The University of Manchester

MANCHESTER 1824

The University of Manchester
The John Rylands University Library

Research Computing Services (RCS)

MeRC
Manchester eResearch Centre

# 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

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 4th Paradigm: data-intensitive 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.**

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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”.

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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.

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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

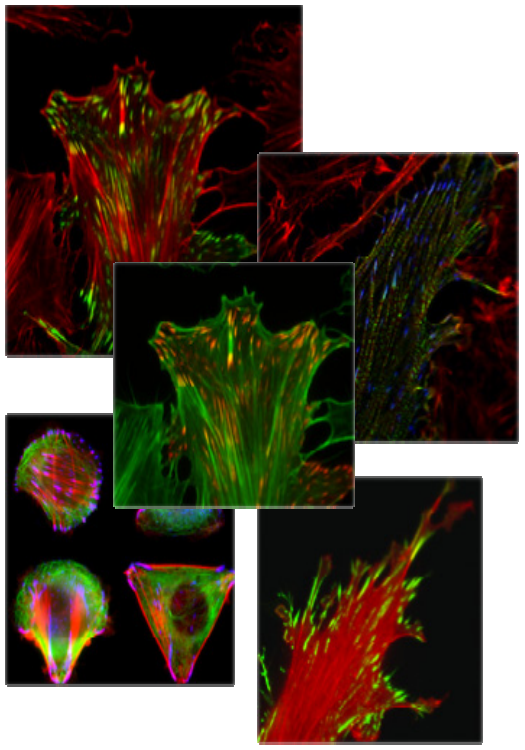
MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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'**

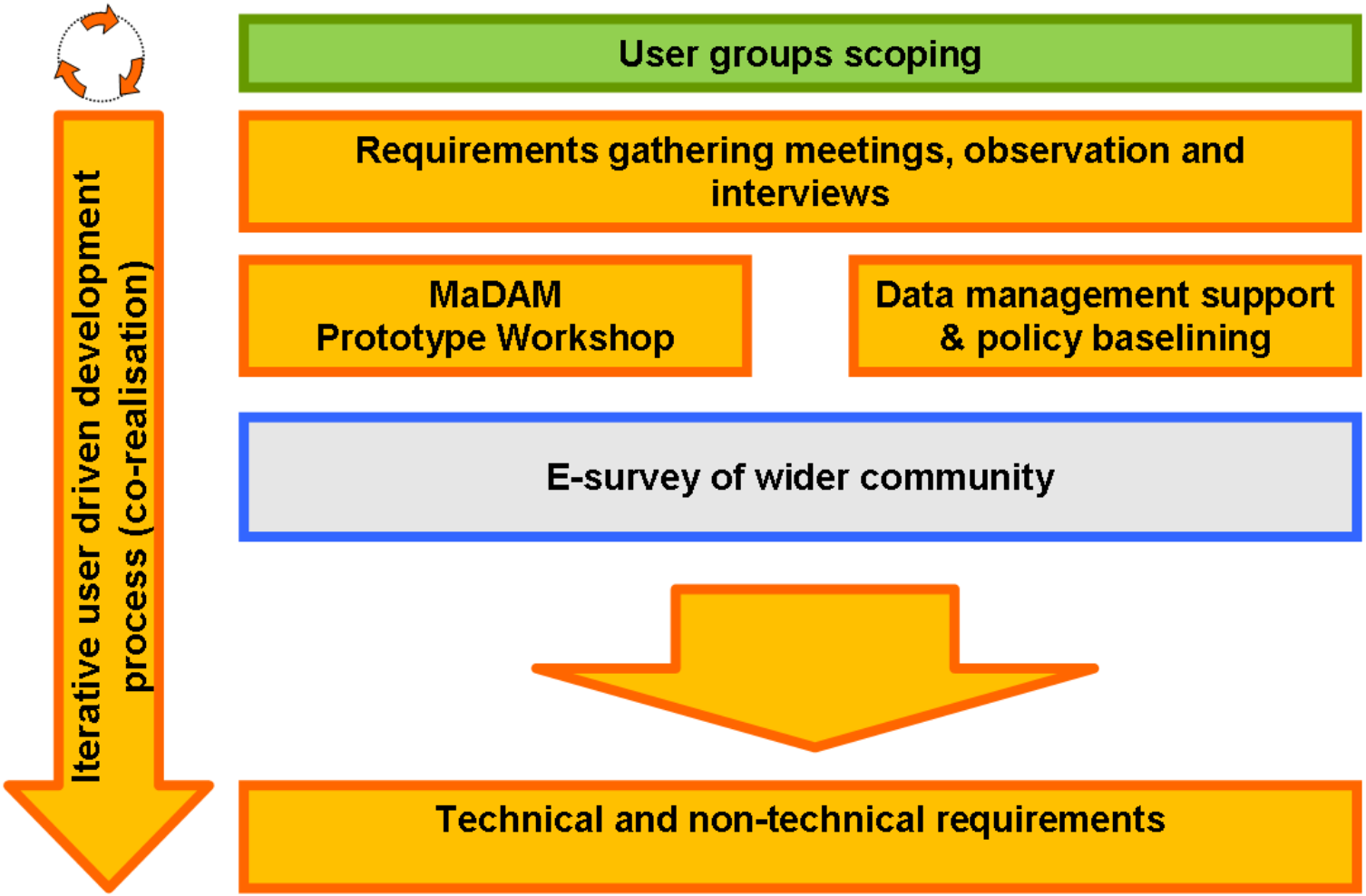
> Microscope samples: single run creates any image set from 1-200 GB
> MRI brain scans: usually one study consists of 20-40 GB

