Research Information and Digital Literacies Coalition
RIDLS

Helping to open up: improving knowledge, capability and confidence in making research data more open

July 2013

Document history

Version	Date	Description of Revision
0.1	21 March 2013	Draft for internal discussion by the team
0.2	22 March 2013	Revised draft following team discussion
0.3	26 March 2013	Initial release for RIDLs steering group
0.4	13 May 2013	Revised draft for team discussion taking into account steering group comments
0.5	20 May 2013	Revised draft for team discussion
0.6	22 May 2013	Revised draft for discussion with RIN
0.6.1	23 May 2013	Minor changes to draft prior to issue to RIN
0.6.2	23 may 2013	Fixed broken cross-references
1.0	18 June 2013	Release version, including minor amendments
1.1	24 June 2013	Same as above, but topped and tailed for final approval from RIDLs steering group
1.2	8 July 2013	FINAL VERSION – minor amendments from version 1.1

Preamble

Rationale for this work

1 This report was commissioned by the Research Information and Digital Literacies Coalition (RIDLs). RIDLs exists to promote collectively the value of training and development in information skills, and enabling activities to facilitate the development of this knowledge and skills. It provides an environment where all those with an interest and stake in the furtherance of information knowledge in higher education, in the academic library world and beyond, may consider practical ideas for advancing the development of information-handling training for HE researchers.

2 In the context of its programme of activities, RIDLs considered the case for focused, small-scale empirical research aimed at advancing the evidence based and supporting the development of policy relating to information and data literacy skills in the higher education sector. Following discussions in September-October 2012, RIDLs agreed that there was a strong case for a study about the knowledge and skills associated with the increasingly important requirements of data sharing and open data. This was felt to be particularly timely in the light of the publication, in June 2012, of the Royal Society report on *Science as an open enterprise*, and of the Government's White Paper on *Open Data – Unleashing the Potential*. These and other developments provided a timely opportunity to investigate the means that are or ought to be deployed to ensure that researchers have the knowledge, confidence and ability to allow for the greatest possible openness for the research data that they create.

3 On that basis, in December 2012, following a tendering exercise, RIDLs awarded a contract to Curtis+Cartwright Consulting Ltd (Curtis+Cartwright), working in association with Max Hammond technology strategy, to carry out a study that would address the above broad aim. Work for this started in January 2013, and a final report was delivered in June.

Acknowledgements

4 RIDLs is grateful to Dr Geoff Curtis[1] and to Dr Max Hammond[2] for the work which they undertook these past few months, and for the production of this report. Thanks are also due to the members of the RIDLs expert sub-group which oversaw the study: Dr Patrick Brindle (Sage Publications), Joy Davidson (Digital Curation Centre and University of Glasgow), Stéphane Goldstein (Research Information Network), Dr Simon Hodson (Jisc) and Wendy White (University of Southampton).

[1] ORCID: 0000-0003-1279-8454
[2] ORCID: 0000-0002-8548-8580

Executive summary

Background

1 Recently there has been increased recognition of the importance of access to data as set out, for example, in *Science as an open enterprise* (The Royal Society, 2012) and the Government's White Paper on *Open Data* (The Cabinet Office, 2012). The need for a move to greater openness is also recognised in Higher Education (HE); the RCUK common principles on data policy provide an overarching framework for the individual Research Council policies (Research Councils UK, 2012).

2 These documents give few details of how the skills and experience will be acquired to implement the broad vision set out. Accordingly, it is timely to investigate the means that are, or ought to be, deployed to ensure that researchers have the knowledge, confidence and ability to allow for the greatest possible openness for the research data that they create, and thus support the goal of an open data culture.

What is open data?

3 This study defines research data as "the data, records, files or other evidence, irrespective of their content or form (*eg* in print, digital, physical or other forms), that comprise a research project's observations, findings or outcomes, including primary materials and analysed data." (Monash University, 2010). Open research data is research data which is accessible, intelligible, assessable and usable (The Royal Society, 2012).

Scope

4 Although the move to opening data derives from trends in science disciplines, open data is now a key aspect of all research disciplines, and this study covers all disciplines.

Approach

5 The study undertook a scoping and literature review, to identify existing sources of evidence and to help set the context for the rest of the project. Further evidence was gathered through interviews with study steering group members and individuals from institutions working in Research Data Management (RDM) and training together with an environment scan to understand the breadth of provision. The evidence was carefully analysed to synthesise the answers to the questions set out in the ITT. The focus at all stages was on *opening* research data, rather than *managing* research data.

Findings

6 Opening data builds on recent requirements and trends in HE for RDM. It is simply one potential goal for managing data. To support opening data, openness should be put at the heart of RDM and RDM training: an "open by default" position. This is a small extension and a change in tone, rather than any fundamental change.

7 While there is little training specifically intended for opening data, there is an increasing number of training courses and materials available to support RDM. Much of this training is at a level that provides only basic working knowledge and an awareness of the issues of RDM, and whereas "open data" may be mentioned, it is only addressed superficially.

8 Few training materials are available for intermediate level skills and none for training experts in opening data. Some courses are available to train information professionals (*ie* library or IT staff) who will subsequently have a role in training others or providing support. The individuals who currently have responsibilities for managing research data are in many cases responsible for working out how to do this: however, there is not yet a significant body of expertise and best practice.

9 None of the training examined was aimed at a particular instrument or dataset. However, such training would normally either be given within the research group, by an instrument supplier, or by a research collaboration.

10 In many disciplines, it is still early days for opening research data. Although the drivers exist, the culture of "open by default" is not yet established, and there is no clear view on the type and level of support in opening data needed by researchers. There is also no common view of the skills required by such support staff or how such support should be delivered.

11 Perhaps the greatest challenge to opening data appears to be meeting the needs of researchers across different disciplines, with strongly differing needs. Sections 2-6 describe a framework for how to address this challenge when designing training and support for opening data, within the broader questions of RDM.

Recommendations

12 Objectives for improving training and support provision for opening research data are set out below with recommendations for achieving these objectives. These objectives reflect a change of emphasis for RDM: nothing dramatically new is required, but open data – opportunities, risks, benefits, and practice – should be integrated within RDM training and support more strongly.

Put opening data at the heart of policy

Recommendation 1: Encourage funders to put 'opening data' at the heart of RDM. Funders should review their RDM policy and practice and ensure that: there is sufficient focus on opening data where appropriate; their requirements for data openness are explicit; and should consider incentives and sanctions to drive compliance. (RIDLs, RCUK, Jisc and BIS).

Recommendation 2: Encourage institutions to put 'opening data' at the heart of RDM. Institutions should review their RDM policy and practice and ensure that there is sufficient focus on opening data where appropriate. Institutional policies must be clear about the role of open data, and must be backed by infrastructure and training to enable their implementation (RIDLs, RCUK, Jisc and BIS).

Recommendation 3: Encourage disciplinary bodies and relevant regional and national organisations to adopt policies that support opening data (RIDLs, RCUK, Jisc and BIS).

Put opening data at the heart of training

Recommendation 4: DCC and UKDA in particular, but other organisations too, should put 'opening data' at the heart of the relevant parts of their websites (DCC, UKDA).

Recommendation 5: Encourage institutions to implement 'opening data' where appropriate, including any necessary culture change (RIDLs, DCC, and UKDA).

Recommendation 6: Encourage institutions to put 'opening data' at the heart of updated training materials and resources (RIDLs, DCC, and UKDA).

Deepen and broaden the training

Recommendation 7: Encourage the development of training materials covering opening data beyond the awareness level and for intermediate and expert levels.

Recommendation 8: Encourage the development of training materials and resources for supporting the culture change aspects of implementing opening data (*eg* awareness training for senior research managers) (Jisc, DCC).

Recommendation 9: Prioritise those disciplines/subjects that are most resistant to change and encourage and seek funding to develop appropriate opening data materials in these areas (RIDLs and Jisc).

Recommendation 10: Establish and disseminate a catalogue of discipline-specific examples of RDM or opening data training (Jisc and DCC).

Recommendation 11: Establish best practice for capturing and managing tacit or informal knowledge of research processes and datasets. It may not be feasible to address this in all cases, but it should be considered (Jisc and DCC).

Identify and disseminate best practice in opening data

Recommendation 12: Best practice is not yet clear for training and support for opening or managing research data. Good practice is emerging, and the DCC should continue collating this (DCC).

Recommendation 13: Develop best practice for an institution's approach to implement opening data (*eg* approaches to culture change, the stakeholders that must be involved, *etc*) (Jisc and DCC).

Develop institutional and community support

Recommendation 14: As experience grows, identify different approaches taken for supporting researchers in opening data and make this available (DCC).

Recommendation 15: Consider whether there is a need to establish or support a community of practice for research data specialists. A number of organisations, including CILIP, the DCC, Jisc, SCONUL, RLUK, and UCISA could potentially establish or support such a community of practice (RIDLs).

Recommendation 16: Engage with other stakeholders (*eg* BIS) in developing and adopting a career structure for research data specialists (RIDLs, BIS, RLUK).

List of contents

List of abbreviations

CPD	Continuing Professional Development
CRIS	Current Research Information System
DCC	Digital Curation Centre
DMP	Data Management Plan
DPA	Data Protection Act
EPSRC	Engineering and Physical Sciences Research Centre
FOIA	Freedom of Information Acts (UK and Scotland)
HE	Higher Education
IP	Intellectual Property
MRD	Managing Research Data (Jisc programme)
Nesta	Formerly National Endowment for Science, Technology and the Arts (NESTA)
PI	Principal Investigator
RASCI	Form of Responsibility Assignment Matrix
RCUK	Research Council UK
RDM	Research Data Management
RIDLs	Research Information and Digital Literacies Coalition
RIN	Research Information Network

1 Introduction

1.1 General

1.1.1 This report has been prepared by Curtis+Cartwright Consulting Ltd (Curtis+Cartwright) working with Max Hammond under contract to the Research Information Network on behalf of the Research Information and Digital Literacies Coalition (RIDLs). It is the final report of a study into 'improving knowledge, capability and confidence in making research data more open. This version of the document (V1.0) is for release and has been approved by the RIDLs study manager.

1.2 Background

General

1.2.1 Recently there has been increased recognition of the importance of access to data. June 2012 saw the publication of two important reports:

- The Royal Society report on *Science as an open enterprise* (The Royal Society, 2012). This is an ambitious and well-argued appeal for scientific research data to be treated as an immensely valuable public good. But the paper has relatively little to say about how training and skills can contribute to increasing such openness, and how legitimate concerns about confidentiality and privacy can be addressed.
- The Government's White Paper on *Open Data* (The Cabinet Office, 2012). This highlights the Government's commitment to ensure the greatest possible access to data (including research data) as part of its transparency agenda.

1.2.2 The need for a move to greater openness is also recognised in Higher Education (HE) and the RCUK common principles on data policy provide an overarching framework for the individual Research Council policies.(Research Councils UK, 2012). Each of these documents is concerned with strategy rather than details of the 'how'. The consequence is that there is little mention of how the skills and experience will be acquired to implement the broad vision set out in these documents.

1.2.3 Other recent reports regarding research in HE have also highlighted the potential represented by open data and have given consideration to the 'how'. These include but are not limited to:

- the 2009 consultative report on *open science at web-scale* (Lyon, 2009), looked at the skills development aspects which called for the embedding of relevant skills, particularly for library and information services personnel;
- the RIN/Nesta report *Open to all?* (RIN, 2010), which garnered views from researchers themselves on the benefits of open behaviour and consequently recommended the development of training and skills in open working.

1.2.4 It is now timely to investigate the means that are, or ought to be, deployed to ensure that researchers have the knowledge, confidence and ability to allow for the greatest possible openness for the research data that they create, and thus support the goal described in the Royal Society report of realising an open data culture.

