

Development of a Pilot Data Management Infrastructure for Biomedical Researchers at University of Manchester – Approach, Findings, Challenges and Outlook of the MaDAM Project

Meik Poschen, June Finch, Rob Procter, Mhorag Goff
Manchester eResearch Centre (MeRC), University of Manchester

Mary McDerby, Simon Collins
Research Computing supported by IT Services for Research,
University of Manchester

Jon Besson, Lorraine Beard, Tom Grahame
The John Rylands University Library (JRUL), University of Manchester

Funded by the

JISC

+ University of Manchester
Contribution

MaDAM Project Overview

Aim: To produce a technical & governance solution based on researchers' requirements with flexibility to meet needs across multiple research groups / disciplines and taking into account the institutional landscape and its policies (October 2009 – June 2011).

Rationale:

- Researchers need to be supported to manage their data well (day-to-day) and comply with legal and funder policies.
- Funders want to ensure public money spent on research is maximised → this means ensuring research data is preserved for reuse.
- Potential future value in data assets needs to be preserved.

Background: No existing institutional repository or strategy for management of research data – BUT the MaDAM Pilot became part of a wider endeavour at University of Manchester to develop such.

The MaDAM Solution will..

- Provide trusted secure storage to reduce risks of data loss and to adhere to funders' (new) retention policies
- Make metadata visible and searchable - enable annotation of data including ad hoc context and 'notes to self'
- Facilitate easier, more secure owner-controlled data sharing
- Reduce redundancy: structured space, enable linking
- Maintain media and format accessibility for long term reuse
- Ensure that technical and non-technical solutions for managing and sharing data will fit in with the research& data lifecycle, diverse working practices, cultures and disciplines

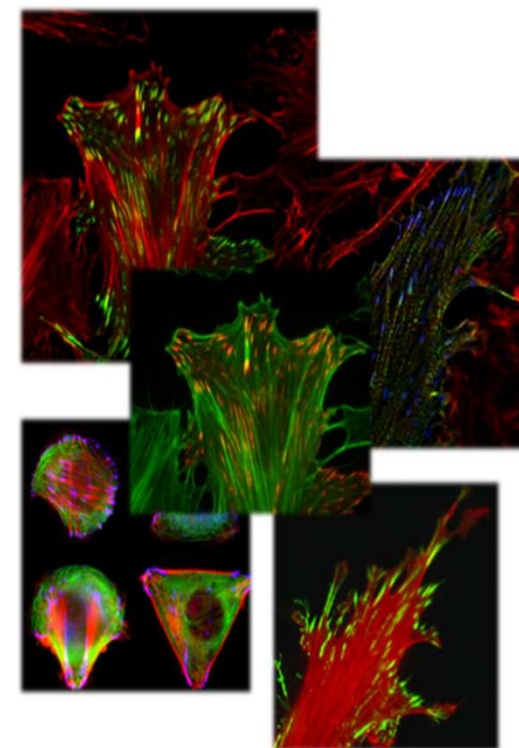
MaDAM Domains & Pilot User Groups

Biomedical Domain at University of Manchester

1. Life Sciences Electron and Standard Microscopy: 4 groups with 8 active core users plus occasional users
2. Medical Science MRI Neuropsychiatry Unit: 1 group/5 users

Images as main Research Objects, but also other data types (text docs, metadata, statistical and output data)

The work with the pilot user groups was further complemented by information/requirements gathered from additional researchers and PIs within the domain, IT and experimental officers as well as research and data policy managers.