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# MaDAM 'Method-flow'



MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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**

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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**.

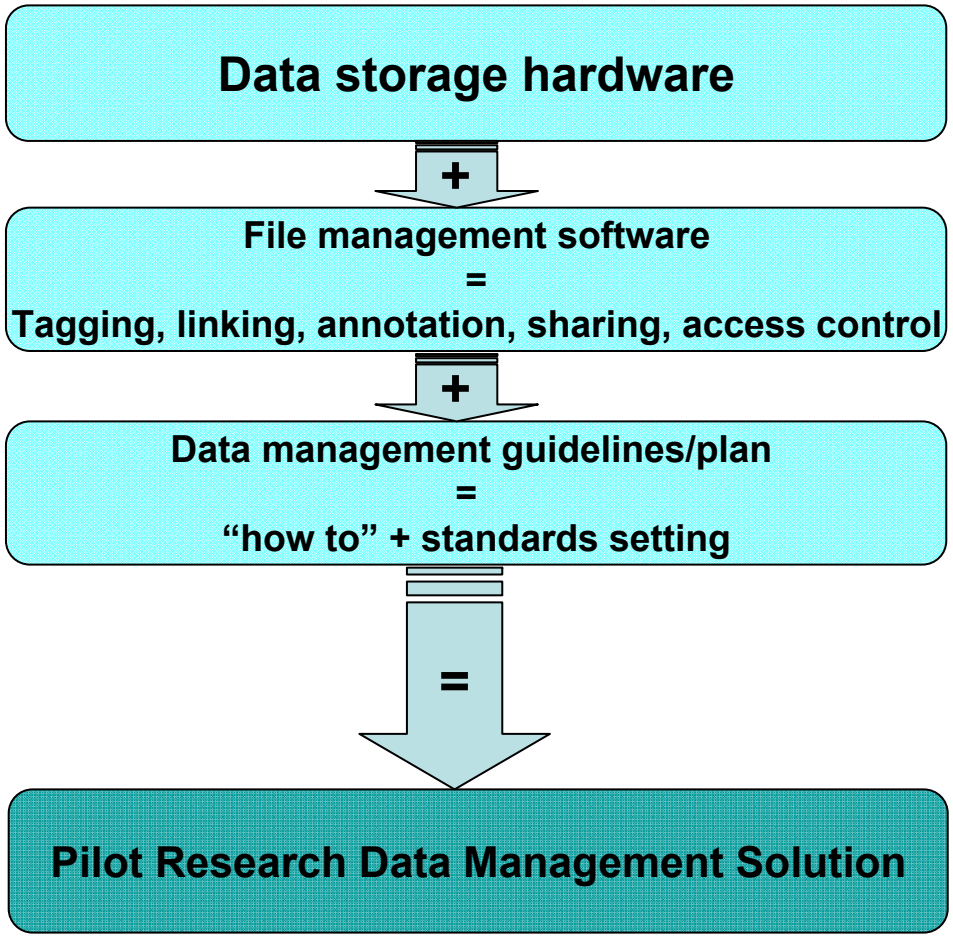
MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# MaDAM Pilot Overview

**Aim: Pilot Research Data Management Solution**




Many angles to cover:

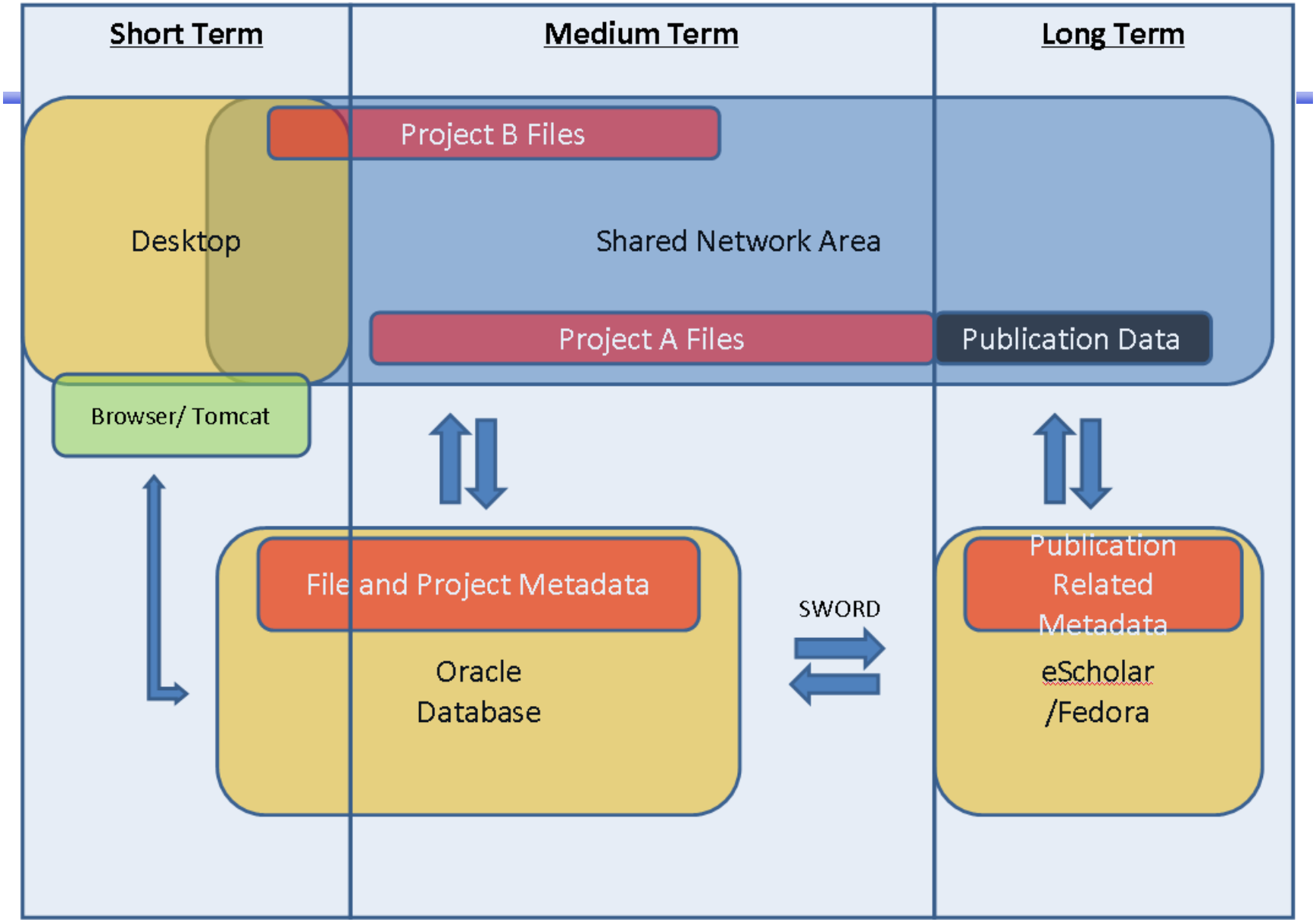
- Research Practice
- Discipline/Domain
- Technical Solution
- Policies/Procedures
- Institutional Settings (Stakeholders & Infrastructure)
- Funding Landscape
- Cost-Benefit Analysis

