
MaDAM Pilot Data Management Infrastructure for Biomedical Researchers at University of Manchester

Project Team

MeRC : **Meik Poschen**, June Finch (PM), Mhorag Goff, Rob Procter

RC : Simon Collins, Mary McDerby

JRUL : Jon Besson, Tom Grahame, Lorraine Beard (PI)

Project Sponsor

Jan Wilkinson (JRUL)

User Representatives

Simon Hubbard (Life Sciences), Alan Jackson (Medical School)

Funded by the

JISC

+ University of Manchester
Contribution

4th Paradigm: data-intensive research

- Challenge: the “remarkable growth of data-intensive research in all knowledge domains” (Blue Ribbon Task Force report, 2010) over the last years and the need “to do better at producing tools to support the whole research cycle – from data capture and data curation to data analysis and data visualization” (Jim Gray, 2007).
- This means taking into account the multitude of data types and formats and ensuring that technical and non-technical solutions for (collaboratively) managing and sharing data will fit in with the **research lifecycle**, diverse **working practices**, **cultures** and **disciplines**.

Funding Stream & Landscape

MaDAM runs from Oct 2009 until June 2011 and is one of 8 projects funded under the Infrastructure Strand of the JISC Managing Research Data Programme (overall about 30 projects funded, including 3 support projects):

JISCMRD: “Higher Education Institutions are coming under increasing pressure to manage the research data generated by their researchers that cannot be curated by subject-based data centres - and many are unsure how to proceed given the absence of clear good practice.”

Overall Research Councils in the UK also recognise the need for better data curation procedures, the US NSF similarly calls such a “scientific necessity”.

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.

Rationale:

- Researchers need to be supported to manage their data well 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.

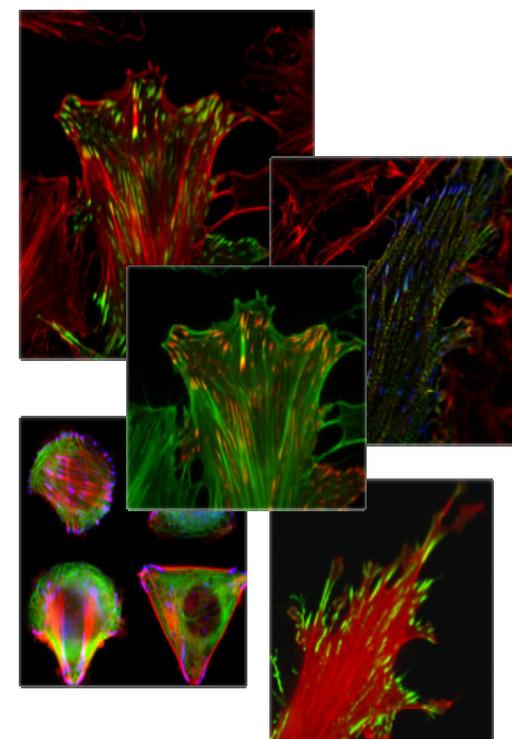
Background: No existing institutional repository or strategy for management of research data – BUT the MaDAM Pilot is part of a wider endeavour at University of Manchester to develop such.

The MaDAM Solution will..

- Provide trusted secure storage to reduce risks of data loss and to adhere to funder's new retention policies (WT: 20 years!!)
- Make metadata visible and searchable
- Facilitate easier, more secure owner-controlled data sharing
- Enable annotation of data including ad hoc context and 'notes to self'
- Reduce redundancy by enabling 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 lifecycle, diverse working practices, cultures and disciplines