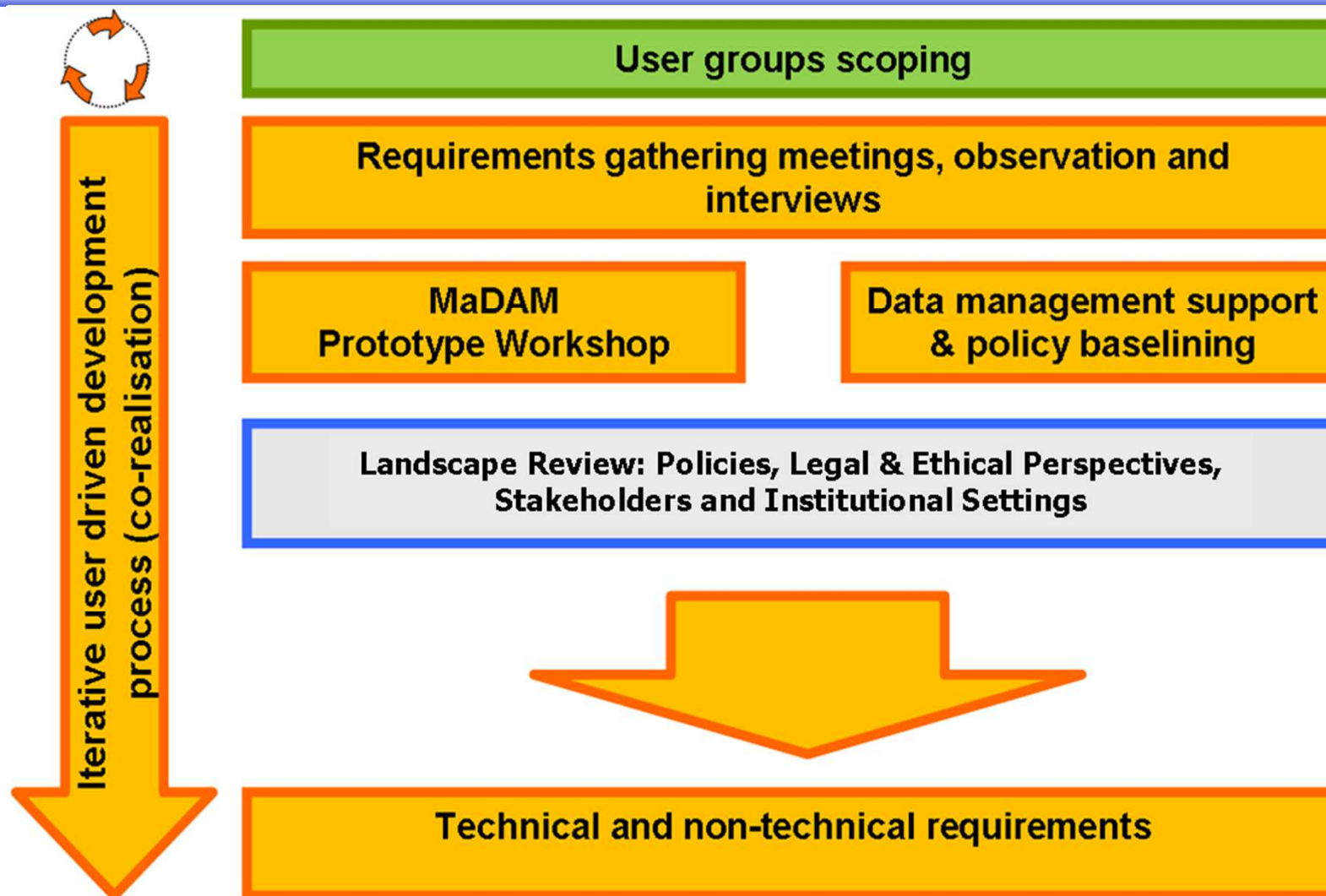


Up to 12 different file types
From 0.5MB to 17GB/file

'Raw data'

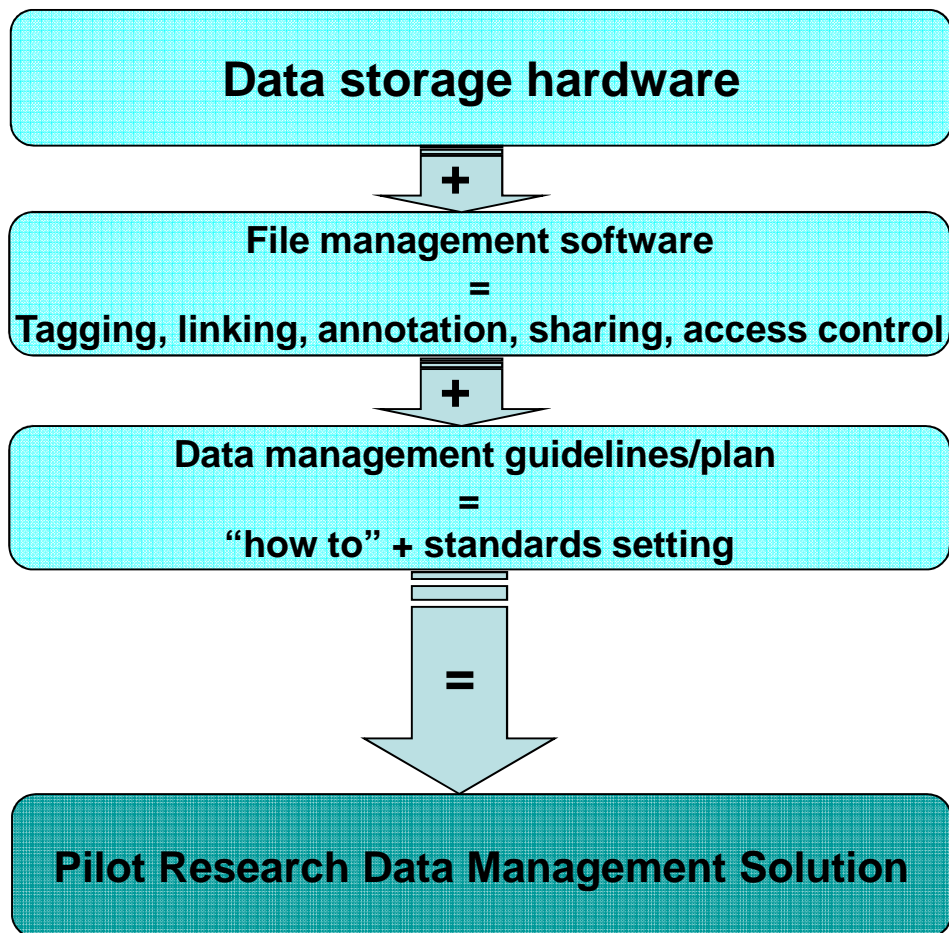
- > Microscope samples: single run creates any image set from 1-200 GB
- > MRI brain scans: usually one study consists of 20-40 GB

MaDAM 'Method-flow'



MaDAM Pilot Overview

Aim: Pilot Research Data Management Solution



Many angles to cover:

- Research Practice
- Discipline/Domain
- Technical Solution
- Policies/Procedures
- Institutional Settings
(Stakeholders & Infrastructure)
- Funding Landscape
- Cost-Benefit Analysis

Pilot Users: Findings

- No official backup policies to protect against loss of data
 - Decentralized & fragmented storage (USB sticks, optical disks)
 - Limited ability to share data internally or externally
 - High levels of redundant data (duplicate copies)
 - No structured annotation of data
 - Limited search capabilities
 - Limited means to disseminate data
 - No archiving policies to guarantee long term curation
- waste of time – risk of data loss – finding, reuse & sharing difficult – clogging of valuable storage space**

Main Requirements

- Generic need for trusted, structured central storage with auto-back up and improved capabilities for reuse, sharing, searching and overall management of data files.
- The prototype provides a navigation structure based on researchers' projects and experiments, centralized and backed up data storage, access rights, linkage and annotation of research data and a search function.
- Need for good practices in data management and digital curation policies to tie in with researchers' actual research practice, institutional settings and cultures.

