existing work in the disciplinary areas and (for non-STFC researchers) identifying candidate research groups and centres, and then seeking their commitment to participate in a case study. This involved a number of visits to relevant data centres, workshops and conferences, e.g. AW visited the UK Data Archive and attended a data management workshop organised by the Wellcome Trust Clinical Research Facility in Edinburgh; CN attended the European Conference on Architecture and Digital Archives, Paris, 8-10 November 2007, organised under the framework of the Gaudi European Project.

Immersion in Practice: Access and Skill Issues

SCARP adapted to the evolution of “good practice” in digital curation during the project as expressed for example in the stages, processes and activities set out in the DCC Digital

Curation Lifecycle Model². This increasingly framed the engagement with research practitioners, serving as an 'ideal type' around which current practices and improvements could be discussed.

The level of immersion first envisaged, where researchers would spend much of their time on site over a period of several months spent implementing curation solutions, was not realised however. This aim presumed high levels of access and 'curation readiness' on the part of host groups, and a level of understanding of their practice that was not achievable in the projected timescales.

The participating groups had distinctive and clear ambitions for their data and it was not difficult to relate these to the DCC Curation Lifecycle. However they generally were interested in a specific aspect of curation (e.g. data quality, integration, secure storage) rather than the more holistic view espoused by DCC, and their practices reflected this. In engaging with research groups, the methods used had to reflect the demands on the time of busy specialists. Many of the research groups consisted of individuals with multiple roles (e.g. research, teaching, clinical practice), but with no individual specifically funded to perform data management.

The 'immersive' approach called on the SCARP researchers to possess or obtain sufficient understanding of each host research community to identify specific curation problems, assess the situation that gave rise to these problems, and then build on the host research group or community's capacity to make progress.

The case studies therefore demanded a hybrid set of skills combining knowledge of the targeted domain (or sub-discipline) with specialist knowledge of data management as currently practised in it. This was a significant challenge since the SCARP researchers had, for the most part, information management or librarianship skills with general knowledge of curation, and were required to apply tools and resources that were still under development, to specialist data management settings they had limited prior knowledge of. The studies produced as a result were intended to interest other researchers in those specialist fields, as well as the curation and research policy communities. This was accomplished with varying degrees of success; impacts from the studies and lessons from the methodology are described later under 'Outcomes'.

Outputs and Results

Synthesis study results

The synthesis study commissioned by the Digital Curation Centre from Key Perspectives Limited is a comparative review based on sixteen case studies (include the SCARP case studies) covering the areas of data sharing, data discovery, access and reuse, long term care and preservation of data. The synthesis study report Key Perspectives (2009) "Data dimensions: disciplinary differences in research data sharing, reuse and long term viability" is available on the DCC web site.

A central finding is that institutional repositories responsible for curating data produced by their own research community will need to develop domain-specific strategies since a generic approach to data curation will not be sufficient to cope with the different data-related needs and expectations of researchers working in different disciplines.

² See <http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf>

The report compiles, synthesises and presents the evidence, across all of the case studies, for disciplinary differences in production, care and use of research data. The factors and reasons for the disciplinary differences, given in the conclusion to the report, include:

- the heritage and practices of niche research communities
- the type and quantity of data they produce
- the uniqueness of those data and their potential value in terms of reuse
- the propensity of each community to create, adapt or adopt common data formats, metadata schema and other relevant standards
- willingness to share data in a world where competition for funding looms large
- the policies of funding bodies in relation to data management, sharing and preservation
- the provision of storage infrastructure including national data centres and effective discovery systems

Overview of SCARP case studies

The seven case studies each looked at aspects of data sharing, curation, archiving, re-use, preservation and the infrastructure for these in different disciplinary fields. Each makes specific recommendations for improvement of practice in that field.

- [SCARP Case Study No. 1 - Curating Brain Images in a Psychiatric Research Group: Infrastructure and Preservation Issues](#) 
14 November 2008 | Angus Whyte
This study involved the Neuroimaging Group in University of Edinburgh's Division of Psychiatry. It combined an assessment of risks to the long-term value of the research group's datasets with field work to understand current data practices in their context. A 5-page [Summary and Recommendations](#) [PDF, 170KB] is available. An annex [Neuroimaging Data Landscapes](#) [PDF, 1.23MB] provides background on the development of imaging, the nature of the data collected for neuroimaging studies in psychiatry, data repository and curation resources available, and legal and ethical constraints on data exchange.
- [SCARP Case Study No. 2 - Curating Atmospheric Data for long term use: Infrastructure and Preservation Issues for the Atmospheric Sciences community](#) 
2 June 2009 | Esther Conway
This study engaged with a number of archives, including the British Atmospheric Data Centre, the World Data Centre Archive at the Rutherford Appleton Laboratory and the European Incoherent Scatter Scientific Association (EISCAT). We developed a preservation analysis methodology capable of identifying and drawing out discipline specific preservation requirements and issues. We present the methodology along with its application to the Mesospheric Stratospheric Tropospheric (MST) radar dataset, which is currently supported by and accessed through the British Atmospheric Data Centre. We suggest strategies for the long-term preservation of the MST data and make recommendations for the wider community.
- [SCARP Case Study No. 3 - Clinical Data from Home to Health Centre: the Telehealth Curation Lifecycle](#) 
29 June 2009 | Tasneem Irshad, Jenny Ure
This study looks at the data curation lifecycle in Telehealth research. Telehealth, or telecare, is an emerging sub-domain of eHealth, and the report profiles current practices in several telehealth pilot projects. Data curation is at an embryonic stage but can draw

on related eHealth initiatives and clinical data management practices, and the report considers the infrastructure needed for data curation in this field of research and practice.