Domains & Pilot User Groups

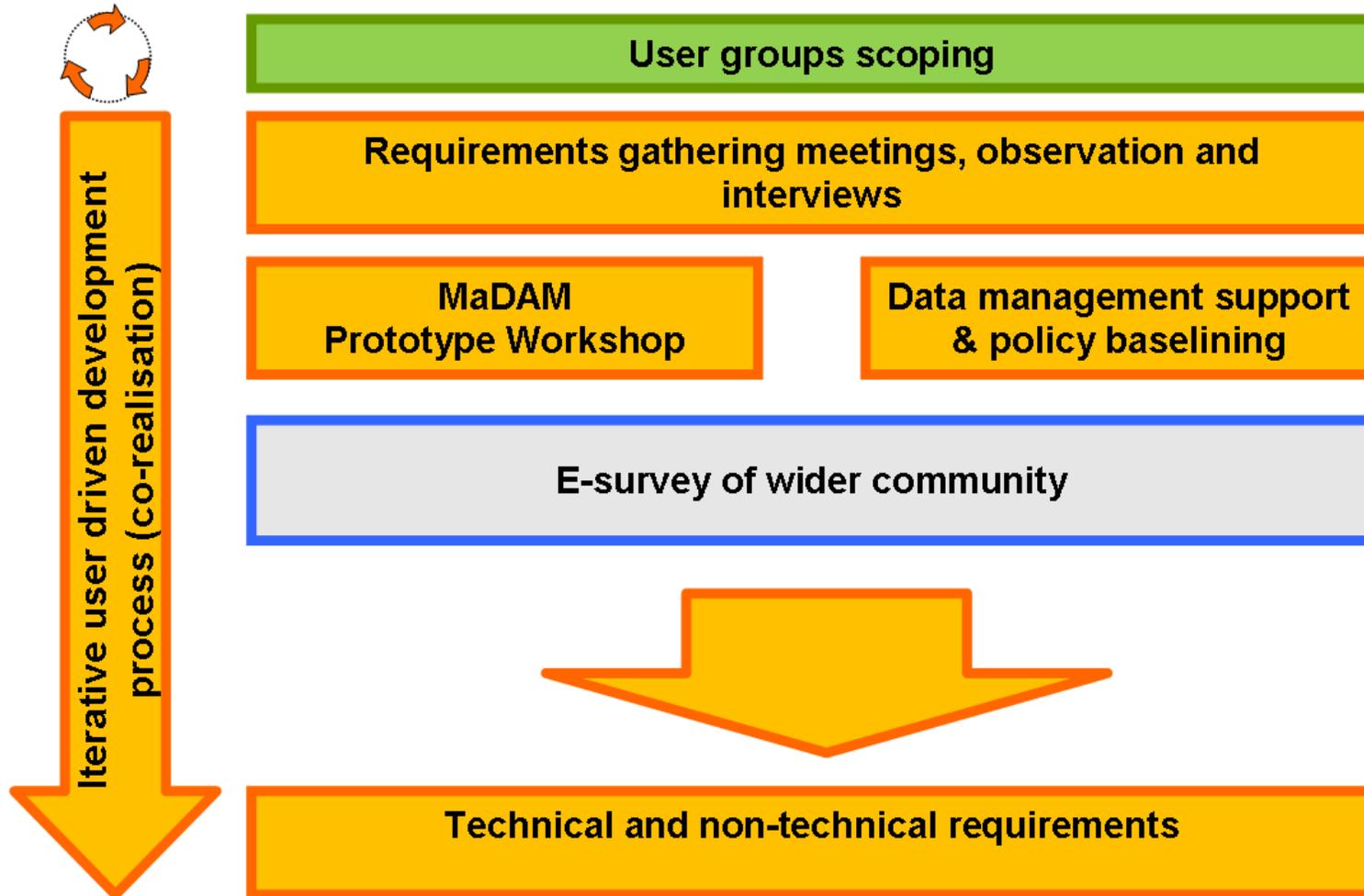
- Biomedical Domain at University of Manchester with user groups from a) Life Sciences Electron and Standard Microscopy (4 groups with 8 active core users plus occasional users) & b) Medical Science MRI Neuropsychiatry Unit (1 group/5 users)
 - Images as main Research Objects in diverse formats, resolutions, sizes coming from a number of instruments (microscopes, brain images from MRI scanners); also other data types (text docs, metadata, statistical and output data)
 - The work with the pilot user groups is 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'



Local Data Management Practices Found

- Cleaning & preparing raw data for analysis
- Identifying and selecting good quality data to work on BUT time investment is needed to develop it
- Use of traditional lab books to record experiment metadata BUT not easy to search
- Sharing data for discussion, feedback, expertise exchange and workflow management
- Use of portable devices for transferring, sharing and flexible temporary storage
- Multiple copies of data needed to explore analysis pathways including potential 'dead-ends'
- Redundancy necessary to organize and find data BUT exacerbates storage capacity issues and also ironically discovery
- Retention of data even from failed experiments BUT much old data is rarely revisited and poor preservation practice means its hard to reuse

Findings

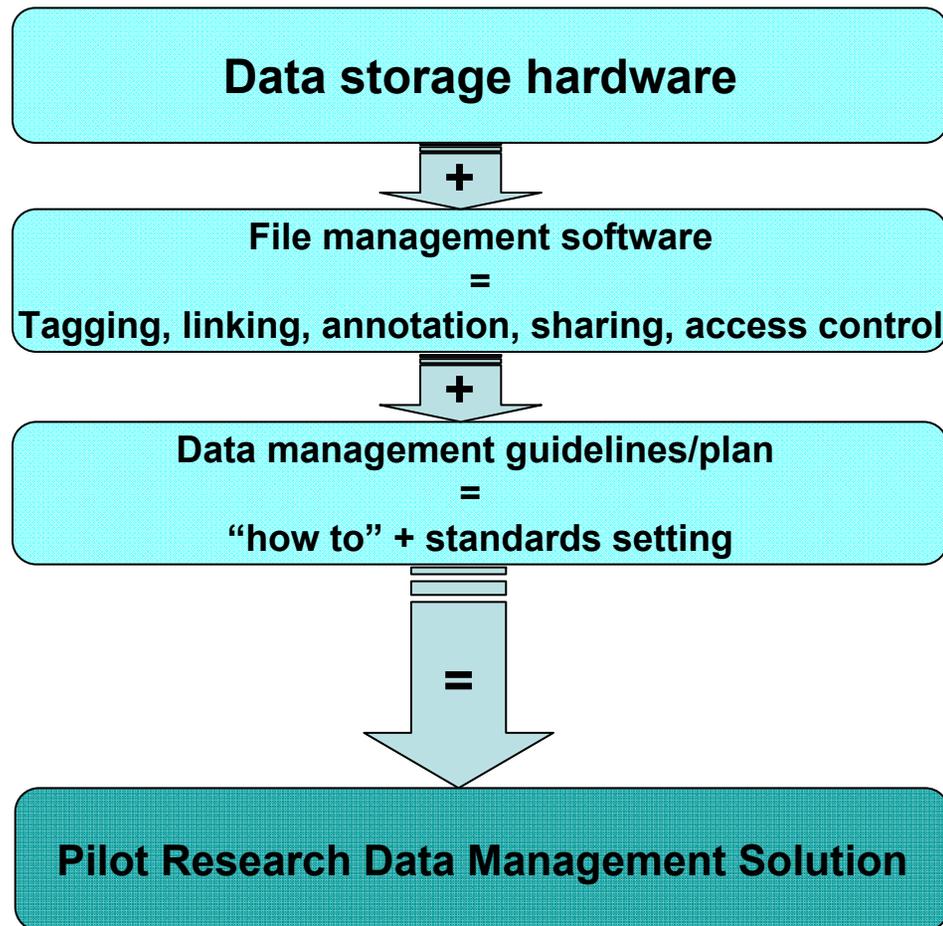
- No official backup policies to protect against loss of data
 - Decentralized & fragmented storage (USB sticks, optical disks)
 - Limited ability to share data internally or externally
 - High levels of redundant data (duplicate copies)
 - No structured annotation of data
 - Limited search capabilities
 - 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**.