Final Web-based MaDAM Pilot System

MANCHESTER 1824

Search Welcome: Meik Poschen Madam Group Logout

Data | Setup | Help

Home Page Search Calendar Bookmarks Latest Data Recycle Bin Feedback

Checkout > PubsA

Explorer

- Checkitout (W)
- BenefitsCaseStudy
- BusinessCase
- ProjectReports
- PubsA
 - draft = 0
 - internal doc = 1
 - published = final
- MaDAM-BusinessCaseFull-draft...
- PubsConf
- Madam Project (W)
- V4 Testing (W)

Upcoming reminders (next 2 weeks)

No review dates upcoming

PubsA

[Update](#) [Delete](#)

Name PubsA

Comments project docs

Creator Meik Poschen **Owner** Meik Poschen

Create Date 17-MAY-2011 11:07 **Review Date**

Status Available **Folder Type** Publication

Folder Size .09M **File Count** 1 (0)

Folder Details

Files

Name	Created On	Mime Type	Size / MB	Created By
MaDAM-BusinessCaseFull-draf...	17-MAY-2011 11:32	application/msword	0.09	Meik Poschen

1 - 1

Sub Folders

Attributes

Keywords

Notes

Help

Page Help

Create

Create "Folder"

Folder Actions

- Upload Files
- Download Folder
- Create Note
- Email Folder Link
- Bookmark Folder
- Create Shortcut
- Move Folder
- File View

Attributes

- Add Template
- Add Adhoc

MaDAM Pilot: Thumbnails

[Home Page](#) | [Search](#) | [Calendar](#) | [Bookmarks](#) | [Feedback](#)

File List

- ☐ FRAPb1Series40_t07_ch00.tif
- ☐ FRAPb1Series40_t01_ch01.tif
- ☐ FRAPBleachSeries38.xml
- ☐ FRAPBleachSeries38_t1_ch01.tif
- ☐ FRAPb1Series40_t19_ch00.tif
- ☐ FRAPBleachSeries38_t2_ch01.tif
- ☐ FRAPb1Series40_t06_ch01.tif
- ☐ FRAPb1Series40_t15_ch01.tif
- ☐ FRAPb1Series40_t19_ch01.tif
- ☐ FRAPPreSeries39_t1_ch00.tif
- ☐ FRAPb1Series40_t14_ch00.tif
- ☐ FRAPb1Series40_t06_ch00.tif
- ☐ FRAPPreSeries39_t3_ch00.tif
- ☐ FRAPb1Series40_t11_ch01.tif
- ☐ FRAPb1Series40_t00_ch00.tif
- ☐ FRAPb1Series40_t20_ch00.tif
- ☐ FRAPb1Series40.xml
- ☐ FRAPb1Series40_t16_ch00.tif
- ☐ FRAPb1Series40_t10_ch01.tif
- ☐ FRAPb1Series40_t17_ch00.tif
- ☐ FRAPPreSeries39.xml

Thumbnails - Number of Columns

Thumbnail grid showing 15 images in a 3x5 layout. The images are fluorescence microscopy images showing cellular structures in red and green. Each thumbnail has a timestamp in the bottom right corner.

MaDAM Pilot: Metadata

MANCHESTER 1824

Search q Welcome: MSASSMP6 Meik Poschen Madam Group Logout
Data | Setup | Help

My Details Projects Metadata

Create / Modify Attributes Created / Available Folder Types Created / Available Templates Groups I belong to

Cancel Delete Save

<input type="checkbox"/>	Name	Units	List of Values	Creator	Group
<input type="checkbox"/>	Chemical	mols	1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Country		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Decay Weight		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Description		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Dig Site		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Est Date		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Find Type		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Found by		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Genre		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Geometric Mean		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Goeyness	yucks	1 2 3	Meik Poschen	Madam Group
<input type="checkbox"/>	Instrument		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Journal		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Location		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Method		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Sample type		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Shelf Level		1 2 3	Simon Collins	Madam Group
<input type="checkbox"/>	Shelf Life		1 2 3	Simon Collins	Madam Group

Help
Page Help

Folder Type
Create Type

Templates
Create Template

Group
Create Group

MaDAM and eScholar








Manchester eScholar Services have the mission to

- “sustain and enhance the research reputations of individuals and organisations affiliated with The University of Manchester”
- “enhance the global research community's ability to access The University of Manchester's research outputs”

