6th International Digital Curation Conference, 6-8 Dec 2010, Chicago, USA – Participation & Practice: Growing the Curation Community through the Data Decade

# Understanding Research Practice & Enhancing Digital Curation in the Biomedical Domain: Requirements for the Participatory Development of a Pilot Data Management Infrastructure

Mhorag Goff, Meik Poschen, Rob Procter, June Finch, Simon Collins, Mary McDerby, Jon Besson, Tom Grahame, Lorraine Beard

The MaDAM project is implementing a pilot data management solution for research data at the University of Manchester as part of the JISC "Managing Research Data" programme.

For more details & contact please see http://www.merc.ac.uk/?q=MaDAM

**Aim**:
To produce a technical & governance solution based on researchers' requirements with flexibility to meet needs across multiple research groups / disciplines.

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

**Background**
- 18 month project time frame starting Oct 2009
- No existing institutional repository or strategy for management of research data.

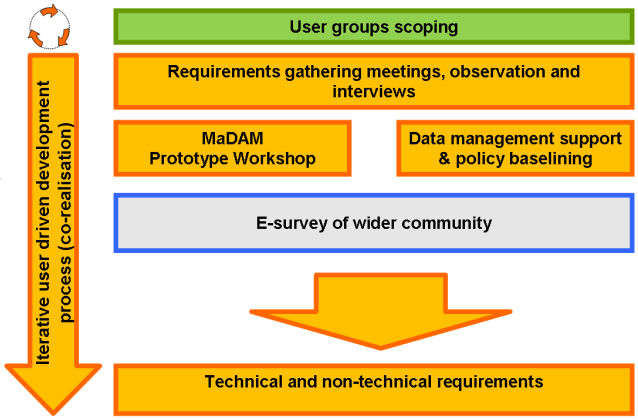
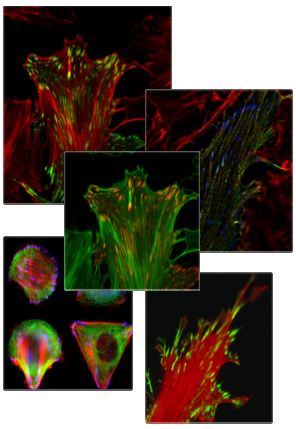



Fig. 1 Requirements Methodology



Up to 12 different file types
From 0.5MB to 17GB/file
**'Raw data'**

**Methodology & Requirements Capture**
- Bottom-up approach (no top-down requirements)
- User driven iterative development process
- Emphasis on researchers' concrete work practice & data lifecycles (co-realisation)
- Qualitative methods including interviews, observation, workshops
- Focus on a variety of imaging related research

**User Communities**
- Life Sciences & Medical Sciences researchers at University of Manchester
- Mainly imaging data generated on a range of instruments
- Large and diverse data sets
- Complex and evolving processing and analysis pathways
- Issue of managing confidential data for Medical Sciences

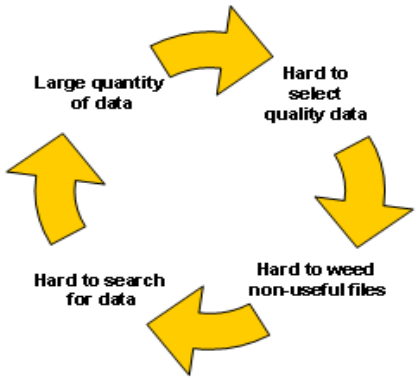



Fig.2 Data 'deluge' cycle

**Curation Practices**
Local data management practices for storing, sharing, preserving, refining and adding value to data in the context of large volumes of data.
- 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

**MaDAM Solution will..**
- Provide trusted secure storage to reduce risks of data loss
- 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

Funded by: JISC
Project Partners: MeRC Manchester eResearch Centre, Research Computing Services, The University of Manchester The John Rylands University Library