Managing Research Data: a pilot study in Health and Life Sciences

The UWE Case Study

Introduction
The University of the West of England (UWE) is one of the largest post-1992 universities. While primarily a teaching and learning institution, it also has a flourishing research community, strong links with business, and a growing reputation for excellence.

Having already successfully introduced an Open Access Research Repository for research outputs, there is a growing understanding within UWE that data curation is an essential part of the research cycle. The need for a usable research data management system is also recognised as all the more important in the current research environment where funders are increasingly stipulating that data management be properly addressed, and research data is made accessible.

The university is now seeking to build on the success of the Research Repository and to develop a coordinated approach to research data management; essentially wishing to close a gap in the research cycle at UWE where outputs are readily available, but supporting data is not.

In particular, this project aims to develop a research data management infrastructure which is transparent and usable by teacher and practitioner researchers, and which is appropriate to a post 1992 institution. To date, we have assessed and built upon previous JISC MRD projects and DCC documentation, and drawn on their outcomes and guidance to develop models of requirements for our own researcher community. Such input has informed our understanding of UWE researcher data management activities and aided the development of initial gambit processes and protocols fit for UWE. These will form the initial building blocks of test cycles of RDM processes and their fit with institutional practice, including process flows, development of an institutional policy, and the expansion of the research repository to include research data outputs.

Case study
A key component of the UWE project is to share our experience and deliverables with the wider JISC and HE communities. This case study summarises our experiences so far, and outlines the key stages in the process that might be adopted or adapted by similar institutions. We have also sought to capture lessons learned, recommendations for replication, and recommendations for further investigation – many of which are inherent in our outputs. In the spirit of the programme, we have taken an open view to sharing as much of our work as possible, and links are provided to all available outputs, with the caveat that information currently sensitive to UWE may not be available at this stage.

Key stages
Through a process of continual reflection, we have identified seven key stages in our journey thus far. Applied chronologically, these are:

1. **Approach**
2. **Outputs (early)**
3. **Stakeholder engagement**
4. **Outputs (interim)**
5. **Stakeholder engagement**
6. **Outputs (v1 models)**
7. **Next steps**
Each of these stages are outlined in further detail within this narrative, encased by a seven-stage roadmap (which evolved as part of our output planning process) to visually outline our direction of travel. Only the final outputs, which mark the cumulation of this phase, will travel with us into work package three to enable us to take only the most appropriate, UWE fit, elements of others’ and our own work into the next phase of the project.

**Approach**

The UWE work is made up of six work packages, of which Dissemination and Project Management run through the entire project lifecycle. The current milestone marks the completion of work packages one and two: the development of Models of data management requirements and Institutional processes which have run concurrently. Work package 3, Test Cycles, is the heart of the project during which we will apply and build on the outputs from our work to date, assessing iteratively their fitness for purpose.

We do not see the outputs at this stage as complete and polished, but as an informed initial baseline against which to test, challenge and support the processes and practices highlighted for replication and further investigation. It is important for the learning and understanding of opportunities and challenges around research data management within UWE to continue to grow in order to support the wider organisation in its development towards greater RDM maturity.

**Research project selection approach**

The project’s journey began with the selection, recruitment and engagement of researchers within two pre-selected research centres. The work is focused on the data capture and curation needs of the Centre for Research in Biosciences (CRIB) and the Centre for Health and Clinical Research (CHCR), both within the Faculty of Health and Life Science (HLS). The two research centres capture a range of data types, issues and challenges, and their involvement builds on interest and commitment already expressed by the centre Directors.

Working with groups within HLS is expected to afford the opportunity to look at a variety of data management issues, which will assist in scoping the challenges for an institution in the early stages of establishing data management policy, protocols and systems. These are expected to include matters relating to ethics and confidentiality, a wide range of data formats, working with external partners, and issues of intellectual property ownership.

In summary, research projects were enlisted through a process of matrix composition; interrogation of the university electronic Project Approval Support System (PASS); application of the matrix; shortlist; refinement of shortlist; invitation to participate; further refinement of shortlist; and finally researcher briefings.

An initial matrix was devised by a team of senior Research and Knowledge Exchange Librarians and the Deputy Head of Research and Development, and populated with projects from the central university Project Approval Support System to produce a short list of potential projects for inclusion. This was reviewed by the centre Directors and recommendations were received for which researchers to approach.
Factors such as individual researcher’s workloads, known difficulties with a project, or complexities such as wide geographical spread of the project, led to some projects being replaced by others. While the criteria identified on the matrix guided the process, the research centre Directors, with their detailed knowledge of the projects and the researchers involved, were the final arbiters of which projects were included.

