

Dealing with Data: One Year On

Dr Liz Lyon, Director, UKOLN, University of Bath, UK
Associate Director, UK Digital Curation Centre

Max Planck eScience Seminar, Gottingen, June 2008.



This work is licensed under a Creative Commons Licence
Attribution-ShareAlike 2.0

UKOLN is supported by:



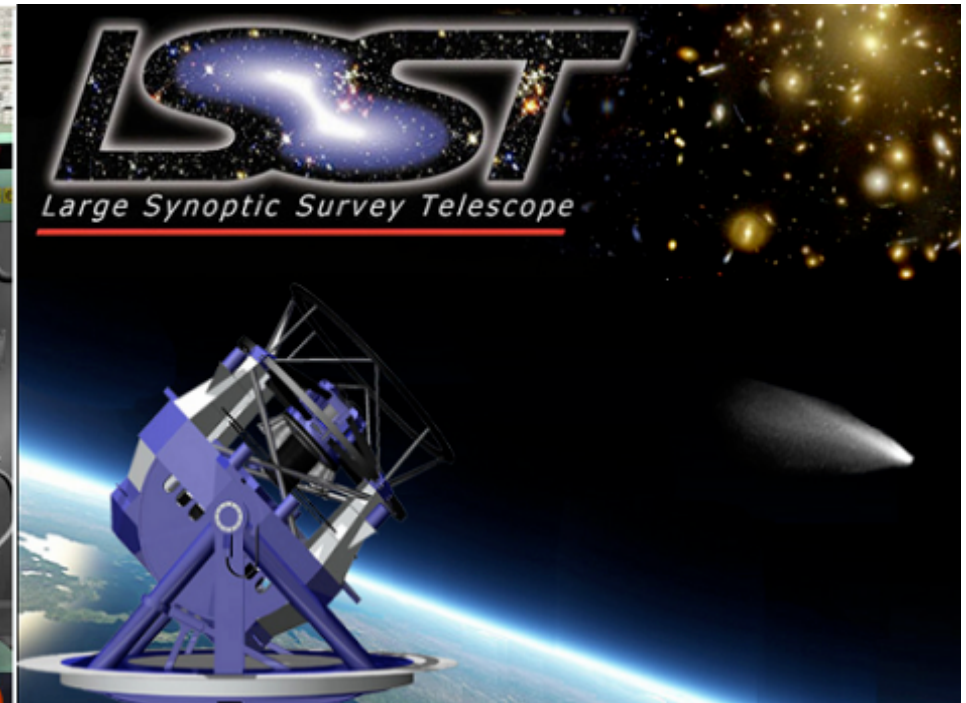
www.ukoln.ac.uk

A centre of expertise in digital information management



Overview

1. Open Science : a changing landscape
2. *Dealing with Data*: What has been achieved?
3. Reflections and future challenges



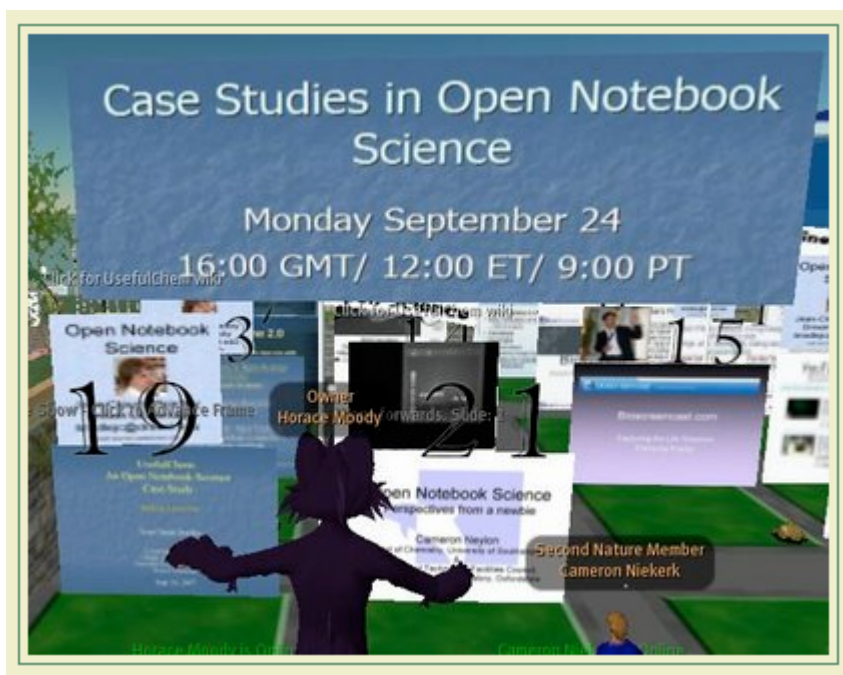


Open Science



....is happening now

- Blogging of results data
- Community “repositories” for data
- Open Notebook Science (ONS)
- Open grant proposals
- Drexel Grand Challenge bid to Gates Foundation



explore

visualizations
data sets
comments
topic hubs

participate

register
create visualization
upload data set
create topic hub

learn more

quick start
visualization types
data format & style
about Many Eyes
FAQ

Data Sets : Global city data



Pre-publication research and preliminary findings

Home | Browse by subject | Advanced search | Search: All documents

Browse by subject | Welcome to Nature Precedings



Home | Site Map | Download | FAQ | Links | Genome Browser

Community Annotation of Rice Gene Families

article | talk | edit | history

ChIP-Chip E. coli

This protocol is in developement and currently under investigation in the lab. Please contibute and watch changes in the nex weeks!

Contents [show](#)

Curators

Torsten Waldminghaus 06:29, 17 March 2008 (CDT)

Anyone should feel free to add themselves as a curator for this consensus protocol. You do not need to be a curator in order to contribute. The OpenWetWare community is currently [discussing the idea of protocol curators](#). Please contribute.

GALAXY ZOO.org

[Welcome](#)[Home](#)[The Science](#)[How to Take Part](#)[Galaxy Analysis](#)[Forum](#)[Press](#)[Blog](#)[FAQ](#)[Links](#)[Contact Us](#)[Login](#)[Register](#)

Dear Galaxy Zoo users,

Thanks for making Galaxy Zoo such a success!

With your help, we've been able to collect millions of classifications, with which to do science faster than we ever thought possible. We are currently preparing the first science papers for submission to peer-reviewed journals and we will keep you posted on the progress of the papers on the [BLOG](#) and the [FORUM](#).

From now on, if you classify galaxies on the [ANALYSIS](#) page, your classifications will continue to be recorded and will be part of the public release, but it won't be part of the first round of papers.

But we still need you! Galaxy Zoo 2 will go live in the near future featuring a much more detailed classification system, while further off we plan GalaxyZoo 3 with lots of exciting new data. We'll notify all of you via the newsletter when we're able to start these two new endeavours.

[Log In](#)

User Name:

Password:

Citizen science: Scientists collaborating with the public



Collective intelligence

- Today: “rate and recommend”, aggregations, comments, tags, annotations, ratings, reviews, opinion
- Tomorrow: “collective intelligence” to analyse, assess, mine, extract, evaluate....

We need to ensure that this collective intelligence is preserved in the long-term

The image shows a screenshot of the ThoughtMesh website. The header features the ThoughtMesh logo and the tagline "Tag your writing. Join the conversation." Below the header are four navigation buttons: "what", "how", "why", and "who". The main content area is divided into two columns. The left column contains text about the platform's purpose and instructions on how to navigate essays and tag them. The right column features a conceptual map titled "Recently meshed spaces" showing connections between "Picasso", "Cubism", and "1900". The map includes several text boxes with descriptions of the concepts and their relationships.