For the MaDAM project eScholar will

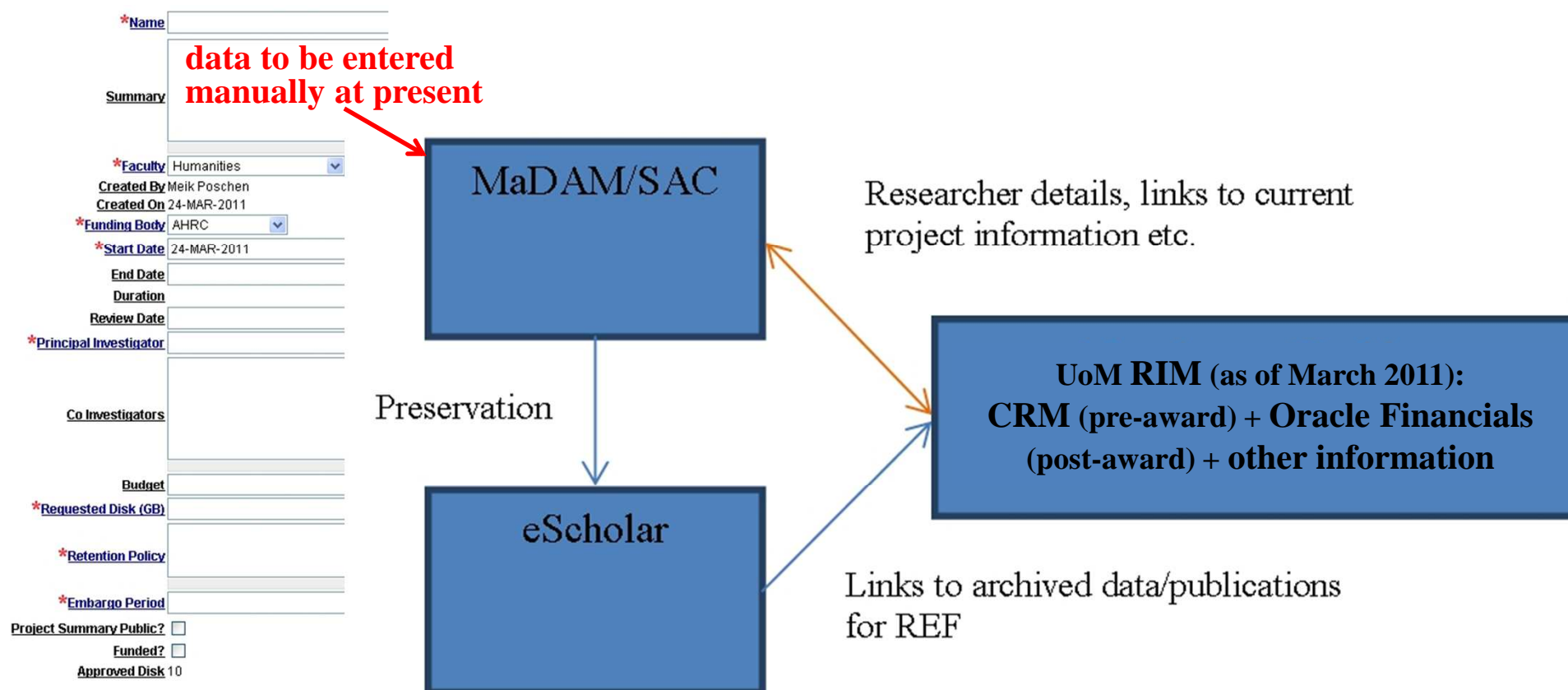
- provide a resolvable end point for publishing of data to the wider research community
- be a searchable archive for MaDAM data allowing the University to meet its retention commitments

My eScholar login

-  **store, manage and preserve** your journal articles, books, working papers and other intellectual assets in a digital form
-  **deposit fulltext files and comprehensive metadata** to maximise the impact and potential of your research findings
-  **capture/import metadata** from external databases e.g. PubMed, EndNote, Reference Manager
-  **"Lite-cite" submit** your publication details for later editing and display
-  **display and maintain a bibliography** of your publications on your personal and other websites, *coming soon!*
-  **access, cite, bookmark and share** your scholarly work, *coming soon!*
-  **create a list of publications** for a CV, a report for a funding body and other administrative tasks, *coming soon!*

Integrating Research Information Management Data

- MaDAM is currently exploring the integration of UoM RIM data (auto-retrieval)
- UoM's RIM environment itself is in the process of being linked more seamlessly



Challenges & Observations (1)

Existing institutional and faculty support for researchers – including IT Services, Research Offices and people managing the core facilities and scanners – directly and indirectly contribute to research data management.

Engagement of these support structures will be essential to policy development and are critical to sustainability in terms of both buy in and the potential for capacity building in their services.

A cultural change is/might be needed for the proper support of domain specific data management plans, research practices and research management policies in general, and this, inevitably, will take time (and won't be easy!). High level institutional support is crucial, too!

Challenges & Observations (2)

- Making the best use of pilot users' limited time
- Managing the expectations of UoM and external interested users
- Ensuring that solutions would fall inline with working practices
- Dealing with a diverse & fragmented landscape (policies, funders)
- Engagement of institutional support structures is essential
- The use of Research Data Management within the actual Working Practice is evolving – emerging patterns and behaviour:

How much storage will research groups/researchers need over what time – how long has data to be kept in an active or easy accessible state for reuse or sharing?

How will the relationship between new policies and research practices develop?

How will dissemination practices and hence Scholarly Communications develop or change?