1.2.5 There are other research environments that overlap with academia (*eg* the strictly regulated environment of the pharmaceutical industry) that may have interesting lessons for how to train researchers in opening data. On the other hand, the strong regulation means that there is less

need to persuade researchers of the benefits: they are simply required to manage their data well. Addressing the implications here is outside the scope of the current work.

1.3 What is open data?

Definition of research data

1.3.1 There is no single, simple definition for research data, and it is possible to view this question from a range of perspectives (Australian National Data Service, 2011). This study uses the definition adopted by Monash University (Monash University, 2010): research data is "the data, records, files or other evidence, irrespective of their content or form (*eg* in print, digital, physical or other forms), that comprise a research project's observations, findings or outcomes, including primary materials and analysed data." Although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.[3]

1.3.2 Importantly, this definition is equally applicable across all domains of research. Although "data" often forms associations with the sciences, it is clear from this definition that the evidence that underpins scholarship in the arts and humanities, and the outputs from these studies, are also research data.

1.3.3 Although "data" is typically associated with the natural and social sciences, most research is in fact data based, including that within the arts and humanities. For example, visual and performance information,[4] and sound and music data.[5]

Definition of open data

1.3.4 There are many definitions of open data, sometimes based on technical properties of the files, sometimes based on licensing requirements, and sometimes more broadly based. This report generalises the *Science as an open enterprise* definition of "intelligent" openness to express what is meant by "open data".[6] This definition requires that data must be accessible, intelligible, assessable and usable as follows:

a) **Accessible**: Data must be located in such a manner that it can readily be found. This has implications both for the custodianship of data and the processes by which access is granted to data and information.

b) **Intelligible**: Data must provide an account of the results of work that is intelligible to those wishing to understand or scrutinise them. Data communication must therefore be differentiated for different audiences. What is intelligible to a specialist in one field may not be intelligible to one in another field. Effective communication to the wider public is more difficult, necessitating a deeper understanding of what the audience needs in order to understand the data and dialogue about priorities for such communication.

c) **Assessable**: Users of the data need to be able to make some judgment or assessment of what is communicated. They will, for example, need to judge the nature of the claims that are made. Are the claims speculations or evidence based? They should be able to judge the competence and reliability of those making the claims. Are they from a competent source? What was the purpose of the research project and who funded it? Is the communication

3 This definition is extended from that used by EPSRC to include all research disciplines. See (EPSRC, 2013c)
4 For example, (Project Cairo, 2011).
5 For example, (Queen Mary University of London, 2013).
6 The changes emphasise that "data" is not just limited to the natural and social sciences.

influenced by extraneous considerations and are these possible sources of influence identified? Assessability also includes the disclosure of attendant factors that might influence trust in the research. For example, medical journals increasingly require a statement of interests from authors.

d) **Usable**: Data should be able to be reused, often for different purposes. The usability of data will also depend on the suitability of background material and metadata for those who wish to use the data. They should, at a minimum, be reusable by other researchers.

1.3.5 These goals go beyond the simple acts of providing access to research data, and in fact describe the entire process of communicating the results of research. Nonetheless, the study team believes that they provide a useful definition of open data, based on what users can do with the data.

Open data, shared data?

1.3.6 This project is concerned with open data, rather than all data management. There is clearly a continuum from "closed" data, through data which are available to select people, on request, through data that are publicly listed but only available to specific users, to data which are available in full with no restrictions on use (and many other specific variations between these points).

1.3.7 This report focuses on the "open" end of this continuum – data which are fully open, or data which are easily discoverable by anyone and accessible based on clear criteria (such as being accredited by the UK Statistics Agency as an approved researcher). This is discussed in more detail at section 3.3.

Benefits of open data

1.3.8 This report will not cover in detail the rationale and perceived benefits for opening data: this has been done elsewhere.[7] A good summary of the perceived benefits for researchers was developed by the Geospatial data community, in that open access to research data:[8]

- reinforces open scientific inquiry;
- encourages diversity of analysis and opinion;
- promotes new research and new types of research;
- enables the application of automated knowledge discovery tools online;
- allows the verification of previous results;
- makes possible the testing of new or alternative hypotheses and methods of analysis;
- establishes a broader base set of data than any one researcher can hope to collect, thereby providing a greater baseline of factual information for the research community;
- supports studies on data collection methods and measurement;
- facilitates the education of new researchers;
- enables the exploration of topics not envisioned by the initial investigators;
- permits the creation of new data sets, information, and knowledge when data from multiple sources are combined;

7 See for example (The Royal Society, 2012) §1.5; (Xu, 2012), (SpaceRef, 2013), (Houghton, 2011) Section 5.
8 (GEOSS, 2012) Annex 2, Section 5.

- helps transfer factual information to and promote development and capacity building in developing countries;
- promotes interdisciplinary, inter-sectoral, inter-institutional, and international research;
- generally helps to maximize the research potential of new digital technologies and networks, thereby providing greater returns from the public investment in data collection and research.

1.4 The data curation lifecycle

1.4.1 The report uses the terminology of the DCC curation data lifecycle model (Digital Curation Centre, 2013b), which is reproduced at Figure 1-1. This model "provides a graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt through the iterative curation cycle. [It] can be used to identify atomic curation activities and subsequently to identify skills needed to undertake activities."

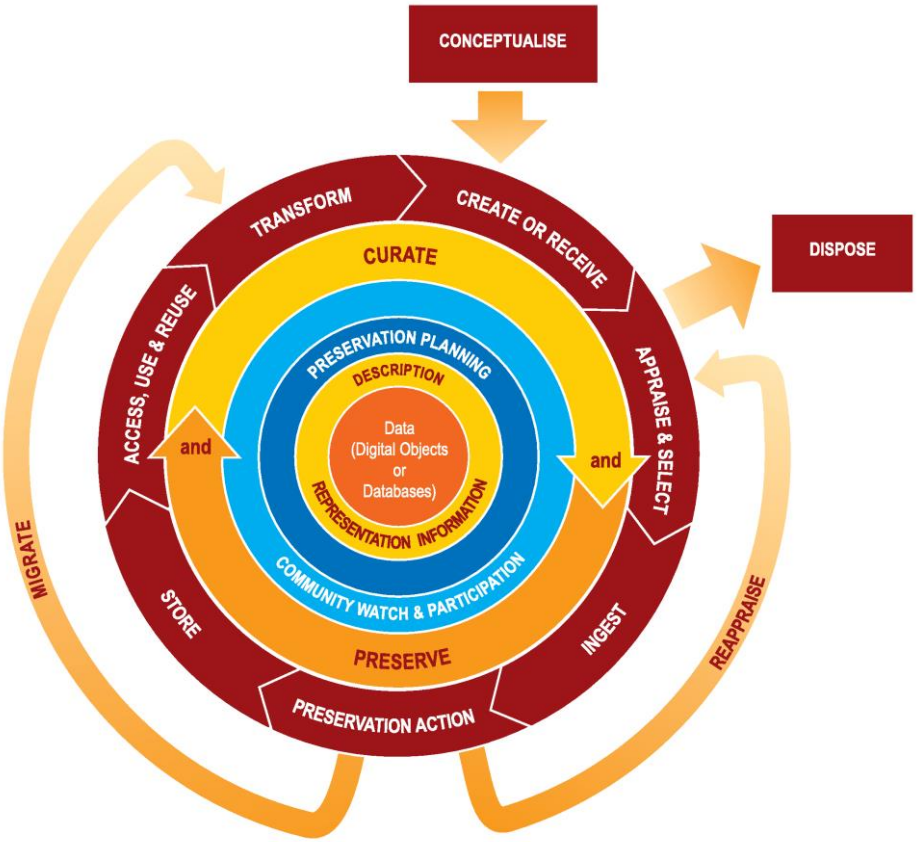


Figure 1-1: the DCC curation data lifecycle model

1.5 Opening data and research data management[9]

Concept of research data management

1.5.1 Managing research data is now a key issue for many HEIs. Good Research Data Management (RDM) practice allows reliable verification of results and permits new and innovative research built on existing information. Moreover, some research data are unique and cannot be replaced if destroyed or lost. All these aspects need to be managed to ensure the full value of public investment in research is realised.

1.5.2 RDM has always been part of the research process – without data, it is impossible to conduct research. However, the responsibility of research organisations for managing research data has been brought to the fore, for instance recently, by the Engineering and Physical Sciences Research Council's (EPSRC's) policy framework (EPSRC, 2013a). In particular, this identifies the institutional responsibility for ensuring that EPSRC-funded research data should be securely preserved essentially for a minimum of ten years after last use with effective data curation being provided throughout the full data lifecycle. EPSRC set out that it wanted all those it funds to have developed a clear roadmap to align their policies and processes with EPSRC's expectations by 1 May 2012, and expects them to be fully compliant with these expectations by 1 May 2015.

1.5.3 Other research councils in both the sciences and in arts and humanities have similar policies with an essential difference being EPSRC's focus on the institution, rather than a Principal Investigator (PI), having responsibility. As well as the research councils, many leading journals require underlying datasets also to be published or made accessible as part of the essential evidence base of a scholarly article. (EPSRC, 2013b) recommends that researchers and institutions can get training and support through the Digital Curation Centre (DCC), Jisc and the Information Commissioner's Office.

1.5.4 Many of the challenges of managing research data are essentially similar between the sciences, arts and humanities, but there are differences. Whereas the sciences often generate very large volumes of data, this data is often precisely specified and relatively simple to manage. Many humanities subjects generate data about living people, which is subject to specific handling and release requirements, and many arts subjects generate data that is much less structured (*eg* archaeological data containing text, images and other multimedia).

Relationship with opening data

1.5.5 An interesting question for this study is the precise relationship between opening data and RDM. This relationship is addressed at subsection 3.3, but in essence the study contends that RDM is a set of skills, experience and best practice that should be applied to data, whether it be open or not. Opening data is not the only reason to manage data, but managing data is an essential prerequisite for effectively opening data. Even within a closed community, there is a compelling case to ensure data are "opened" so that data are accessible, intelligible, assessable, and usable for all community members.

1.5.6 It is important to note that there are limits to openness. There may be legal, ethical, or commercial concerns that restrict what data can be made open, or to whom it may be opened. For some data, these concerns may completely preclude openness, whereas for other data a more-controlled sharing approach may be possible.[10] A recurring theme within this report is to

9 Adapted from (Digital Curation Centre, 2013e) and (Jisc, 2013a).

10 See, for example,(UK Data Service, 2013).

focus on open as a default position – but there are clear exclusions. This is expressed by the RCUK guidance supporting their policy on Open Access (Research Councils UK, 2013):

- *"As part of supporting the drive for openness and transparency in research, and to ensure that researchers think about data access issues, the policy requires all research papers, if applicable, to include a statement on how underlying research materials, such as data, samples or models, can be accessed. However, the policy does not require that the data must be made open. If there are considered to be compelling reasons to protect access to the data, for example commercial confidentiality or legitimate sensitivities around data derived from potentially identifiable human participants, these should be included in the statement."*

Although this language emphasises when data need not be made available, the implication is clear: data must be available unless there is a compelling reason otherwise.

1.6 Aim and scope

Aim

1.6.1 The aim of the study was to investigate the training available to support improved knowledge, capability and confidence in making research data more open, identifying gaps and exemplars.

1.6.2 The objectives for the study were to:

1) investigate the current landscape in training for researchers to support the development of openness of research data.

2) identify gaps in provision in the context of skill/attribute coverage, stakeholder involvement and disciplinary/institutional engagement.

3) investigate how improvements to training provision might be effected, in the context of different disciplinary settings and institutional environments.

