The University of Manchester Library

IT Services

MeRC Manchester eResearch Centre

Towards a Research Data Management Service at University of Manchester

The MaDAM and MiSS Projects

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

Mary McDerby (PM), Robin Pinning, Mike Jones, Simon Collins
IT Services, University of Manchester

Jon Besson, Kathleen Haigh-Hutchinson, Phil Butler, Tom Grahame, John Blunden-Ellis, Chris Gibson, Lorraine Beard (PI), Jan Wilkinson (Sponsor),
The University of Manchester Library

Funded by the JISC + University of Manchester Contribution

Faculty User Champions:
James Nazroo (Humanities), **Phil Withers** (EPS), **Alex Henderson** (EPS & Life Sciences), **David Knight** (Life Sciences), **John Ainsworth** (Medical School)

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

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, The Fourth Paradigm, 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 & data lifecycle, diverse working practices, cultures and disciplines.

Funding Landscape: New Approaches

MaDAM ran from Oct 2009 to June 2011 and was one of 8 projects funded under the Infrastructure Strand of the JISC Managing Research Data Programme (JISCMRD; overall over 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.

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

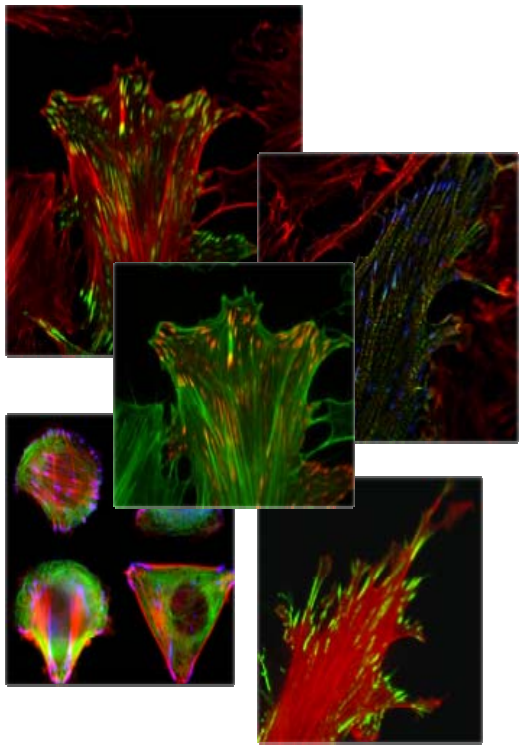
MANCHESTER 1824

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'

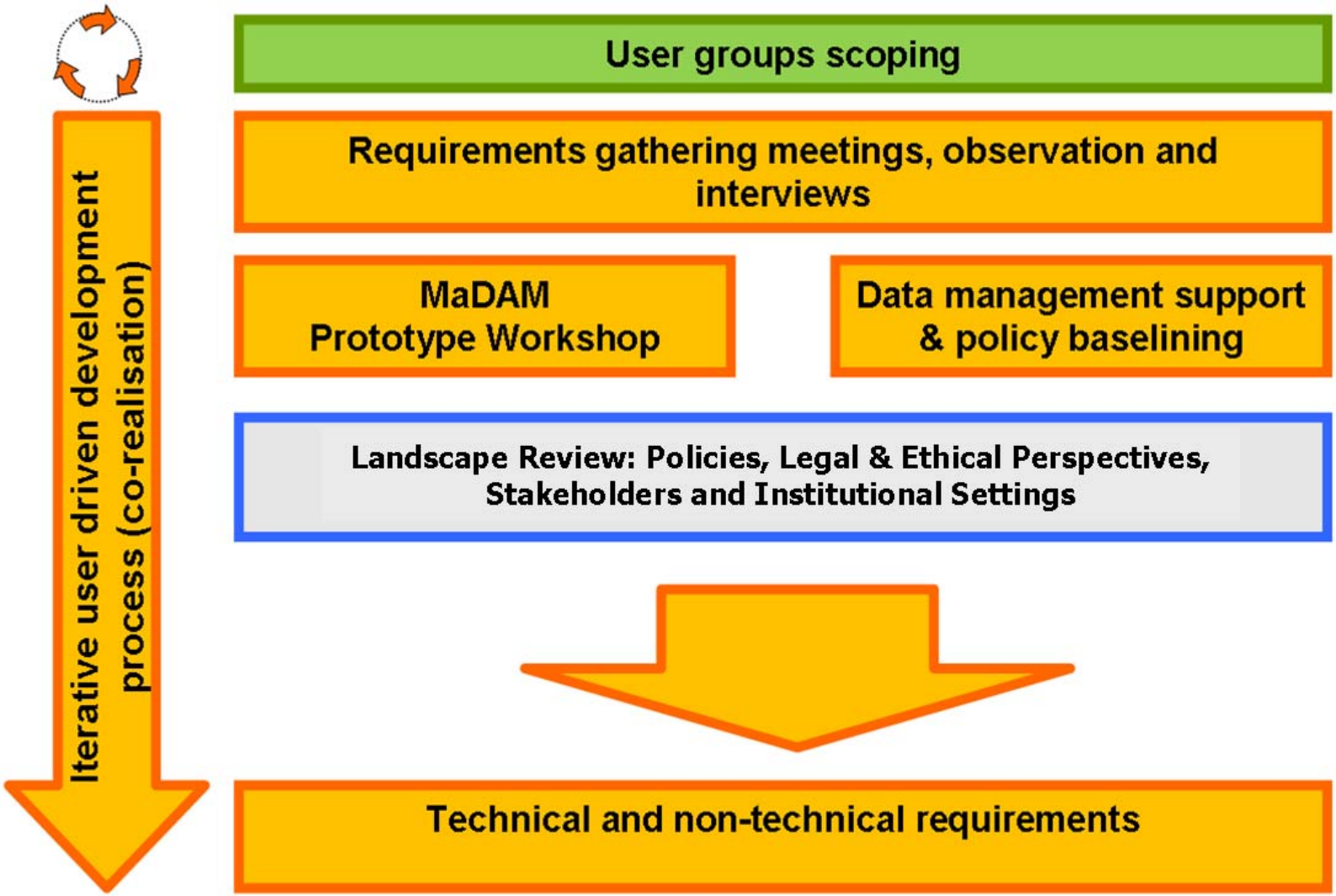
> Microscope samples: single run creates any image set from 1-200 GB
> MRI brain scans: usually one study consists of 20-40 GB