Transition: from MaDAM to MiSS

- MaDAM: successful in addressing the needs of its user groups and in developing a pilot infrastructure, which is live, maintained and actively utilised by its pilot user base.
- Madam's outputs & findings (researchers' benefits) together with being part of an initiative for a sustainable University-wide Research Data Management Service helped secure funding for the successor project MiSS (MaDAM into Sustainable Service).
- MiSS (MaDAM into Sustainable Service) will be building on MaDAM, although it is more a transitional project than a continuation which will move the pilot into a sustainable service within the University's new technical framework at the end of its lifetime in March 2013.

Picking up from MaDAM..

..in MiSS on endeavours started but not completed/evaluated within pilot's lifetime/remit:

- Researchers' dissemination practices...
- ...and full integration of eScholar (MaDAM: proof of concept)
- Automatic population of research information data (accounting, grants)
- Enabling better metadata ingestion from different sources in various disciplines (via community input)

Delivering MiSS

MiSS will be delivering a service which will include

- 1) Rebuilding and integrating the MaDAM technical service infrastructure, making it more generic but tailorable (domain/discipline specific plug-ins/plugin points),
- 2) providing a Research Data Management Policy (incl. DMPs), along with a supporting Service, and
- 3) integrating with the necessary human infrastructure, addressing needs across UoM.

MiSS User Community

For the domain specific user community the project includes five **research groups** covering all four faculties.

Academic Champions in

- Life Sciences
- Engineering and Physical Science
- Medical and Human Sciences
- Humanities/applied quantitative social research

We will furthermore set up a **user committee** open to all research disciplines at UoM to balance specific with generic needs.

MiSS Challenges

- Bridge between and cater for generic and specific needs, making the RDML easy usable but open enough for specific tools and automated data ingestion by providing 'plug-in points'
- Integration with the Manchester Technical IT Services infrastructure, which is evolving concurrently (Manchester Working Environment, MWE)
- Balance researcher, internal and external research data management needs and policies (research cultures & work practices, University structure with research offices & faculties and funder requirements)
- Beware scope-creep and keep all stakeholders in line towards delivering a RDML Service in March 2013

Many Thanks!

MaDAM

<http://www.merc.ac.uk/?q=MaDAM> (outputs)

<http://www.library.manchester.ac.uk/aboutus/projects/madam/>

MiSS

<http://www.manchester.ac.uk/miss/>
@MISS_RDM

Meik Poschen

meik.poschen@manchester.ac.uk

<http://www.merc.ac.uk/?q=Meik>

DMP Online and DMPTool: Different strategies towards a shared goal



Martin Donnelly
University of Edinburgh
martin.donnelly@ed.ac.uk
@mkdDCC

Andrew Sallans
University of Virginia
als9q@virginia.edu
@asallans



International Digital Curation Conference
Bristol, England – 7 December 2011



Running order

1. Impetus for DMP Online development
2. Background to the US-UK collaboration & impetus for DMP Tool
3. Differences in requirements
4. Tabular comparison of tools and functionalities
5. Partnerships (DMP Online | DMP Tool)
6. Outcomes achieved so far (DMP Online | DMP Tool)
7. Next steps (DMP Online | DMP Tool)
8. Joint next steps, shared goals and shared experiences

1. Impetus for development

- DMP support developed to meet funders' increasing requirements
- Growth of DMP Checklist, and feedback received
- Need for smarter solution than a paper checklist
- Desire to provide more tailored guidance at point of need

2. Background to the US-UK collaboration & Impetus for DMP Tool

- Announcement of NSF mandate
- Conversations at IDCC in Chicago (December 2010)
- Experimentation with DMP Online code – not suitable
- Decision to develop a separate tool

3. Differences in requirements

- Mapping to a checklist vs direct answering of funder requirements (philosophical)
- Institutional 'ownership' vs national service (cultural)
- Authentication processes (technical / organisational)
- Tabular comparison...

4. Comparison of tools' functionalities

Feature	DMP Online (v3.0)	DMP Tool (v1.0)
Authentication	Username and password	Shibboleth or username/password
Associated guidance	Can be customised for each template	Funder + Institution
Templates	Multiple	Funder + Institution
Licensing	Free to use. Hope to make Open Source in time.	Open Source (Bitbucket)
Logic / approach	Funder requirements mapped to a generic Checklist (rules-based)	Direct answering of verbatim requirements
Export	PDF, HTML, CSV, TXT, XML, RTF, DOCX	PDF, TXT, RTF, URL
Sharing	Granular permissions to enable sharing / transfer of plans	Read-only sharing via URL
Hosting	Virtual environment at Edinburgh. (JANET in future?)	Hosted by CDL, but anyone can run their own instance (minus the database content)
Liaison with funders	Close to endorsement from multiple funders	Viewed positively by funders

5a. DMP Online partnerships

- Checklist and interface developed via iterative community consultation process
- Recommended and supported by JISC
- Generic guidance developed with the UK Data Archive
- More tailored guidance being developed with the funders / UKDA. (Endorsement by funders also being sought.)
- Disciplinary guidance and worked examples developed with JISC MRD projects (DMT Psych, DATUM for Health...)
- More JISC MRD projects adapting the tool for institutional contexts
- Partnering with DMP Tool, and involved in conversations with European Commission, Australia...