4) make evidence-based recommendations.

Scope

1.6.3 Although the move to opening data derives from trends in science disciplines, open data is now a key aspect of all research disciplines, and this study covers all disciplines.

1.6.4 This study has looked at the issues surrounding opening data. It has not looked in any detail at the training needs and training available to support users of open data. The study team considers that while there are many similarities between opening and using data, there may be some areas of difference that would lead to additional training needs. While these are outside the scope of this study, the issues are discussed in outline at subsection 4.2.3.

1.7 Approach

1.7.1 The approach used was to carry out:

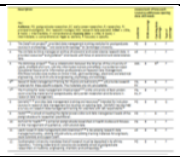
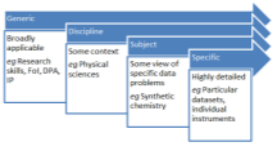
- a scoping and literature review, to identify existing sources of evidence and to help set the context for the rest of the project.

- information gathering, including:[11]
  - discussions with stakeholders;
  - an initial steering group meeting to understand the members' views and finalise the study scope;
  - organising and holding 12 depth interviews to understand in depth the range of training and support currently available, and to identify current gaps and strategic issues with the provision of such training and support;
  - conducting an environment scan to understand the breadth of provision;
- synthesis and analysis to bring together the sources of evidence to answer the questions set out in the ITT;
- reporting.

1.7.2 The focus at all stages was on *opening* research data, rather than *managing* research data.

1.8 How to use this report

1.8.1 Developing training and support for opening research data requires input from many parts of an organisation. There is no single "answer", and the study team believe it is too early to identify best practice in this area (see Section 6). This report views the challenge from several different perspectives, which together form a framework for deciding how best to provide support.

Section 2 sets training in opening data in the context of policy, infrastructure and culture and practice. This context is important to understand how these elements work together to promote open data. Training and support are necessary but not sufficient.	
Section 3 describes the training need for opening research data. This sets out the knowledge and skills necessary to open data. These are complex and cover a wide range of topics: it is unlikely that most researchers need them all – see Section 4.	
Section 4 discusses who to train. Different individuals within a research institution have different responsibilities toward research data, and these responsibilities can change over the lifecycle of research data. This section describes the key roles from the data perspective, and discusses approaches to training and support for each.	
Section 5 documents the current provision of training and identifies gaps. A range of training materials are available – this section gives an overview. These materials could be re-used and adapted by organisations that intend to deliver training.	
Section 6 addresses best practice regarding training for opening data. There is no single best approach for training for opening data. This section describes two key aspects: how generic or specific training can be, and the factors likely to lead to success of a project to establish training and support for opening data.	
Section 7 sets out the key findings and recommendations from the study. These are focused on organisations such as RIDLs, Jisc and RCUK rather than on individual institutions, and consider a national-scale view on how the training and resource provision should be developed.	
Annex A lists the references and interviewees.	

11 A list of interviewees is provided at Annex A.

2 Training in context

2.1 Introduction

2.1.1 Knowledge and skills in opening data, and training that would build these skills, cannot be seen in isolation. They are developed and used in the context of other aspects that are necessary to open data. These aspects are shown at Figure 2-1 and comprise: Infrastructure (both technical and human), Policy, and Culture & Practice. In addition, opening research data can be driven by benefits to the researchers (most obviously, career progression).

2.1.2 This project was focused on the role of training and other support, so the study team has only considered these other aspects in outline – but it is clear that any training programme must be developed in context, and any other intervention must consider training and support needs. This Section provides a very brief overview of these aspects, and suggests some points where they interact with the provision of training and support.

2.1.3 An organisation that intends to promote opening of research data will choose which of these aspects are relevant, and how they should interact. For example, a research council may not provide technical infrastructure, instead relying on the research organisations to provide this infrastructure, and setting policy to drive this development. On the other hand, a diffuse and decentralised university may not be able to create policy to mandate opening data, but may opt to provide a technical infrastructure with associated training and support, to promote open data.

2.1.4 It is not necessary for all aspects to be provided by the same organisation. For example, much research data could be sensibly stored in a subject-specific repository, rather than an institutional one. Although subject-specific repositories are not provided by the research organisation, they form part of the technical infrastructure available to and used by the researchers.

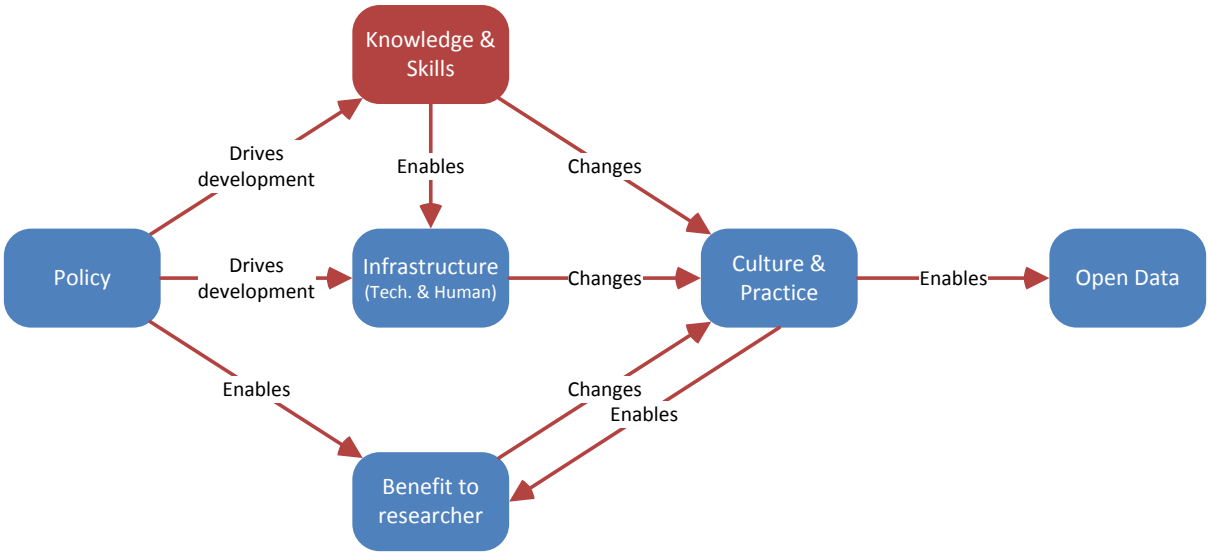


Figure 2-1: the context for opening data

2.1.5 Communicating the benefits of opening data to researchers is essential, but how to create these benefits is out of scope for this work. Benefits will not be discussed further in this section.

2.2 Policy

2.2.1 In many cases, policy changes are driving the move toward open research data. Policy can be created at a range of levels, including:

- Legislation, which may require or prohibit the release of certain types of data (*eg* Data Protection Act (DPA), Freedom of Information Act (FOIA), institutional policies, *etc*);
- Government policy that affects strategic direction;
- Disciplinary codes of practice;
- Research funder policy, which may well set broad requirements for opening data;
- Institutional policy, which may be specific about processes and requirements;
- Departmental policy, which may expand on institutional policy;
- Publisher policy, which may require or encourage opening of research data.

2.2.2 Particularly at the institutional and departmental level, it is necessary to link policy to the infrastructure and skills available to implement the policy.

Questions for training & skills

- What training is necessary for researchers to know what the policy is?
- What skills will be necessary to implement this policy, in practice? How will they map onto our human support infrastructure?

2.3 Infrastructure & Tools

2.3.1 The infrastructure available to support opening data is clearly a significant factor in opening data. Infrastructure in this context refers to technical systems (for example data repositories), and to human infrastructures (for example the availability of a data librarian or research data specialist to support the development of RDM plans).

Technical infrastructure

2.3.2 Technical infrastructures are the sets of tools and systems available that actually provide the mechanism to store and share data. Repositories are the tools within the infrastructure that are most often considered – they are the direct interface through which data are opened.

2.3.3 However, a wider range of systems are involved in opening data, including:

- The primary data stores where information is initially collected;
- Storage and distribution technologies intended to assure long-term preservation of data;
- Platforms to enable access to preserved data (including repositories and other digital asset management systems);
- Website content management systems;
- Current Research Information System (CRIS) that maintain information on the research being undertaken, which can provide metadata to describe research outputs, funders *etc*;
- Identity Management, Authentication and Authorisation systems that underpin controlled access to data;
- A plethora of tools such as Dropbox, Google Drive, Microsoft Skydrive, Amazon S3 and Glacier, email, USB Sticks, portable hard disk drives, and many more. Although these are often not provided by the research institution (and this report does not seek to comment about their quality or usefulness), they form part of the everyday infrastructure that supports the creation, management, and sharing of data.

2.3.4 The technical infrastructure includes all of these elements, as well as subject-specific and institutional repositories.

Questions for training & skills

- Which aspects of the infrastructure need specific training?
- How can training courses or materials be kept up to date with changes in the technology?
- Who should be trained to use which specific[22] systems (including the roles of researchers and support staff)?
- How can researchers be trained to use the infrastructure as a system, as well as the individual components within it?
- How can knowledge be captured, shared and developed from researchers creating and using new tools and workflows?

Human infrastructure

2.3.5 The human infrastructure is the arrangement of skills and responsibilities that is available to a researcher to support them with opening data. The requirements and nature of the infrastructures needed for opening data are still unclear, but drawing on previous work on the support of advanced ICT within HE (Hammond, et al., 2010), it seems likely that the following roles will have some importance:

- Library staff, who may have responsibility for repositories, data librarianship or research data management;
- IT services staff who may have responsibility for the technical infrastructure, as well as responsibility for developing new tools;
- Facilitator/Internal Consultant roles that may support the interface between researchers and library and IT services;
- The research office, who may have a compliance role to ensure that grant proposals include a realistic and deliverable RDM plan which is aligned with institutional policy, and which is costed;
- Specialist trainers;

- Internal experts, who may be based in any part of the organisation, and who have responsibilities to support parts of the process;
- External experts, including centres of expertise such as the DCC, as well as subject specific repositories that may have their own helpdesks or training material;
- Other researchers.

Questions for training & skills

- Which parts of the human infrastructure will require training?
- Which parts will provide training?
- What Continuing Professional Development (CPD) or on-going support is necessary for these personnel to keep up-to-date?
- How will the skills be distributed across the organisation? Should the expertise be focused in a few individuals, across all individual researchers, or a mix? (See Sections 3 and 4);

2.4 Culture & Practice

2.4.1 The culture and practice of researchers are important, and should not be underestimated in the context of opening data. In practice, there is little that any organisation can do to change the culture amongst researchers directly – but many things that they can do to promote shifts in culture and practice. For example, there can be reluctance to change the habits of a research lifetime and embrace electronic data and sharing approaches.

2.4.2 Importantly, some disciplines have established cultures and standard practices for opening research data. This culture is driven not by the institutions, but by the community of researchers itself. It is necessary to consider these existing cultures when planning support.

2.4.3 Training and support can be seen as enablers of culture change – unless researchers are aware of the reasons they might want to open their data, and have the skills to decide how to do so, opening data will not become part of their accepted practice.

Questions for training & skills

- What is the existing culture of data sharing and data openness among the researchers to be trained?
- What are the perceived barriers to sharing? Is there a role for more information, consultancy, and technical support to overcome them?
- Can culture be best changed by focusing on established researchers, or the next generation?
- What role might disciplinary communities play in changing culture?

3 What to train?

3.1 Introduction

3.1.1 The details of who needs what training in what format depends on the organisational context (Section 2), and the organisational structure and responsibilities (Section 4). Training (and other support) can be provided at a range of levels, from generic to very specific (Section 6). This Section sets out the full range of skills requirements identified during the study, to be used in combination with the other sections of this report to understand training and support in a holistic way.