Overall, the main barriers to this exercise were: ethical clearance delays; researcher workload; time commitment; nature of individual projects (e.g. content, geographical spread); lack of understanding of research data management issues; lack of clarity on the part of researchers about the project remit/purpose; maturity of the individual research projects; weighting of early career researchers; retention of RDM skills post-project.

Final coverage of data types includes text based documents, spreadsheets, databases, images, audio and video. Issues touching on the use of human tissue, commercial partnerships, ownership within a sponsorship framework, and potential conflicts between funder and sponsor requirements were raised.

An assessment of funders’ requirements confirmed a combination of data expected to be output to central data stores, and data with no obvious final resting place (orphan data), which is within the scope of the proposed UWE data repository.

It is important to note at this stage that the project is not anticipating major technical developments, rather to extend the existing institutional (Eprints) repository. A fundamental principle of the project approach is to apply existing good practice, and not to do things twice! As such, the scope of the university data repository is data outputs not otherwise accommodated; the repository is intended as the final resting place for completed data that supports a research output. Issues of discoverability of data accommodated elsewhere will be addressed. In addition, while not a data.bris type working storage solution, development of the Data Repository will need to address issues around storage of large datasets to assist in scoping the challenges for an institution in the early stages of establishing coordinated data management processes.

### Building on previous JISC MRD projects

The second key element of our project approach was to assess and build on the outcomes of previous JISC MRD projects. The objective of understanding what has gone before was to inform the shape of our own requirements analyses, in terms of both collection methods and content.

In analysing the requirements of UWE researchers we drew on the outputs of projects within the JISC MRD programme 2009-11 to develop a questionnaire for the researcher participants (which in turn fed into an interview schedule). As part of this process, we identified a number of common concerns particular to the two projects felt to have particular relevance to the selected UWE research centres (MaDAM and Incremental). These were shaped into potential barriers for further investigation within UWE, and became four sections of the questionnaire exploring the barriers the researchers had experienced in managing their research data. Two attitudinal questions were added to explore reasons for restricting access.
In addition to an investigation into the barriers our researchers experience in managing their data we wished to develop an understanding of the data held by the researchers; the policies and guidelines influencing their data management practices; and the advice and support researchers use and need to assist them.

The first is central to the project, the second and third were topics in the JISC MRD Incremental, IDMB and Sudamih projects which led to key findings.

Whilst drafting the questionnaire, and in an attempt to understand the complexities of RDM, we developed a mind map of the components or facets to conceptualise lessons learned. This is included in our outputs as something potentially of use to other institutions in the early stages of their RDM thinking.

A detailed description on our assessment of previous MRD projects as building blocks for our own work and for replication within the UWE project is available; also see the questionnaire itself.

Digital Curation Centre (DCC) tools
In addition to the development of models for assessing the research data management requirements of the pilot researchers, we were mindful of the challenges of institution-wide engagement. The need to evolve the hearts and minds within the university was apparent from very early on in the work, together with the role of the pilot project to challenge the current university ethos around RDM. Not only is it essential for the project to drive a sea change in sympathy with the “UWE way”, it has a key role in the development of strategic thinking well beyond the project team and steering group. We have learned that early engagement is likely to be pivotal to the sustainability of the project’s RDM protocol and practices offerings.

Initially we developed, mostly from DCC materials, a matrix of benefits of effective research data management against the risks if research data management is not developed effectively. It soon became clear that a two dimensional matrix was inadequate in explaining the differing benefits and risks accruing to the wide range of stakeholders over time, so we added a third dimension using colours and positioning to represent various stakeholders. The matrix has been, and will continue to be, used with a wide range of stakeholders towards wider institutional engagement and understanding.

Also based on DCC outputs, we derived an enterprise maturity model (or indicator of the institution’s current data management activity) applicable to the UWE way, taking as a starting point the DCC mini cardio quiz, or pulse check. The resulting model (affectionately known as the as-is-inator) is simply a two-sided set of statements (best presented in A3) on a scale against eight key research data management challenges covering Risks associated with poor data management; Research funder policies, Institutional RDM policies; Training, support and guidance for researchers; IT infrastructure, back-up and storage; Institutional repositories; Funding; and Staff skills. The statements are intentionally emotive to engage and, hopefully, virally infiltrate wider discussion and are re-useable in the context of a Position Statement.

In addition to tangibly benchmarking the as-is, and outlining what the ideal future UWE RDM status might look like, this model forms the basis of a Position Statement and subsequent aspirational, high level Strategy including a vision and principles of commitment. The model is not scientific (i.e. no weighting even though some middle ground is much further developed than others in statement terms) but does offer a rudimentary benchmarking measure for use with an unlimited range of stakeholders, at any level, here at UWE.

As part of this modelling, we have looked at collating the data electronically, with support from DCC who are able to customise the original quiz for specific target audiences should we require. However, we do not want
to lose the granularity of the continuous measure that paper affords, and we are finding that there is mileage in having a paper version as some stakeholders tend to find this more accessible. We will, none-the-less, be trialling it as a discrete measure with PowerPoint Turning Point (voting buttons) over the coming months with some of our senior managers.

**Outputs (early)**
The early outputs of the work were briefing tools. In essence, we have taken informed guidance (in) and formed models to investigate its potential application to UWE (out).

- Launch poster
- Researcher information sheet
- Researcher briefing presentation
- Building on previous JISC MRD projects
- Research Data Management mind map
- Benefits matrix (stakeholder analysis)
- Cardio maturity model
- Researcher questionnaire

**Stakeholder engagement**
With the research, planning and preparation complete, the team were ready to engage on a one-to-one level with the selected researchers. The objective of this was to assist sample researchers to identify data management activities based on their requirements.

An information sheet and a consent form for researchers were drafted, giving an outline of the project. However, the process stalled at this point as centre Directors felt unable to distribute the information sheet or consent form until university ethical clearance had been achieved. This had been overlooked as a result of university Services, which include the library, falling outside of the parameters of Faculty ethical processes (a lesson learned for later process modelling). Invitations to participate, and the information forms, were only distributed once the HLS Faculty ethics process had been applied to this project, and ethical clearance gained.

The dissemination at this point generated some further discussion among researchers about the appropriateness of some projects being involved; issues included the stage of development of some projects, the time commitment for researchers and a fundamental lack of clarity in one case around the purpose of the data management project, and the importance of data management as an issue. This resulted in further refinement of the shortlist; from which seven research projects across the two centres were selected.
**Researcher briefings**

Researchers were invited to initial **briefing sessions**, where the purpose of the project was presented in more detail. One briefing session was attended by researchers from all but one of the projects. A separate meeting was held for researchers in the remaining project. Fortuitously, this enabled the team to address particular concerns to a single project about workload and data ownership.

Following the main researcher briefing, the project team did consider whether the projects involved (some of which were represented by fixed term researchers and research students) represented a possible lack of commitment from senior researchers in one centre. On consideration, it was decided that the benefits of enlisting early career researchers (propensity to hold orphan data, RDM inexperience, diversity of data types, time to commit to the project) outweighed the risks (retention of RDM expertise, senior staff commitment perceptions); assurances from the centre Director who, as a supervisor to one of the projects, would keep a watching brief on developments with these projects and enable outcomes to be integrated into the research culture of the research centre, resolved this issue. Touch points with centre Directors have now been incorporated into the project communications plan.

Following successful briefings, the result was seven researcher projects becoming involved, ranging from DNA sequencing of oak tree bacteria; through technologies for animal health and food quality traits; genetics of pre-term infants; visual scanning training in Occupational Therapy for patients with visual search deficits following stroke; Cognitive-behavioural approaches in routine care (Rheumatology); Facilitating Activity and Self-management for Arthritis; to breast feeding in traveller and gipsy communities.

**Key questionnaire findings**

The key objectives of the questionnaire were to inform the researcher interviews to ensure relevance and depth of information collection, and to enable an indicative requirements outline to be initiated. The questions had drawn on the lessons learned from previous JISC MRD projects, and replicated issues requiring further investigation at UWE. The **questionnaire** was distributed electronically to the researchers immediately after the briefings.

From the responses, an analysis of researcher data management activities, together with technical and administrative RDM activities was drafted. What started as lists developed into a **toolkit presentation** (in order to embrace all strands in a single file) of technical, administrative – and also process – requirements identified to date to support the interviewer in discussion with researchers. Training was added as a further category during the process as it was apparent that the guidance and training work package would be led by a clear set of requirements identified at this stage. Other projects had, at the JISC MRD policy workshop on 12/13 March, discussed applying a service / infrastructure model to their RDM processes. This is something that the team may consider.

At this stage it also became clear that understanding of existing UWE processes was vital to the development of the process strand of work. This work formed part of the wider UWE discussions (below). The key elements have since been addressed via a v.1 UWE RDM administrative processes flow chart.

The **interview schedule** itself is a reflective summary of the key questionnaire findings, and this is supported by a **brief summary of the attitudinal findings**, which revealed some interesting differences between researchers in the two centres that might prove useful in due course and would otherwise have been lost.
Many of the issues reported by the researchers touched, as expected, upon technical issues outside of the immediate scope of the project but potentially within the project’s frame of influence. Responsibility for intermittent availability of server space, data sharing and transfer, long term data storage and back-up of files probably defer to a central IT function, however it is acknowledged that further investigation of other MRDII work, such as the University of Oxford’s DataFlow project, or data.bris, may offer interim or longer term solutions.

Common and specific problem areas included the wide range of storage media used, confusion over data ownership, lack of support, advice and guidance on managing data, a need for clarification of university guidelines in relation to how long data should be preserved, and recognition within the funding bid process that data management takes time. These are all areas for further investigation when considering guidance and training and will be addressed in work package four.

Responses to the questionnaire were forthcoming, prompt, and detailed; a likely reflection on a combination of the research and preparation, the high levels of researcher engagement, and the questionnaire being administered using Survey Monkey which team members had successfully used in previous projects.

**Wider UWE discussions**

Wider UWE discussions outside of the pilot framework took place concurrently, involving members of the project team, research administration staff, and meetings between the Project Manager and the research centre Directors. Opportunities identified by research centre Directors included examples of good practice research governance at UWE, such as the new Peer Review College and its potential future links with PASS. Links with industry and the potential for our current research administration tools to measure effectively research success and impact at centre and university level were also discussed. On reflection, this provides a wealth of material for further investigation and replication.

Discussions regarding the new policy framework on research data published by the EPSRC were instrumental to this process, responsibility for which lies within UWE’s Research and Business Innovation Service (RBI). These compounded the importance of the link with RBI for sustainability, longevity and, not least, implementation of our RDM models; a significant lesson learned. We will work closely with RBI’s emerging Research Governance Task Group going forward to effect change directly, timely and centrally.

These interviews, together with additional meetings with key research administration staff, have also enabled an initial understanding of existing UWE processes.

We took guidance available from DCC documentation (Mini Cardio) to enable us to start to develop *appropriate to UWE* protocols and processes. Understanding the UWE way has been a two-pronged approach combining these meetings with the application of the enterprise maturity model outlined above.

**Project team Yammer communications**

The JISC MRD project here at UWE is very modest in size with just one FTE dedicated team member. We learned quickly that the small and part-time nature of the project team presented a real risk of single point of failure should the Project Officer be wholly responsible for continuity. To mitigate this risk, we have introduced a closed team Yammer group, to which we have also invited senior library management for overview purposes. This has massively reduced the practice of round-robin emails and provides a central source for informal discussion and update. Key documents are loaded onto the group, and pages are used to brainstorm ideas and leads centrally. This is working well for the project, although we did experience a few
temporary teething problems (e.g. inadvertent posting outside of the group).

**Outputs (interim)**
The interim outputs of the work were, in essence, preparation for the interviews.

- **Questionnaire analysis (attitudinal)**
- **Researcher interview schedule**
- **Discussion toolkit**

**Stakeholder engagement**

**The researcher interviews**
Arranging the researcher interviews themselves was, again, a really positive reflection on researcher engagement, with invitations and options to attend posted via Doodle with an almost immediate complete response. This suggests that the project’s early consideration of researcher engagement, together with an appreciation of the challenges researchers are likely to face, and a clear understanding of the risks and benefits to them, was successful.

At this face-to-face interaction, consent forms were formally completed to ensure ethical compliance. The Project Officer was responsible for all seven interviews, which were tracked using a grid to ensure accurate coverage against the devised schedule, and recorded for ease of analysis (with explicit researcher permission including caveats surrounding the management of the data relating to them as human participants). The Discussion Toolkit was not shared, per se, with the researchers at interview so as not to overwhelm them but interviewer knowledge of its content was crucial to the success of the interviews.

The interviews revealed that research data management at UWE is frequently a matter of experience and personal practice; with personal methods developed for file naming; storage media selected due to personal preferences; back up self managed by researcher on irregular schedule, and sometimes in addition to UWE server back up; data management plans developed ad hoc and informally, with little help and no review; learning from experience and own mistakes and little formal training in data management.