- [SCARP Case Study No. 4 - Curated Databases in the Life Sciences: The Edinburgh Mouse Atlas Project](#) 
13 July 2009 | Elizabeth Fairley and Sarah Higgins
This study scopes and assesses the data curation aspects of the Edinburgh Mouse Atlas Project (EMAP), a programme funded by the Medical Research Council (MRC). The principal goal for EMAP is to develop an expression summary for each gene in the mouse embryo, which collectively has been named the Edinburgh Mouse Atlas Gene-Expression Database (EMAGE).
 - [SCARP Case Study No. 5 - Roles and Reusability of Video Data in Social Studies of Interaction](#) 
19 October 2009 | Angus Whyte
The study reviews the curation landscape in several interdisciplinary fields that use video analysis in studies of human interaction. A 5 page [Summary and Conclusions](#) [PDF, **136KB**] is also available. The study primarily focuses on uses of video in ethnographic studies and in eye movement research, and is based on interviews and field study.
 - [SCARP Case Study No. 6 - Digital Curation approaches for Architecture](#) 
30 October 2009 | Colin Neilson
This study highlights choice in how to provide for appropriate care of digital objects, choice in digital curation treatments, as a means of promoting more effective current and future architectural practice and research. The digital assets produced by use of digital tools and from digital methods of working in the teaching, learning, research and practice of Architecture require appropriate curation treatment if the full value of the assets is to be realised.
- SCARP Case Study No. 7 - Curation of research data in the disciplines of Engineering
20 January 2010 | Alex Ball and Colin Neilson
The study examines approaches to data deposit, sharing and reuse in engineering research fields within the UK Higher Education sector.

Outcomes

The case studies detailed some of the realities of research 'on the ground' in the domains concerned, provided insights into how curation is currently practiced, and conveyed how practitioners in those domains want their curation to be supported. The studies offer detailed portraits of the steps taken in a variety of research domains to curate and preserve research data.

Case study outcomes to date are identified below; firstly in terms of building the curation capacity of the researchers and groups participating in the study, secondly by their contribution to DCC activities, thirdly through wider take-up, and fourthly in terms of increased community awareness. This is followed by some comments on the methodology.

Building Curation Capacity

- *Study 1 Neuroimaging* was a catalyst for the research group to form a Data Working Group to formalise its data curation approach and progress the study recommendations. The study also featured in the subsequent RIN/BL 'Patterns of information use and exchange: case studies of researchers in the life sciences' report quoting group members on SCARP's impact as "now we manage our data".

- *Study 2 Atmospheric Science* piloted a visual tool for preservation analysis in the British Atmospheric Data Centre.
- *Study 3 Telehealth Data* was co-authored by a research group member and identified a range of curation approaches likely to be transferable to the telehealth domain from other eHealth domains.
- *Study 4 Curated databases in the Life Science* is cited on the Emage web site as in-depth report about their data curation approach
http://www.emouseatlas.org/emage/about/data_curation.html
- *Study 5 Video Data on Social Interaction* led to a joint bid to the AHRC from Edinburgh Datashare and one of the participating projects to develop a curated video data collection. Other participants identified new tools, metadata schemas and format recommendations useful to them.
- *Imagestore* –The initial work on curation requirements for legacy analogue and ‘born digital’ scientific images carried out at the Image Bioinformatics Research Group of Oxford University informed subsequent development in the JISC ADMIRAL and SHUFFL Projects. Although not written up as a case study the work is recorded in a wiki at http://imageweb.zoo.ox.ac.uk/wiki/index.php/The_ImageStore_Project

DCC Organisational Learning

- *Study 1 Neuroimaging* was used in Curation 101 training courses to illustrate application of the DCC Curation Lifecycle Model, and presented at the Research Data Management Forum.
- *Study 2 Atmospheric Science* contributed to the development of preservation analysis methods in the FP6 project CASPAR.
- *Study 5 Video Data on Social Interaction* provided an iterative approach to using the Curation Lifecycle Model alongside risk management
- SCARP researchers contributed to drafts of the DCC Data Management Plan Template, and a Legal Briefing on Sharing Medical Data based on their case study experiences.
- The experience of the SCARP work contributed to DCC working as a partner with the Institute for the Study of Science, Technology and Innovation (ISSTI) on the BL/RIN ‘Disciplinary case studies in the Life Science’ project (<http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/disciplinary-case-studies-life-sciences>)
- Progress on SCARP was discussed at each of the DCC face-to-face meetings in the project period.

Wider Take-up

- *Study 1 Neuroimaging* was used to seed an annotated list of studies for a ‘Data Curation Solution Community’ on Fedora Commons³
- *Study 4 Edinburgh Mouse Atlas* was used as a curation report template by researchers working on the Developmental Gene Expression Map, an FP6 funded project.

³ Data Curation Solution Community:

<http://www.fedoracommons.org/confluence/display/FCCWG/Data+Curation>

- *Study 5 Video Data on Social Interaction* led to collaboration with JISC Digital Media on a DCC Curation Manual chapter on curating video data.

Increasing Community Awareness

- *Study 1 Neuroimaging* was presented at the IASSIST 2008 conference, in poster sessions in two other international conferences, a workshop organised by the University of Oxford EIDCSR Project, and an article in the International Journal of Digital Curation.
- *Study 2 Atmospheric Science* was used at the wePreserve DPE/Planets/CASPAR/nestor Joint Training Event (Prague, October 2008) as the basis of a presentation on Preservation Analysis Workflow and Preservation Descriptive Information. Also for iPres (Sixth International Conference on Preservation of Digital Objects, San Francisco, October 2009) presentation on 'Curating Scientific Research Data for the Long Term: A Preservation Analysis Method in Context.'
- *Study 3 Telehealth* is presented in a paper submitted to the International Journal of Digital Curation, and is to be presented at the 2nd Dec. 2009 SCARP workshop.
- *Study 5 Video Data on Social Interaction* is presented in a paper for submission to IJDC or elsewhere.
- SCARP was also highlighted in presentations by Project Director Liz Lyon at iEXPO Conference, Paris, June 2009; Research Data Management Workshop, Oxford, June 2008; Max Planck eScience Seminar, Goettingen, June 2008; NERC Data Management Workshop, October 2007.