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

MaDAM 'Method-flow'



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

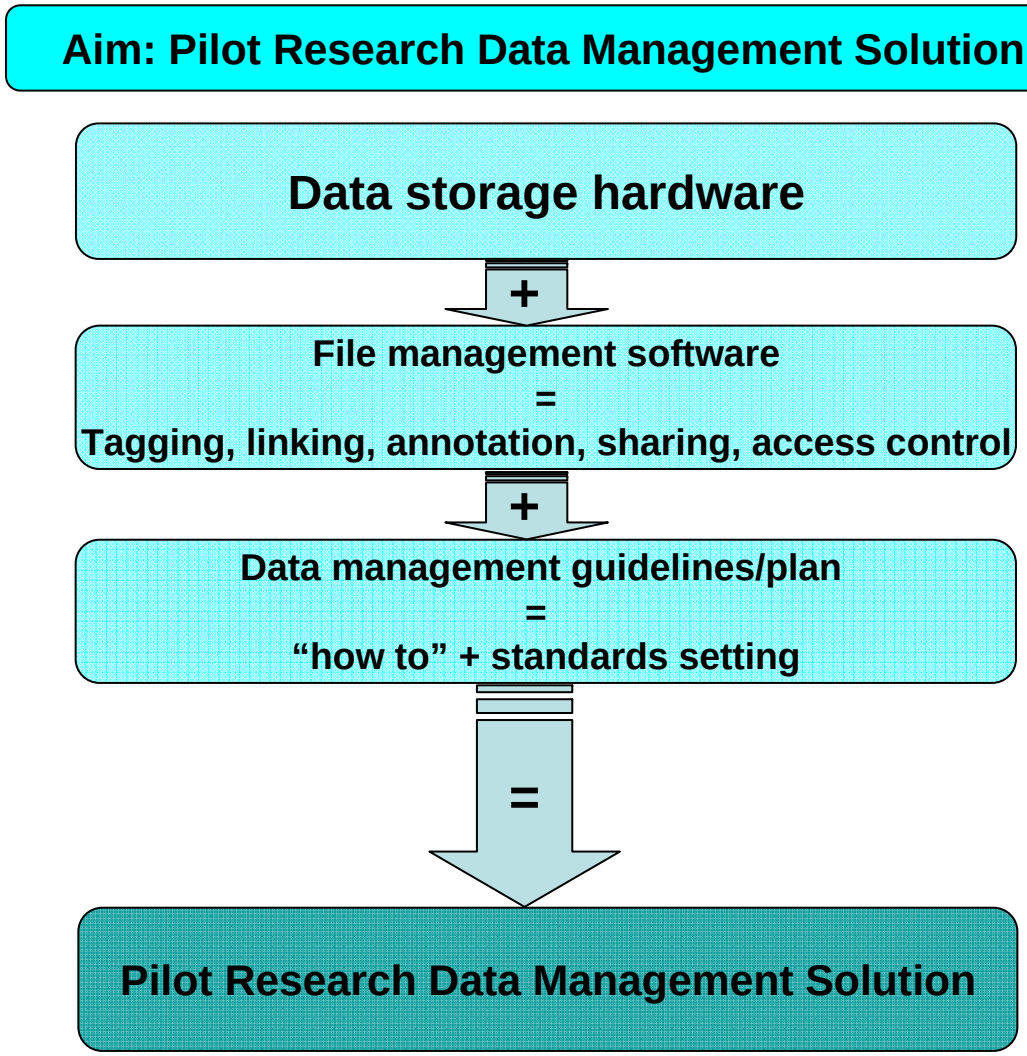
The University of Manchester

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.

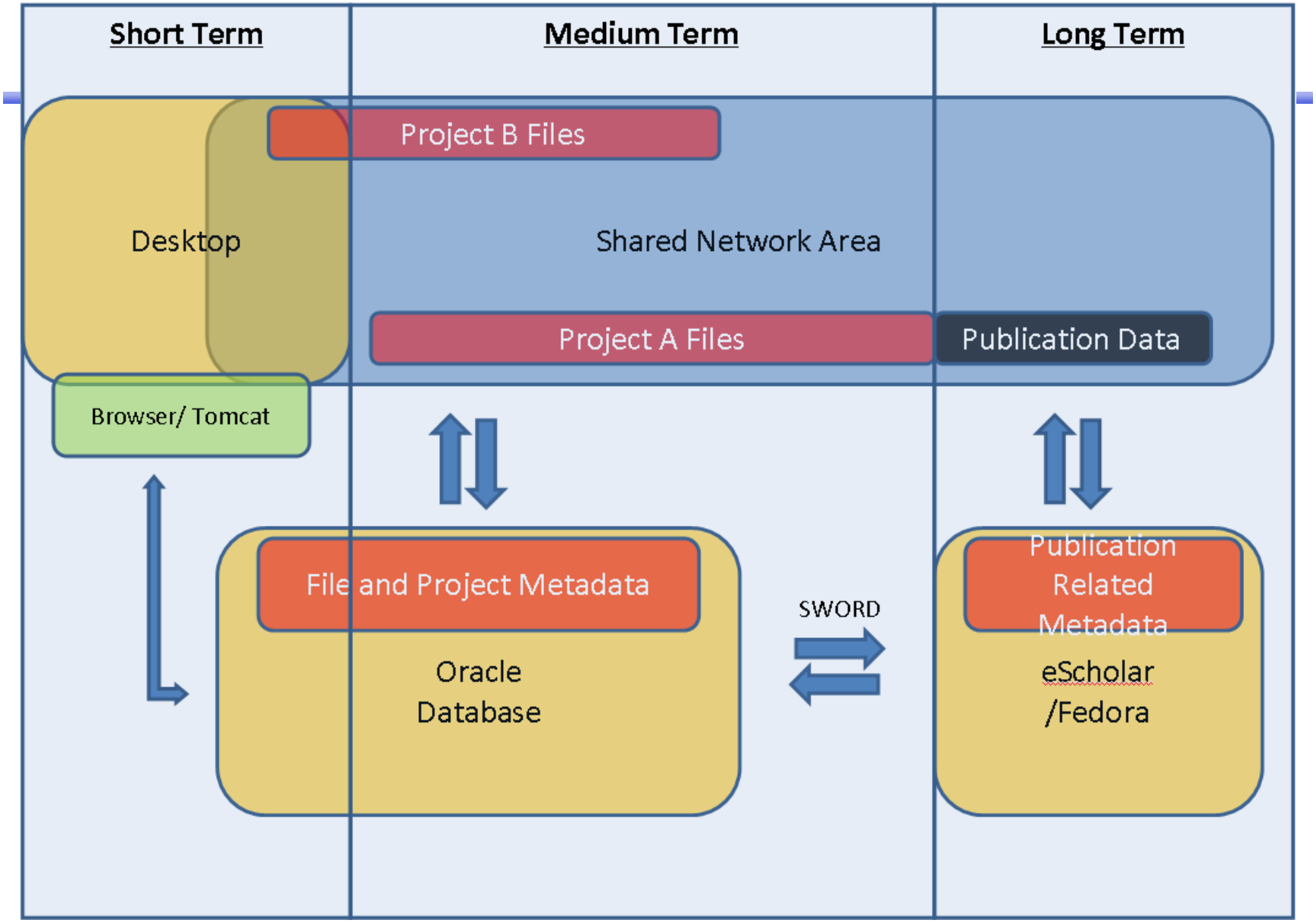
MaDAM Pilot Overview



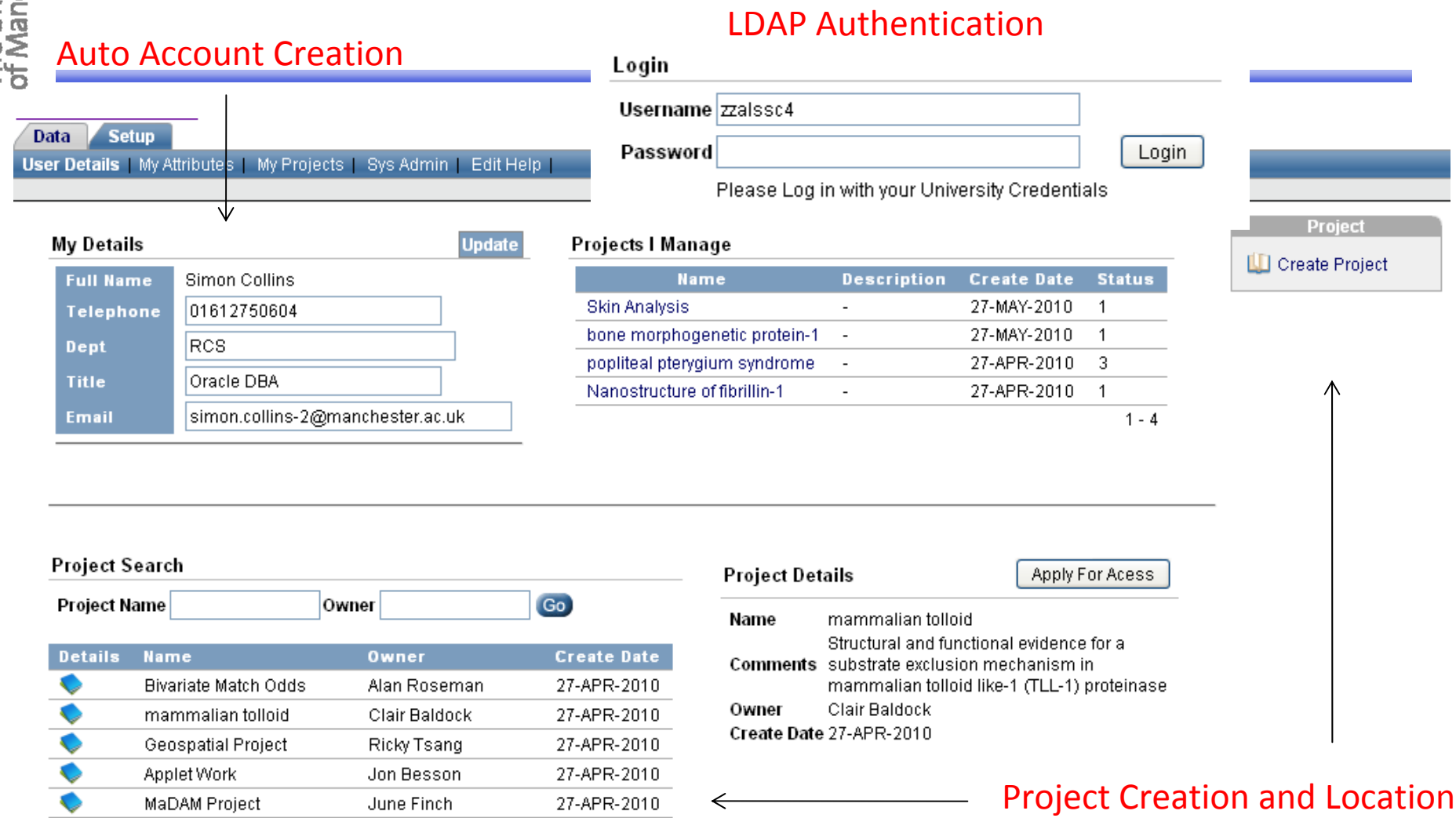
Many angles to cover:

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

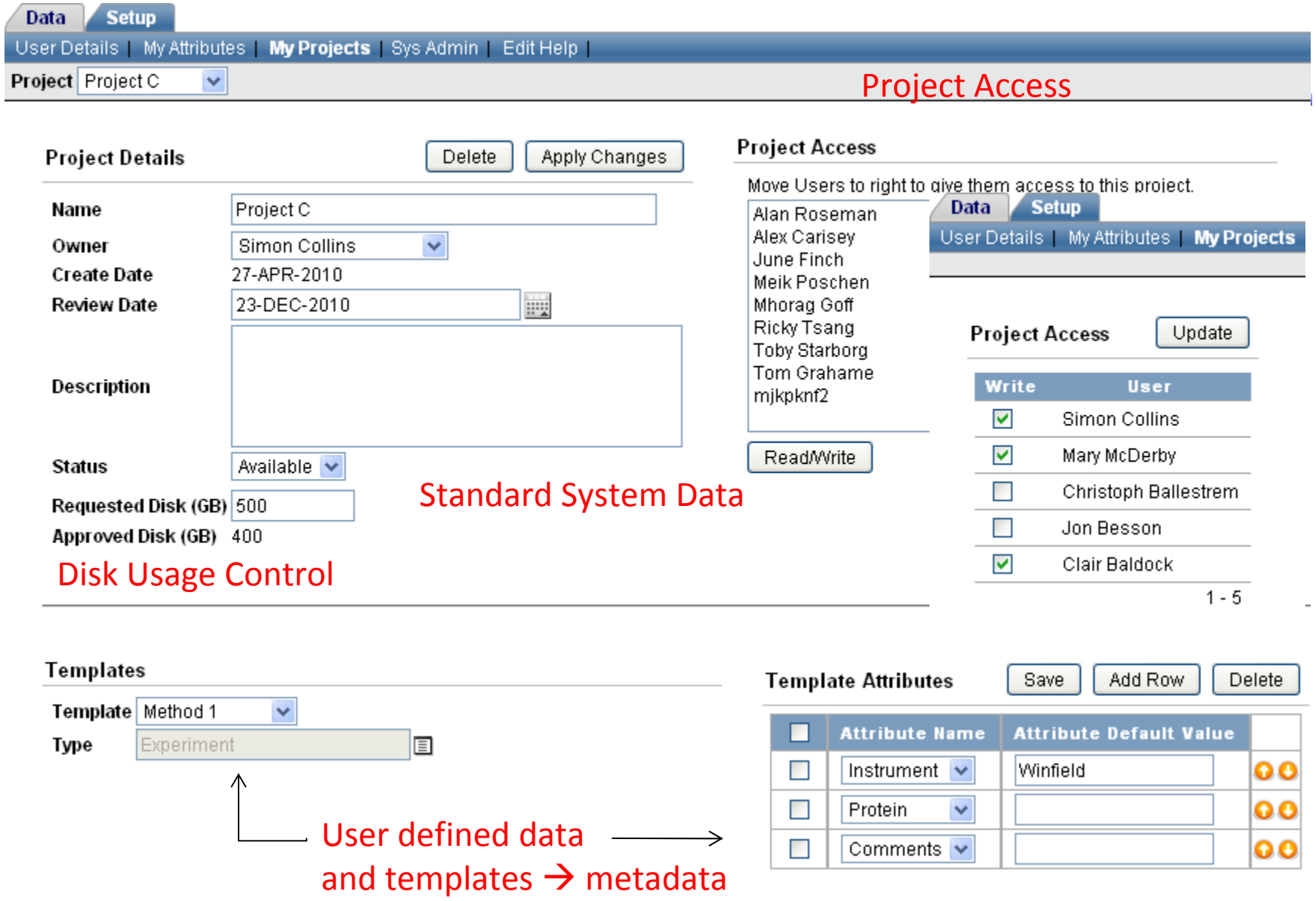
MaDAM Pilot Storage/Architecture



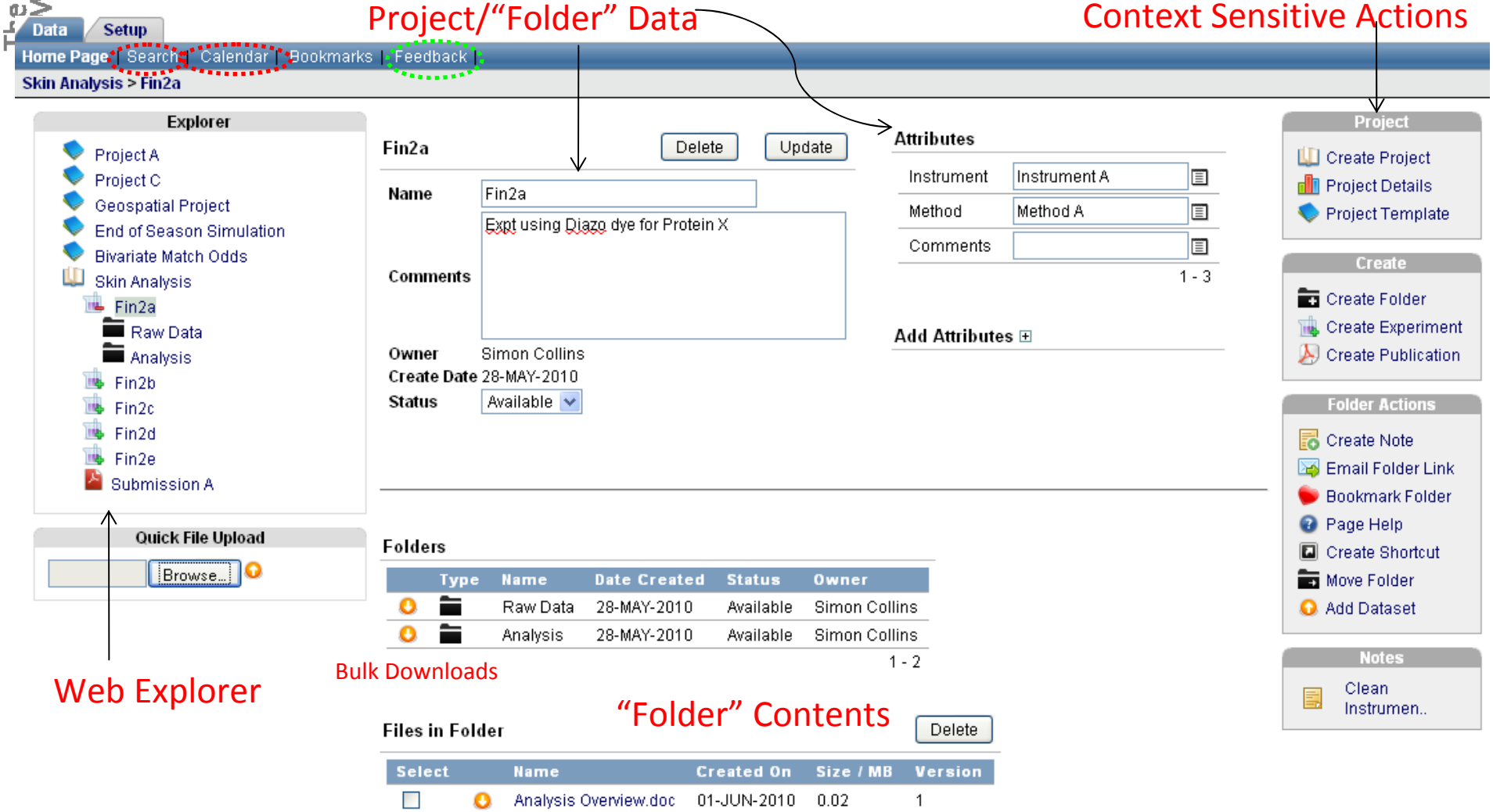
MaDAM Pilot: Authentication/Accounts



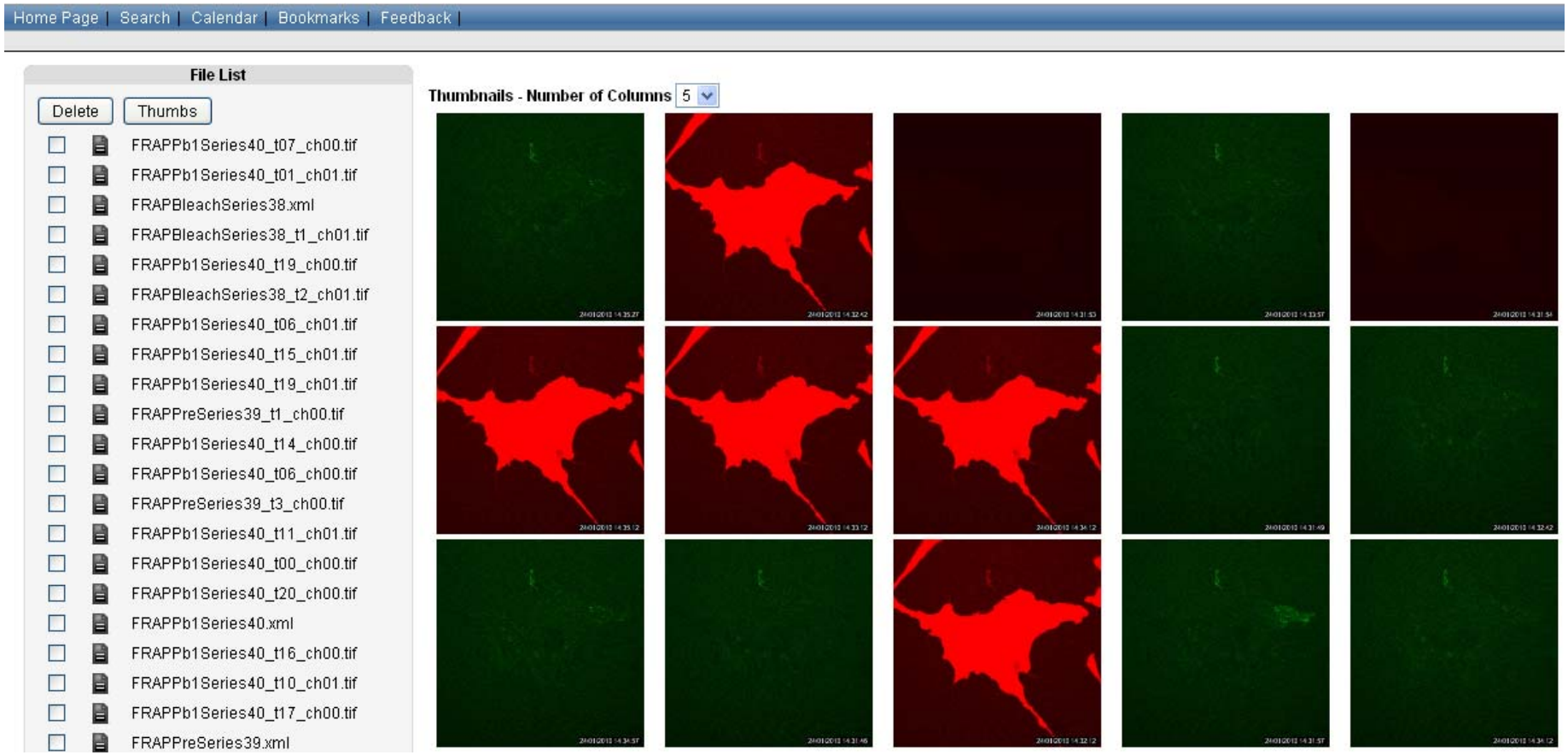
MaDAM Pilot: Project Setup/Access



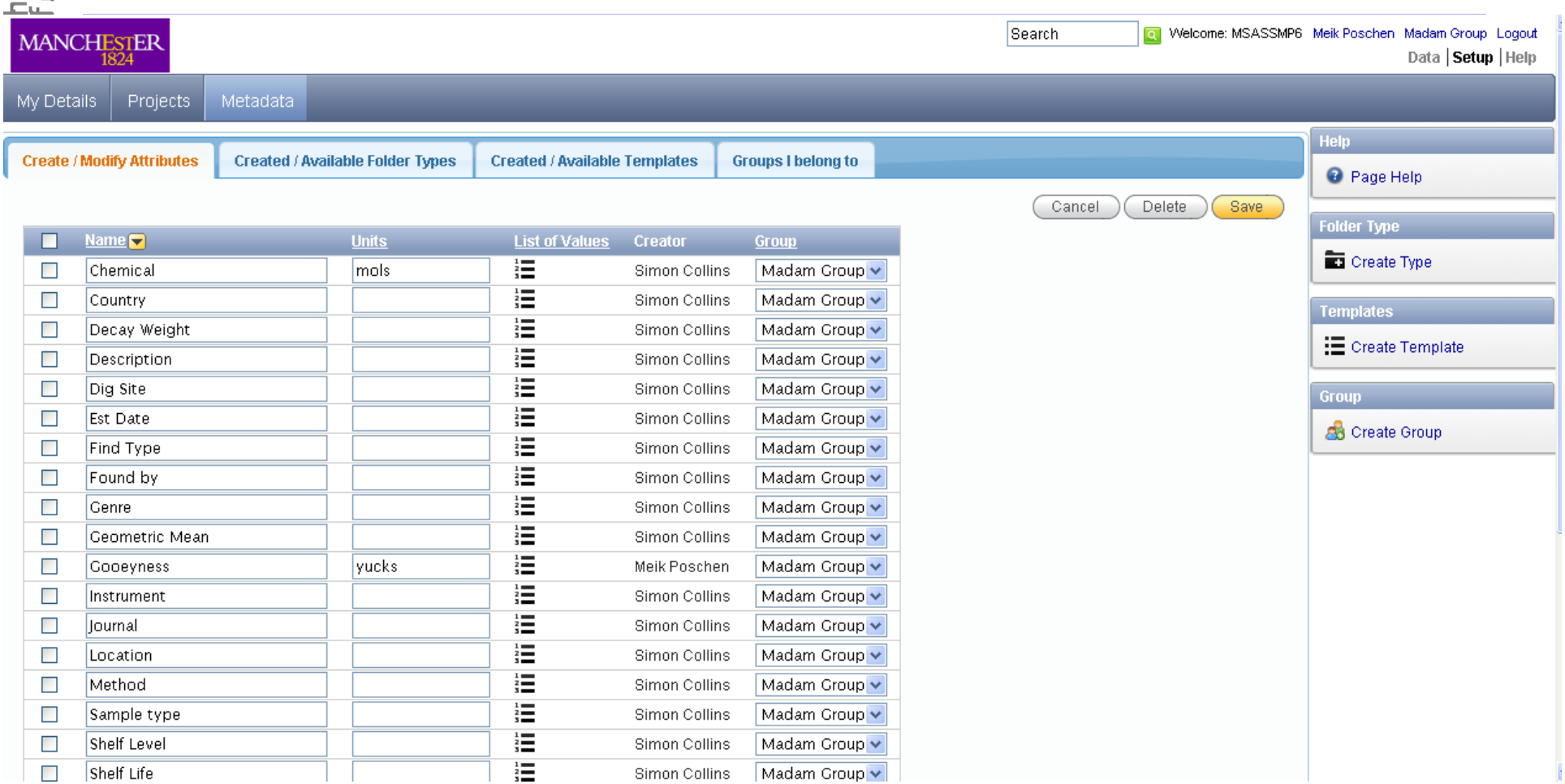
MaDAM Pilot: Data Management



MaDAM Pilot: Thumbnails



MaDAM Pilot: Metadata



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

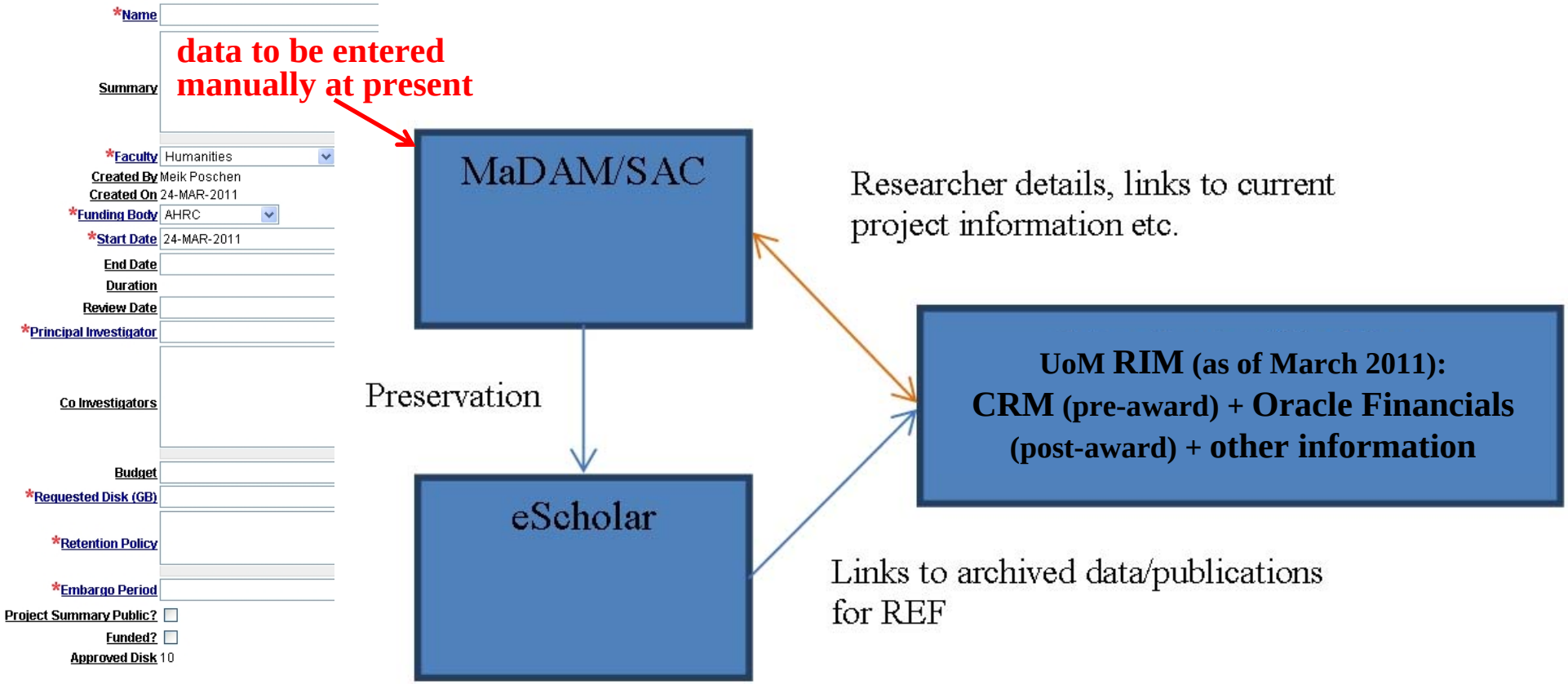
MeRC
Manchester eResearch Centre

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

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.

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!). High level institutional support is crucial, too!

- **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)

Challenges & Observations (3)

- 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
- Recognising a significant future challenge in managing storage capacity by balancing the facility to store research data centrally with initiation of a review process in which researchers will evaluate and re evaluate their data at appropriate intervals to instigate disposal of unwanted data whilst data of continuing usefulness is kept.

Some 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

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.
- Proof-of-concept integration of UoM Research Information Management Data (grant/accounting systems), DMPs and eScholar on the dissemination end.
- MaDAM 's outputs & findings (researchers' benefits) together with being part of an initiative for a sustainable University-wide Research Data Management Service led to funding the successor project

MiSS (MaDAM into Sustainable Service)

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

MiSS (MaDAM into Sustainable Service)

- MiSS is building on MaDAM, although it is more a transitional project than a continuation which will establish a sustainable service within the University's technical framework at the end of its lifetime in June 2013.
- Funded under JISC's Managing Research Data 2 strand, plus significant University of Manchester contribution.
- The RDMI will support researchers to manage (day-to-day), discover, store, archive and disseminate their data and comply with UoM, legal and funder policies and requirements.
- It will address the risk of data loss, potential re-use of data, preservation of future value in data assets, (open & community) standards, different research practices & data life cycles.

Delivering MiSS

MiSS will be delivering a service which will include

1) Rebuilding the MaDAM technical service infrastructure, making it more generic but tailorable (domain/discipline specific plug-ins/plug-in points), integrating it into the Manchester Working Environment (MWE),

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.

+ This core service will further integrate with the infrastructure holding administrative data (grants/financials) and eScholar.

Technical Framework Approach

- Integration/complying with MWE: standard set of software to create MWE-centric services and applications, including SharePoint - .NET Framework 4 - Microsoft IIS web server
- Shared Storage Project: automatic storage tiering – but potential cost implications for faster (= frequently accessed) storage → sustainability question for long term data storage/ retention
- Managing & linking external data: for data not stored in the main MiSS storage the RDMI will have to keep a record of that data and/or link to it
- RESTful API - SQL Server - ASP .NET (GUI) - C#

MANCHESTER 1824

The University of Manchester

MeRC
Manchester eResearch Centre

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) into DMPs
- Enabling better metadata ingestion from different sources in various disciplines (via community input)
- Setting up a financial model for the service (costing of storage, support, sustainability)

MiSS User Community

For the domain specific user community the project includes five research groups covering all four faculties with Academic Champions in

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

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

In the end: for the RDMS eventually all researchers/ academics at UoM will become the user community…

MiSS Baseline Requirements (1)

- Avoid duplication of data & enable linkage, create data records when data has not to be moved, especially if it is big.
- Try to keep the effort of (additional) data entry as low as possible, again avoid the duplication of effort on e.g. the creation of metadata entries (and DMPs).
- The openness of the RDMI (API, plug-in points) is crucial, not only to provide linkage to other systems, databases and repositories but also to prepare for automatic ingestion of metadata, e.g. from instruments or files.
- Metadata functionality is a base requirements, although some domains have a stronger need for enhancing discoverability and search than others.

MiSS Baseline Requirements (2)

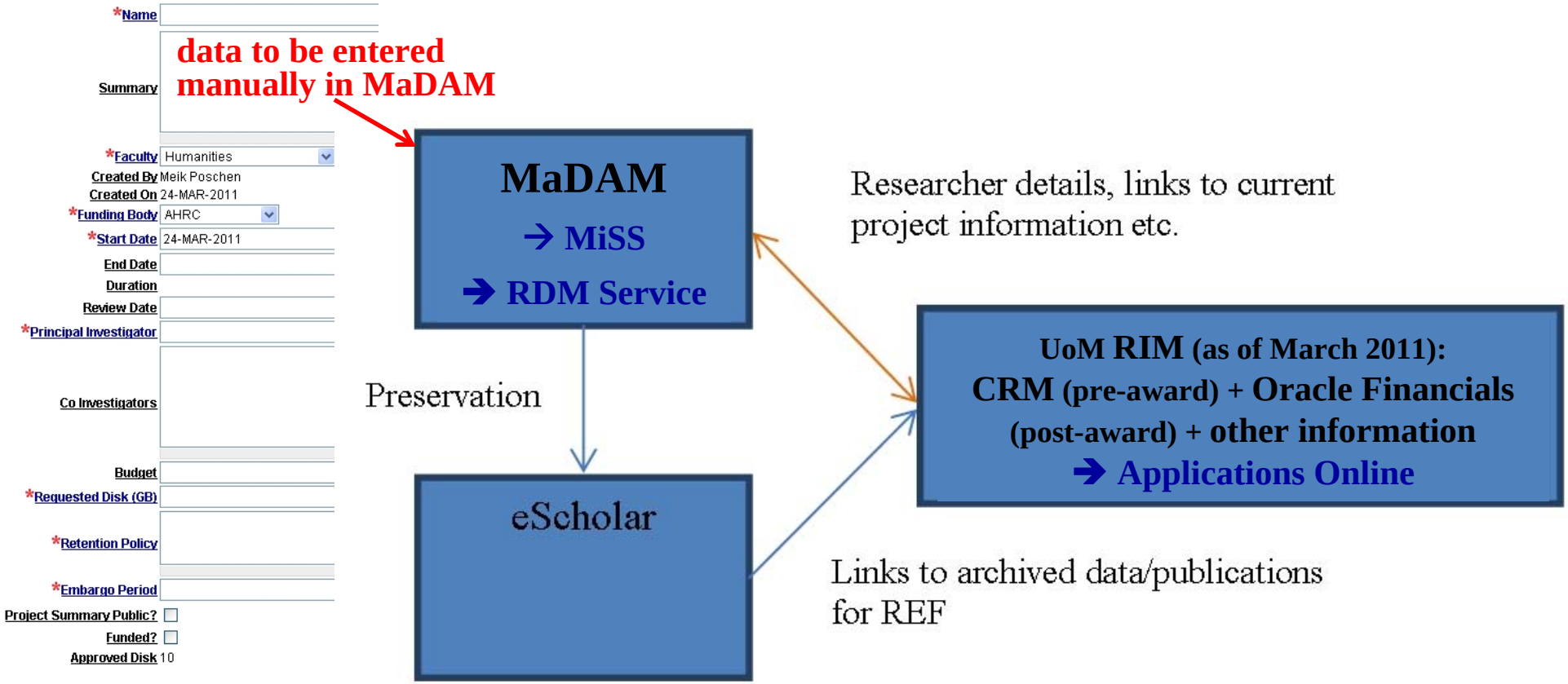
- Data access, version control, security and integrity models and policies are widely seen as integral for the RDMI, providing an audit trail or transparent path for data sets throughout the data lifecycle will be important.
- Unique Digital Identifiers (DOI & URI) will foster sharing and make linking data and publications easier.
- Transparent financial models are needed for storage and the service in general.
- The RDMI system has to be easy usable – and a tailorable interface would be 'nice to have'..

MiSS Challenges

- Bridge between and cater for generic and specific needs, making the RDMI easy usable but open enough for specific tools and automated data ingestion by providing ‘plug-in points’ (open API)
- Integration with the Manchester Technical IT Services infrastructure, which is evolving concurrently (Manchester Working Environment, MWE) – but different time scales
- Balance researcher, internal and external research data management needs and policies (research cultures & work practices, University structure with research offices & faculties and funder requirements)
- Engagement of existing support/admin structures at UoM is essential – high level institutional support is crucial

Integrating RIM Data: MaDAM - MiSS - RDMS

- MaDAM started exploring the integration of UoM RIM data (auto-retrieval → proof of concept)
- UoM's RIM environment itself was in the process of being linked more seamlessly



eDMP Framework

In the RDMS the electronic Data Managemnt Plan (eDMP) will be a living document – constantly updated throughout the lifetime of a project.

- Proposal Writing: “Planning” (→ Applications Online)
- During Project: “living document” with review stages (→ additional fields to fill in and update; → ingestion of Ethics & IPR data; → reporting mechanism)
- Final: stored as snapshot alongside research data (→ archiving)

Putting the ‘Service’ into Research Data Management 1/2

The RDM Service is being developed through a phased approach involving continuous in-depth consultation with academics across all disciplines.

The first phase involves the launch of a basic RDM Service in January/February 2013, which will evolve towards the full RDM Service over the year – this will include:

- A single point of contact for RDM enquiries.
- A website to provide a single location for RDM information and resources.

Putting the ‘Service’ into Research Data Management 2/2

- Support materials and resources for Data Management Planning, such as templates and examples of good data sharing statements to align with what funding bodies are demanding – developed in consultation with researchers.

- Support staff who can give advice and training on RDM.

- Procedures & guidance on the RDM policy – what researchers need to do and when (*UoM has ratified a high level RDM policy in May 2012*).

Research Data Management Support Network

- University of Manchester Library – providing the 'front of house' service, technical support for the RDM System, and raising academic awareness.
- IT Services – providing core IT infrastructure and support.
- Research Support Services (RSS) – providing funder mandate support and adherence.

We are continually raising awareness, meeting with stakeholders (researchers, support staff etc.) and are conducting a series of focus groups and training events in the build up to the full service.

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