Such clear articulation of training needs at this stage has enabled the project team to outline guidance and training needs towards the development of UWE specific guidance and training. Work on these will include an evaluation of existing models of delivery with a view to consolidate disparate sources of institutional guidance and support through a central dashboard, or hub. This can potentially include all related existing governance policies (such as ethics, data protection). It may be possible to adapt and rescope some of the existing guidelines to apply to RDM. Identification of all existing policies and guidelines forms part of the Research Governance Task Group’s terms of reference and we will be able to utilise this forum.
We have also been able to establish a draft administration process flow chart for RDM. This forms the basis for a pilot plan for integrating data management and curation procedures into existing university processes and cultures. We expect to challenge and support existing processes, to identify opportunities for short and longer term gains, and to develop and formalise protocols and process including written documentation.

**Maturity Benchmark and Target Operating Model**

All stakeholders with whom we have currently engaged have been asked to provide both an as-is and aspirational assessment against the maturity model. By asking people to identify their perception of our current position, and our aspirational position as a university, we have established some tangible as-is and target operating model measures, by stakeholder group where appropriate. This has currently only been used a little for some early baselining – on the understanding that we need to understand the as-is to seek to change behaviours – but we will continue to gather data throughout the life of the project and to refine the direction of travel. As part of the process we carefully selected phrases from the project brief so that we also have a project trajectory.

When conducting the maturity as-is analysis we learned quickly that a separate analysis was needed for each cohort, to avoid double counting in any scoring algorithm. A rudimentary algorithm was used to take the mean score of a cohort on a 1-5 scale, where a mark anywhere within the box scored at that integer. Only where respondents had deliberately placed a mark between two boxes was any granularity added (at 0.5). Analysis by cohort revealed differences between stakeholder groups, with those with formal responsibilities for research administration generally scoring more favourably. The only area that everyone measured to date agreed on was the current funding model, suggesting that among those closest to the work, there is no ambiguity as to the reliance on the JISC MRD Programme to support this work at UWE.

This analysis is not available at present but may be released at a later date. Similarly, the resulting Position Statement, summarising the current status-quo and used to inform policy discussions, may be released later.

The current baseline of what the ideal future UWE RDM status might look like, our Target Operating Model, is appended (and is visually similar to the as-is). On a five-statement scale, and based on the algorithm outlined above, Staff skills scored the highest. The least commitment was to IT infrastructure, back-up and storage followed by the general Risks associated with poor data management and Institutional RDM policies. Institutional repositories, Research funder policies, Training, support and guidance for researchers and Funding all ranked at a high 4.8.

The findings will be incorporated into a draft institutional policy (with the Position Statement outlined above as an opener) starting with a Strategy (vision) to give some momentum to institution-level buy in and to seek institutional commitment from which to build the necessary infrastructure platform. This will, in turn, feed a set of enterprise research data management principles of commitment (models of which have successfully been applied in HE Data Strategy projects with which team members have had experience, school Assessment for Learning models and, of course, The University of Edinburgh).

**Researcher Yammer communications**

Building on the success of the project team Yammer group, a separate closed Yammer group has now been set up for the project researchers to communicate with each other (and us, and vice-versa). Uptake has been slow, with two researchers signed up at present. We have recently emailed them to suggest that they might like to post some reflections on how the questionnaire and interview has helped them with their personal research data management to date as this was a common feedback at interview. The researchers’ Yammer
group is now in operation instead of the initially proposed closed Facebook group as a result of concerns around intellectual property.

**Outputs (v1 models)**
The final outputs from work packages one and two represent the models of research data management that we will take forward into work package three for testing, the heart of the project. These have been baselined at v.1 as they will be expanded and developed hereon.

- UWE RDM administrative infrastructure
- Eprints requirements
- Guidance and training needs
- Maturity benchmark
- Target operating model (TOM)
- Positioning paper

Note that the Eprints requirements have, to date, been derived in isolation (with an eye on JISCMAIL discussions) based on UWE’s existing Eprints Research Repository. Work package three will lead with technical developments, with a technical solution available early in the testing cycle which will be developed in iterations.

**Next steps**
Work package three, Test Cycles, is the heart of the project. The model outputs outlined in this case study will form the basis of consultation and development. The next steps are as follows:

- Pilot plan (for data management for two Research Centres)
- Timetable for work packages
- Iterative (as part of a wider enterprise “push”) maturity measuring and TOM refinement.
- High level RDM Strategy: vision and principles.

We would welcome feedback on this case study, including views on its suitability as a guide to universities attempting to address the RDM void.

*Stella Fowler, Project Manager*