Publications mentioning SCARP

The DCC SCARP project has been mentioned in several publications, including the following:

Conway, E., Dunckley, M., & Giaretta, D. (n.d.). Curating scientific research data for the long term: A preservation analysis method in context. In *iPres 2009*. Oakland: California Digital Library. Retrieved from http://epubs.stfc.ac.uk/bitstream/4095/iPRES2009_Conway_preservation_A.pdf.

Friend, F. (2008). Working For A Research-Friendly IPR Framework In The UK. In B. Fitzgerald, *Legal Framework for e-Research: Realising the Potential*. Sydney: Sydney University Press. Retrieved from <http://www.austlii.edu.au/au/journals/SydUPLawBk/2008/42.html>.

Fry, J., Lockyer, S., Oppenheim, C., Houghton, J., & Rasmussen, B. (2008). Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes. *Victoria*. Bristol. Retrieved from http://ie-repository.jisc.ac.uk/279/2/JISC_data_sharing_finalreport.pdf.

Harvey, R. (forthcoming). *Digital Curation: A How-To-Do-It Manual*.

Higgins, S. (2008). The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, 3(1), 134-140. Retrieved from <http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48>.

Lyon, L. (2007). Dealing with Data: Roles, Rights, Responsibilities and Relationships Consultancy Report. Bristol. Retrieved from http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf.

Lyon, L. (2009). Open Science at Web-Scale: Optimising Participation and Predictive Potential Consultative Report. Bath. Retrieved from <http://www.jisc.ac.uk/media/documents/publications/research/2009/open-science-report-6nov09-final-sentojisc.pdf>.

Pryor, G. (2008). Code, Analysis, Repository & Modelling for e-Neuroscience. Edinburgh. Retrieved from <http://www.dcc.ac.uk/docs/publications/case-studies/carmen.pdf>.

Treloar, A. (2009). Design and Implementation of the Australian National Data Service. *The International Journal of Digital Curation*, 4(1), 125-137. UKOLN. Retrieved from <http://www.ijdc.net/index.php/ijdc/article/viewFile/107/83>.

Williams, R., & Pryor, G. (2009). Patterns of information use and exchange: case studies of researchers in the life sciences. London, UK. Retrieved from <http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/disciplinary-case-studies-life-sciences>.

Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. (2009). Constructing Data Curation Profiles. *The International Journal of Digital Curation*, 4(3), 93-103. UKOLN. Retrieved from <http://ijdc.net/index.php/ijdc/article/viewFile/137/165>.

Whyte, A., Job, D., Giles, S., & Lawrie, S. (2008). Meeting Curation Challenges in a Neuroimaging Group. *International Journal of Digital Curation*, 3(1), 171-181. Retrieved from <http://www.ijdc.net/ijdc/article/view/74>.

A number of project proposals to JISC Programmes have mentioned DCC SCARP as influential, including the successful EIDCSR project at Oxford.

Sheer Curation and Immersive Studies- SCARP Methodology

The phrase 'sheer curation'⁴ was coined early in the project, and refers to the principle of integrating curation into normal workflows, for example by automatically capturing metadata from routine data management actions, or building curation into other activities that are an accepted part of research work. While other projects share these goals, in-depth case studies in data curation are rare enough to deserve some comments on their suitability for investigating curation needs.

SCARP's approach to case studies offers lessons to others attempting to engage with research practitioners in order to design or improve data services and policies. Any such dialogue will involve methods of (a) understanding current practices and their context (b) assessing the situation to identify issues and problems, and (c) building on the host research group or community's capacity to make progress in addressing them. The 'learning cycle' shown in Figure 2 is adapted from Figure 1, so that .

⁴ For a description of "sheer curation" see http://en.wikipedia.org/wiki/Digital_curation

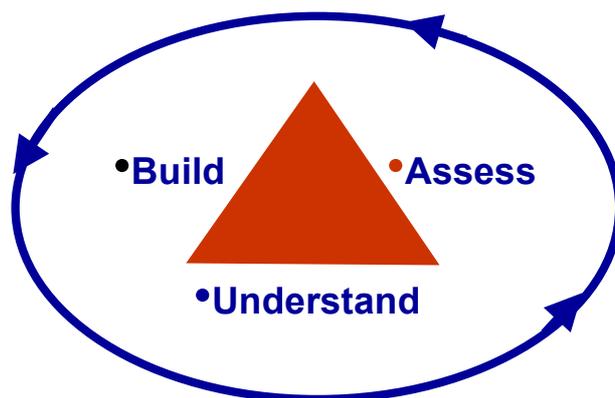


Figure 2 SCARP learning cycle

In the 'assess' strand, the DCC Curation Lifecycle Model was used either alone or alongside a risk assessment or preservation needs analysis. Usage scenarios would also have been appropriate, had there been the opportunity to assess a range of alternate solutions for software or service development beyond the exploratory stage that SCARP worked in. The 'build' strand consisted of identifying feasible and desirable next steps, recommending standards, schemas, methods, guidelines or risk mitigation steps that the host groups saw as relevant to current practice.