5b. DMP Tool Partners

- CDL/UC3
- UCSD
- UCLA
- Univ of Illinois
- Smithsonian Institution
- Univ of Virginia
- DataONE
- Digital Curation Centre



6a. Outcomes achieved so far (DMP Online)

- At December 2011, we have just under 800 registered users
- Wide differential in creation of plans for different funders
- DCC's data management planning resources are cited in the STFC data policy and supported by other funders
- Checklist has been adapted by institutions on both sides of Atlantic

6b. Outcomes achieved so far (DMP Tool)

- At December 2011, we have 548 registered users
- 14 funder templates in place
- 10 Additional institutions with Shibboleth configured
- DataONE workshop for scientists (Santa Fe, May 2011)
- NSF data management workshop position paper (Princeton, July 2011)
- Beta testing at ESA meeting (Austin, TX, August 2011)
- Other presentations at DLF, IDCC, CNI
- Online webinar organized by CDL (November 2011), linked to tool launch

7a. Next steps for DMP Online

- DMP Online v3.0 will launch in early 2012
 - Overlaying multiple templates
 - Granular sharing of plans
 - Endorsement from funders (TBC)
 - Several other new features
- Closer integration with a greater number of systems, including RCUK's J-eS and the JISC MRD projects
- Standards development?

7b. Next steps for DMP Tool

- Continue ongoing weekly conference calls
- Gathering feedback and survey results
- Conversation with Funders about data management plans and the integration of the DMP Tool
- All hands meeting January 2012
 - Discussion of future development priorities
 - Identify resources for priorities (programmers & funding)
 - Discussion of potential organizational and project models

8. Joint next steps, shared goals and shared experiences

- Continuing collaboration and alignment of effort, keeping channels of communication open to share good practice
- Expansion to include other countries and funders
- A commitment to trying to address local/national needs while also recognizing that research is increasingly global and that researchers have to face DMP requirements on many fronts
- Recombining the tools is a ongoing future goal...

Thank you



<http://www.dcc.ac.uk/dmponline>



<https://dmp.cdlib.org/>



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 UK: Scotland License.

To view a copy of this license, (a) visit <http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/>; or (b) send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.





... because good research needs good data

Data Management Skills Support Initiative (DaMSSI)

This work jointly funded by the Research Information Network (RIN)
and Joint Information Systems Committee (JISC)



Laura Molloy

Humanities Advanced Technology and Information Institute (HATII)
University of Glasgow



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 UK: Scotland License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/2.5/scotland/>; or,
(b) send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.



... because good research needs good data

The Data Management Skills Support Initiative ('DaMSSI')

- JISC Managing Research Data programme (04/2010 call)
- JISC Research Data Management Training Materials ('RDMTrain') strand – support project
- 5 RDMTrain projects – postgraduate training materials – 6 disciplines:
 - 'CAiRO' at Bristol (creative and performing arts)
 - 'MANTRA' at Edinburgh (social and political science, clinical psychology and geosciences)
 - 'DATUM for Health' at Northumbria (health studies)
 - 'DMTpsych' at York (psychological sciences)
 - 'DataTrain' at Cambridge (social anthropology and archaeology).



... because good research needs good data

Background

- Research data management skills not currently widely embedded in UK postgraduate training
- Agreement needed on what constitutes a basic set of postgraduate data management skills
 - Consistent descriptions of skills
 - Framework that supports progression of skills over time
- Two recently developed models may be able to help:
 - Vitae's Researcher Development Framework
 - 'Research lens' for SCONUL's Seven Pillars model



... because good research needs good data

Key work areas

- Support for RDMTrain projects
- Mapping of projects' training materials
- Comparison / synthesis of mappings from the projects
- Recommendations for Vitae and SCONUL
- Career profiles
- Recommendations for professional bodies and LIS course providers



... because good research needs good data

Findings

- Usefulness of discipline-specific definitions, examples and exercises;
- Students like DMPs – with discipline-specific interpretation;
- Data management / information-handling language a barrier for researchers;
- Balance needed: discipline-specific detail / brief and concise training;
- Timing matters: early (but not too early); embedding helps;
- Face-to-face delivery popular;
- Benefits of RDM must be foregrounded.

Refined into 19 recommendations: in project report.



... because good research needs good data

Achievements

- Researcher career development models will incorporate data management skills
- Value of generic and discipline-specific data management skills in successful training is established
- Engagement with professional bodies

Future activities

- Engagement with professional bodies
- RIN Information-Handling Working Group
- Digital Curation Centre
- DPOE and DigCurV



... because good research needs good data

DaMSSI recommendations and career profiles available online at:

Digital Curation Centre:

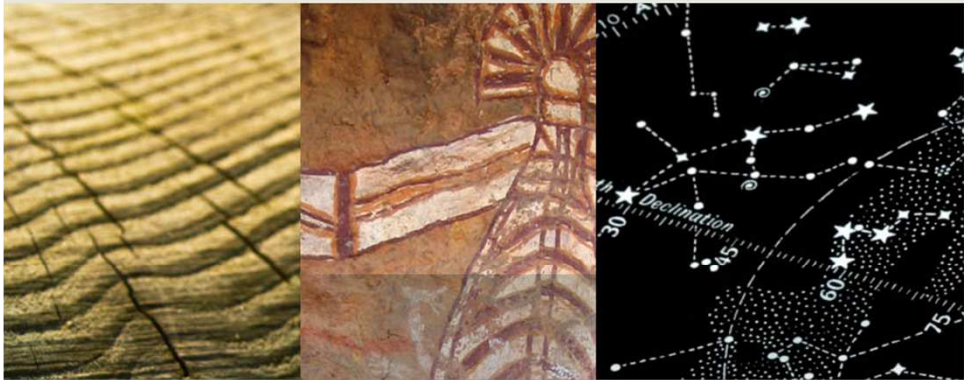
<http://tinyurl.com/DaMSSI-DCC>

Research Information Network:

<http://tinyurl.com/DaMSSI-RIN>

Thank you for listening!