ThoughtMesh beta
Tag your writing.
Join the conversation.

what how why who

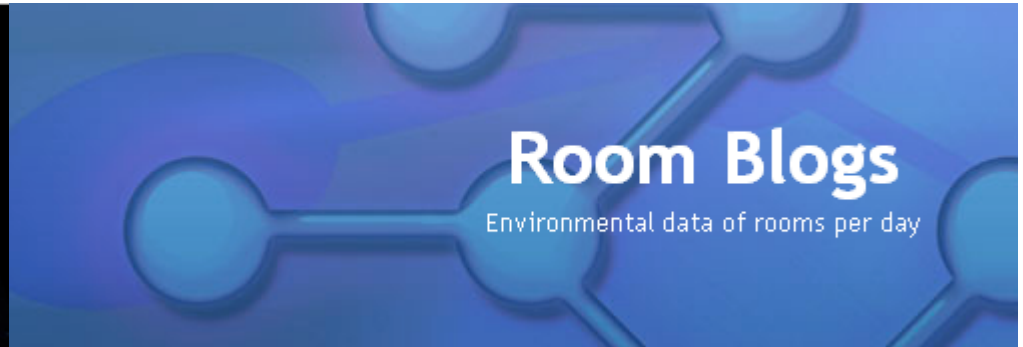
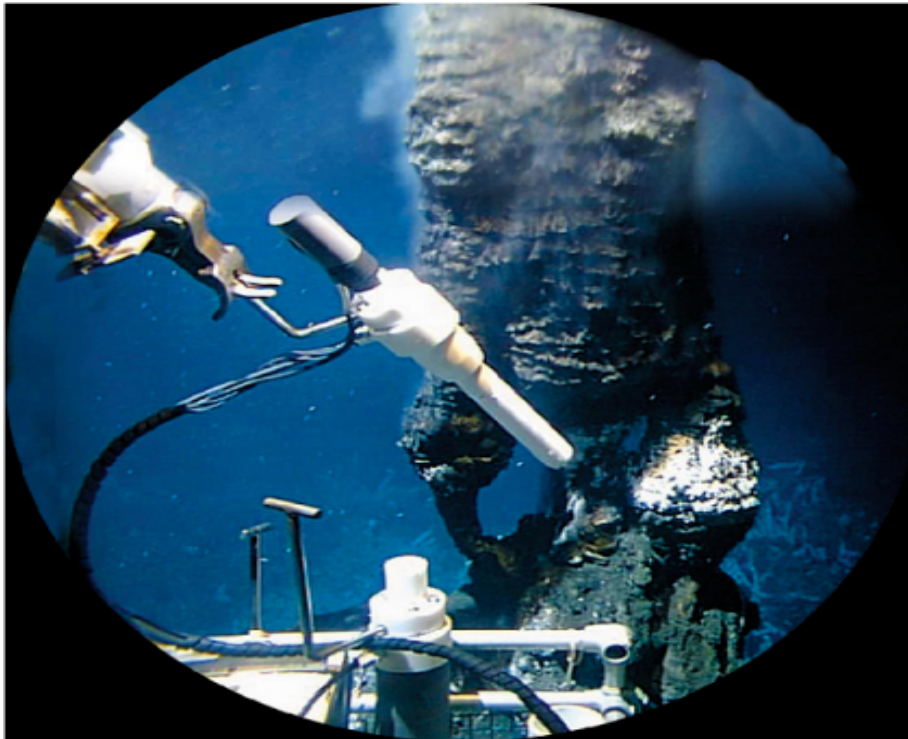
ThoughtMesh is an trusted mode for publishing and discovering scholarly papers online. It gives readers a tag-based navigation system that uses keywords to connect excerpts of essays published on different web sites.

How to navigate essays
View a **dermu** or how to use tags to navigate through a single essay and across the mesh (opens in a new window).

How to tag an essay
View a **dermu** or how to add your own writing to the mesh (opens in a new window).

Recently meshed spaces

Picasso
Cubism
1900



27:1043 - 2008-4-7

8th April 2008 @ 00:00

Auto blog

Data collected from 27:1043 on 2008-4-7

Video journalist and lifecaster [Sarah Meyers](#) with [Justin.tv](#) founder Justin Kan in photo by Brian Solis at [DoubleClick's](#) April 26, 2007 [ad:tech party](#) in San Francisco. Wearing the capcam, Kan was lifecasting at that event.

Sensors

Capcam

Blogs

Capture & 'cast...



*“National Nuclear Security
Administration (NNSA) announces
5 new Centers of Excellence
focussing on the emerging field of
Predictive Science.”*

US DoE \$17million grant to each Center
Simulations of hypersonic flight, supernovae....

7th March, 2008

Content as infrastructure

- Today: primary data, images, text
- Tomorrow: digests, simulations, models
- Today: discovery to delivery
- Tomorrow: mine & model, simulate & synthesise
- Today: statistics
- Tomorrow: Predictive science

We need data verification & validation methodologies to ensure data quality in trusted archives to enable predictive science....



“London Polyclinic” Imperial College / Nat Phys Lab

Mixed reality environments

- Research and learning applications
- Opportunities for participative exploration
- Rich test-bed for experimentation
- Mimic, innovate and extend
- Immerse and experience
- Ubiquitous? Pervasive? Persistent?

We need to ensure that these virtual worlds are curated and preserved.....

Data Curation and Preservation choices?

1. Disciplinary data centre
2. Institutional / departmental / lab repository
3. Repository federation or network
4. National library or national archive
5. “Public” data repository or service
6. Web archiving services
7. Commercial data store - Amazon S3
8. Ecosystem of hosted lifebits services (Jon Udell)
9. None of these?
10. All of these?



EMBL-EBI



About BAS

Antarctic Environmental Data Centre

What has been achieved?



Dealing with Data: Roles, Rights, Responsibilities and Relationships Consultancy Report

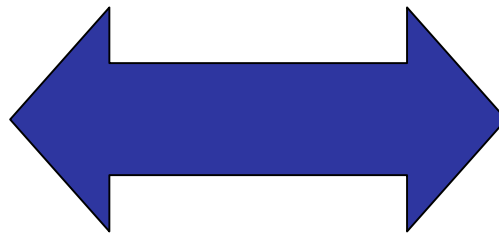
Document details

| | |
|----------------|--|
| Author: | Dr Liz Lyon, UKOLN, University of Bath |
| Date: | 16 th June 2007 |
| Version: | V1.0 |
| Document Name: | data-consultancy-report-final.doc |
| Notes: | |

UKOLN Liz Lyon June 2007

35 Recommendations for JISC

Roles, Rights, Responsibilities,
Relationships: scientist,
institution, data centre, user,
funder, publisher



Stewardship of digital research data: a framework of principles and guidelines

Responsibilities of research institutions
and funders, data managers, learned
societies and publishers

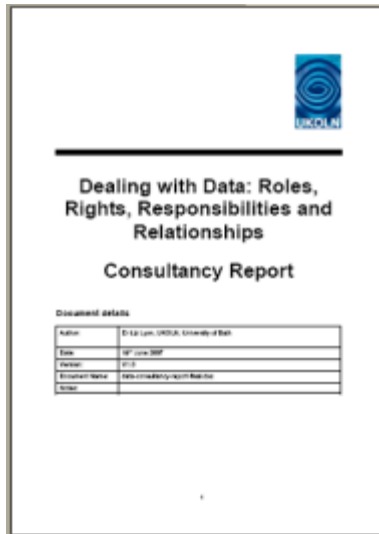
*Research Information Network
RIN January 2008*

5 Principles: Roles &
responsibilities, standards &
QA, access, usage & credit,
benefits & cost-effectiveness,
preservation & sustainability

January 2008



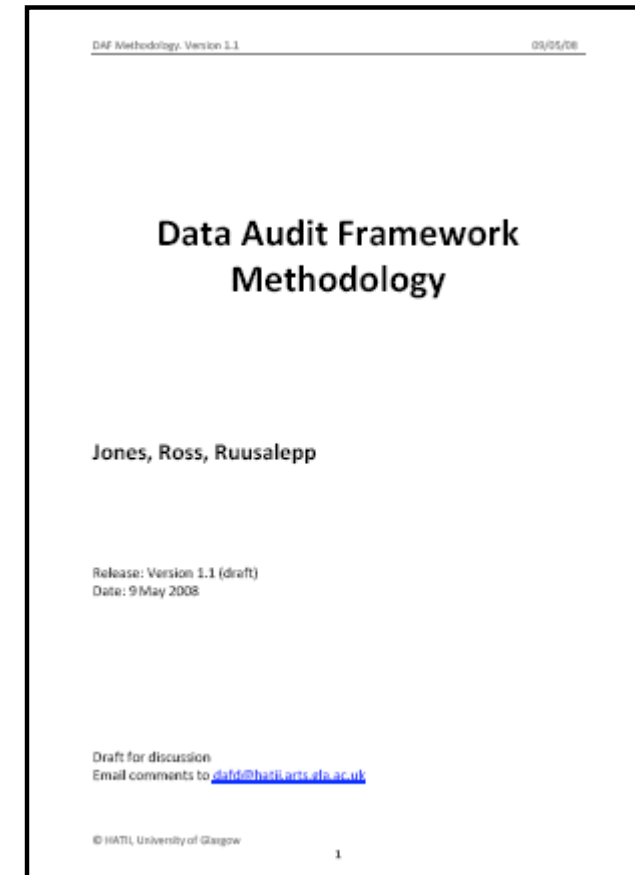
Report Recommendations 1



- *DataSets Mapping and Gap Analysis (UK)*
- *Data Curation & Preservation Strategy (UK)*
- ***Rec 4 Data Audit Framework (HE Institutions)***
- *Institutional Data Management, Preservation & Sharing Policy*
- *Data Management & Sharing Policy (Funders)*
- *Data Management Plan (Projects)*
- *Data Networking Forum (People)*

Data Audit Framework (DAFD)

- JISC funding
- HATII, University of Glasgow
- Draft Methodology V1.1
- Audit case studies:
 - Univ Glasgow (archaeology)
 - Univ Bath (engineering)
 - Kings College (bio-informatics)
 - Univ Edinburgh (geosciences)
- Pilots:
 - Univ Edinburgh
 - Imperial College
 - Kings College
 - UCL
- Online tool development

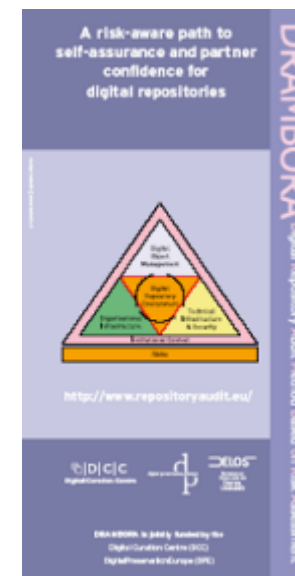
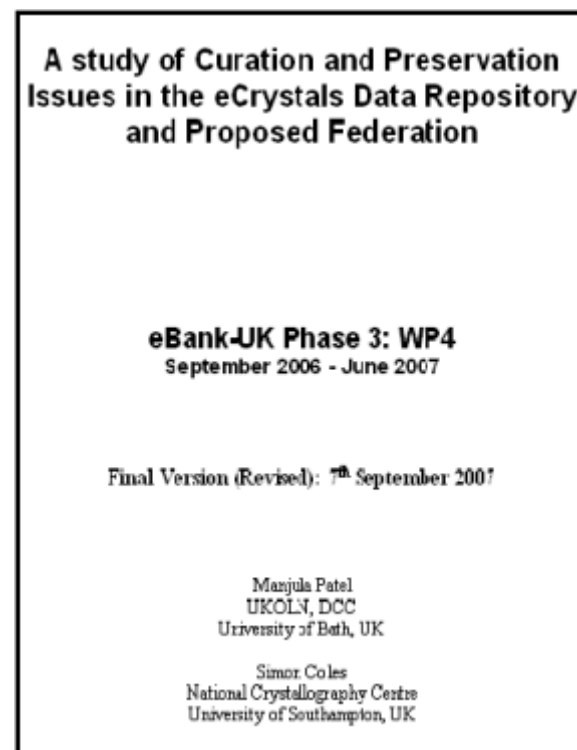


eCrystals Curation & Preservation Study



Working with the Digital Curation Centre
Examined four main areas

1. Audit and certification (TRAC, DRAMBORA, NESTOR, ISO International repository audit and certification BOF Group)
2. The Open Archival Information System (OAIS) and Representation Information (RI)
3. eBank-UK application profile and preservation metadata
4. ePrints.org repository platform



[http://www.ukoln.ac.uk/projects/ebank-uk/curation/eBank3-WP4-Report%20\(Revised\).pdf](http://www.ukoln.ac.uk/projects/ebank-uk/curation/eBank3-WP4-Report%20(Revised).pdf)

Recommendations

eCrystals Federation: Preservation & sustainability Recommendations

Data repositories

- Use DRAMBORA Interactive for self-assessment
- Add PREMIS preservation metadata
- Collect eCrystals representation information
- Examine repository platform conformance to OAIS Reference Model
- Survey partner preservation policies

DRAMBORA
Digital Repository Audit Method Based on Risk Assessment

Logged in as: admin
Last Login: 02 Oct 2007
[Log Out](#)

DRAMBORA Online Tool :: Audit Centre :: Identify Risks

[Home](#) | [Select Functional Classes](#) | [Define Mandate](#) | [Define Objectives](#) | [Identify Constraints](#)
[Identify Activities, Assets and Owners](#) | [Identify Risks](#) | [Assess Risks](#) | [Manage Risks](#) | [Identify risks](#)

Select existing risk
Define new risk

Risk Name:

Risk Description:

Corresponding Objectives:

Corresponding Activities:

Functional Classes:

Identified risks

Interactive

- Management: Fair and...
- Loss of trust or reputation...
- Activity is overlooked or el...
- Business objectives not met...
- Repository loses mandate...
- Community requirements change...

...or

***Digital Curation Centre
partnership***

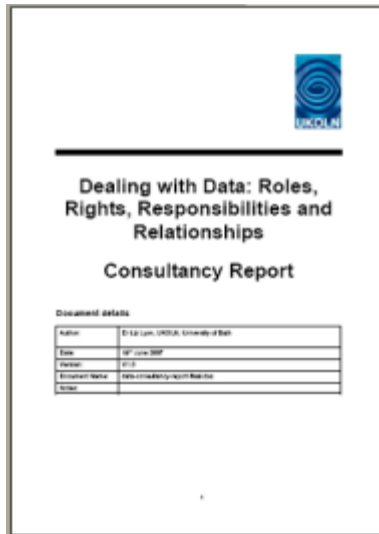


Shared Research Data Service Feasibility Study



- HEFCE award £255K via Research Libraries UK and Russell Group Universities IT Directors to SERCO
- Objectives:
 - Develop understanding of UK's current and future research data service needs
 - Work with other UK stakeholders to identify priorities for action
 - Develop a number of scenarios/options for the shared service from “do nothing” to a managed national service
 - Develop a detailed business plan for the preferred option(s)
 - Include assessment of costs and benefits in options appraisal
 - Indicate both scale of investment required & an estimate of likely ROI
 - Present outline governance and management proposals for the preferred option(s)
- 4 case study “volunteers”: Bristol, Leeds, Leicester and Oxford
- Report January 2009

Report Recommendations 2



- *DataSets Mapping and Gap Analysis (UK)*
- *Data Curation & Preservation Strategy (UK)*
- *Data Audit Framework (HE Institutions)*
- *Institutional Data Management, Preservation & Sharing Policy*
- *Data Management & Sharing Policy (Funders)*
- *Data Management Plan (Projects)*
- ***Rec 5 : Data Networking Forum (People) linked to RIN Framework Principle 1***



Research Data Forum



- March 2008, Manchester *<http://www.dcc.ac.uk/data-forum/>*
- Joint DCC – RIN event
- Data centre managers, IR managers, funders & policy makers
- Aims & Objectives:
 - Improve data acquisition, management, analysis, validation, archiving and dissemination
 - Increase awareness of national & international data policies and standards
 - Facilitate co-operation between organisations and individuals
 - Exchange experience and best practice
- Next meeting in November in Birmingham, UK (tbc)

Heard at the Forum....

“protected by PDF”

“Rembrandt in the attic”

“Don’t forget the researcher!”

“stuff isn’t getting done”

“demand outstrips supply...”

“careers developed more by luck than judgement”

“Data managers as failed scientists”

“need to sit down and write the manual”

“teeth and sticks and carrots”

“professionalising data management”

“Data is not just about eScience/eResearch”

“we need services not projects!”

4th International Digital Curation Conference
"Radical Sharing: Transforming Science?"
1-3 December 2008
Hilton Edinburgh Grosvenor Hotel
Edinburgh, Scotland

Keynotes: David Porteous,
Generation Scotland, John
Wilbanks, Science Commons,
Martin Lewis, RLUK, Malcolm
Atkinson, NeSC

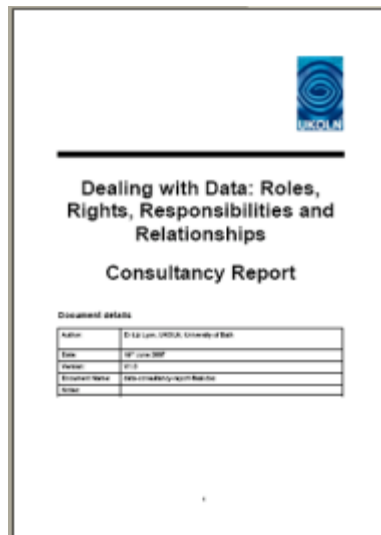
Sessions: Sustainability,
Privacy issues, collaborative
approaches to data sharing

Call for Papers: submit now!

Developing the curation community



Recommendations 3 Digital Curation Centre



- *Co-ordinated advocacy programmes*
- ***Rec 33 Co-ordinated training programmes***
- *Disciplinary Data Case Studies (SCARP)*

Roles, Rights
Responsibilities
Relationships:

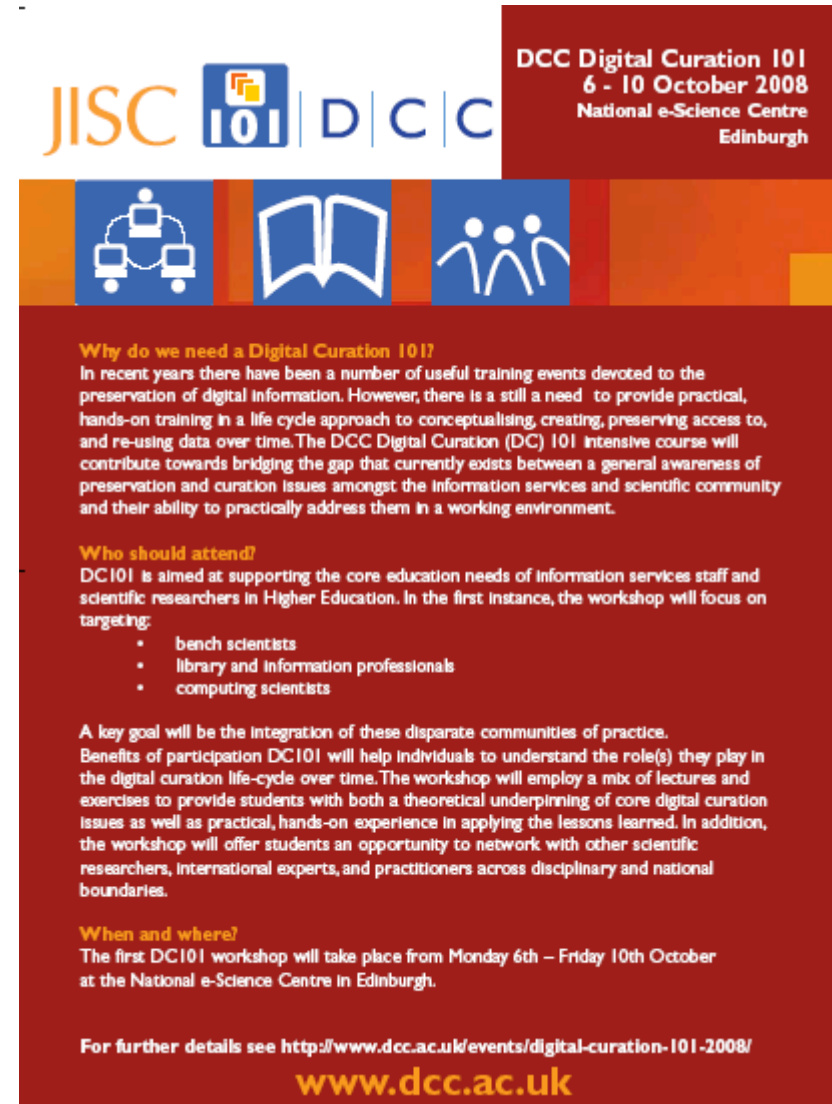
- *scientist*
- *institution*
- *data centre*
- *user*
- *funder*
- *publisher*

DEALING WITH DATA

| Role | Rights | Responsibilities | Relationships |
|--|---|--|--|
| Scientist: creation and use of data | Of first use. To be acknowledged. To expect IPR to be honoured. To receive data training and advice. | Manage data for life of project. Meet standards for good practice. Comply with funder / institutional data policies and respect IPR of others. Work up data for use by others. | With institution as employee. With subject community. With data centre. With funder of work. |
| Institution: curation of and access to data | To be offered a copy of data. | Set internal data management policy. Manage data in the short term. Meet standards for good practice. Provide training and advice to support scientists. Promote the repository service. | With scientist as employee. With data centre through expert staff. |
| Data centre: curation of and access to data | To be offered a copy of data. To select data of long-term value. | Manage data for the long-term. Meet standards for good practice. Provide training for deposit. Promote the repository service. Protect rights of data contributors. Provide tools for re-use of data. | With scientist as "client". With user communities. With institution through expert staff. With funder of service. |
| User: use of 3 rd party data | To reuse data (non-exclusive license). To assess quality metadata to inform stability. | Abide by license conditions. Acknowledge data creators / owners. Manage derived data effectively. | With data centre as supplier. With institution as supplier. |
| Funder: withdrawal to public policy driven | To implement data policies. To require those they fund to meet policy obligations. | Consider wider public-policy perspective & stakeholder needs. Participate in strategy co-ordination. Develop policies with stakeholders. Participate in policy co-ordination, joint planning & fund service delivery. Monitor and enforce data policies. Resource post-project long-term data management. Act as advocate for data curation & fund expert advisory service(s). Support workforce capacity development of data curators. | With scientist as funder. With institution. With data centre as funder. With other funders. With other stakeholders as policymaker and funder of services. |
| Publisher: maintain integrity of the scientific record | To expect data are available to support publications. To request pre-publication data deposit in long-term repository. | Engage stakeholders in development of publication standards. Link to data to support publication standards. Monitor & enforce public standards. | With scientist as creator, author and reader. With data centres and institutions as suppliers. |

DCC Digital Curation 101

- Digital Curation Centre
- 6-10 October 2008
- National eScience Centre, Edinburgh
- Intensive course
- Lectures + hands-on
- Target participants: bench scientists, LIS professionals, computational scientists
- Survey questionnaire
http://www.dcc.ac.uk/jisc/data_projects_questionnaire/

The poster for DCC Digital Curation 101 features a red background with a horizontal band of icons (network, book, people) in the center. Text is arranged in columns, providing details about the event's purpose, target audience, and logistics. Logos for JISC and DCC are at the top.

JISC **DCC** **DCC Digital Curation 101**
6 - 10 October 2008
National e-Science Centre
Edinburgh

Why do we need a Digital Curation 101?
In recent years there have been a number of useful training events devoted to the preservation of digital information. However, there is still a need to provide practical, hands-on training in a life cycle approach to conceptualising, creating, preserving access to, and re-using data over time. The DCC Digital Curation (DC) 101 intensive course will contribute towards bridging the gap that currently exists between a general awareness of preservation and curation issues amongst the information services and scientific community and their ability to practically address them in a working environment.

Who should attend?
DC101 is aimed at supporting the core education needs of information services staff and scientific researchers in Higher Education. In the first instance, the workshop will focus on targeting:

- bench scientists
- library and information professionals
- computing scientists

A key goal will be the integration of these disparate communities of practice. Benefits of participation DC101 will help individuals to understand the role(s) they play in the digital curation life-cycle over time. The workshop will employ a mix of lectures and exercises to provide students with both a theoretical underpinning of core digital curation issues as well as practical, hands-on experience in applying the lessons learned. In addition, the workshop will offer students an opportunity to network with other scientific researchers, international experts, and practitioners across disciplinary and national boundaries.

When and where?
The first DC101 workshop will take place from Monday 6th – Friday 10th October at the National e-Science Centre in Edinburgh.

For further details see <http://www.dcc.ac.uk/events/digital-curation-101-2008/>
www.dcc.ac.uk

JISC-PoWR

Preservation of Web Resources: a JISC-sponsored project

[HOME](#)[ABOUT](#)[CONTACT](#)[WORKSHOPS](#)[TEAM AREA](#)

PoWR Themes

[Digital Preservation](#)[Events](#)[Preservation](#)[Project News](#)[Records Management](#)[Selection](#)[Web 1.0](#)[Web 2.0](#)[Workshops](#)

Recent Posts

[Whose Responsibility is Web Resource Preservation?](#)

[When Domains Go Missing](#)

[Workshop 1 - Preservation of Web Resources: Making a Start](#)

Whose Responsibility is Web Resource Preservation?

Posted by [Marieke Guy](#) on June 6th, 2008

It is possible that one of the reason why so little is being done about Web resource preservation is that everybody feels it is somebody else's responsibility. It might be very easy for us all to avoid the issue by standing back and waiting for someone else to tackle what, we have already explained, is a very **complex problem**. However taking this approach may mean that nobody does anything and we all lose out.

So *whose* responsibility is Web resource preservation then?

There are a number of parties who may have an interest in the preservation of Web resource. These range from the international institutions down to the individual.

Individual Approach

- The author of the resource - It may be important for individuals to retain a copy of papers, articles and other written works for future use, CVs etc.

Institutional Approach

Workshop 1 - Preservation of Web Resources: Making a Start

Posted by [Marieke Guy](#) on 2nd June 2008

Bookings are now open for the first JISC-PoWR workshop to be held in London on Friday 27th June 2008. The workshop entitled *Preservation of Web Resources: Making a Start* is free and open to Web, information and records managers working in HE/FE Institutions.

For information on how to reserve a place see the [workshop 1 page](#).

Calendar

June 2008

| M | T | W | T | F | S | S |
|----|----|----|----|----|----|----|
| | | | | | | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | | | | | | |

[« May](#)

Partner blogs

[Opencontentlawyer](#)
[UK Web Focus](#)
[ULCC DA Blog](#)

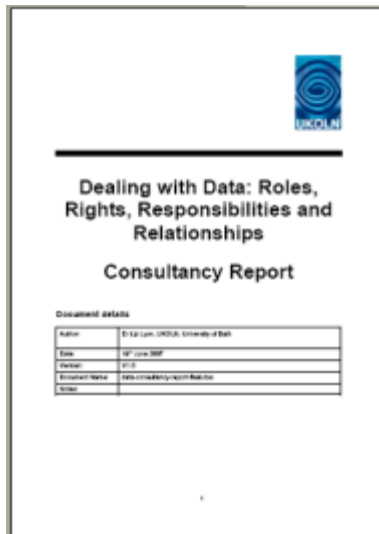
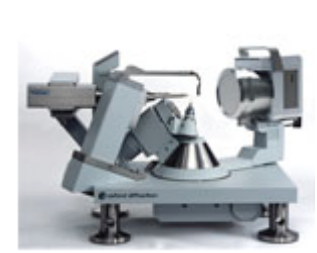
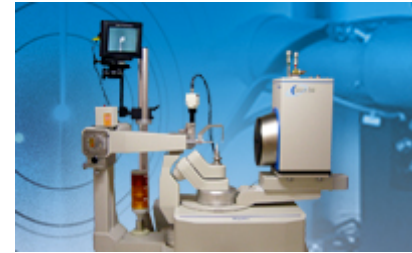
Partners

[JISC IIE](#)
[UKOLN](#)
[ULCC](#)

JISC-PoWR

<http://jiscpowr.jiscinvolve.org/>

Report Recommendations 4



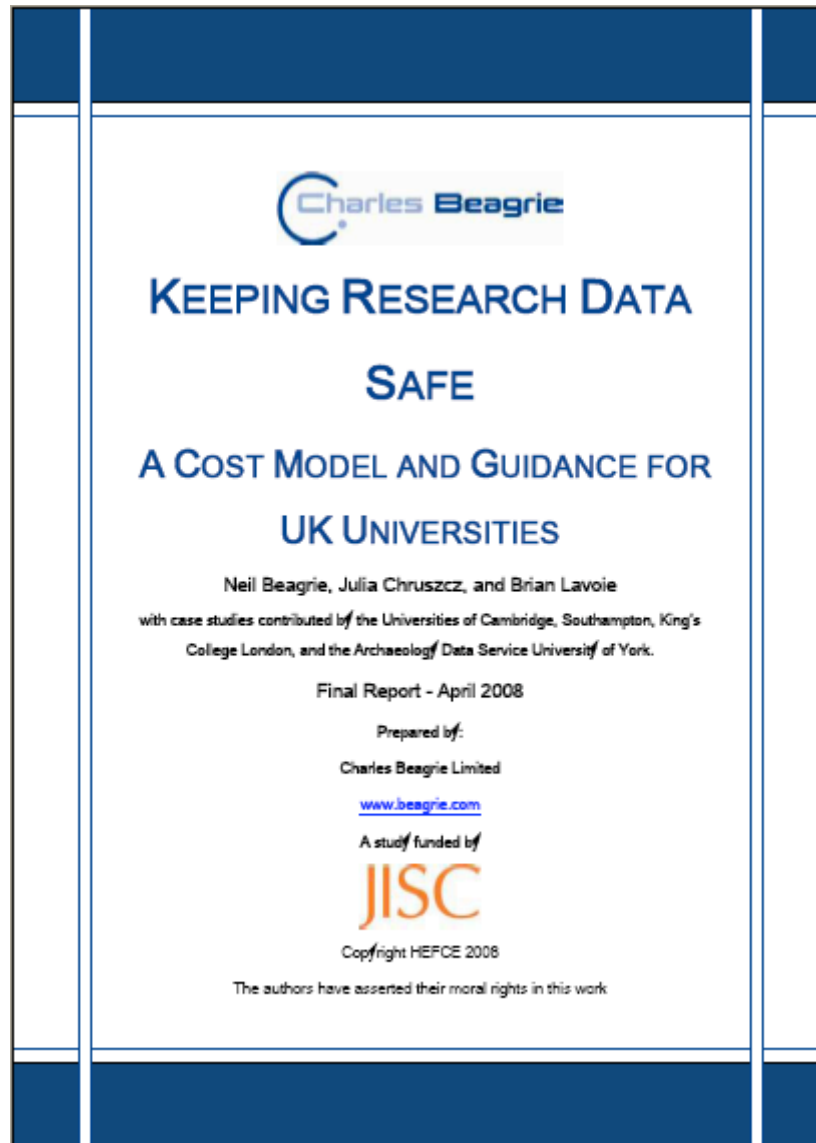
- *Instrumentation and laboratory equipment*
- *Dataset re-use: significant properties*
- *Versions, identifiers, citation*
- *Robust bi-directional linking*
- *IPR and model licences for data*
- ***Rec 34 Careers, specialist skills, capacity***
- ***Rec 35 Data curation in the curriculum***
- ***Rec 30 Cost-benefits of data curation***



JISC Curation Careers study

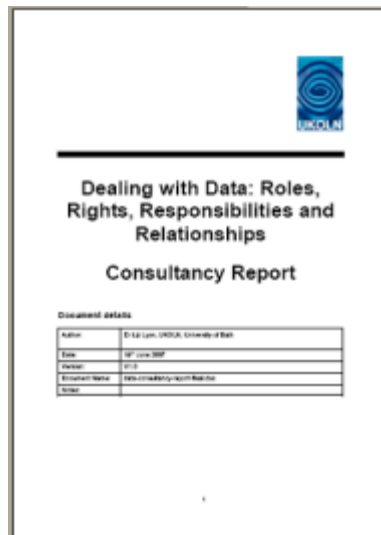
- Key Perspectives (Alma Swan)
 - Skills, role and career structure of data scientists and curators: an assessment of current practice and future needs
1. Training: LIS and informatics schools curricula
 2. Career structures, pathways, rewards: interviews / focus groups with scientists, data scientists, Survey questionnaire: collaborating with DCC and Curation 101 Programme.
 3. Establish skills needed: 2 case studies (rural & economic land use and systems biology), interviews with academic librarians, research funder reps, data centre managers

JISC Preservation costs study



- Neil Beagrie, Julia Chruszcz, Brian Lavoie, April 2008
- Overview of benefits, issues and service models
- Costing framework
- *Presentation to follow*

Recommendations 5 Digital Curation Centre



- *Co-ordinated advocacy programmes*
- *Co-ordinated training programmes*
- **Rec 11 Disciplinary Data Case Studies (SCARP)**

Roles, Rights
Responsibilities
Relationships:

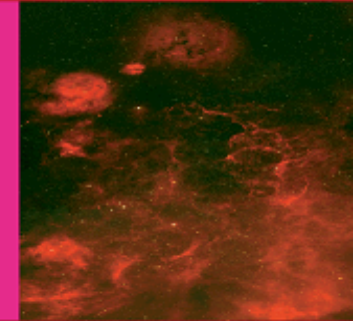
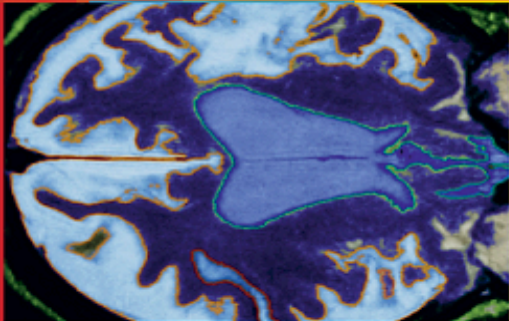
- *scientist*
- *institution*
- *data centre*
- *user*
- *funder*
- *publisher*

DEALING WITH DATA

| Role | Rights | Responsibilities | Relationships |
|--|---|--|--|
| Scientist: creation and use of data | Of first use. To be acknowledged. To expect IPR to be honoured. To receive data training and advice. | Manage data for life of project. Meet standards for good practice. Comply with funder / institutional data policies and respect IPR of others. Work up data for use by others. | With institution as employee. With subject community. With data centre. With funder of work. |
| Institution: curation of and access to data | To be offered a copy of data. | Set internal data management policy. Manage data in the short term. Meet standards for good practice. Provide training and advice to support scientists. Promote the repository service. | With scientist as employer. With data centre through expert staff. |
| Data centre: curation of and access to data | To be offered a copy of data. To select data of long-term value. | Manage data for the long-term. Meet standards for good practice. Provide training for deposit. Promote the repository service. Protect rights of data contributors. Provide tools for re-use of data. | With scientist as "client". With user communities. With institution through expert staff. With funder of service. |
| User: use of 3 rd party data | To reuse data (non-exclusive license). To assess quality metadata to inform stability. | Abide by license conditions. Acknowledge data creators / owners. Manage derived data effectively. | With data centre as supplier. With institution as supplier. |
| Funder: withdrawal to public policy driven | To implement data policies. To require those they fund to meet policy obligations. | Consider wider public-policy perspective & stakeholder needs. Participate in strategy co-ordination. Develop policies with stakeholders. Participate in policy co-ordination, joint planning & fund service delivery. Monitor and enforce data policies. Resource post-project long-term data management. Act as advocate for data curation & fund expert advisory service(s). Support workforce capacity development of data curators. | With scientist as funder. With institution. With data centre as funder. With other funders. With other stakeholders as policymaker and funder of services. |
| Publisher: maintain integrity of the scientific record | To expect data are available to support publications. To request pre-publication data deposit in long-term repository. | Engage stakeholders in development of publication standards. Link to data to support publication standards. Monitor & enforce public standards. | With scientist as creator, author and reader. With data centres and institutions as suppliers. |



| D | C | C



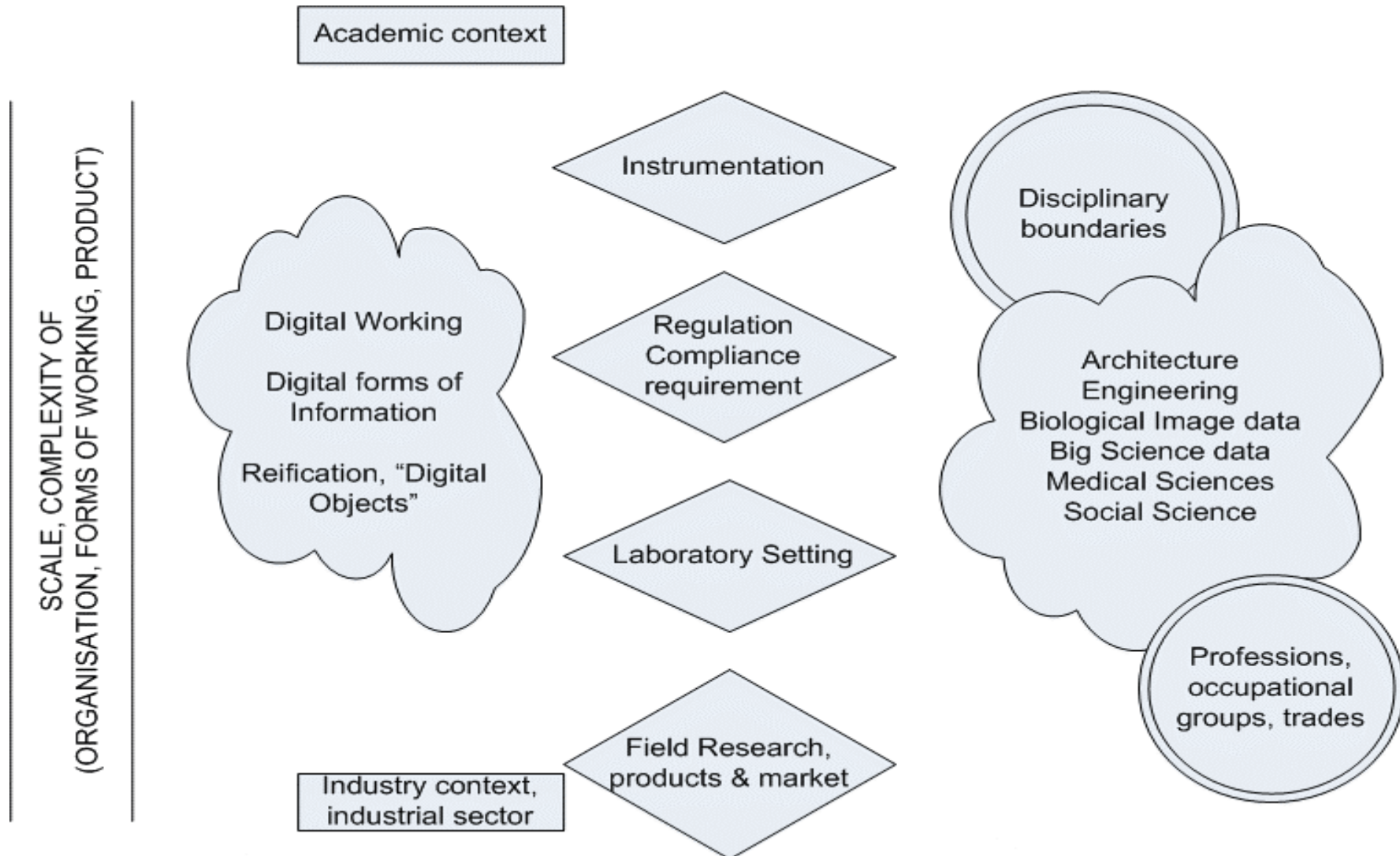
DCC SCARP PROJECT

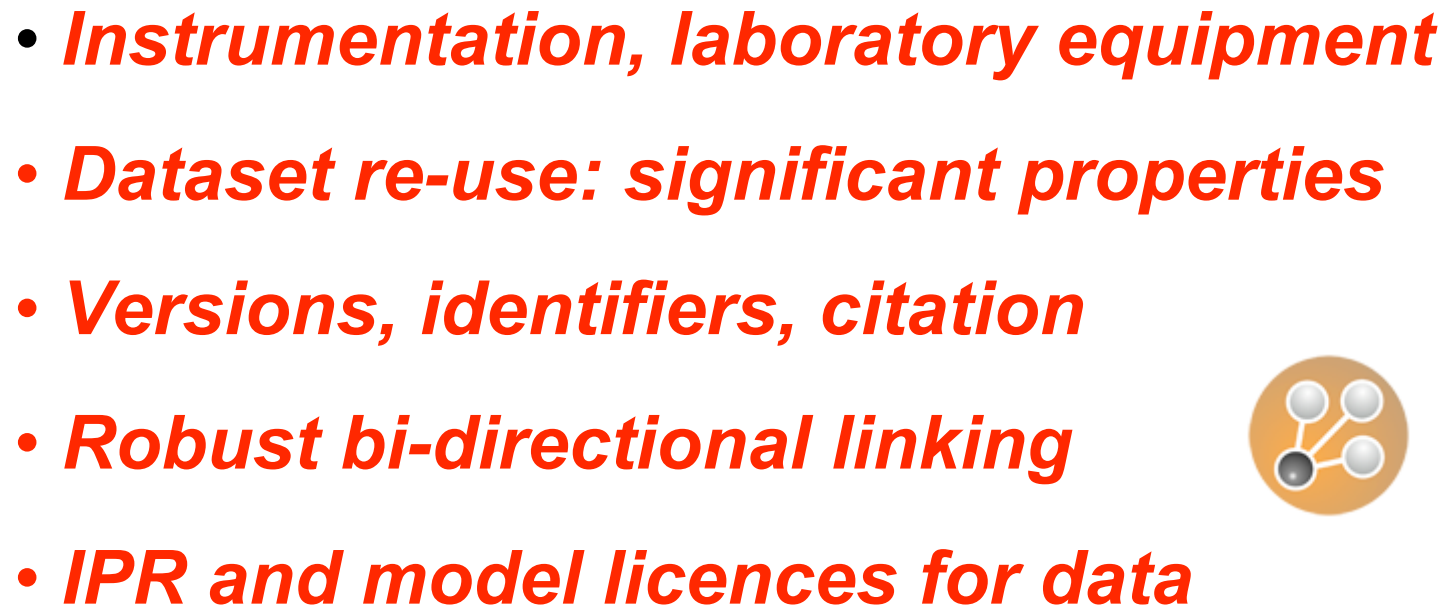
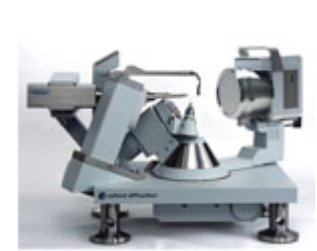
- Immersive approach to case studies
- Disciplinary factors in curating Architectural Research (Colin Neilson)
- Curating Brain Images in a Psychiatric Research Group (Angus Whyte)
- Curating earth observation data (Esther Conway)
- www.dcc.ac.uk/scarp/



D | C | C

Factors looked at by SCARP





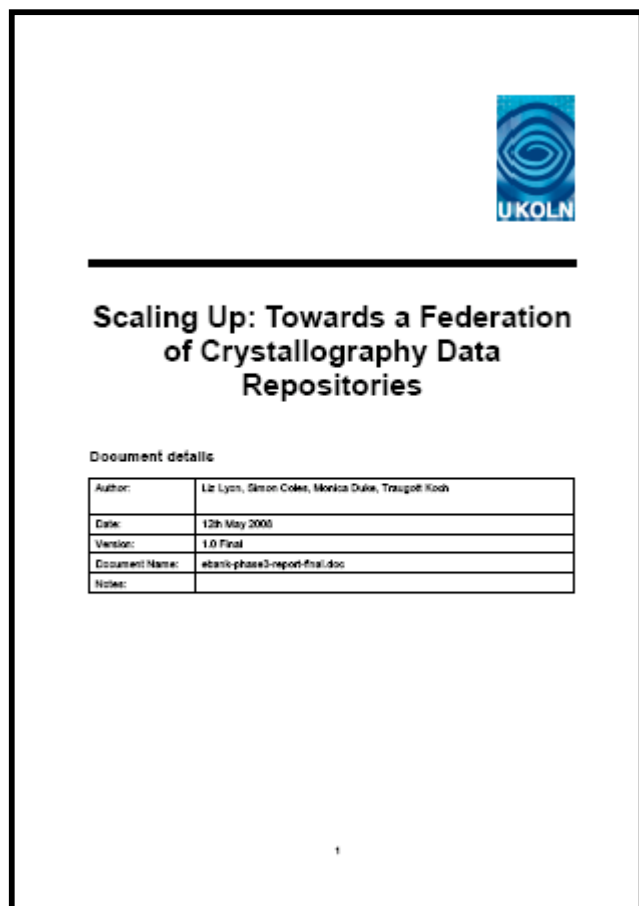
Scaling Up Report

May 2008 UKOLN and
University of Southampton

Interviews & analysis of a discipline:
crystallography

Synthesis: IR Policy & Practice,
Laboratory Practice & Workflows,
Technical Interoperability &
Standards, Metadata Schema &
Application Profiles, Semantic
Interoperability, Data Citation,
Identifiers & Linking, Federation
Architectures & Third Party Services,
Rights & Licensing, Data Quality &
Validation, Preservation, Curation &
Sustainability

Recommendations (7), commentary



Scaling Up Report

Findings:

Diverse lab practice

LIMS and proprietary formats

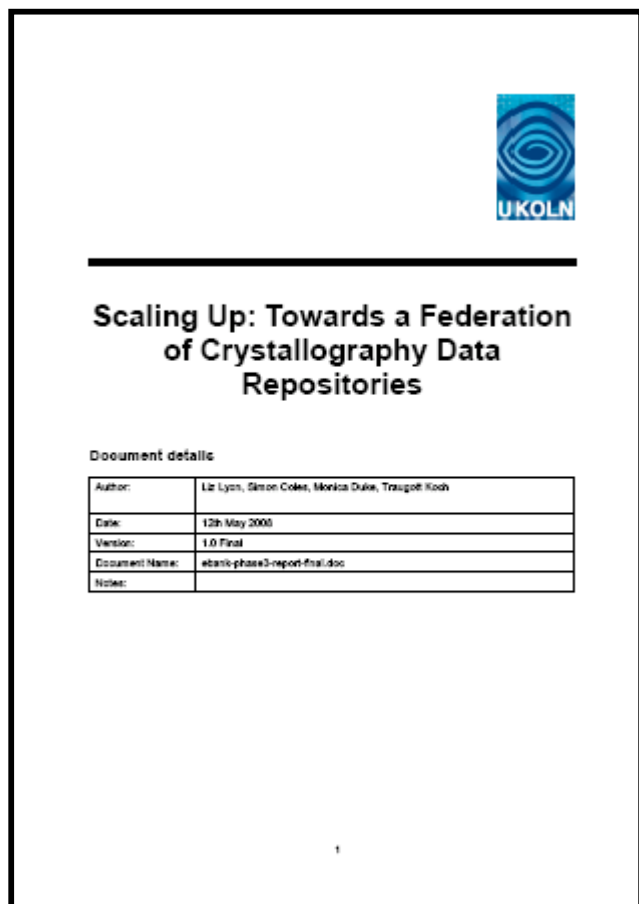
Data policy should reflect lab practice & institutional model

Data quality criteria/validation

“Prior publication” problem

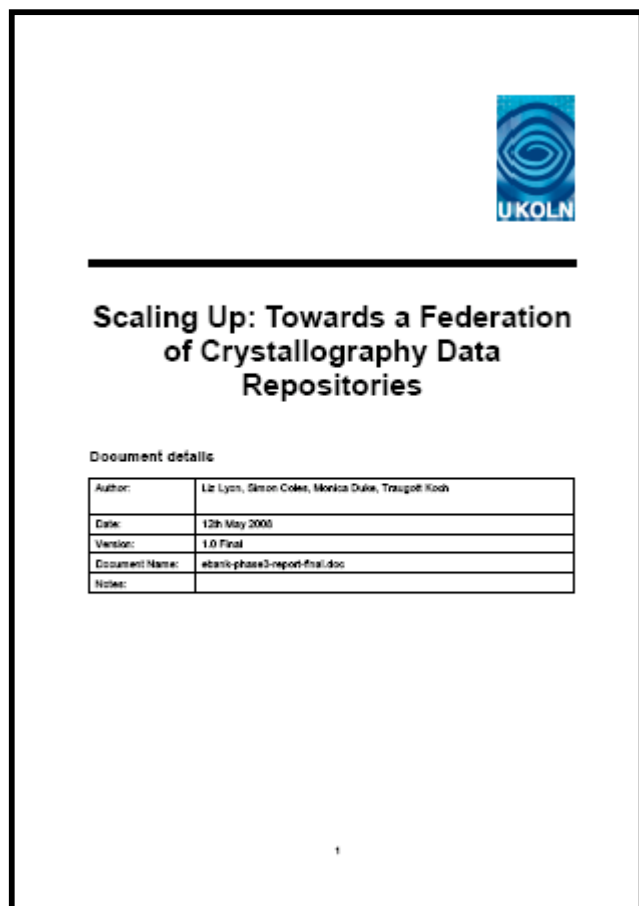
We need scalable assignment of “terms” for data discovery

No discipline preservation model



Scaling Up Report

7 Recommendations



Sub-institutional repositories:
departmental, laboratory,
research group

Laboratory informatics + LIMS

Automatic term assignment for
discovery

Open data licence(s)

Data validation and QA

Quantitative criteria for appraisal

Collective intelligence and
repository content & services

Scaling Up Report

Checklist of Community Criteria for Interoperability

| Community Criteria for Interoperability | Crystallography exemplars |
|--|---|
| 1. Involvement of professional bodies and publishers. | Royal Society of Chemistry, <u>IUCr</u> . |
| 2. Development and adoption of a common domain data format standard. | CIF |
| 3. An established data validation mechanism. | <u>CheckCIF</u> . |
| 4. Implementation and adoption of a common domain identifier. | <u>InChI</u> |
| 5. A metadata schema application profile which supplies a common core element set. | <u>eBank-UK</u> schema |
| 6. An existing subject repository, which may operate on a commercial basis. | CCDC |
| 7. A degree of homogeneity and co-ordination in disciplinary research practice. | CIF and COMCIFS |
| 8. An established service ethic and associated policies, which drives research practice for the common good. | NCS or CCDC or CDS |

Disruptive effects

diverse lab practice

instrument lock-in

limited data-sharing culture

lack of m2m interfaces

fragmented strategy and planning

Research data application profiles

- JISC-funded scoping study
- UKOLN (Alex Ball)
- To assess feasibility, validity and functionality of application profiles for research data
- Consider disciplinary requirements and data models
- Define and validate usage scenarios
- Scope a community uptake strategy
- Identify key stakeholders and any barriers to adoption
- Timescale to complete Autumn 2008

To Share or not to Share



- Research Information Network Report by Key Perspectives
- June 2008
- Interviews > 100 researchers, data managers, data experts
- Data sharing attitudes and practice
- Six areas: astronomy, chemical crystallography, classics, climate science, genomics, social & public health sciences, systems biology, rural economy & land use

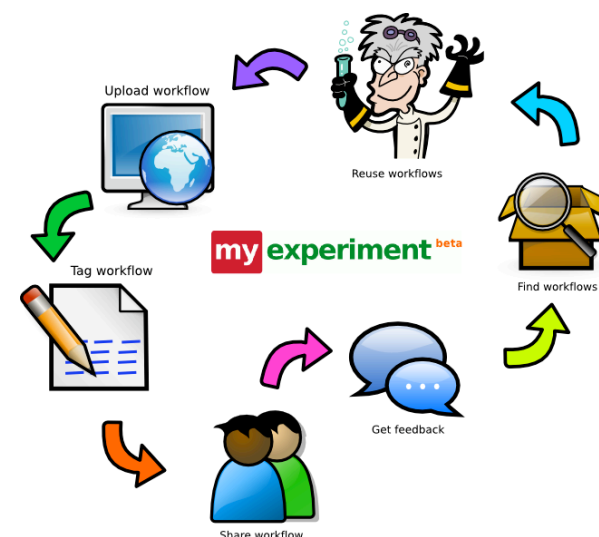
To Share or not to Share



- Convention to share derived or reduced data: access to raw data is rare
- Funder policies & research practice not perfectly matched
- Small-scale projects most at risk
- Centralised data centres cannot accept all data produced
- Shortage of local expertise
- Lack of career rewards on data creation & sharing is a major constraint on publishing
- + lack of time, resources & skills

Practice challenges...

- Data management plans?
- Preservation beyond “data”: workflows, blogs, discourse?
- Appraisal: what data do we keep?
- Data provenance: audit, tracking?
- Citation & versions: persistent IDs?
- Granularity: cite dataset or value?
- Instrumentation, proprietary formats
- Data validation and reproducibility
- Adding value by linking data: across disciplines & sectors

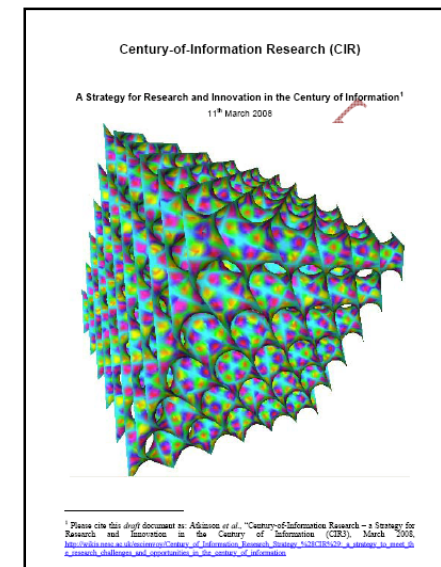


...Work needed: at UK level

- ***To co-ordinate strategic planning – leadership?***
- ***To align policies and monitor implementation***
- ***To invest in infrastructure – who pays?***
- ***To build capacity – incentives and rewards?***
- ***To provide high-level advocacy – funders?***



Global join-up



Slides will be available at :

<http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/presentations.html>



www.ukoln.ac.uk

A centre of expertise in digital information management



The eCrystals Federation