The 'understand' strand was more characteristic of SCARP and the case studies invariably involved semi-structured interviews. These followed a topic guide influenced by related studies⁵ and typically involved a succession of in-depth follow-up interviews with key contacts in the host groups. Alongside this several studies involved participant observation of research group meetings, and of study participants 'walking through' some aspect of their normal research work. These were extremely valuable opportunities to get an insight into the 'host' group's normal working life, those taken for granted routine accomplishments that researchers would not feel worth mentioning in interviews but, once remarked on, had some significance for curation. In-depth understanding of this 'tacit knowledge' ideally requires prolonged observational ethnographic study, but walkthroughs and limited periods of observation are likely to be more feasible in studies intended to develop curation capabilities.

Observation is not practical in some contexts; it requires (obviously) some form of combined verbal and physical activity that can be readily observed in a given place, and while interdisciplinary labs with a strong collaborative culture lend themselves to that, other forms of research work do not. In some domains, architecture being a case in point, much research work (and therefore data management) involves a great deal of solitary work or is unconfined to the lab or office. In such cases diary methods may be more appropriate (see for example the RIN project Case Studies in the Life Sciences).

⁵ Project StORe <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/store.aspx> ; CASPAR <http://www.casparpreserves.eu/> ; Corti, L. and Wright, M. (2002); Lord, P. & Macdonald, A. (2003).

Final Workshop: Common Issues

The final SCARP workshop, held in December 2009 at the International Digital Curation Conference in London, focused on issues that affect all disciplines, and the impact on the DCC.

There was concern that the strong agenda for openness (not applicable in all disciplines due to legal, ethical and other considerations) should not detract from the near-universal need for better curation and management of research data. Uncertainty on intellectual property issues was also identified as a common issue, and more work is needed on supporting data owners in appropriate licences (whether these be CC0, public domain dedication, restrictive licences such as the UKDA “special” licence, etc).

It was also suggested that “data bargaining” might be widely applicable, particularly in those disciplines where open sharing is not feasible or not the community norm. Models for “data sharing agreements” may be helpful. In many areas, help on ethical clearance may be useful.

In many areas help may be needed in terms of data collections rather than datasets (the latter can be a problematic term). It’s also clear that “selection criteria” continue to represent challenges for those who take on research data for others.

The overall message is that curation is about doing better research. Better curation and management will help the researcher even in the reasonably short term, and will have longer-term benefits.

Conclusions

SCARP project has demonstrated that it is possible to engage with research practitioners in an at least partially immersive manner, to better understand how digital curation capacity and support structures may be developed effectively. Further and broader studies are essential.

The goal of researchers is to increase understanding in their fields producing research results that highlight research problems and their possible solutions. Digital Curation, can serve to enable this through for instance appropriate sharing or data reuse methods. However formal approaches to digital curation need to demonstrate their relevance to helping practitioners to exploit the data they create or curate and realise benefit from their own work. Case studies are a useful tool when used as part of a strategy for engaging with researchers, especially when possible routes to improved methods are exemplified and where the researchers own use of best practice is highlighted⁶.

The effort needed to scale up and study a larger number of groups should not be underestimated however, and would be best approached within the context of an overall engagement strategy where endeavours to understand current practice are coupled with a range of specific curation tools or methods that may be tailored to fit requirements, or require evaluation of their fitness for purpose.

⁶ For instance in the SCARP project we were able to make use of the DCC/RIN Research Data Management Forum to present results and engage with Research Data Managers. <http://data-forum.blogspot.com/>

Interdisciplinary working was a key factor in the SCARP studies for two significant reasons. Firstly research groups engaged in interdisciplinary projects were interested in supporting case studies, as a means to take forward their curation practice. Secondly, their interdisciplinary work engendered curation requirements; particularly in the areas of data integration, schema development, quality assessment, and shared storage. Changes in digital data capture technologies towards higher resolution/fidelity digital were a further driver of curation requirements.

The 'data assemblages' or boundary objects that mattered most to researchers and sustained their interdisciplinary research were often visual representations, assembling data from diverse sources within the common boundaries of an 'ideal type' model (e.g. image atlas) or a spatial-temporal map (e.g. multi-tiered video annotations along a shared timeline), and thereby enabling a common reference point for detecting patterns in data. By contrast there was little interest in the concept of centralised institutional data repositories.

Research training was also a key factor in a number of studies. Informal fora working within and across research groups afforded early and mid-career researchers opportunities for data sharing, networking and collaborative problem solving. Sharing and the infrastructure to support it were motivated by this concern for joint pursuit of interesting lines of enquiry or shared problems, as much as by a desire to meet the legal and ethical obligations associated with reputable research.

Case study approaches can act as a bridge between research practitioners and the policy bodies responsible for providing direction in relation the infrastructure and funding required to support research. Organisations working in a digital curation advocacy role, such as the Digital Curation Centre, need to be able to give voice to research practitioners' views on care of data and achievement of benefits from research, in a bottom up manner. Advocates need to avoid being seen solely as a top down conduit of policy from service providers and funders. An engagement strategy to support uptake of digital curation techniques should be founded on learning from researchers about the appropriateness of curation approaches according to the particular circumstances of research groups and domains.

The benefits to a research group in taking part in the SCARP studies, and engaging with the Digital Curation Centre, included:

- promotion and dissemination of their own work,
- external analysis of their data working methods with possible suggestions for improvements in relation to management and exploitation of data
- a check, on a self-audit basis, on best practice for data management in their field
- a possible route to influence funders, host institutions and infrastructure providers by increasing awareness of the particular needs of research groups.

Immersive case studies are resource intensive; partly as not all those attempted are likely to be successful. The SCARP project has shown it is possible to produce useful case studies which help inform the development of digital curation tools and the appropriate uptake of useful techniques across research domains.

Implications

Case studies of research data practices in context are essential for the development of services to support and develop practice. They should therefore be of continued interest to JISC. However they are also essential to efforts to improve understanding of how scholarly communication is shaped by the new opportunities to reuse published data, and of the appropriate policy measures. Data-focused studies of research and curation practices and

their implications for policy and collaborative tool development are at an early stage, and will interest a broad range of funders with established interests in (for example) Science and Technology Studies, Computer Supported Collaborative Work, and Information Studies.

Such studies, especially those aiming like SCARP to improve the participating groups' curation practices, demand appropriate skills and organisational models. The organisation and deployment of skills in SCARP would probably have been more effective had a 'generalist hub - specialist node' model been used rather than a 'network of generalists'. Identifying and building curation solutions takes the time of specialists in the data management techniques of the domains (and research groups) concerned, perhaps best obtained by paying those specialists to work collaboratively with an informatics or curation generalist, the latter conducting the 'understand and assess' strands and acting as project manager. As well as being more efficient, such a model would also be more effective since time otherwise spent gaining entry and access to case study settings would then be freed towards the analysis and reflection essential for case studies to be of value.

The focus of SCARP was on establishing 'disciplinary differences' in curation practice. Some common aspects were notable across disciplines; researchers' protectiveness towards data they create, and an orientation towards reuse as a means to advance their current research. There was also an indication that early to mid-term researchers are agents for greater use of web technologies to support informal sharing as a means for career development. Aside from complying with data deposit commitments, which were in any case limited, researchers were making data publicly available online as a step towards further opportunities for inter-disciplinary collaboration.

Differences at disciplinary level have been summarised in a separate 'synthesis report', but what characterises SCARP more are the diverse differences found at the sub-disciplinary level. These have several general implications; they indicate the need to establish requirements for curation by engaging specific research communities, rather than presuming a general set of disciplinary characteristics. They also suggest that research communities are better equipped with the contextual understanding of their data that is needed to curate it than centralised institutional data repositories could possibly be.

The importance of interdisciplinary working as a driver of researchers' interest in and need for curation, demonstrates the relevance of curation to "team science"; the trans-disciplinary collaborations that have been correlated with high-impact research⁷, and which are expected to benefit from online collaborative tools. SCARP studies accord with recent findings that 'team science', even in well-established hybrid disciplines such as neuroimaging and video interaction analysis, depends on specific contextual factors to be effective including 'collaborative readiness', trust and a perceived mutual benefit from sharing⁸. Several studies (neuroimaging, video interaction analysis) were only feasible because these factors were already in place at the relatively local level; in one case a departmental research group and the other an informal network of researchers with a shared methodological orientation.

The SCARP Project attempted with some qualified success to apply DCC tools (e.g. the Curation Lifecycle Model) in the context of research 'at the coal face', and provide useful lessons for the tools further development. Case studies in the DCC's Phase 3 need to draw on a range of such tools, or focus on certain aspects of methods such as the Data Audit Framework (DAF). As institutional and cross-institutional services to support research data management are further developed, alongside models and methodologies to underpin that

⁷ Wuchty, S., Jones, B. and Uzzi, B. (2007)

⁸ Stokols, D., Misra, S., Moser, R, Hall, H., and Taylor, B., (2008).

development, these will require assessment of their fit to real-life contexts. As already noted curation needs are likely to be prominent in transdisciplinary cross-institution collaborations, and cross-site case studies should be part of a strategy for engaging research groups. Case studies of Institutional research data service providers themselves would be a valuable source of learning for DCC and the wider curation community, providing an understanding of the challenges, opportunities and constraints that other forms of engagement (surveys, for a) are less likely to yield.

Recommendations

Synthesis

The SCARP Synthesis Study identified the following broad range of recommendations.

JISC

- 1) The JISC Managing Research Data Programme is an important contributor to developing practice in data curation throughout the sector. We strongly recommend that JISC regard this change management initiative as a priority to be developed further as soon as funding permits, and take specific account of discipline issues.
- 2) JISC should commission a study to map a wider range of discipline (or sub-discipline) curation requirements in depth, to inform this further work.

Higher Education and Research funders

- 3) Mandates for data sharing and curation are important and valuable in the cultural change process: such policies should be developed by research funders in consultation with researchers in their various disciplines.
- 4) Formal recognition of the value of curated data as research outputs is necessary and would be an important signal to institutional leaders and researchers. We recommend that all those involved in the assessment of research should be explicit about the position and value of curated data in such assessments.
- 5) We strongly recommend that national-scale research data support services be defined and supported: for example, a development from the UKRDS Feasibility Study, or a service similar in scope to the Australian National Data Service. Such developments need to take specific account of discipline-related issues.

Publishers and Learned Societies

- 6) Publishers and Learned Societies should be more specific about their requirements for lead authors/researchers to be responsible for ensuring the retention of their data, and for making it available for re-use under appropriate conditions.
- 7) Publishers should ensure that Supplementary Data are openly available in forms and under conditions appropriate to the disciplines, and suitable for re-use.
- 8) Publishers should promote direct citation of datasets used in research underlying the articles they publish.
- 9) Learned Societies should use their key position in relation to researchers, as agents for change, and support their disciplinary requirements.

HEIs and Research Institutions

- 10) Institutions have a vital role to play in supporting curation of data produced by their researchers, and in managing the medium to long term retention of some of those data. It is critical that institutions take a lead in this area: all institutions with a research focus have a responsibility to address these issues.
- 11) Institutions should identify how data curation will be supported for their researchers. This could be through local services (such as the Library and IT services), through regional

partnerships, or through external services such as data centres. In taking these decisions, institutions should take account of discipline needs and resources.

- 12) Institutions should explicitly identify funding to support data curation development, infrastructure, staffing and training and make full use of the resources provided by the DCC.

Researchers and Scholars

- 13) Researchers should work with colleagues and their discipline community to develop selection/appraisal criteria to identify priority data.
- 14) Researchers have a responsibility to curate their data; where data supports another research output such as a journal article, persistence may be assured by transferring data to an appropriate data store or archive .
- 15) Researchers should seek to reach consensus on appropriate standards and formats for their discipline or sub-discipline in partnership with relevant professional bodies, in order to enhance data discoverability and re-use.

Case study recommendations

The individual case studies make over 100 recommendations that are mostly specific to the fields or projects they concerned. Four examples are highlighted here because they relate specifically to DCC's ongoing work.

Training, Advice, Self Assessment

- More curation training and advice should be tailored to specific domains, for example in health/medical areas around formulating *data sharing* plans.

Community development/ Advocacy

- Research aims pursued through multidisciplinary collaboration drive data integration and reuse. DCC should be instrumental in transferring good practices between disciplines, bringing together groups with similar research issues and solutions in e.g. de-identifying, licensing, linking, cleaning, documenting, standardising, preserving and visualising data.

Curation resources

- DCC and data support services looking for ways to improve curation practices should consider the existing 'human infrastructure'; for example how research communities' use data for learning and professional development, in what settings and on what conditions.
- DCC should proactively identify research groups with common curation interests and issues, through awareness of developments in instrumentation for data creation and of the tools used to integrate data across disciplinary boundaries. The 'engines of curation' include tools for visualising and browsing data that share a common timeline or occupy the same space.

References

Beagrie, N., Chruszcz, J and Lavoie, B. (2008) Keeping Research Data Safe: a Cost Model and Guidance for UK Universities Higher Education Funding Council for England.
<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>

Corti, L. and Wright, M. (2002). Consultants' Report to the Medical Research Council on the MRC Population Data Archiving and Access Project, UK Data Archive, University of Essex, UK.

Digital Curation and Preservation: Defining the research agenda for the next decade, Warwick Workshop 7/8 November 2005 *Curation Services and Technologies Session Report*. http://www.dcc.ac.uk/training/warwick_2005/Warwick_Workshop_report.pdf

Fry, J. et. al. (2008). Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes. JISC. <http://www.jisc.ac.uk/publications/documents/databenefitsfinalreport.aspx>

Karasti, H. (2001) Bridging work practice and system design - Integrating systemic analysis, appreciative intervention and practitioner participation. *Computer Supported Cooperative Work –an International Journal*, Vol. 10. No 2, pp. 211-246. <http://www.ingentaconnect.com/content/klu/cosu/2001/00000010/00000002/00332819>

Lord, P. & Macdonald, A. (2003). e-Science Curation Report-Data Curation for e-Science in the UK: An Audit to Establish Requirements for Future Curation and Provision. Prepared for the JISC Committee for the Support of Research (JCSR). Twickenham, UK, The Digital Archiving Consultancy Limited. http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf

Lyon, L. (2007). Dealing with Data: Roles, Rights, Responsibilities and Relationships - Consultancy Report. JISC. http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf

Nelson, B. (2009). Data sharing: Empty archives. *Nature* 461, 160-163. <http://dx.doi.org/10.1038/461160a>

RCUK (2009) RCUK Policy and Code of Conduct on the Governance of Good Research Conduct. <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/reviews/grc/goodresearchconductcode.pdf>

Stokols, D, Misra, S, Moser, R, Hall, H, and Taylor, B., (2008). 'The Ecology of Team Science Understanding Contextual Influences on Transdisciplinary Collaboration' *American Journal of Preventive Medicine* 35(2S):S96–S115)

Suchman, L. and Trigg, R. (1991). Understanding Practice: Video as a Medium for Reflection and Design. In J. Greenbaum & M. Kyng (Eds.), *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 65-89

Whyte, A. Conway, E., Neilson, C and Miles, A. (2008) *Methods and Tools for SCARP : SCARP Deliverable B4.2*. Digital Curation Centre.

Wuchty, S, Jones, B and Uzzi, B (2007) 'The Increasing Dominance of Teams in Production of Knowledge' *Science* <http://dx.doi.org/10.1126/science.1136099>

Appendixes

Appendix 1. Interview Topic Guide

The topic guide outlines the scope of interview topics for the Edinburgh case studies. The topics are organized according to 5 main themes. Interviews were semi-structured, and questions varied depending on particular cases and backgrounds of the interviewees. They begin with a set of 'background' topics for first meetings with the host research team.

1. Background

Interviewee's role in research team. Research team's role in organization. Disciplinary background and research experience. Overview of data management activities associated with curation, associated issues & priorities.

1. Could you describe your role in relation to the rest of the team?
2. What about other groups – do you have to liaise with clinicians, or patients, or managers?
3. Could you say a little about your background and how you got involved in the project?
4. What to you are the main aims of the pilot?
5. What is the time span for the project?
6. Who are the main funders?

2. Actual data collection process

Have you heard of 'digital curation'? What do you think it involves, what are the issues?

1. What kinds of *electronic primary data* do you create and/or work with?
2. Could you walk me through what happens to the data from when it is first recorded by the patients – where does it go first, and what happens to it at different stages?
3. What sort of things can affect the data between doing the measurement and the doctor reading it?
4. What kinds of *secondary data* do you work with e.g.
5. Do you reuse data from previous studies?
 - If so, from what sources? For what purposes e.g. meta-analysis, use in teaching materials?
 - If not, is it something you are intending to do in future?

3. Stewardship practices

How do research teams develop shared practices of data curation, sharing, reuse and preservation; and to what extent may similarities and differences in these practices be explained in terms of researchers' alignment with disciplines or domains?

1. Do you regularly use online resources on telecare or COPD? What online resources do you use to find *relevant studies or other literature*? E.g. Bibliographic sources? Websites? Email exchanges with personal contacts? Email lists and newsgroups?
2. What *policies and standards* relating to data management, data sharing or preservation does the research here have to comply with that you know of?
3. Overall, what factors affect your need to curate and preserve research data?
4. What do you see as the *main challenges* to improving how primary data is managed for current and future needs?
5. (For each 'challenge' identified, ask...)
 1. What happens now that needs to change and how is that being addressed?
 2. Who is involved in addressing the situation?
 3. Who would benefit or be otherwise affected (stakeholders)?
 4. What is driving change, or helping changes to go ahead?
 5. What are the main barriers if any?