MaDAM Pilot Overview

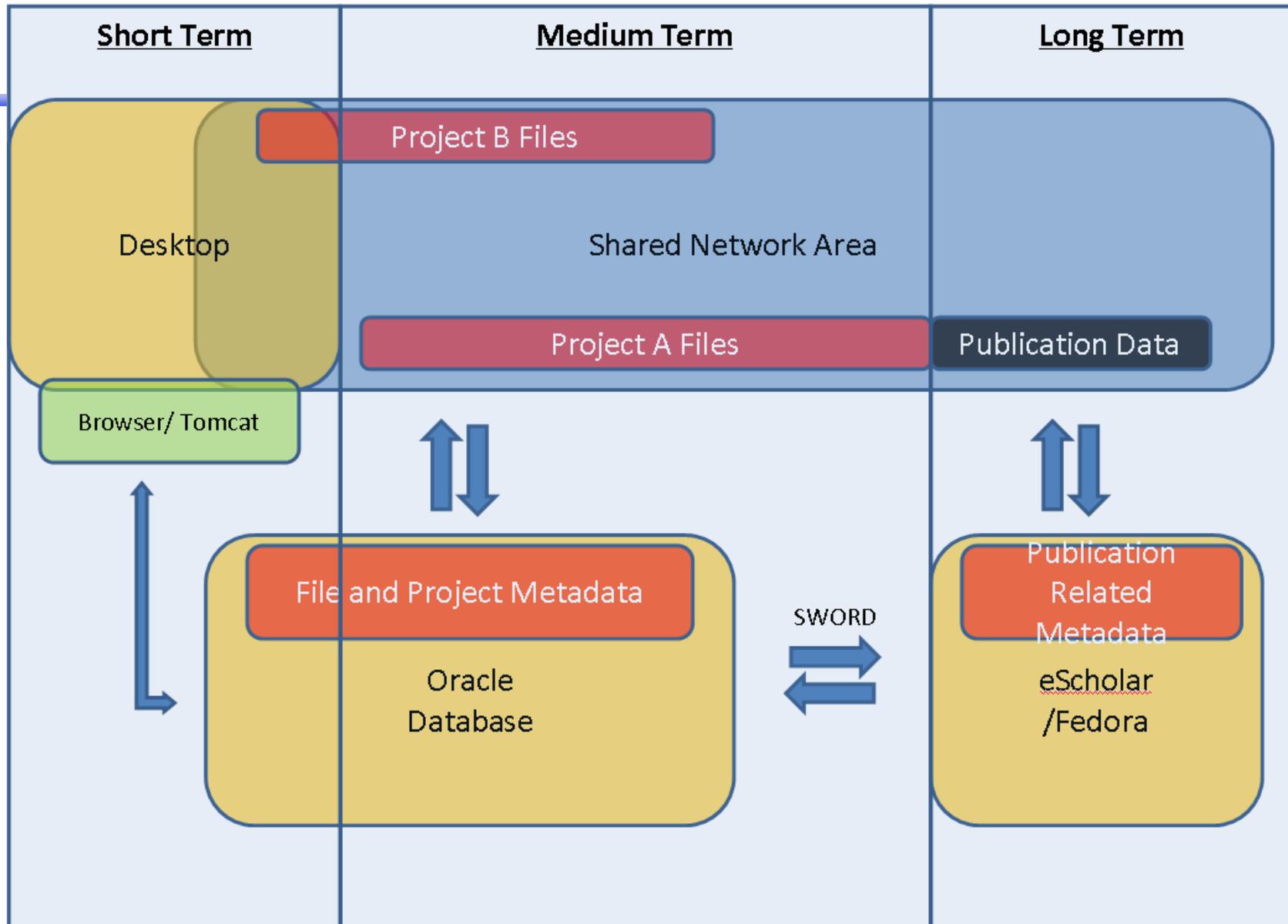
Aim: Pilot Research Data Management Solution



Many angles to cover:

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

MaDAM Pilot Storage/Architecture



MaDAM Pilot: Authentication/Accounts

Auto Account Creation

Data **Setup**

User Details | My Attributes | My Projects | Sys Admin | Edit Help

My Details Update

Full Name	Simon Collins
Telephone	01612750604
Dept	RCS
Title	Oracle DBA
Email	simon.collins-2@manchester.ac.uk

Projects I Manage

Name	Description	Create Date	Status
Skin Analysis	-	27-MAY-2010	1
bone morphogenetic protein-1	-	27-MAY-2010	1
popliteal pterygium syndrome	-	27-APR-2010	3
Nanostructure of fibrillin-1	-	27-APR-2010	1

1 - 4

LDAP Authentication

Login

Username

Password

Please Log in with your University Credentials

Project

Project Search

Project Name Owner

Details	Name	Owner	Create Date
	Bivariate Match Odds	Alan Roseman	27-APR-2010
	mammalian tolloid	Clair Baldock	27-APR-2010
	Geospatial Project	Ricky Tsang	27-APR-2010
	Applet Work	Jon Besson	27-APR-2010
	MaDAM Project	June Finch	27-APR-2010
	Mary's Exciting Project	Mary McDerby	27-APR-2010

Project Details

Name mammalian tolloid

Comments Structural and functional evidence for a substrate exclusion mechanism in mammalian tolloid like-1 (TLL-1) proteinase

Owner Clair Baldock

Create Date 27-APR-2010

Project Creation and Location

MaDAM Pilot: Project Setup/Access

Data Setup

[User Details](#) | [My Attributes](#) | **[My Projects](#)** | [Sys Admin](#) | [Edit Help](#)

Project Project C
Project Access

Project Details

Delete Apply Changes

Name Project C

Owner Simon Collins

Create Date 27-APR-2010

Review Date 23-DEC-2010

Description

Status Available

Requested Disk (GB) 500

Approved Disk (GB) 400

Project Access

Move Users to right to give them access to this project.

Data Setup

[User Details](#) | [My Attributes](#) | **[My Projects](#)**

- Alan Roseman
- Alex Carisey
- June Finch
- Meik Poschen
- Mhorag Goff
- Ricky Tsang
- Toby Starborg
- Tom Grahame
- mjkpknf2

Read/Write

Project Access

Update

Write	User
<input checked="" type="checkbox"/>	Simon Collins
<input checked="" type="checkbox"/>	Mary McDerby
<input type="checkbox"/>	Christoph Ballestrem
<input type="checkbox"/>	Jon Besson
<input checked="" type="checkbox"/>	Clair Baldock

Standard System Data

Disk Usage Control

1 - 5

Templates

Template Method 1

Type Experiment

User defined data and templates → metadata

Template Attributes

Save Add Row Delete

	Attribute Name	Attribute Default Value	
<input type="checkbox"/>	Instrument	Winfield	↑ ↓
<input type="checkbox"/>	Protein		↑ ↓
<input type="checkbox"/>	Comments		↑ ↓

MaDAM Pilot: Data Management

The screenshot shows the MaDAM Pilot interface with several annotations:

- Project/"Folder" Data:** Points to the 'Name' and 'Comments' fields of the 'Fin2a' project.
- Context Sensitive Actions:** Points to the 'Project' and 'Folder Actions' panels on the right.
- Web Explorer:** Points to the 'Explorer' sidebar on the left.
- Bulk Downloads:** Points to the 'Files in Folder' table at the bottom.
- "Folder" Contents:** Points to the 'Files in Folder' table at the bottom.

Navigation: Data | Setup | Home Page | Search | Calendar | Bookmarks | Feedback | Skin Analysis > Fin2a

Explorer: Project A, Project C, Geospatial Project, End of Season Simulation, Bivariate Match Odds, Skin Analysis (Fin2a, Raw Data, Analysis), Fin2b, Fin2c, Fin2d, Fin2e, Submission A

Quick File Upload: Browse...

Fin2a Details:

- Name: Fin2a
- Comments: Expt using Diazo dye for Protein X
- Owner: Simon Collins
- Create Date: 28-MAY-2010
- Status: Available

Attributes: Instrument (Instrument A), Method (Method A), Comments (1 - 3)

Project Actions: Create Project, Project Details, Project Template

Create: Create Folder, Create Experiment, Create Publication

Folder Actions: Create Note, Email Folder Link, Bookmark Folder, Page Help, Create Shortcut, Move Folder, Add Dataset

Notes: Clean Instrumen..

Folders Table:

Type	Name	Date Created	Status	Owner
Folder	Raw Data	28-MAY-2010	Available	Simon Collins
Folder	Analysis	28-MAY-2010	Available	Simon Collins

Files in Folder:

Select	Name	Created On	Size / MB	Version
<input type="checkbox"/>	Analysis Overview.doc	01-JUN-2010	0.02	1

MaDAM Pilot: Thumbnails

Home Page | Search | Calendar | Bookmarks | Feedback |

File List

- FRAPPb1Series40_t07_ch00.tif
- FRAPPb1Series40_t01_ch01.tif
- FRAPBleachSeries38.xml
- FRAPBleachSeries38_t1_ch01.tif
- FRAPPb1Series40_t19_ch00.tif
- FRAPBleachSeries38_t2_ch01.tif
- FRAPPb1Series40_t06_ch01.tif
- FRAPPb1Series40_t15_ch01.tif
- FRAPPb1Series40_t19_ch01.tif
- FRAPPPreSeries39_t1_ch00.tif
- FRAPPb1Series40_t14_ch00.tif
- FRAPPb1Series40_t06_ch00.tif
- FRAPPPreSeries39_t3_ch00.tif
- FRAPPb1Series40_t11_ch01.tif
- FRAPPb1Series40_t00_ch00.tif
- FRAPPb1Series40_t20_ch00.tif
- FRAPPb1Series40.xml
- FRAPPb1Series40_t16_ch00.tif
- FRAPPb1Series40_t10_ch01.tif
- FRAPPb1Series40_t17_ch00.tif
- FRAPPPreSeries39.xml

Thumbnails - Number of Columns

MaDAM Pilot: Metadata

MANCHESTER 1824 Search 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

<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

Cancel Delete Save

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!*

The ‘Storage, Archiving, Curation’ (SAC) Project (1)

Originated from the ‘Computational Science Review’,
Recommendation 6:

“Although it was beyond the scope of this review, there is a clear requirement for a University-wide strategy for data storage, archiving and curation.”

“The University IS Strategy Board should develop a strategy for data storage, archiving and curation which joins up the institutional repository with a concerted data storage and management activity.”