Laura.Molloy@glasgow.ac.uk



Building human and infrastructure capacity through a national approach: the ANDS experience

David Groenewegen

Dr. Andrew Treloar

What is ANDS?

- ANDS is supported by the Australian Government
- Began in 2009, currently funded to June 2013
- Collaboration between Monash University, CSIRO and the Australian National University
- Staff in 6 cities across the country





Our assumptions:

- Research data is a first class output of research
 - *So it should be published*
- Research outcomes are enhanced by better access to richer data
 - *It's just like getting access to a more powerful microscope*





So better data is a good idea!

ANDS enables transformation of:

Data that are:

-  Unmanaged
-  Disconnected
-  Invisible
-  Single use

To Structured Collections that are:

-  Managed
-  Connected
-  Findable
-  Reusable

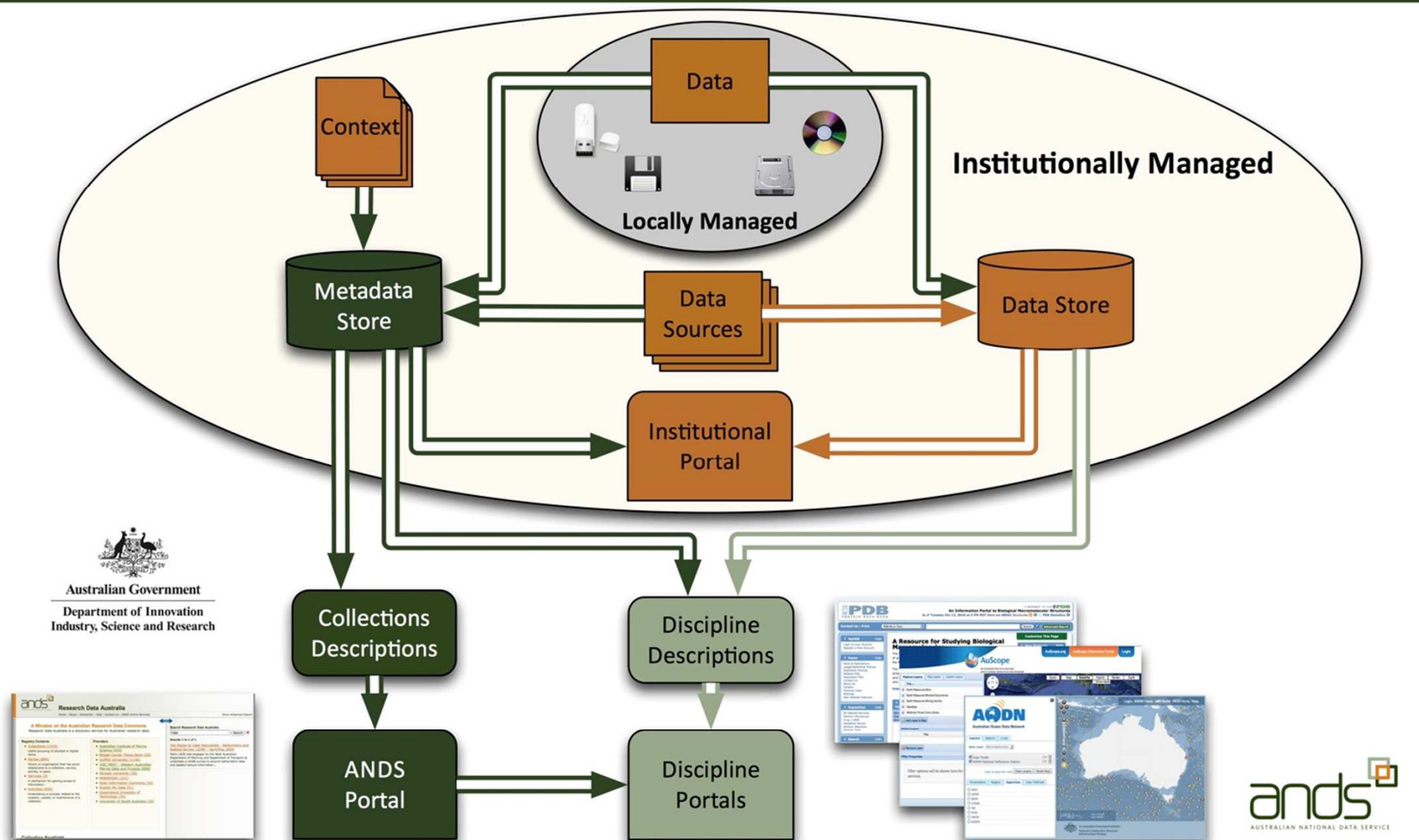
so that Australian researchers can easily publish,
discover, access and use research data.