4. Tools and infrastructure

How are practices of data curation, sharing, reuse and preservation supported with tools and infrastructure, and how might they be better supported?

How does the following affect your handling/use of data?

1. Regulatory compliance
2. Statutory Compliance
3. Educational / Research value
4. Institutional requirements
5. Risk Management
6. Evidential Value
6. Historical value
7. Administrative value

5. Preserving context

What aspects of the context in which data is created and annotated are relevant to preserving its value for future research or learning, and how may this be better supported?

How do you...

1. Manage digital information from its point of creation
2. Promote the reuse of and adding of value to digital information
3. Ensure the long- term accessibility and re-usability of digital information
4. Perform archiving activities such as selection, appraisal and retention
5. Ensure that the authenticity and integrity are maintained over time
6. Perform preservation activities such as migration or emulation

7. Maintain hardware components to enable data to be accessed and understood over time
8. Maintain links between digital information, annotations, and other published materials

6. Policy enablers and barriers

What are the enablers and barriers to adopting the principles, standards and concepts advocated in UK research institutions' data policies and guidelines; and how are those policies and guidelines being informed by current research practice?

1. Does your most recent or usual funding body have policies or guidelines on *data management and preservation* that you are expected to follow, for example when applying for grants or depositing data?
 - a. If so, which policies/guidelines?
2. What do you see as the main obstacles to applying these policies/ guidelines?
3. Who or what do you find helpful in applying them?
 - a. Are mandatory policies helpful in your view to ensuring research data are usable now and in the future?
 - b. How should the work of those involved in managing and preserving data be credited and rewarded?
4. Is it normal practice for projects to be scrutinised by an *ethics committee*? If so, what are they typically concerned with ensuring for the kind of research you are involved in?
5. Do ethical considerations affect data curation and preservation in your view? If so, how?
6. How are *DPA and FoI* issues handled for your research team?
 - a. Can you identify any issues that typically arise with DP or FoI that you would like more support with? If so which, and what kinds of support would you like to see?
7. Are you expected to follow *other policies or guidelines* affecting the use of data– for example from your organization or a professional body?
 - a. What do you see as the main obstacles to applying these policies/ guidelines?
 - b. Who or what did you find helpful in applying them?
8. Are there any external pressures or incentives, e.g. from funders or your institution, to carry out any form of *risk management* of your research data? If so, what form do they take?
 - a. Have you been involved in applying risk management approaches to research data?
 - b. Do you have any views on whether such approaches are helpful or not?
9. Do the *research team/ department* have a written policy on *data quality*?
 - a. If so please briefly describe what it covers and who is responsible for maintaining it (e.g. file formats, description standards, anonymisation, IPR clearance)?

10. What criteria are used and how were they established?
11. What kinds of factor in your experience have helped to *get agreement* on data management/ quality policies?
12. Have other disciplines' practices or perspectives helped, hindered or made no difference to agreeing? Can you give an example?

How will you be able to tell...?

1. That the data you use comes from an appropriate source?
2. That the data has been produced by appropriate methods?
3. What software has been used to process it or analyse it?
4. At what stage in a project would you consider it appropriate to describe the contents of your data to allow others to find it and understand it from a data base
5. From the data you collect, what do you currently record in the data base (if anything)?
6. Who else documents/ indexes your data? (or who ideally should?)
7. At what stage in a project would you consider it appropriate to describe how and why that data has been produced?
8. What do you currently record that would tell others that the data has come form an appropriate source?
9. What software has been used to process or analyse it if any?

At what stage in a project would you consider it appropriate to record administrative details about your data, e.g. about: -

1. Terms of the consent obtained?
2. Confidentiality and anonymisation rules used?
3. Security rules e.g. access controls applicable?
4. Version control?
5. Copyright/ IPR?

Are there any other aspects of the context (e.g. people, documents, events) it is important to record for future reference on why or how the data was produced?

1. If so which aspects?
2. What aspects would you look for in a secondary dataset?
3. Considering the data you are personally responsible for, do you store all of it or a selection/sample?
4. If you select, what criteria do you normally use?
5. Have any standard criteria been documented for appraising and selecting datasets? If so please describe and if possible provide any documentation?
6. What are the backup procedures for your data?

Publishing & Reusing Data

1. Are there particular problems with storing or re-using the kind of telecare data you collect? (e.g. sheer scale of sensor and other data)
2. Are there particular benefits for research or care? (e.g. long term tracking of individual change in response to treatments)
3. Are there issues of consent for using it later?
4. Are there any pressures to reuse? (e.g. from research councils, GPs, managers?)
5. What do you see as the benefits of making it easier to share and reuse telecare data?
6. What would you say are the main barriers to re-using primary data, either for research or teaching and learning?
7. What would you find helpful in the way of support for this process in the future?
8. What needs to happen for telecare data to be shared, in your opinion
9. Have you used any repository service to make your publications available, through an institutional repository (e.g. Edinburgh Research Archive)?
10. Have you any preference for 'institutional repositories' compared with 'domain depositories' as an outlet for your published work? If so which and why? (e.g. institutional focus, local contacts/service vs. specialist knowledge of datasets & analytic techniques, professional networking).
11. Are there any reasons for you to provide an 'audit trail' to link data you have analysed to publish conclusions based on it?
12. If so what reasons are they?
13. Does your current practice include providing such an 'audit trail'?
14. Source repositories contain primary research data. If a standard feature of such repositories was the ability to identify and link to the publications that had been developed from these data, how advantageous would you find it?
15. How advantageous to you would it be if it were possible to go directly from within an online publication (electronic journal article or other text) to the primary source of data from which that publication was developed?
16. Can you take any paper/manuscript you have recently worked on, and say a little about the sources you have cited?
17. What is the basis for their relevance? (e.g. empirical, theoretical, methodological, author reputation, affiliation with research centres etc.)
18. Would you use similar or different criteria to look for datasets?