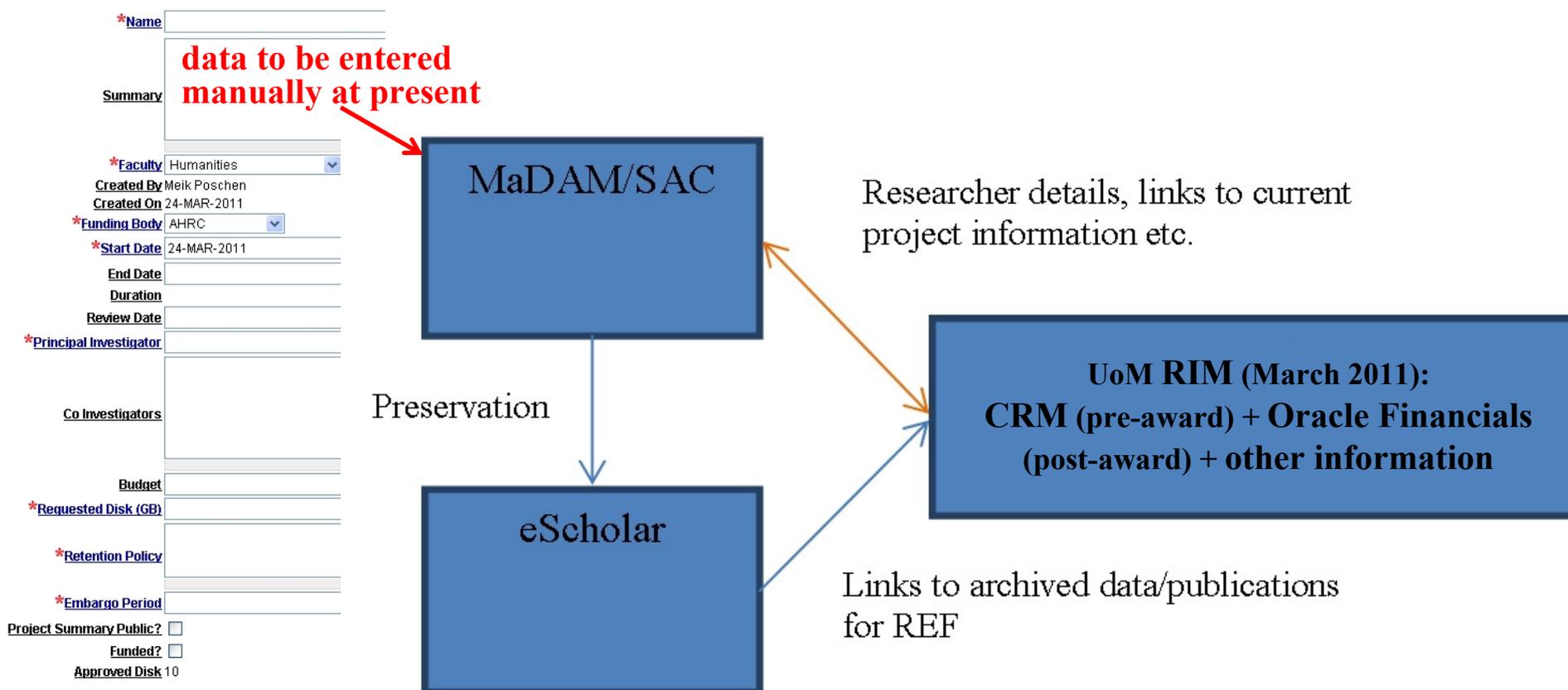
SAC is championed by Manchester Informatics (Mi) and the John Rylands University Library (JRUL)

The 'Storage, Archiving, Curation' (SAC) Project (2)

- The SAC project has produced a proposal for a wider Research Data Management Service (RDMS) at the University of Manchester, with the aim to roll out this service incrementally, adding research groups sequentially – starting with MaDAM
- MaDAM is used as a demonstrator and its results are being fed into the SAC proposal
- This opens the possibility of a sustainability route for MaDAM after the initial project's lifetime

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)

- **Current approaches by researchers to long term preservation are underdeveloped** because their basic needs for secure, trusted storage (and back-up) to support the research lifecycle are not yet being met.
- **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.

Challenges & Observations (2)

- **Good progress:** establishing the functional requirements for the prototype data management infrastructure & technical support and sustainability is being addressed through Cost-Benefit Analysis and financial modelling.
- **BUT:** A cultural change is 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!).
- **Sustainability:** The MaDAM pilot will be part of the assessment of the further development of a data management and digital curation strategy for the wider University in Manchester ('Storage, Archiving and Curation' (SAC) proposal for a Research Data Management Service at the University of Manchester)

Some Final Remarks

- There still remain open questions at this point, e.g.
 - How much storage will each research group or researcher 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?
- For the bottom-up approach of MaDAM this means further observing, evaluating and documenting evolving and emerging patterns and behaviour of actual research practice.

Many Thanks!

MaDAM

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

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

Manchester Informatics (Mi)

<http://www.informatics.manchester.ac.uk/>

Meik Poschen

meik.poschen@manchester.ac.uk

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