ANDS is helping build the Australian Research Data Commons:

A meeting place for researchers and data:

- The set of data collections that are shareable
- The descriptions of the collections
- The relationships between the data, the researchers, the problems, the instruments and the institutions
- The infrastructure that enables populating and exploiting the commons

Australian Research Data Commons



ANDS role:

- Establish several national services in support of research data
- Help populate the ARDC with data collections that are managed, connected, and discoverable and reusable
- Partner with institutions to establish coherent institutional research data infrastructure
- Improve the ability of the Australian research system to exploit its research data through the lens of the ARDC using tools, policy and human capability

ANDS as a funder

- ~\$30 million
- Funding primarily provided at institutional level
- Infrastructure programs
 - Metadata Stores
 - Data Capture
- Human capacity programs
 - Seeding the Commons
 - Capabilities

Data Capture

- So far
 - Funded 76 projects supporting better capture of data and metadata from a range of instruments
 - Completed 13 of them
 - Benefit for researchers: making the right thing easier
 - Benefit for their institutions: fixing the future
- What's next
 - Moving from development to wider deployment

Metadata Stores

- So far
 - Funded a range of institutional solutions for collections metadata
 - ReDBoX (IR-based), VITRO (RDF-based), Tardis (RDBMS), ORCA(RDBMS)
 - Plus five deployments at different institutions
 - Connector to ResearchMaster CRIS for info about people and projects
- What's next
 - Funding to 22 institutions for development/deployment/configuration of solutions that support a standard set of deliverables

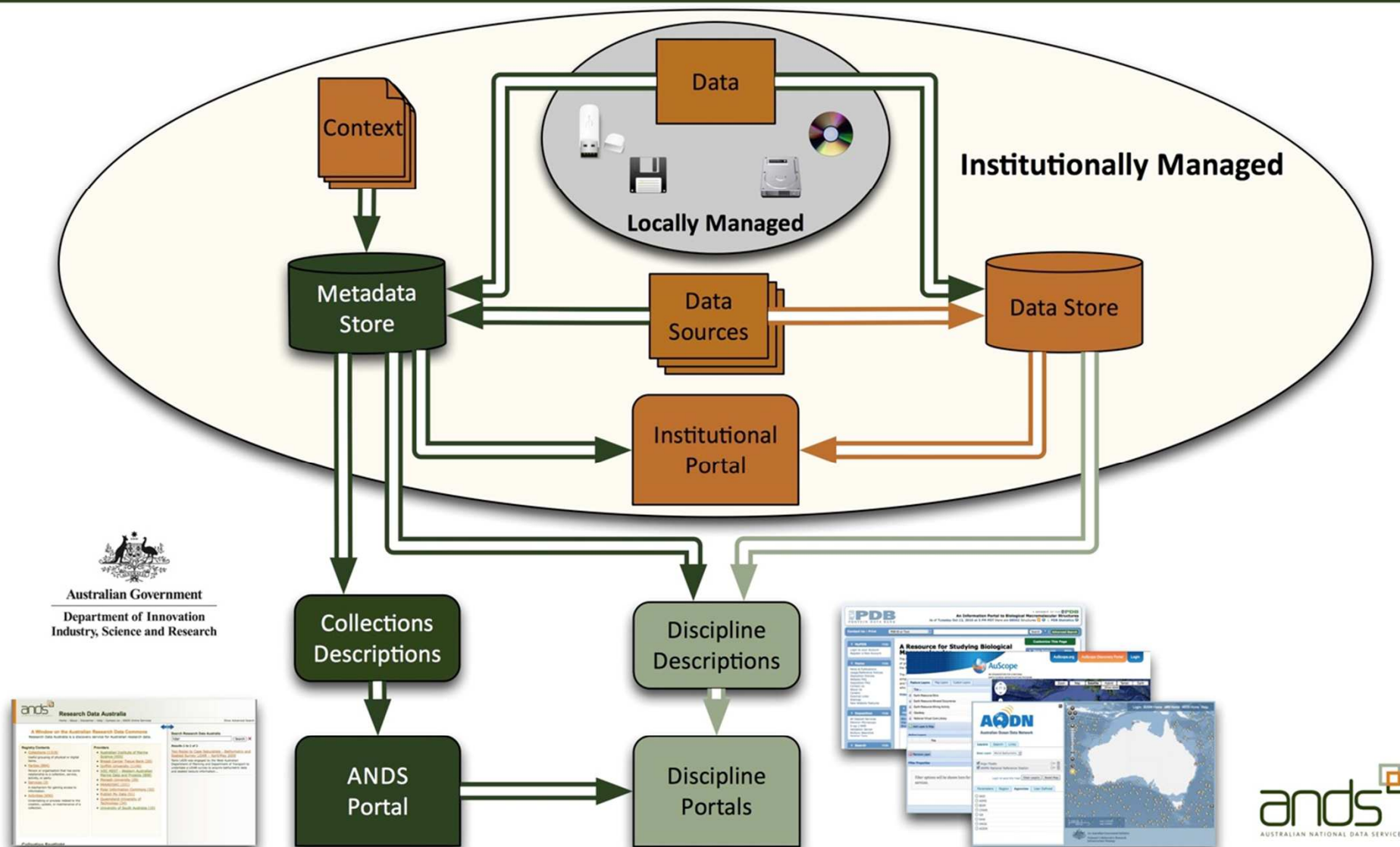
Seeding the Commons:

- So far
 - 8 of 34 projects completed
 - 19 institutions with data management policies and planning in progress
 - 19 supplying records to Research Data Australia
- What's next
 - About to fund another 7
 - Building greater coherence within and across universities

Coherence

- Within institutions:
 - Connections between projects, across institutions
 - Formation of data management steering groups across areas
 - Excellence in Research Australia (ERA) has played a big role
 - New data management positions
- ANDS role:
 - National services
 - Provision of advice, support, tools and infrastructure

Australian Research Data Commons



Thank you

- Questions?

David.Groenewegen@ands.org.au

contact@ands.org.au

ands.org.au

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative