



Second EUDAT Conference, October 2013 Data Management Plans and Certification

Motivation: increasing importance of Data Management Planning

Simon Lambert
Scientific Computing Department
STFC Rutherford Appleton Laboratory
UK

Why data management planning?

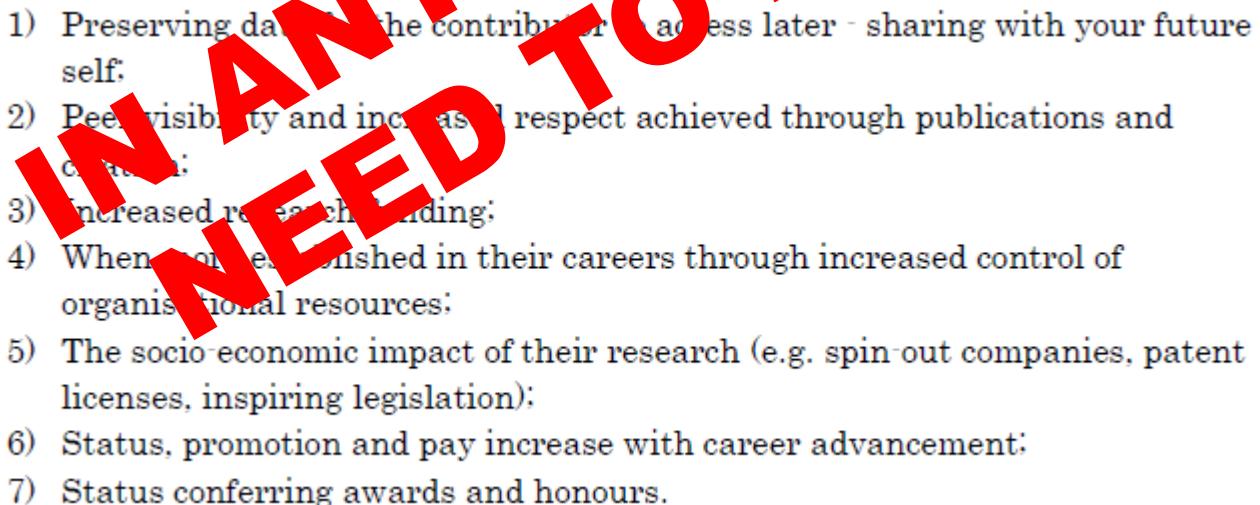
- Data management follows from openness of data
- Meaningful openness of data requires management
- Principles, policies and mandates
 - Royal Society, “Science as an open enterprise”
 - G8 Open Data Charter
 - European Commission Recommendation on access to and preservation of scientific information
 - RCUK Data Principles
 - Funders’ requirements



Science & Technology
Facilities Council

Why data management planning?

- But it's not only about obligation
- The ODE study: model of drivers and barriers to data sharing
- Individual contributor incentives

- 
- 1) Preserving data for the contributor to access later - sharing with your future self;
 - 2) Peer visibility and increased respect achieved through publications and citations;
 - 3) Increased research funding;
 - 4) When researchers progress in their careers through increased control of organisational resources;
 - 5) The socio-economic impact of their research (e.g. spin-out companies, patent licenses, inspiring legislation);
 - 6) Status, promotion and pay increase with career advancement;
 - 7) Status conferring awards and honours.



Science & Technology
Facilities Council

Related areas

- Data management and preservation
 - Reuse now, reuse in future
 - Provenance and authenticity
- Data management and repository audit/certification
 - But data management does not (necessarily) mean running the repository!
- Data management and open access (to publications)



Science & Technology
Facilities Council

Domain differences

- Different scientific domains, different needs and approaches
 - What kind of science do you do?
 - What is the culture?
 - What is the existing infrastructure?
- But there is one common question: who pays beyond the lifetime of the grant? (Berman & Cerf)



Science & Technology
Facilities Council

The case of “Big science”

- MaRDI-Gross project
 - Funded by JISC in UK
 - Final report August 2012

DMP Planning for Big Science Projects
MaRDI-Gross project
v1.0, 2012 August 17

Project: Managing Research Data Infrastructures for Big Science (MaRDI-Gross, <http://eurl.org/na/projects/mardi-gross>)
Release: v1.0, 2012 August 17
URL: <https://eurl.org/na/projects/mardi-gross/report>
Authors: Juan Bicarregui¹, Norman Gray², Rob Henderson³, Roger Jones², Simon Lambert¹, and Brian Matthews¹
¹E-Science Centre, STFC, UK; ²School of Physics and Astronomy, University of Glasgow, UK; ³Department of Physics, University of Lancaster, UK
Distribution: Public

Abstract

This report exists to provide high-level guidance for the strategic and engineering development of **Data Management & Preservation (DMP)** plans for ‘Big Science’ data.

Although the report’s nominal audience is therefore rather narrow, we intend the document to be of use to other planners and data architects who wish to implement good practice in this area. For the purposes of this report, we presume that the reader is broadly persuaded (by external fact if nothing else) of the need to preserve research data appropriately, and that they have both sophisticated technical support and the budget to support developments.

The goal of the document is not to provide mechanically applicable recipes, but to allow the user to develop and lead a high-level plan which is appropriate to their organisation. Throughout, the report is informed where appropriate by the OAIS reference model.

This report was funded by JISC in 2011–2012, as part of the RDMP strand of the JISC programme **Managing Research Data**. For document history, see p. [2](#).

Copyright 2012, University of Lancaster

The starting point

- “The demand for principled data management and data sharing is a reasonable one”
- “A reasonable framework for at least approaching the problem already exists in Open Archival Information System (OAIS)”
- “There is a bounded set of resources which, if mastered by the reader, will allow them to produce a project DMP plan which is practically acceptable to the project, and discharges the principled demands of the funder and of society”



Science & Technology
Facilities Council

Some issues for big science

- What is the designated community?
- What are the preservation goals?
- What data need *not* be kept?
- What is the data release schedule?



Science & Technology
Facilities Council

The position of STFC

- STFC has data principles, derived from the RCUK principles
 - Data = raw + derived + published
 - “Data management plans should exist for all data within the scope of the policy ... Proposals for grant funding should include a data management plan ... Data management plans should follow relevant national and international recommendations for best practice.”
 - “Data resulting from publicly funded research should be made publicly available after a limited period, unless there are specific reasons (e.g. legislation, ethical, privacy, security) why this should not happen.”



Science & Technology
Facilities Council

STFC's recommendations for good practice

- Formulated following guidance provided by the Digital Curation Centre
- Normally expect data to be managed through an institutional repository (in broad sense)
- Cover all data expected to be produced, from “raw” to “published”
- Specify which data are to be deposited in a repository, where and for how long, with justification
- Make sure the data is accompanied by sufficient metadata to enable reuse