MANCHESTER 1824

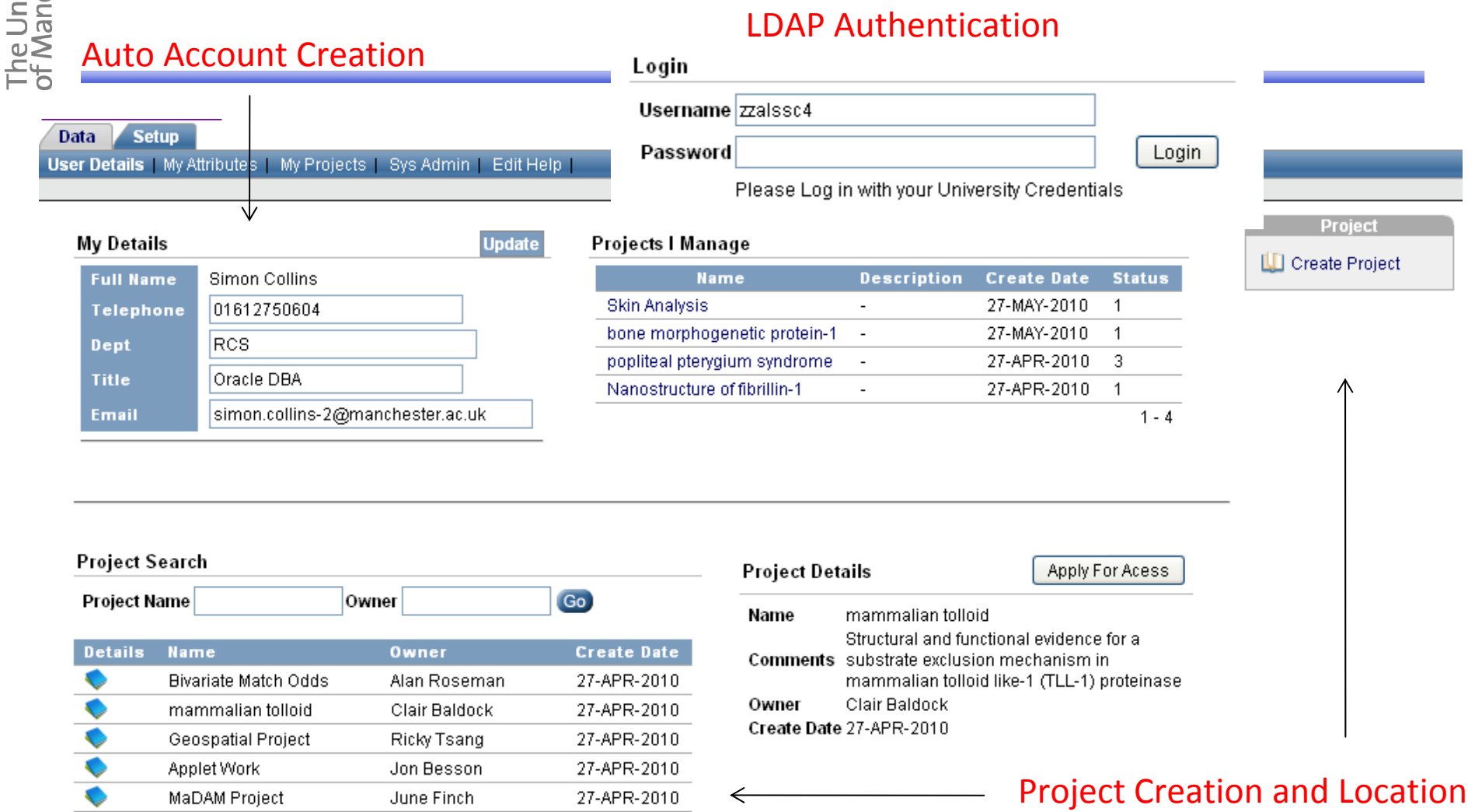
The University of Manchester

MeRC
Manchester eResearch Centre

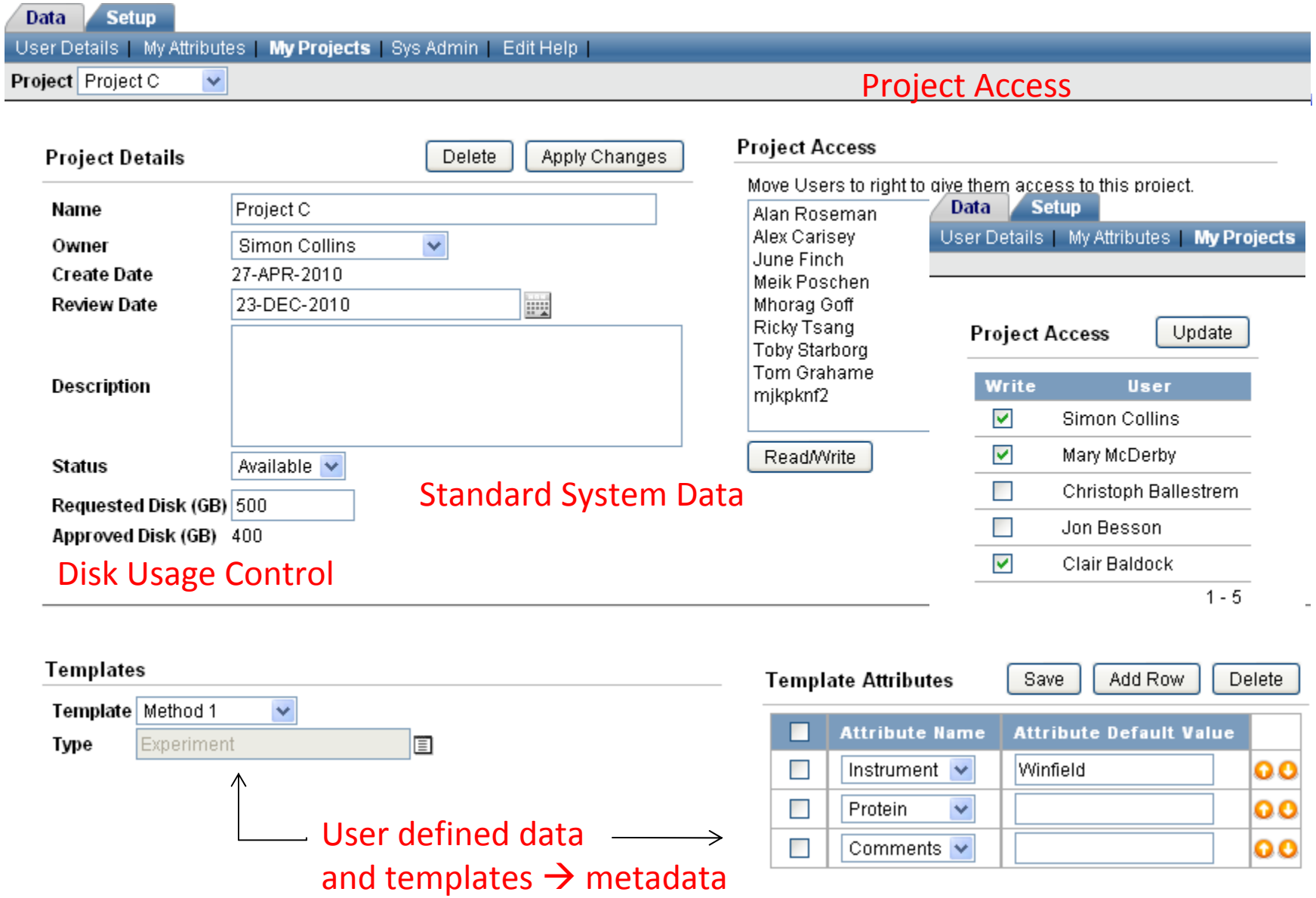
# MaDAM Pilot Storage/Architecture



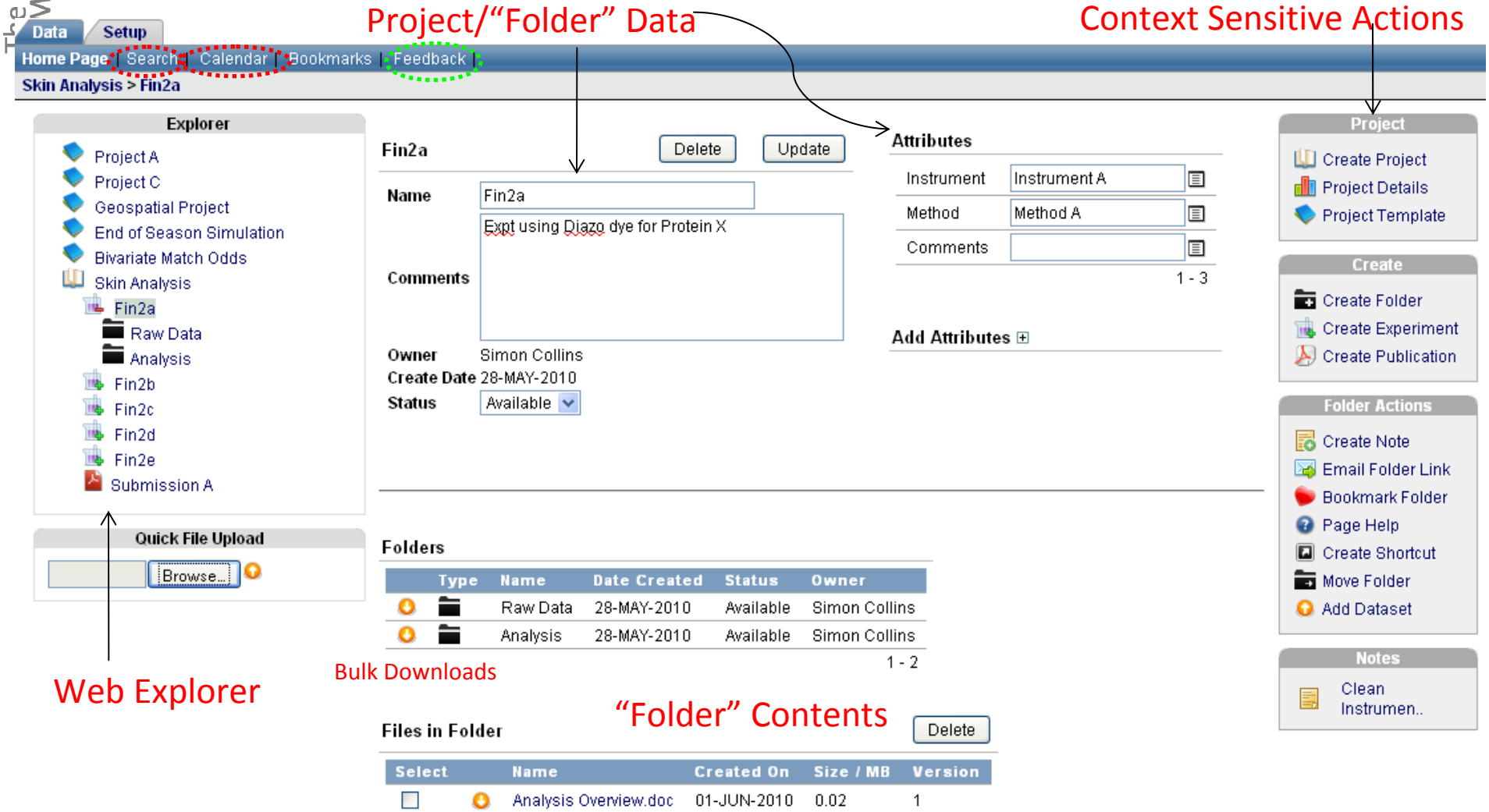
# MaDAM Pilot: Authentication/Accounts



# MaDAM Pilot: Project Setup/Access



# MaDAM Pilot: Data Management

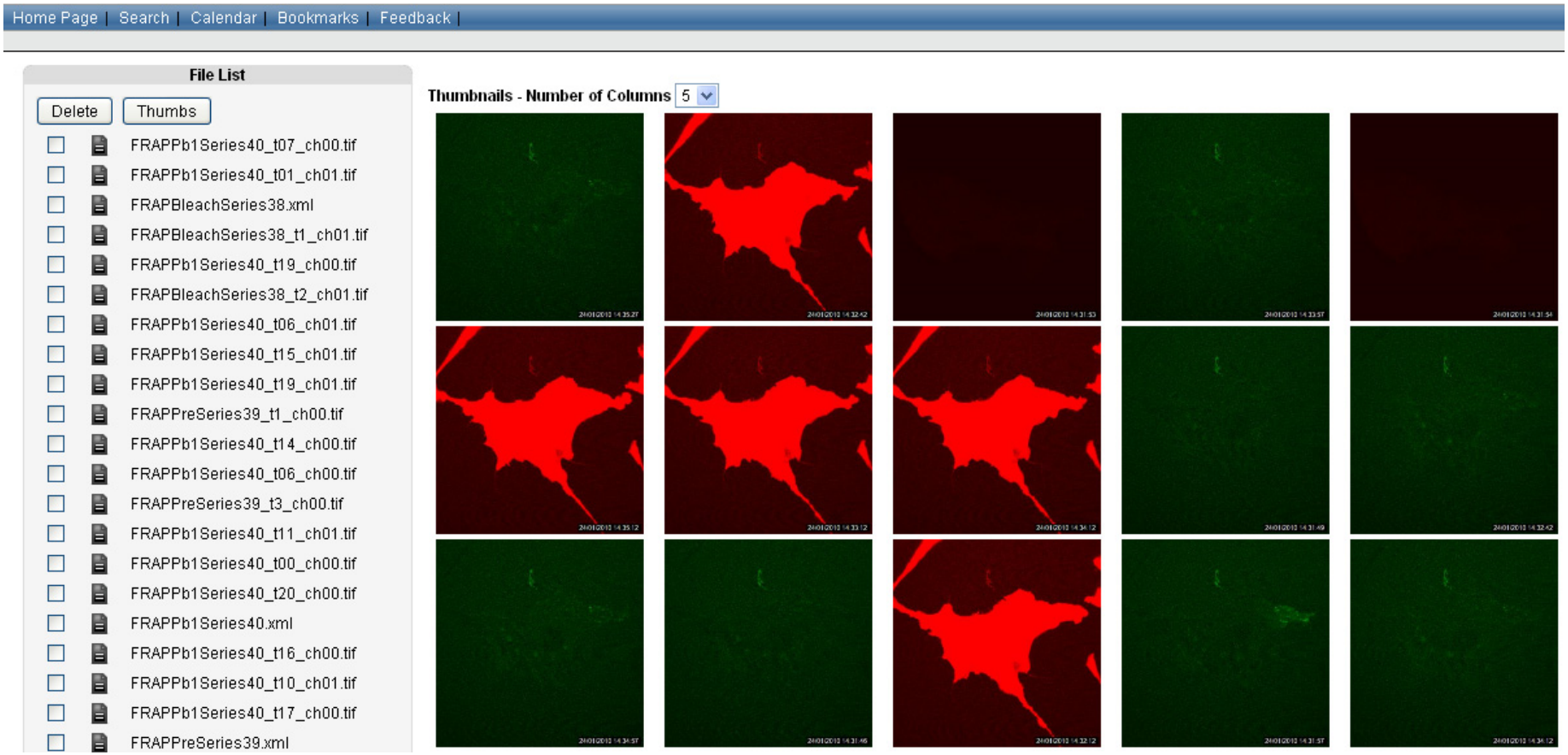


MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# MaDAM Pilot: Thumbnails

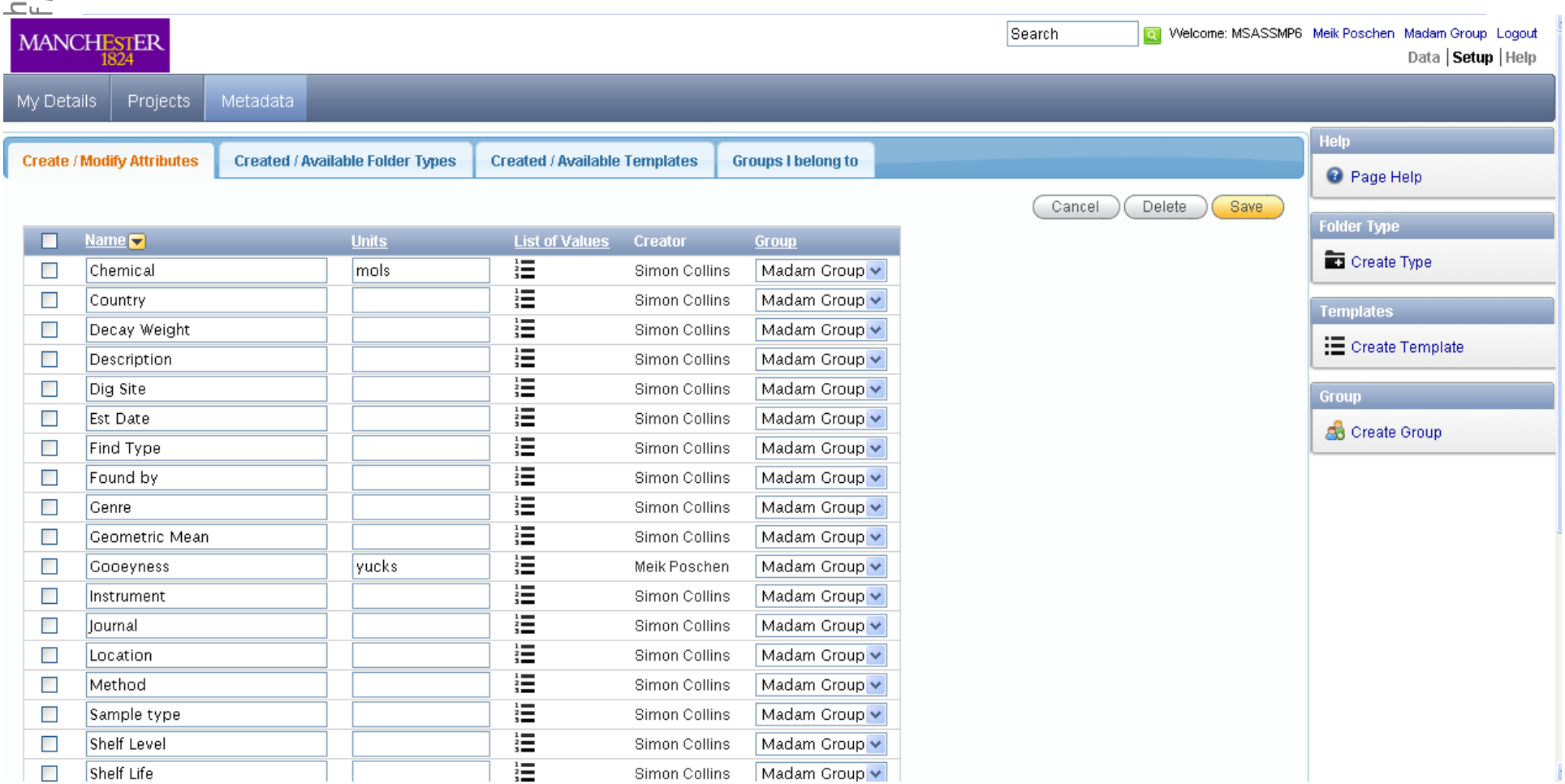


MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# MaDAM Pilot: Metadata



MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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 it's retention commitments

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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 (M*i*) and the John Rylands University Library (JRUL)

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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

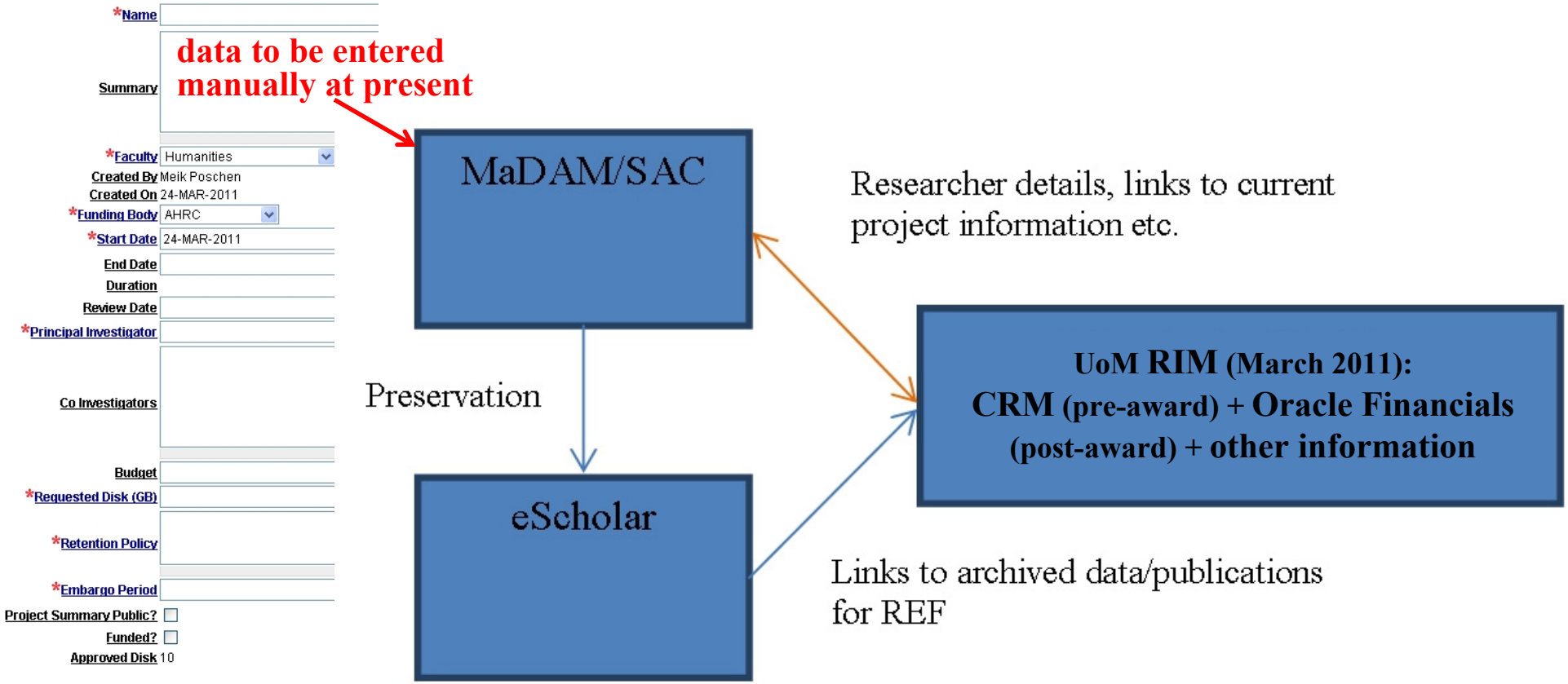
MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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



MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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.

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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)

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# 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.

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

# Many Thanks!

**MaDAM**

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

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

Manchester Informatics (M*i*)

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

Meik Poschen
meik.poschen@manchester.ac.uk
http://www.merc.ac.uk/?q=Meik