3.1.2 The required skills break down into the following areas, which are described at subsection 3.2.

- Background, rationale and context for opening research data;
- Opening research data principles and practice, covering:
  - How to manage research data;
  - The factors that may constrain releasability;
  - What standards might be appropriate and when and how to use them;
  - What storage and backup processes are most appropriate to active data;
  - What documentation is appropriate;
  - What metadata is required and how this helps discoverability;
  - What to do about data preservation;
  - What types of licence might be appropriate and how to plan for and use licences;
  - What legislation might apply and what needs to be done to comply;
- RDM planning.

3.1.3 Note that these skills do not directly align with the DCC data curation lifecycle. The lifecycle is a process model, which requires sets of these skills at various stages.

3.1.4 The relationship of opening data with RDM (see subsection 1.5) is analysed at subsection 3.3.

3.2 Skills requirements

3.2.1 This subsection outlines the skills that are required to open research data with some ideas for content. This list is drawn from the courses analysed at, Table 5-1 supported by the interviews listed at Annex A. The details of who needs what training in what format depends on the organisational context (Section 2), and the organisational structure and responsibilities (Section 4). The study team does not suggest that all researchers or support staff need all these skills.

Background, rationale and context for opening research data

3.2.2 This covers skills related to awareness of the context and issues surrounding opening data. This includes the context for opening data, the associated benefits, the types of issue that will need to be addressed (*eg* use of standards, confidentiality and privacy, *etc*), best practice principles for opening research data and an introduction to RDM plans. These skills apply not just to researchers and support staff but also relate to managers. Awareness needs to include:

- Research data and open research data;

- the context for open research data, including funders' policies (*eg* (Research Councils UK, 2012), and the benefits of well-managed, open research data;
- the different environments for RDM (*ie* disciplines/subjects, institutions, research groups, research collaborations, *etc*) and the differences in practice between them to help highlight the differences between generic and subject/discipline practice;
- opening and managing research data principles and practice (see paragraph 3.2.3), including an overview of:
  - organising, storing, curating and sharing and managing research data;
  - principles and practice of managing research data, including the research data lifecycle;
  - how to make the data discoverable, including considering the audiences for data, metadata standards, *etc*;
  - licensing research data, commercial considerations, and managing compliance;
  - confidentiality, privacy and relevant legislation and policy (*eg* DPA, FOIA, institutional policies, *etc*);
- Developing RDM plans (see paragraph 3.2.12);
- where to find out more, what generic and discipline/subject training is available, and where to get support and advice;
- how to contribute practical experience of opening data to the research community.

Opening and managing research data principles and practice

3.2.3 This area of understanding covers all aspects of opening and managing research data, licensing, confidentiality, integrity, privacy and related legislation. These can be strategic or practical requirements, depending on role (see Section 4).

Management of research data

3.2.4 This area covers the processes and practice for managing research data effectively, including:

- data ownership and management responsibilities;
- funders' and institutional requirements for managing research data;
- RDM lifecycle;
- general principles of managing research data, including use of RDM plans;
- specific processes and practice for the institution, discipline/subject or research collaboration;
- where to find out more, and to get support and advice.

Releasability

3.2.5 The releasability topic addresses the issues that need to be considered when planning the sharing or opening of data and includes the limits of openness and how the various constraints on releasability should be applied, including:

- possible ethical issues that might arise when sharing or opening data (*eg* inclusion of personal information that might allow identification of an individual);
- Intellectual Property (IP) and issues regarding exploitation;

- policies and codes of practice (legal, funder, institutional, discipline-specific) on ethics and releasability of research data (*eg* embargo periods) and the institutional decision making process (see also legislation topic below).

Standards

3.2.6 This topic describes how to use standards for data sharing, including:

- introduction to formats, international and *de facto* standards and how these are used to support opening or sharing data;[12]
- details of specific international (*eg* JPEG, MP4) and *de facto* (*eg* instrument output) standards for sharing data for the research being undertaken.

Documentation, metadata and discoverability

3.2.7 This covers:

- the need for good and useful documentation for the data;
- best practice for metadata, including specific consideration of schemas for data and metadata;
- how to make the data discoverable;
- how to make the data citeable;
- institutional, discipline/subject or research collaboration policies and practice.

Storage

3.2.8 The data storage topic covers:

- the types of storage and when they might be applicable, covering institutional and discipline/subject repositories, cloud storage, *etc*;
- preparing data for curation (*eg* 'cleaning' data);
- how to maintain integrity of the data, including guidelines regarding the need and timing for back-up;
- naming and labelling data, use of URLs, *etc*.

Data preservation

3.2.9 Data preservation techniques need to be considered as part of ensuring that research data is sustainable for its lifetime. This covers:[13]

- introduction to data preservation;
- data preservation for specific data types and/or storage technologies;
- data curation.

12 (Digital Curation Centre, 2013c) provides a useful resource.

13 The term 'data preservation' could be seen to cover all of the other aspects of organising, storing and sharing research data. The term is used here to denote all other activities, *etc* to ensure sustainability.

Licensing

3.2.10 Licensing needs to be addressed as part of controlled sharing and re-use of data. The topic covers:[14]

- why data need to be licensed;
- when and how to use a licence, including funders' policies;
- the types of licence (*eg* creative commons licences, bespoke licences, *etc*);
- interaction of licensing with control over re-use and attribution;
- what to do if the licence is breached.

Relevant legislation, codes and policies

3.2.11 This topic covers legislation, codes and policies that affect opening data, including:

- relevant UK legislation (*ie* DPA and FOIA);
- ethical and legal standard, requirements and practice associated with legislation, codes and policies, particularly with regards to data containing personal and/or confidential information;
- the general situation in other countries (*eg* Europe, the US and the rest of the world), from the perspective of research collaboration;
- institutional policy on confidentiality and privacy, including DPA and FOI;
- policies of commercial partners or funders;
- what you need to do and what do users of the data have to do to comply.

Research data management planning

3.2.12 A data management plan provides a useful means of documenting the decisions made (and to be made) at the outset of a research project. Creating a data management plan (DMP) is good practice for all RDM, and can facilitate opening data (for example, considering openness upfront may lead to a project capturing less data, or structuring data differently, to allow release). Creating a DMP is essentially an application of the specific skills discussed above, and a documentation of the result. DMPs would typically consider:[15]

- how data will be selected to be made open (and what will legitimately and necessarily be discarded); the criteria for this process;
- who owns the data;
- how the data will be managed and securely stored for its planned lifetime;
- the standards that will used for sharing the data;
- the documentation that will be produced, including metadata;
- how the data will be made discoverable;
- what licenses will apply;
- how relevant legislation (*ie* DPA and FOIA), funders' and institutional policies will be addressed;

14 Adapted from (Digital Curation Centre, 2012)

15 (Digital Curation Centre, 2013a) provides a good introduction to the development and use of data management plans.

- any decisions that are still needed and who will make them;
- who will review/authorise the data management plan?

Support for intelligent openness

3.2.13 Together, these areas of understanding support the desired Intelligent Openness of research data as illustrated at Table 3-1. Data management planning is an underlying process requirement which underpins all areas.

Aspect	Intelligent openness (paragraph 1.3.4)	Relevant skills areas
Accessible	Data must be located in such a manner that it can readily be found and in a form that can be used.	Management of research data. Releasability. Standards. Storage. Documentation, metadata and discoverability; Data preservation. Licensing. Relevant legislation.
Intelligible	Comprehensive for those who wish to scrutinise something. Audiences need to be able to make some judgment or assessment of what is communicated. They will need to judge the nature of the claims made. They should be able to judge the competence and reliability of those making the claims.	Management of research data. Releasability. Standards. Documentation, metadata and discoverability. Data preservation. Licensing.
Assessable	In a state in which judgments can be made as to the data or information's reliability. Data must provide an account of the results of work that is intelligible to those wishing to understand or scrutinise them. Data must therefore be differentiated for different audiences. Assessability also includes the disclosure of attendant factors that might influence public trust.	Management of research data. Releasability. Documentation, metadata and discoverability. Data preservation. Relevant legislation.
Useable	In a format where others can use the data or information. Data should be able to be reused, often for different purposes, and therefore will require proper background information and metadata. The usability of data will also depend on those who wish to use them.	Releasability. Standards. Documentation, metadata and discoverability. Data preservation. Licensing. Relevant legislation.

Table 3-1: skills areas mapped to aim of intelligent openness

3.3 Managing research data, or opening research data?

3.3.1 There is clearly a continuum from "closed" data, through data which are available to select people, on request, through data that are publicly listed but only available to specific users, to data which are available in full with no restrictions on use (and many other specific variations between these points).

3.3.2 This project was scoped to opening research data, rather than the broader questions of managing research data, and as such this report focuses on the "open" end of this continuum – data which are fully open, or data which are easily discoverable by anyone and accessible based on clear criteria (such as being accredited by the UK Statistics Agency as an approved researcher).

3.3.3 It is possible to see open research data as the top of a stack of key processes associated with research data (see Figure 3-1), based on the ability to create and collect research data, then to

manage that data, then to share that data in a restricted manner, and then finally to make research data open. Each layer in the stack requires the skills and processes of the layer below, but to a higher degree – and all of these layers should be underpinned by good planning.

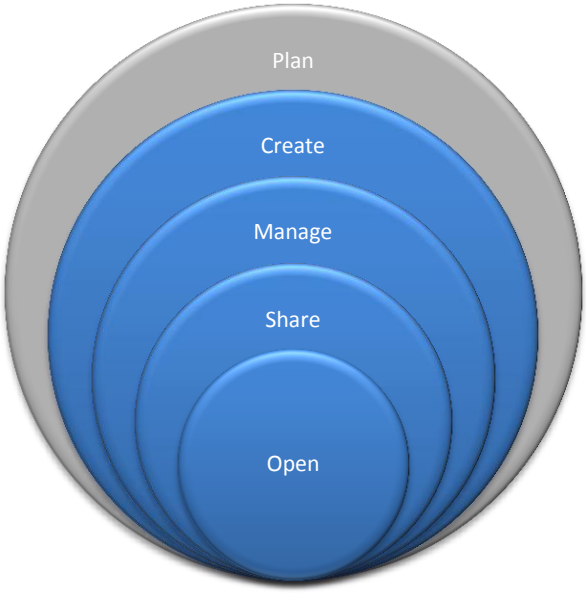


Figure 3-1: the open data "stack"

3.3.4 Effective RDM underpins, and is a pre-requisite for effectively opening research data. The study team considers that the distinction between the two is essentially artificial – opening data is simply one potential goal for managing data. There is an increasingly well-established approach to RDM with considerable training resources available, see for example (Digital Curation Centre, 2013d). It is helpful to consider what, if anything, is different for opening rather than sharing data.

3.3.5 Opening data may require particular focus on some elements of the data management process – but they are small extensions, rather than any fundamental change. Table 3-2 highlights these small differences, against the identified skills needs.

3.3.6 Several interviewees described open research data as a natural aim and progression for managed research data – and that opening data should simply be seen as a part of managing data. The study team is inclined to agree.