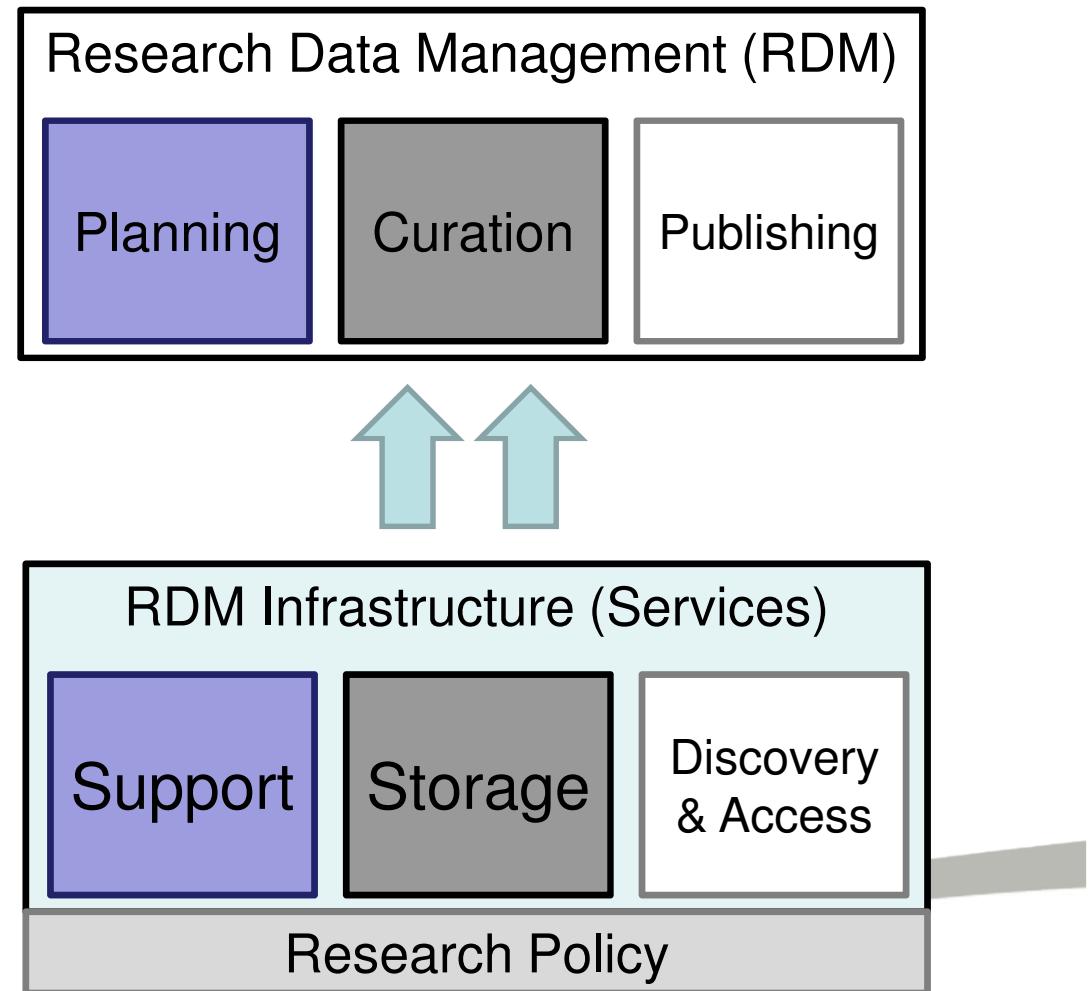
Science & Technology
Facilities Council

An example from Australia

University of New
South Wales

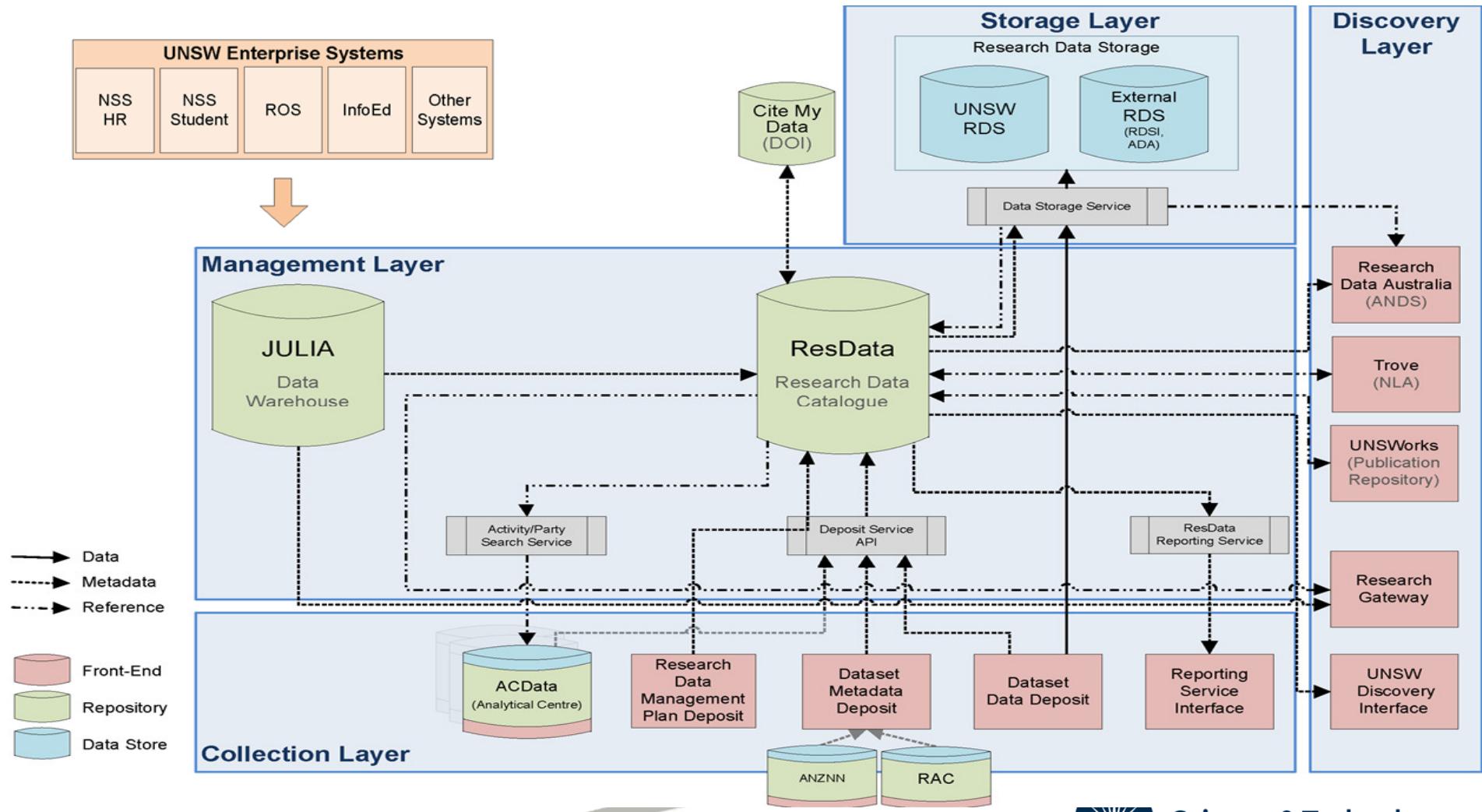
“An Institutional
Infrastructure for
Facilitating Research
Data Management”

Arif Shaon et al.,
eResearch
Australasia 2013,
Brisbane



Science & Technology
Facilities Council

A full data management solution



Science & Technology
Facilities Council

DMP related to repository audit/certification

- Data management planning for a project is not the same as running a trustworthy digital repository ...
- ... but there are things in common.
- Standard “Audit and Certification of Trustworthy Digital Repositories” (ISO 16363, CCSDS 652.0-M-1)
- Three areas covered:
 - Organisational infrastructure
 - Digital object management
 - Infrastructure and security risk management



Science & Technology
Facilities Council

DMP related to repository audit/certification

- Formulated in terms of metrics

- 3.1.3 The repository shall have a Collection Policy or other document that specifies the type of information it will preserve, retain, manage, and provide access to.
- 3.3.1 The repository shall have defined its Designated Community and associated knowledge base(s) and shall have these definitions appropriately accessible.
- 4.1.1 The repository shall identify the Content Information and the Information Properties that the repository will preserve.



Science & Technology
Facilities Council

The end

- ODE www.ode-project.eu
- APARSEN www.aparsen.eu
- MaRDI-Gross <http://mardigross.jiscinvolve.org/wp/>
- PaN-data www.pan-data.eu
- Repository audit/certification
www.digitalrepositoryauditandcertification.org



Science & Technology
Facilities Council