Appendix 2. Questionnaire Example

The question text below was used in 'Roles and reusability of video data in social science', and delivered via www.surveymonkey.com to University of Edinburgh academics in the Schools of Health in Social Science; Law; Business and Economics; Philosophy, Psychology and Language Sciences; Social and Political Studies; and Education.

Introduction: The Digital Curation Centre's SCARP project is currently conducting a short study to profile current uses of digital video as social research data, i.e. video that has been captured or transferred onto digital files, to be used in the research process. The study explores through desk research, interviews, this survey and a follow-up workshop, how researchers are managing this data and considering its long-term reusability. Digital video material is playing an increasing role in social research across many disciplines, for example in studies of human interaction. The richness, volume and complexity of video data have implications for how UK researchers can be supported to manage this data effectively, maintain access to it and add to its value. The short study explores these issues and broad implications for institutions, research councils, and providers of tools, services and infrastructure, though it is not intended to directly affect UoE provision.

If you have used *any* digital video in you research we would greatly appreciate your response to this questionnaire, which is targeting academics in selected schools in the College of Humanities and Social Sciences. It should take only 10-15 minutes to complete, and you have the opportunity to win a £50 Amazon token. Results will be used to identify the range of practices and views represented; a summary will be published in a SCARP study report, and emailed to respondents if requested. Please note that no data identifying you or your research group will be used to analyse the questionnaire responses, or disclosed without your permission to anyone outside the University of Edinburgh. You can find further information on SCARP at (link to www.dcc.ac.uk/scarp). If you have any queries about the study or problems completing it please email a.whyte@ed.ac.uk.

Your research background

Please identify

1. School
 - (pick from list)
2. What domain or discipline do you associate your research with?
 - (comment)

Video as research data

3. For how long has your research team been using video?
 - More than 3 years, 1-3 years, just starting
4. How important is video data to fulfilling your current research aims?
 - Very important, quite important, useful/interesting but of minor importance
5. Which of the following aspects of your research does video have a role in:-
 - a. Record/ review research practice
 - b. Qualitative description/ content analysis
 - c. Quantitative analysis
 - d. Mixed qual/quant analysis
 - e. Presentation of research results
6. What kinds of other data do you use with video in your analysis?
 - Transcripts or scripts, Codes variables or keywords, Notes or comments, Logs/ time-based data from other sources
 - (other)
7. How often do you use video data from a source outside your own research team?
 - Frequently, occasionally, rarely, never

8. Please briefly describe what is usually depicted in the video you use
 - (comment)
9. Why is video of value to your research, i.e. what properties of video are interesting to you?
 - (comment)

Publishing and reusing video research data

10. What potential benefits would most encourage you to share and reuse video data with other researchers? (please choose as many as apply)
 - a. Describing situations of interest to my/others' research
 - b. Re-analysis using different perspectives or methods
 - c. Scrutiny of the links between data and analysis
 - d. Comparative analysis across studies
 - e. Potential for follow-up studies
 - f. Reflection on/ development of the practices recorded
 - g. Value for teaching and learning
 - h. Raise my research profile/ funding prospects
 - i. Other (please specify)

11. What potential risks or costs would most discourage you from sharing your video data for reuse by others? (please choose as many as apply)
 - a. Loss of ownership
 - b. Data protection and other confidentiality constraints
 - c. Consent of subjects/participants and other ethical constraints
 - d. Representing details of the context of the video data
 - e. Representing details of the research methods
 - f. Subversion of intellectual property rights, including copyright:
 - g. Time/effort required to share video
 - h. Lack of infrastructure for sharing video
 - i. Premature broadcast of research findings
 - j. Other (please specify)

12. Which of these are challenges to using and managing video research data that you are most interested in following up? (please choose as many as apply)
 - a. Planning or managing the research data lifecycle
 - b. Using metadata standards to describe or tag video content
 - c. Selecting file formats and media
 - d. Linking video with other data for analysis
 - e. Planning for change and obsolescence in video capture technology
 - f. Archiving video data for post-project access
 - g. Data protection, confidentiality and secure access
 - h. IPR e.g. copyright and licensing issues
 - i. Web sharing/publication: making video data accessible and searchable

Please indicate how far you agree with each of the statements below

(Strongly agree, Agree, Neither, Disagree, Strongly disagree, Don't know)

13. Better infrastructure is needed for shared access to video datasets across HE institutions
14. I would consider re using video data shared online
15. I would consider sharing video data on a website aimed at promoting its reuse for research purposes.

16. Would you be interested in attending a workshop in the next 2-3 months, on issues in video research data curation and reuse?

- Definitely/ Possibly/ Not at this stage

17. Are you willing to take part in a short follow-up interview about your research process? (subject to your availability, to be confirmed by e-mail)

- Yes/ Not at this stage

18. Would you like to be emailed a copy of the study report when completed?

- Yes/ No

Prize draw for £50 Amazon.com tokens.

If you would like to be entered please complete the sentence below. The winner will be drawn from the most insightful or interesting!

“Digital curation” means... ? (comment box)

Please provide your name and email address if

- you have indicated we can contact you again, or
- you have entered the £50 token draw, or
- you would like a summary of the results

Contact details will not be used for any other purpose or disclosed to non-UoE staff without your further explicit permission.

First name

Second name

Email address

Thank you very much for your time.