	Factor	What is different for opening research data
1	Background, rationale and context for opening and managing research data	RDM underpins the use of research data during a project, and is a prerequisite to making data open. Managing research data is an essential prerequisite for opening research data. However, managing research data can sometimes be an internal activity, whereas opening research data is primarily an external activity. Internal data management is an obvious local benefit; the benefits from opening data are more diffuse.
2	Opening research data principles and practice	
	– Management of research data	No differences.
	– Releasability	Should be considered for all managed data, must be considered for open data. Planning to release data may influence which data are collected in the first place.
	– Standards	Standards should play a part in all RDM, whether the data is being opened or not. Standards are particularly important to enable re-use of open data.
	– Documentation, metadata and discoverability	While good quality documentation is a normal part of managing research data, there may be a greater emphasis on quality of documentation for opening data. Open data may well be stored in repositories without the context in which it was developed. There is greater emphasis on discoverability.
	– Storage	While externally accessible storage may be required as part of managing data, there is likely to be more emphasis on externally accessible storage for opening data.
	– Data preservation	No significant differences, but if data is made available openly there may be an expectation of long-term availability.
	– Licensing	Much greater focus on licensing for open data.
	– Relevant legislation, codes and policies	Legislation applies even if data not opened. Different practical issues especially for confidentiality and privacy of releasable data. For example, better data management planning is likely to be needed (*eg* to ensure collected data is releasable with minimum data cleaning).
3	Research data management planning	The same process is required, but planning for openness needs to be included within the overall plan – it should not be an add-on.

Table 3-2: comparison of opening data and RDM skills requirements

4 Who to train?

4.1 Introduction

4.1.1 One of the obvious considerations for developing skills in opening data is to decide who to train. The key distinction is their role within the data management ecosystem. These roles describe the decision-making authority that an individual has over the data, and this may not be aligned with an organisational role or responsibility.

4.1.2 Who fulfils these roles may vary between different organisations, departments, research groups, and datasets. In some cases, one individual may hold more than one of these roles. Researchers working alone will be the data creator and data controller, some data controllers will be research data specialists, and some researchers may be research data specialists. Many research managers will be data controllers.

4.1.3 Importantly, roles regarding data can change with time. In particular, the role of data controller can vary for a specific dataset – although the PI is likely to be the data controller during the design and conduct of a research project, this responsibility may be handed over to a repository manager following conclusion of the project.

4.1.4 The key roles are described at Table 4-1.[16]

Role	Definition	Description/Examples
Research Manager	An individual responsible for strategic choices about the management of research.	Depending on perspective, could be a member of an organisation's senior management team, a head of faculty or department, or a research group manager.
Data Controller	An individual responsible for decisions about a particular set of research data.	Within a research organisation, responsibility is typically assigned to an individual directly involved in the creation of the data - in funded research this will be the PI. Other possibilities include the grants team or research administrator. After a research project is completed, the responsibility for the data may be transferred to another individual, for example a repository manager. The ownership in terms of intellectual property in the data is not implied here – IP is generally held by the research institution or by a commercial organisation funding the research. Research Data Controllers may or may not be Data Controllers under the meaning of the Data Protection Acts.
Data Creator	An individual actually generating research data.	Researchers and students. Potentially also technicians operating instruments.
Research data Specialist	A specialist who supports the management of research data.	This includes a wide range of individuals, including Library and IT staff, repository managers, research data managers, consultants, and other experts. In some cases, the data controller may be a research data specialist.
Data user	An individual who wishes to access/exploit the research data.	This includes research managers, data controllers or data creators and the general public.

Table 4-1: roles in an open data ecosystem

[16] Other work has considered aspects of these roles, including (Lyon, 2012),(Pryor & Donnolley, 2009) and (RIN, 2010).

4.2 Organisational structures and skills mapping

4.2.1 How the roles are related within the research data ecosystem or an organisation depends on the organisational structure, and the skills required by each role will also depend on this structure.

4.2.2 How an organisation should structure this ecosystem is out of scope for this report, but Table 4-2 gives an example of how the responsibilities of each role could be mapped against the sequential actions of the DCC Curation Lifecycle model (Digital Curation Centre, 2013b), using a standard RASCI approach.[17] Research Managers are not described, as their roles are only indirectly related to the data – they may be accountable for ensuring that research data is managed, but will have no specific responsibilities here; put another way, they own the matrix. In this example, no one is assigned responsibilities for the transformation of data. Transformations are extremely context dependent, and will depend in many cases on future uses of the data.

Stage	Data controller	Data creator	Research data specialist	Data user
Conceptualise	RA	S	C	-
Create or Receive	A	R	-	-
Appraise and Select	AC	R	S	-
Ingest	A	S	R	-
Preservation Action	A	-	R	-
Store (short-term)	A	R	S	-
Store (long-term)	A	I	R	-
Access, Use and Reuse	A	I	R	R
Transform	-	-	-	-

Table 4-2: an example organisational responsibility assignment matrix for RDM

4.2.3 Mapping skills to this kind of responsibility matrix will guide development of training and support, based on the skills necessary for the individual's role. In practice, researchers are unlikely to identify themselves as primarily as data creators or data controllers, so training is much more likely to be designed based on organisational roles (PhD student, PI, *etc*) – but the exercise of mapping data lifecycle against roles is worthwhile to think through what skills the actors in this system should have.

4.3 Training for data roles

4.3.1 This subsection discusses how the generic roles discussed above could be mapped to organisational roles and training approaches that could be applied for each. There is currently limited evidence to show what works in practice. The Vitae Researcher Development framework provides a useful context for this training. [18] Subsection 6.2 sets out the types of training needed by different levels of specificity (*ie* generic, discipline, subject, specific).

17 R = Responsible - actually does a task. A = Accountable – ultimately answerable for a task. S = Supporting – supports the task being undertaken. C = Consulted about a task. I = Informed about a task or the outcome of a task.

18 Examination of the Researcher Development Framework (Vitae, 2013) indicates that this training covers the broad aspects of A1: knowledge base (information literacy and management), C1: professional conduct (legal requirements, IPR and copyright); C2: research management (research strategy, project planning and delivery, risk management); D1: working with others (collaboration); D2: communication and dissemination (publication); D3: engagement and impact (enterprise).

4.3.2 Each role is described in outline, along with key observations on the kinds of training and support that are likely to be most appropriate. Paragraph 4.3.3 describes the training approaches in more detail.

Training approaches

4.3.3 Training and support can be delivered with a wide range of different approaches. To avoid discussing the specific types of intervention that may be most useful (see also Section 6), the study team has assigned training and support needs to two axes: focus and delivery. Note that these are continuous axes, rather than discrete choices.

- **Training focus** describes the degree to which training concentrates on strategic or practical aspects.[19]
  - **Strategic** aspects include why open data can be beneficial, the issues and concerns that surround it, the policy and funding landscape, and high-level understanding of the data lifecycle.
  - **Practical** aspects include details of how to manage and open data, including metadata schema, deposit tools, workflows: the "nuts and bolts" of how to process and manage research data.
- **Training delivery** describes the types of activity that are undertaken.
  - **Personal** approaches include mentoring, networking opportunities, consultancy and support from experts, input from professional societies, and similar very-focused activities.
  - **Impersonal** approaches include online or group courses, guidance material, presentations at conferences, and similar activities. This does not imply that the content is generic (see subsection 6.2), but that the training is less individual.

Research manager

4.3.4 Research managers are responsible for managing research – setting strategy and policy, potentially making recruitment decisions, probably with some level of budgetary responsibility. Depending on perspective, a research manager could be a member of an organisation's senior management team, a head of faculty or department, or a research group manager.

4.3.5 Research managers may be in a position to set overall strategy for opening research data, and will likely be able to influence practice within their sphere of responsibility. Most research managers will be established academics or professional administrators.

4.3.6 Research managers will require strategic or policy-driven support. Based on the study team's experience in supporting culture change in a number of organisations, this should include a case for why opening research data is a sensible choice, and support with addressing concerns about the process. Training and support needs are **Personal** and **Strategic**.

Data controller

4.3.7 Data controllers are responsible and accountable for research data. From the perspective of an organisation, responsibility is typically assigned to an individual directly involved in the creation of the data - in funded research this will be the Principal Investigator (PI).

[19] In practice, a training course will need to reflect the appropriate mix of strategic and practical aspects.

4.3.8 After a research project is completed, control of the data may be transferred to another individual, for example a repository manager. In this case, the new data controller will be a research data specialist, and the analysis below for that role is more relevant for them.

4.3.9 The data controller role is challenging – these individuals have broad responsibilities, and need to understand the entire research data lifecycle. They are accountable for ensuring that the data is well managed and compliant with any requirements from their research funders or their organisation. They make decisions about which data to share, and decide how to share it.

4.3.10 Two of the key challenges in training data controllers are:

1) Many are senior academics with busy schedules. They may not make time for training activities.

2) Some may not recognise how much input is required from them to make this system work. There may be a view that decisions about storage, management, and openness should be made elsewhere in the organisation.

4.3.11 Training for support needs for data controllers are likely to be broad, covering both **Strategic** and **Practical** aspects, and delivered in a **personal** manner.

4.3.12 There is an alternative (or additional) longer-term approach to training data controllers: train the students and early-career researchers who will go on to become data controllers. This should ensure that the skills, and perhaps more importantly the mindset, required for opening data are more strongly embedded in research culture as these individuals mature in their careers.

Data Creator

4.3.13 Data Creators are the individuals actually generating research data. Depending on the context in which the research is undertaken, these individuals are often research students or postdoctoral research staff – but in some contexts, the data creator will be a more established academic (for example when contributing to but not leading a large-scale research project, and in disciplines where more-senior researchers still undertake hands-on research themselves). In some situations, data could be created by the operator of a research instrument, despite that individual having no responsibility for the research being undertaken.

4.3.14 Data Creators are the only actors in the system who know definitively how specific data items were created – these individuals must be responsible for capturing and describing data as it is created. The amount of further responsibility delegated to them depends on the local situation and in particular on their relationship with the data controller.

4.3.15 The training most applicable to data creators focuses on the **practical** aspects of data management and openness, and can be largely delivered in an **impersonal** manner. Importantly, however, many data creators will go on to become data controllers in their later careers, so it is nonetheless necessary to provide strategic context for the training they receive.

Research data specialist

4.3.16 Research data specialists are experts who support the management of research data. This includes a wide range of individuals, including Library and IT staff, repository managers, research data managers, consultants, and other experts. In some cases, the data controller may be a research data specialist.

4.3.17 Research data specialists typically support a wide range of different types of data from across an institution, although some particularly data-intensive projects or departments may require their own data scientists or other research data specialists.

4.3.18 Research data specialists require both broad and deep knowledge of the aspects of data management that they are responsible for; they are likely to be in a position to support other actors in the system with complex problems, or detailed issues. They need to understand the range of tools and support available, the organisational policies (and perhaps the policies of research funders and other organisations), and be able to guide research managers, and data creators and owners through the opportunities.

4.3.19 Managing research data, or providing services to help manage research data, is likely to be a core part of these roles. This is in contrast to the data controllers, for whom managing research data is increasingly important, but nonetheless secondary to conducting the research itself.

4.3.20 The training and support needed by research data specialists will depend strongly on their specific role. In general, the focus is likely to be **practical**, with both **personal** and **impersonal** training playing important roles. The career structure for research data specialists also needs to be considered.

Data user

4.3.21 Data users are those that wish to access/exploit the research data. Data users might also be research managers, data controllers or data creators.

4.3.22 As this project was scoped to training and support for opening data (rather than for using open data), the requirements of data users will not be considered in detail. In outline, however, data users, in line with the definition of intelligently open data, need to be able to locate and use relevant data, understand how to interpret what it means, understand the constraints on their use of it through licensing, and meeting confidentiality and privacy requirements.

Summary of training needs

4.3.23 Figure 4-1 illustrates the identified training requirements for each role on the axes described.

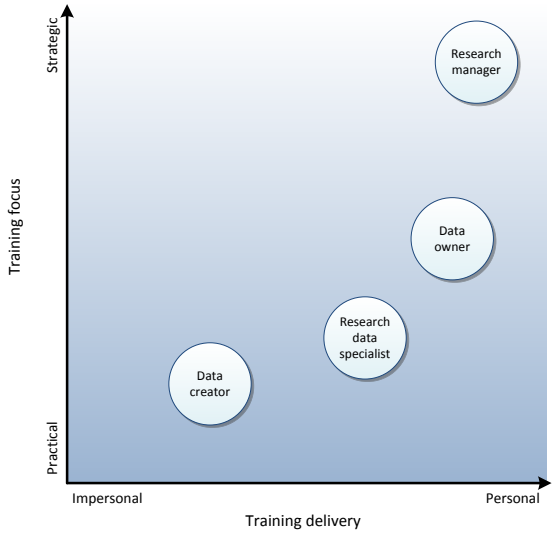


Figure 4-1: training and support needs

5 Existing provision and gaps

5.1 Introduction

5.1.1 This Section summarises available RDM training and resources and assesses its utility for meeting opening data skills needs. It also identifies gaps in training and resources.

5.2 Available training and resources

5.2.1 Table 5-1 identifies training courses and resources and provides the study team's assessment. The list is not intended to be exhaustive; rather it reflects the training courses and resources identified during the environment scan and in discussions with interviewees.

5.2.2 Each identified course was reviewed to identify (based on the study team's judgement):

- the intended audience (postgraduate researcher, early career researcher, researcher, PI, research manager, research data specialist);
- the level of RDM content (basic, intermediate, expert[20]);
- the level of opening data content (basic, intermediate, expert);
- the type of coverage (generic, focused on a discipline/subject,[21] specific[22]).

5.3 Assessment

Overview

5.3.1 While there is little training specifically intended for opening data, there is an increasing number of training courses and materials available to support RDM which address at least some issues associated with opening data. However, with some notable exceptions (see paragraph 5.3.12), most of this training is at a level that provides only basic working knowledge and an awareness of the issues.

5.3.2 The majority of available training is aimed at postgraduate researchers or those on taught postgraduate qualifications but can also be used for all others who need to understand the basics. Some materials and resources are aimed at information professionals/librarians to enable them in providing training and support to researchers.

5.3.3 There are some excellent generic (*eg* (Digital Curation Centre, 2013d)) and focused (*eg* (UK Data Archive, 2013)) resources available. It is noticeable that much of this content has been centrally funded, in particular through the two phases of Jisc's Managing Research Data programmes, ((Jisc, 2013a) and (Jisc, 2013b)) and the support of the DCC.

20 Basic/Intermediate/Expert are the study team's judgement about the target audience for the training. Basic requires no prior knowledge, and provides only a high-level overview. Intermediate may require some prior knowledge, and provides a reasonably comprehensive understanding of the subject. Expert training is for individuals who are, or intend to become specialists in this area of understanding.

21 'Focused on a discipline/subject' means that the training materials and resources are expressed in terms of the language of a specific discipline/subject, and with examples drawn from that specific discipline/subject.

22 'Specific' means that the training materials and resources are intended for, say, a specific research instrument, research collaboration, *etc*.

	Description	Assessment of how each resource addresses opening data skill needs			
	Key: Audience**: PG: postgraduate researcher; EC: early career researcher; R: researcher, PI: principal investigator, RM: research manager, RDS: research data specialist.** RDM coverage**: x: Little; B: basic; I: intermediate; C: comprehensive.** Opening data**: x: Little; B: basic; I: intermediate; C: comprehensive.** Type**: G: Generic; F: focused; S:** specific**.**	Audience	RDM coverage	Opening data coverage	Type of coverage
1	DCC (Digital Curation Centre, 2013d) is a leading centre of expertise in digital information curation, namely maintaining, preserving and adding value to digital research data throughout its lifecycle and provides:				
	– an extensive resource on RDM;	All	C	C	G
	– digital curation 101 training course;	All	B	B	G
	– RDM road shows;	All	B	x	G
	– 'tools of the trade' course.	RDS	B	B	G
2	Research data MANTRA'(University of Edinburgh, 2012) was developed at Edinburgh. It focuses on geosciences, social and political sciences and clinical psychology.	PG, EC	I	B	F
3	The Mantra training materials have been developed subsequently to support liaison librarians (University of Edinburgh, 2013).	RDS	I/C	B	F
4	The DataTrain project provided data management training modules for post-graduate courses in archaeology(University of Cambridge, 2013a) and social anthropology (University of Cambridge, 2013b) at Cambridge University.	PG	B	x	F
5	The UK Data Archive manages an archive of economic and social science research data. It provides resources (UK Data Archive, 2011) and guidance (UK Data Archive, 2013) at all levels with focus on economic and social science data.	All	C	I	F
6	The RDMRose project (RDMRose, 2013) was a collaboration between the libraries of the University of Leeds, Sheffield and York, with the Information School at Sheffield. It provided an Open Educational Resource for information professionals on RDM. RDMRose includes case studies on clinical trials, gastroenterology, electronic and electrical engineering, civil and structural engineering, psychology and sociology.	RDS	B	B	F
7[23]	RDM Training for Physics and Astronomy' (University of Hertfordshire, 2012) will provide research materials for these specific subjects. The materials are not yet available.	PG, EC	B	B	F
8[23]	The Training for Data management (TraD) project' (University of East London, 2013) uses elements of MANTRA and covers training materials for postgraduates, early career researchers and librarians in RDM.	PG, EC, P	B	x	F
9	DaMaRO' (University of Oxford, 2012b) provides data management training and resources (see for example (University of Oxford, 2012a)) intended for induction courses in RDM but touches on opening data. DaMaRO has also held a survey on RDM training for scientists with 193 respondents (University of Oxford, 2013).	PG	B	B	G
10	CAiRO (Jisc, 2013) is aimed at meeting the digital preservation and data management needs of the postgraduate arts researcher-practitioner.	PG	I	B	F
11	DATUM for health (Northumbria University, 2011) is aimed at postgraduate researchers in health studies and focuses on the management of qualitative, unstructured data.	PG	B	x	F
12	Leeds RDM pilot (RoaDMaP) (University of Leeds, 2013) is developing RDM policy, piloting infrastructure, and piloting training materials for engineers, social scientists and support staff.	All	B	B	F
13	Southampton University mandates that all research must be deposited in its ePrints repository. Training material and resources are available aimed at postgraduate researchers in medicine, engineering, chemistry and archaeology (Scott, et al., 2013).	PG	B	x	F

Table 5-1 overview of exemplar training materials and resources

[23] The assessment is based on preliminary outputs; the project is still in progress as at 2 May 2013.

What approaches have been used?

5.3.4 The training courses examined are typically based around a set of expository PowerPoint slides and related materials aimed at postgraduate researchers. The same materials can also be used for training other researchers or research managers who do not have basic RDM or opening data skills. Some centrally funded training has been provided as online courses (*eg* MANTRA and RDMRose), which can be (and have already been) adopted or adapted by other organisations. It is harder to identify the support and training being provided internally by institutions. The study team considers that this is at an early stage in many institutions, and focused on the generic level.

5.3.5 The sample examined is small, and the field is too immature to identify which approaches work "best" (see section 6).

Differences by discipline/subject areas

5.3.6 Much of the available training is focused on a specific discipline/subject. The differences between them are essentially the use of language and examples relevant to the specific discipline/subject. The study team's view is that this reflects the early stage of the development of training for opening data.

5.3.7 There are likely to be many specific discipline/subject areas that are not yet covered by focused training or resources.[24]

5.3.8 No evidence was seen of training being provided across a discipline (*eg* by a professional body) – the training materials may be discipline-specific, but have been created by individual projects rather than any representative of the discipline more generally.

Difference by institutional environment

5.3.9 The only discernible difference between institutions is that institutions may be at different points in the development of training for RDM/opening data. For example, the analysis of training materials supporting Table 5-1 suggests that some are still bringing in RDM training while others have started to address the issues of opening data. However, at this stage the latter training seems somewhat of an add-on.

5.3.10 This project has not analysed institutional policies in detail, but stakeholders suggest that the gaps in training for opening data reflect institutions' RDM policies, which may have been written to address the compliance issue of managing data without focusing on the openness aspect. Anecdotally, some institutional policies refer to opening data, but few emphasise this aspect, or explicitly state that data should be "open by default".

Possible gaps in provision

Depth

5.3.11 One gap is a lack of depth. Many of the training courses designed to support RDM are more concerned with managing the data for the researchers' own use rather than specifically opening the data. Some of the materials examined do cover it but at a basic awareness level.

[24] The *DaMSSI (Data Management Skills Support Initiative) final report* states that 'Postgraduates require training in the basic skills of data management which can be delivered as a set of core generic principles. Discipline-specific examples and references should be included alongside the generic to engage the audience and illustrate relevance and context' (Davidson, 2010).

5.3.12 Few training materials are available for intermediate level skills and none for the explicit training of experts in opening data. However, there are existing resources (especially from the DCC) that can be used to acquire further knowledge and skills as required. There are some courses available to train information professionals (*ie* library or IT staff) who will subsequently have a role in training others or providing support. This may reflect an understanding that opening data (and managing research data) are relatively new concerns within research institutions. The individuals who currently have responsibilities for managing research data are in many cases responsible for working out how to do this; there is not yet a significant body of expertise and best practice.

Breadth

5.3.13 One obvious gap is in breadth – there is strong demand for training materials using language and examples of a specific discipline/subject, materials for only a few disciplines/subjects have yet been developed, and this has been in an *ad hoc* basis.

5.3.14 None of the training examined was intended to be specific (*ie* aimed at a particular instrument or dataset). This is not surprising; such training would normally either be given within the research group, by an instrument supplier, or by a research collaboration. The stakeholders interviewed for this project described this as in many cases being tacit or informal knowledge.

Summary

5.3.15 It is early days for opening data. As yet there is no clear view on the type and level of support in opening data needed by researchers. There is also no common view of the skills required by such support staff or how such support should be delivered, and this is likely to vary between organisations (Jones, et al., 2013). In many ways this is akin to the situation that pertained for advanced ICT (Hammond, et al., 2010). Similarly, there is no obvious way to share experience of practical approaches to opening data.

6 Best practice?

6.1 Overview

6.1.1 One of the aims for this project was to identify any emerging good practice in training and support for opening research data. The consistent view from the stakeholders interviewed is that it is too early to identify good practice for some significant parts of the overall puzzle.

6.1.2 There is developing evidence of which aspects of RDM are particularly important for opening data (see Section 3). Common approaches have also been identified, and a range of projects are deploying these. These go beyond funded projects such as those within the JISC MRD programme: many (probably most) research-intensive institutions have identified RDM and associated questions of open data as important strategic issues, and are tackling the issues locally. All of these projects consider training and support to some extent, although there are broad differences in the approaches taken. The study has not identified any open training materials produced outside a funded project.

6.1.3 The study team does not believe that it is possible at this stage to say which approaches are particularly effective. The evidence is insufficient to justify such an analysis. What is clear is that these projects are identifying similar challenges. The greatest of which appears to be meeting the needs of researchers across different disciplines, with strongly differing needs.

6.1.4 This section describes a framework for addressing this challenge when designing training and support for opening data within the broader questions of RDM.

6.2 Structuring training and support

6.2.1 A particular question that arises in most RDM projects regards differences between disciplines, faculties and subjects. This problem is complex, and the answer is not straightforward. In many cases, the RDM approach can vary strongly even within a single department, given the wide range of research undertaken. As such, the study team believes that it is necessary to view skills and training at several levels of specificity, which have been defined below. This must be seen, at the organisational level, in the context of subsection 4.3 above, which discusses the roles involved in opening data. It may be necessary to provide more-specific training for some roles than others. An organisation intending to develop training and support to open data should consider how to meet the needs of researchers at each level.

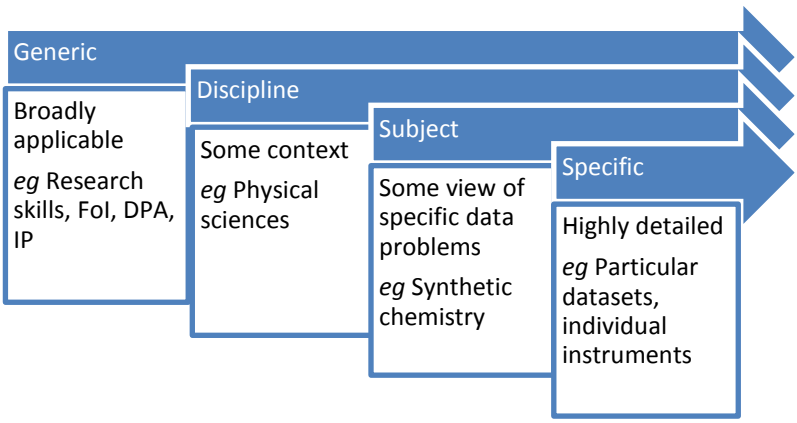


Figure 6-1: training focus

6.2.2 An important point to consider for the practical design of training and support is that generic training can always be incorporated within more-specific training – generic RDM skills could be taught at the subject level. There is strong demand from researchers for training that is specific to them, but the balance to be struck here is again efficiency and cost – whereas individual, specific training is likely to be most beneficial to the recipient, it is far more expensive (financially, and in time and effort) to develop and deliver.

Generic

6.2.3 Generic training and support can be delivered to all or most individuals with similar roles. For example, training in responsibilities under the Data Protection Act or Creative Commons licensing does not depend on specific context.

6.2.4 The key downside is that generic training must be broad, and there is significant demand from researchers for training to be more practice-based, drawing on examples from their own subject. On the other hand, generic training and support can be seen as providing underpinning knowledge, or a broad overview of the issues of RDM and open data. Generic training and support are easy to deliver to large cohorts, and are consequently relatively cheap.

6.2.5 An example of generic training could be a half-hour web-based course titled “The Data Protection Act”, and mandatory for all staff.

Discipline

6.2.6 Discipline-based training can be more focused than generic training. For example, many humanities researchers must deal with personal data generated from living individuals, whereas many physical scientists will deal with instrumental data. At this level, it is possible to discuss generic approaches to specific problems. Discipline-level training is therefore an intermediate approach – more specific than Generic training, but without being able to present very specific examples and experiences.

6.2.7 An example of discipline training could be “Collecting and managing live-person data in the social sciences”, delivered to postgraduate students as three two-hour workshops.

Subject

6.2.8 Subject level training raises an immediate question: what is a subject? The needs of, for example, synthetic chemists are very different to computational chemists. The appropriate scope for this level of training is likely to vary between institutions, faculties, and departments.

6.2.9 A key point for subject level training is that it must be developed with, and potentially delivered by, stakeholders from within the research discipline itself. At the subject level, it is necessary to understand the specific context in some detail. Examples and guidance should be practical, focusing on solving the real problems within the subject.

6.2.10 An example of subject-specific training could be “How to store and manage NMR spectrometer data for synthetic chemists”, delivered as a 4-page guide to postgraduate students and researchers.

Specific

6.2.11 Specific training or support is directly targeted to an individual researcher, dataset, or research project. The nature of this level of support is that it will probably consist either of peer-training

within a research group or department, or consultancy or mentoring approaches delivered by research data specialists. This level of training or support is hard to conceptualise from an organisational perspective. It may be appropriate to decide that the organisation cannot provide training and support to data creators at this level, but that it should be left to data controllers or research managers who are closer to the data to develop their own approaches, with support from the organisation.

6.3 Factors for success

6.3.1 Based on the research undertaken for this project, the study team has identified success factors for the development and delivery of training. There are also other broader factors, which need to be taken into account for the development and delivery of the training, that are likely to lead to the success of opening data initiatives.

6.3.2 Due to the immaturity of RDM and in particular opening research data, these should perhaps best be seen as factors that have or would have made an RDM *project* work well: they are likely to lead to improved data management and openness, but the evidence is not sufficient to demonstrate this unambiguously.

Success factors for the delivery and development of training

6.3.3 As discussed above, it is too early to identify a full range of success factors for the delivery and development of training in opening data. However, one important success factor has been identified. Namely, although the requirements of different disciplines are in many cases similar, there is clear feedback from a range of stakeholders regarding the way that content is delivered. In particular, it is important to ensure that the detailed training materials use language and examples that are relevant to the specific discipline/subject interests of those being trained. Put another way, although researchers may really need only Generic- or Discipline-level training, they respond better to Subject- or Specific-level training.

Success factors for opening data

6.3.4 These success factors will need to be considered in developing and delivering training courses. For example, it might be helpful to ensure that senior managers receive awareness training to help them support opening data activities. These factors are set out below.

Involving the right stakeholders

6.3.5 Any significant change in the processes or culture of an undertaking requires the involvement of a broad group of stakeholders. These all need to be aware of and support opening data. The key stakeholders for opening data are similar to those for managing data, and are likely to include:

- senior research manager (*eg* PVC research);
- senior information manager (*eg* CIO, director of library services, head of IT);
- staff from information management services (including library and IT staff);
- discipline stakeholders (*eg* head of science, head of chemistry, *etc*);
- internal and external experts;
- researchers.

Establishing the right environment

6.3.6 However, the study team's view is that any successful training will be established in a wider context of open data within the institution/discipline and more broadly, including:

- Institution/discipline:
  - Clear institutional policy and processes on opening data and a means for auditing compliance;
  - Awareness of opening data for all involved in research in the institution, including research managers and PIs;
  - Senior institutional/discipline commitment to support opening data, with funding and other resources;
  - Commitment that intelligently open data becomes the norm for research data;
  - Establishing and demonstrating benefits to researchers in opening their data at the institutional/discipline level.
- Broader context:
  - Clear and, ideally, harmonised funders' policies for opening data with responsibility given to institutions to ensure this happens;
  - Funders that audit compliance with their policy, and have meaningful sanctions available.

7 Key findings and recommendations

7.1 Findings

Opening data compared to research data management

7.1.1 This project was scoped to Opening, rather than Managing data. Opening data is radical in that it aims to provide more opportunities for exploitation and re-use of existing data with many consequent benefits. However, opening data itself builds on recent requirements and trends in HE for RDM. Opening data is simply one potential goal for managing data, with opening data being put at the heart of RDM. These are small extensions, rather than any fundamental change.

Current situation for training in opening data

7.1.2 While there is little training specifically intended for opening data, there is an increasing number of training courses and materials available to support RDM. However, there is a lack of depth in currently available training materials. Much of this training is at a level that provides only basic working knowledge and an awareness of the issues of RDM, and whereas "open data" may be mentioned, it is only addressed superficially. Few training materials are available for intermediate level skills and none for the explicit training of experts in opening data. There are some courses available to train information professionals (*ie* library or IT staff) who will subsequently have a role in training others or providing support. The individuals who currently have responsibilities for managing research data are in many cases responsible for working out how to do this: there is not yet a significant body of expertise and best practice.

7.1.3 A further gap is in breadth. There is strong demand for training materials using language and examples of a specific discipline/subject, materials for only a few disciplines/subjects have yet been developed, and this has been in an *ad hoc* basis.

7.1.4 None of the training examined was intended to be specific, aimed at a particular instrument or dataset. This is not surprising as such training would normally either be given within the research group, by an instrument supplier, or by a research collaboration. The stakeholders interviewed for this project described this as in many cases being tacit or informal knowledge.

7.1.5 The sample examined is small, and the field is too immature to identify which approaches work "best".

Support for opening data

7.1.6 It is early days opening data. As yet there is no clear view on the type and level of support in opening data needed by researchers. There is also no common view of the skills required by such support staff or how such support should be delivered.

7.1.7 Perhaps the greatest challenge to opening data appears to be meeting the needs of researchers across different disciplines, with strongly differing needs. Sections 2-6 describe a framework for how to address this challenge when designing training and support for opening data, within the broader questions of RDM.

7.2 Recommendations

7.2.1 Objectives for improving training and support provision for opening research data, including policy objectives are set out below. These objectives reflect a change of emphasis for RDM: nothing dramatically new is required, but open data – opportunities, risks, benefits, and practice – should be integrated within RDM training and support more strongly. Data cannot and should not always be open – but whether to open data needs to become part of the everyday decisions and processes for managing that data. In any case, data must always be managed.

7.2.2 In summary, these objectives are to change the tone of policy and training for RDM, by explicitly adopting an "open by default" posture. This will involve:

- Putting opening data at the heart of policy;
- Putting opening data at the heart of training;
- Deepening and broadening the training available;
- Identifying and disseminating best practice in opening data;
- Developing institutional and community support.

7.2.3 Many of these objectives are outside the direct responsibilities of RIDLs or its members. However, RIDLs is certainly in a position to influence what should happen.

7.2.4 The RIDLs group in consultation with Jisc and the DCC should consider and refine the objectives, develop a plan and implement it. Part of this process will require engagement and agreement with other stakeholders (*eg* Research Councils).

Specific actions

7.2.5 The objectives include suggestions for which organisations might need to be involved or have a key interest in it. While RIDLs cannot effect these objectives by itself (especially items 1 and 2 it certainly is in a position to influence what should happen.

7.2.6 Improving training and support provision for opening data could be achieved through the following broad objectives:

Put opening data at the heart of policy

7.2.7 Many funders and research institutions probably consider that their existing policies promote open data. The view of the stakeholders engaged for this project is that these policies are weak and imprecise, and leave a large margin of manoeuvre to individual data controllers (and therefore PIs). Policy provides an opportunity to drive openness, by stating explicitly that data must be opened unless there is a clear reason not to.

7.2.8 In turn, this policy focus drives the training delivered within an organisation.

Recommendation 1: Encourage funders to put 'opening data' at the heart of RDM. Funders should review their RDM policy and practice and ensure that: there is sufficient focus on opening data where appropriate; their requirements for data openness are explicit; and should consider incentives and sanctions to drive compliance. (RIDLs, RCUK, Jisc and BIS).

Recommendation 2: Encourage institutions to put 'opening data' at the heart of RDM. Institutions should review their RDM policy and practice and ensure that there is sufficient focus on opening data where appropriate. Institutional policies must be clear about the role of open

data, and must be backed by infrastructure and training to enable their implementation (RIDLs, RCUK, Jisc and BIS).

Recommendation 3: Encourage disciplinary bodies and relevant regional and national organisations to adopt policies that support opening data (RIDLs, RCUK, Jisc and BIS).

Put opening data at the heart of training

7.2.9 For training, as for policy concerns, if "open by default" is the goal, then all training should reflect this. In practice, the content will be very similar to RDM training, but by focusing on the goal of openness, the tone and focus may be slightly altered.

Recommendation 4: DCC and UKDA in particular, but other organisations too, should put 'opening data' at the heart of the relevant parts of their websites (DCC, UKDA).

Recommendation 5: Encourage institutions to implement 'opening data' where appropriate, including any necessary culture change (RIDLs, DCC, and UKDA).

Recommendation 6: Encourage institutions to put 'opening data' at the heart of updated training materials and resources (RIDLs, DCC, and UKDA).

Deepen and broaden the training

7.2.10 A range of training materials is already available at national and regional levels, and more are likely to be developed at disciplinary and institutional levels. These recommendations aim to support this development.

Recommendation 7: Encourage the development of training materials covering opening data beyond the awareness level and for intermediate and expert levels.

Recommendation 8: Encourage the development of training materials and resources for supporting the culture change aspects of implementing opening data (*eg* awareness training for senior research managers) (Jisc, DCC).

Recommendation 9: Prioritise those disciplines/subjects that are most resistant to change and encourage and seek funding to develop appropriate opening data materials in these areas (RIDLs and Jisc).

Recommendation 10: Establish and disseminate a catalogue of discipline-specific examples of RDM or opening data training (Jisc and DCC).

Recommendation 11: Establish best practice for capturing and managing tacit or informal knowledge of research processes and datasets. It may not be feasible to address this in all cases, but it should be considered (Jisc and DCC).

Identify and disseminate best practice in opening data

7.2.11 As best practice is not yet clear, it will be necessary to actively monitor developments as RDM and open data mature.

Recommendation 12: Best practice is not yet clear for training and support for opening or managing research data. Good practice is emerging, and the DCC should continue collating this (DCC).

Recommendation 13: Develop best practice for an institution's approach to implement opening data (*eg* approaches to culture change, the stakeholders that must be involved, *etc*) (Jisc and DCC).

Develop institutional and community support

7.2.12 The following recommendations aim to establish approaches to sharing and/or harmonising approaches to supporting opening data.

Recommendation 14: As experience grows, identify different approaches taken for supporting researchers in opening data and make this available (DCC).

Recommendation 15: Consider whether there is a need to establish or support a community of practice for research data specialists. A number of organisations, including CILIP, the DCC, Jisc, SCONUL, RLUK, and UCISA could potentially establish or support such a community of practice (RIDLs).

Recommendation 16: Engage with other stakeholders (*eg* BIS) in developing and adopting a career structure for research data specialists (RIDLs, BIS, RLUK).

A References

A.1 List of interviewees

A.1 This Annex lists the individuals who have contributed input to this report. The study team would like to thank and acknowledge all the contributors for their time and willing participation.

Interviewee	Role
Wendy White	University of Southampton and RLUK
Stéphane Goldstein	RIN
Simon Hodson	Jisc
Joy Davidson	DCC/HATII, University of Glasgow
Robin Rice	University of Edinburgh (MANTRA)
Cameron Neylon	PLOS
Stephen Grace	UEL
Carlos Morais Pires	Scientific Data infrastructure, EC
Chris Fleming	BIS
Dr Matthew Woollard	UKDA
Mark Thorley	RCUK
Brian Collins	UCL
James Wilson	University of Oxford
"IP lawyer"[25]	A major pharmaceutical company

Table 7-1: list of interviewees for the information gathering

A.2 Bibliography

Australian National Data Service, 2011. *What is research data?.* [Online]
Available at: http://ands.org.au/guides/what-is-research-data.html
[Accessed 2 May 2013].

Davidson, J., 2010. *DaMSSI (Data Management Skills Support Initiative) Final Report.* [Online]
Available at: http://www.dcc.ac.uk/sites/default/files/documents/JISCfinalreport_DaMSSI_FINAL.pdf
[Accessed 2 May 2013].

Digital Curation Centre, 2012. *How to Licence Research Data.* [Online]
Available at: http://www.dcc.ac.uk/resources/how-guides/license-research-data
[Accessed 2 May 2013].

Digital Curation Centre, 2013a. *Data Management Plans.* [Online]
Available at: http://www.dcc.ac.uk/resources/data-management-plans
[Accessed 2 May 2013].

Digital Curation Centre, 2013b. *DCC Curation Lifecycle Model.* [Online]
Available at: http://www.dcc.ac.uk/resources/curation-lifecycle-model
[Accessed 2 May 2013].

Digital Curation Centre, 2013c. *Digital curation standards.* [Online]
Available at: http://www.dcc.ac.uk/resources/standards
[Accessed 2 May 2013].

25 This interview was conducted with a guarantee of anonymity to the interviewee.

Digital Curation Centre, 2013d. *DCC.* [Online]
Available at: http://www.dcc.ac.uk/
[Accessed 2 May 2013].

Digital Curation Centre, 2013e. *Why preserve digital data?.* [Online]
Available at: http://www.dcc.ac.uk/digital-curation/why-preserve-digital-data
[Accessed 2 May 2013].

EPSRC, 2013a. *EPSRC Policy Framework on Research Data.* [Online]
Available at: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx
[Accessed 14 March 2013].

EPSRC, 2013b. *Impact, Timescales and Support.* [Online]
Available at: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/impact.aspx
[Accessed 2 May 2013].

EPSRC, 2013c. *Research Data Standards.* [Online]
Available at: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/scope.aspx
[Accessed 2 May 2013].

GEOSS, 2012. *GEOSS Data Sharing Action Plan.* [Online]
Available at:
http://www.earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Rev2.pdf
[Accessed 2 May 2013].

Hammond, M., Hawtin, R. & Davies, C., 2010. *Advanced ICT support for researchers: Review of Models.* [Online]
Available at:
http://www.jisc.ac.uk/media/documents/programmes/researchcommunities/Finalreport_SupportResearchers.pdf
[Accessed 2 May 2013].

Houghton, J., 2011. *Costs and Benefits of Data Provision.* [Online]
Available at: http://ands.org.au/resource/houghton-cost-benefit-study.pdf
[Accessed 2 May 2013].

Jisc, 2013a. *Managing Research Data (JISCMRD).* [Online]
Available at: http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx
[Accessed 2 May 2013].

Jisc, 2013b. *Managing Research Data Programme 2011-13.* [Online]
Available at:
http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx
[Accessed 2 May 2013].

Jisc, 2013. *Managing Creative Arts Research Data (MCARD) training module.* [Online]
Available at: http://www.projectcairo.org/node/9
[Accessed 2 May 2013].

Jones, S., Pryor, G. & Whyte, A., 2013. *How to Develop RDM Services - a guide for HEIs.* [Online]
Available at: http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services
[Accessed 2 May 2013].

Lyon, L., 2009. *Open Science at Web-Scale: Optimising Participation and Predictive Potential—Consultative Report.* [Online]
Available at: http://www.jisc.ac.uk/media/documents/publications/research/2009/open-science-report-6nov09-final-sentojisc.pdf
[Accessed 2 May 2013].

Lyon, L., 2012. The Informatics Transform: Re-Engineering Libraries for the Data Decade. *The International Journal of Digital Curation,* 7(1), pp. 126-138.

Monash University, 2010. *Research Data Management Policy.* [Online]
Available at: http://policy.monash.edu.au/policy-bank/academic/research/research-data-management-policy.html
[Accessed 2 May 2013].

Northumbria University, 2011. *DATUM for Health Training Materials.* [Online]
Available at:

http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/datum/health/materials/?view=Standard
[Accessed 2 May 2013].

Project Cairo, 2011. *Project Cairo.* [Online]
Available at: http://www.projectcairo.org/
[Accessed 2 May 2013].

Pryor, G. & Donnolley, M., 2009. Skilling Up to Do Data: Whose Role, Responsibility, Career?. *The International Journal of Digital Curation,* 2(4), pp. 158-170.

Queen Mary University of London, 2013. *Research Data Management at the Centre for Digital Music.* [Online]
Available at: http://rdm.c4dm.eecs.qmul.ac.uk/
[Accessed 2 May 2013].

RDMRose, 2013. *About RDMRose.* [Online]
Available at: http://rdmrose.group.shef.ac.uk/
[Accessed 2 May 2013].

Research Councils UK, 2012. *RCUK common principles on data policy.* [Online]
Available at: http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
[Accessed 17 January 2013].

Research Councils UK, 2013. *RCUK Policy on Open Access.* [Online]
Available at: http://www.rcuk.ac.uk/documents/documents/RCUKOpenAccessPolicy.pdf
[Accessed 2 May 2013].

RIN, 2010. *Data management, information literacy and DaMSSI.* [Online]
Available at: http://www.rin.ac.uk/our-work/researcher-development-and-skills/data-management-and-information-literacy
[Accessed 2 May 2013].

RIN, 2010. *Open to All? Case studies of openness in research.* [Online]
Available at: http://www.rin.ac.uk/system/files/attachments/NESTA-RIN_Open_Science_V01_0.pdf
[Accessed 2 May 2013].

Scott, M., Boardman, R. P., Reed, P. A. & Cox, S. J., 2013. *Introducing research data.* [Online]
Available at: http://eprints.soton.ac.uk/338816/
[Accessed 2 May 2013].

SpaceRef, 2013. *Subcommittee Discusses Benefits of Open Access to Research Data and Challenges in Implementation.* [Online]
Available at: http://www.spaceref.com/news/viewpr.html?pid=40243
[Accessed 2 May 2013].

The Cabinet Office, 2012. *Open Data White Paper: Unleashing the Potential.* [Online]
Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf
[Accessed 2 May 2013].

The Royal Society, 2012. *Science as an open enterprise.* [Online]
Available at: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf
[Accessed 17 January 2013].

UK Data Archive, 2011. *Managing and Sharing Data: Best Practice for Researchers.* [Online]
Available at: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
[Accessed 2 May 2013].

UK Data Archive, 2013. [Online]
Available at: http://www.data-archive.ac.uk/
[Accessed 2 May 2013].

UK Data Archive, 2013. *Create and Manage Data.* [Online]
Available at: http://www.data-archive.ac.uk/create-manage
[Accessed 2 May 2013].

UK Data Service, 2013. *Secure Access.* [Online]
Available at: http://ukdataservice.ac.uk/get-data/secure-access.aspx
[Accessed 2 May 2013].

University of Cambridge, 2013a. *Data Management for Archaeology.* [Online]
Available at: http://www.lib.cam.ac.uk/dataman/datatrain/archintro.html
[Accessed 2 May 2013].

University of Cambridge, 2013b. *Data Management for Social Anthropology - Overview.* [Online]
Available at: http://www.lib.cam.ac.uk/dataman/datatrain/socanthintro.html
[Accessed 2 May 2013].

University of East London, 2013. *TraD - Training for Data Management at UEL.* [Online]
Available at: http://www.uel.ac.uk/trad/
[Accessed 2 May 2013].

University of Edinburgh, 2012. *Research Data MANTRA project.* [Online]
Available at: http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/mantra
[Accessed 2 May 2013].

University of Edinburgh, 2013. *Do-It-Yourself Research Data Management Training Kit for Librarians.* [Online]
Available at: Do-It-Yourself Research Data Management Training Kit for Librarians
[Accessed 2 May 2013].

University of Hertfordshire, 2012. [Online]
Available at: http://research-data-toolkit.herts.ac.uk/document/rdmtpa-presentation-jiscmrd-progress-workshop-nottingham-2012
[Accessed 2 May 2013].

University of Leeds, 2013. *Leeds Research Data Management Pilot RoaDMaP.* [Online]
Available at: http://library.leeds.ac.uk/roadmap-project
[Accessed 2 May 2013].

University of Oxford, 2012a. *Research Data Management.* [Online]
Available at: http://www.admin.ox.ac.uk/rdm/
[Accessed 2 May 2013].

University of Oxford, 2012b. *Research Data Management Induction Resources.* [Online]
Available at: http://damaro.oucs.ox.ac.uk/induction.xml
[Accessed 2 May 2013].

University of Oxford, 2013. *RDM for sciences training survey results.* [Online]
Available at: http://damaro.oucs.ox.ac.uk/docs/RDM%20for%20sciences%20-%20training%20survey%20results.xlsx
[Accessed 2 May 2013].

Vitae, 2013. *Vitae Researcher Development Framework.* [Online]
Available at: www.vitae.ac.uk/rdf
[Zugriff am 18 June 2013].

Xu, G., 2012. *The benefits of open data - evidence from economic research.* [Online]
Available at: http://openeconomics.net/2012/10/03/the-benefits-of-open-data-evidence-from-economic-research/
[Accessed 2 May 2013].