JISC Final Report

Project Information			
Project Identifier	*To be completed by JISC*		
Project Title	Supporting Data Management Infrastructure for the Humanities		
Project Acronym / Shortname	Sudamih		
Start Date	01 October 2009	**End Date**	31 March 2011
Lead Institution	University of Oxford		
Project Director	Professor Paul Jeffreys		
Project Manager	Dr. James A. J. Wilson, Oxford University Computing Services, 13 Banbury Road, Oxford, OX2 6NN.		
Contact email	james.wilson@oucs.ox.ac.uk		
Partner Institutions	n/a		
Project Web URL	http://sudamih.oucs.ox.ac.uk		
Programme Name	*Research Data Management Infrastructure*		
Programme Manager	Dr. Simon Hodson		

Document Information			
Author(s)	James A. J. Wilson		
Project Role(s)	Project Manager		
Date	10/05/2011	**Filename**	Sudamih_FinalReport_v1.0.docx
URL	http://sudamih.oucs.ox.ac.uk/		
Access	This report is for general dissemination		

Document History		
Version	**Date**	**Comments**
1.0	10/05/2011	Complete report for publication

Table of Contents

1 Acknowledgements

The Sudamih Project was funded by the JISC under the Managing Research Data Programme and supported by the University of Oxford. The core project team consisted of Asif Akram, Dr. Meriel Patrick, and Dr. James A J Wilson, under the direction of Professor Paul Jeffreys and Dr. Michael Fraser. Additional governance and direction was provided by the Sudamih Project Steering Group, chaired by Professor David Shepherd, and the Project Working Group.[1] We are particularly grateful for the involvement of the academic research staff (Professor Andrew Wilson, Dr. Ian Archer, and Dr. Miko Flohr) who have supported the project throughout with their advice and expertise. The involvement of research support staff from various departments has also been greatly appreciated, in particular for helping us better understand how the different parts of the institution can work together to support data management infrastructure.

2 Project Summary

The Supporting Data Management Infrastructure for the Humanities (Sudamih) Project forms part of a programme of activities at the University of Oxford to build an infrastructure capable of managing the research data produced at the University from inception to reuse. The Project developed training and software to assist researchers in the humanities better manage their data. It also considered the institutional support required to maintain these outputs and the costs and benefits of doing so. The key outputs of Sudamih were a suite of training materials and courses to improve researchers' data management skills, and a pilot 'Database as a Service' (DaaS) system enabling researchers to quickly and intuitively create, edit, search, and potentially open up relational databases of research data. Whilst the focus of the project was on the Humanities Division within Oxford, it is hoped that the outputs can in the future be extended to other disciplines, and the findings applied more generally across higher education institutions.

Sudamih was funded by the JISC under the Managing Research Data Programme, so project outputs have been made available to the UK Higher Education community for adaptation and reuse.

Notable findings from the Project include:

- The intellectual value of humanities datasets tends not to depreciate over time
- Humanities scholarship often aggregates to a 'life's work' body of research, with any given researcher often wishing to go back to old notes, sources, or datasets in order to find new information
- Methods of organizing data vary considerably, as does the extent to which researchers succeed in creating and maintaining a well-functioning system.
- Most researchers are willing in principle to share their data with others, but in practice do not regularly do so, for a variety of reasons. In the humanities, issues surrounding the incompleteness of the original data, or the layer of interpretation often required to render it consistent, can lead to reluctance to share, as researchers worry that their 'processed' data may be misinterpreted by others
- Data storage is generally on personally-owned machines and backing-up is generally also to personal devices on an *ad hoc* basis. Knowledge of centrally-provided services is limited and they are seldom used.

[1] The full lists of Steering Group and Working Group members are provided in Appendix A – Composition of Steering Group and Project Working Group

- There is a significant amount of confusion over the ownership of research data. This is exacerbated by complex situations in which multiple people or organizations may have different claims on the same resource.
- Data management training can have a large positive impact in terms of long-term cost savings relative to the near-term costs of running and maintaining courses and learning materials
- Researchers see various benefits of using a centrally-provided database management system such as the DaaS over current popular alternatives
- There is a need in the humanities for very-long-term data sustainability solutions and cost models designed to deal with effectively permanent storage and access

Key recommendations include:

- Base training around actual commonly-faced research problems, whilst also covering organizational principles and strategies
- Make it immediately obvious what aspects of data management any given training covers and who it is intended to benefit. Be careful to use terms that the researchers understand rather than the technical terminology used by data librarians
- Establish a single location or point of contact which researchers can be referred to when dealing with data management issues
- Institutions that are serious about winning research funding should in the future have a specialist technical advisory service which researchers can consult for assistance with the technical aspect of bids
- Universities should clearly disseminate information about central services that support data management to ensure researchers are aware of them
- Different academic departments and institutional service providers should work together to understand who should be responsible for implementing, and sustaining, various aspects of data management training
- Universities should clarify the intellectual property rights that researchers have with regard to their structured data outputs, and in particular their rights when depositing data in a repository or service such as the DaaS
- The HE sector should encourage and enable researchers to treat research data more like research publications, ensuring that they can be preserved, accessed, and cited over the long term

3 Main Body of Report

3.1 Project Context

The Supporting Data Management for the Humanities (Sudamih) project builds upon the data management infrastructure framework developed by an internally-funded scoping study at the University of Oxford[2] and its continuation through the JISC-funded Embedding Institutional Data Services in Research (Eidcsr) project.[3]

The scoping study took the form of an intra-institutional collaborative effort to understand service requirements for managing and curating the research data generated at the University of Oxford. It involved interviewing around 40 researchers as well as conducting a consultation exercise with service units across the University. The project also contributed to the UKRDS feasibility study and the piloting of the Data Audit Framework (DAF)[4] through the JISC funded DISC-UK DataShare project.[5]

The subsequent Eidcsr Project (which concluded at the end of December 2010) sought to assess and develop institutional infrastructure to address the requirements of three collaborating research groups involved in the creation of three-dimensional anatomical models of hearts. The process for doing so included the creation of very large histological and MRI images of hearts which could then be segmented and meshed to produce a model upon which *in silico* experiments could be conducted. The researchers with whom we worked were particularly keen to preserve their data securely so that it could be searched and retrieved later. To this end we developed a basic client that could store the data in the secure long-term filestore at Oxford University Computing Services whilst facilitating the capture of appropriate research metadata describing the data and the experiments that generated it. The metadata could then be held in the 'Databank' service[6] offered by the Bodleian Libraries and (with the development of a suitable interface) searched and retrieved. Eidcsr also involved the creation of a University data management policy and data visualization tools. This was an intra-institutional project involving the cooperation of researchers in the Department of Cardiovascular Medicine, the Department of Physiology, Anatomy and Genetics, the Computational Biology Group, the Libraries, the Research Services, Oxford e-Research Centre, and the Computing Services.

Sudamih has for most of its duration run parallel with Eidcsr, sharing a project manager and technical staff to make more efficient use of resources and ensure full coordination of effort. The infrastructure developed by Sudamih is distinct from that developed by Eidcsr, but intended to be complementary. Whereas Eidcsr sought to establish data preservation and documentation workflows (and tools to assist researchers follow these) alongside policy development, the main strands of the Sudamih Project were to develop training for researchers in order to enable them to better manage their data, and to develop a 'Database as a Service' (DaaS) system that would enable researchers quickly and intuitively to create, edit, and open up research databases via the Web.

3.2 Project Outputs and Outcomes

Output / Outcome Type *(e.g. report, publication, software, knowledge built)*	**Brief Description and URLs (where applicable)**

[2] The scoping digital repository services for research data management project, http://www.ict.ox.ac.uk/odit/projects/digitalrepository/.
[3] The embedding institutional data curation services in research project, http://eidcsr.oucs.ox.ac.uk/.
[4] Data Audit Framework (now renamed 'Data Asset Framework'), http://www.data-audit.eu/.
[5] UK Data-share project, http://www.disc-uk.org/datashare.html.
[6] Databank, http://databank.ouls.ox.ac.uk/.

PROJECT PLAN:	
Supporting Data Management Infrastructure for the Humanities - Project Plan v.2.1	Most recent version of the Sudamih Project Plan. Outlines aims of project and how they will be met. August 2010. http://sudamih.oucs.ox.ac.uk/docs/SudamihPP_2.1_nobudget.pdf
REPORTS:	
Sudamih Researcher Requirements Report v.1.0	Report on findings of researcher interviews. July 2010 http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Researcher%20Requirements%20Report.pdf
Use of the Data Audit Framework within the Sudamih Project v.1.0	Brief report on the effectiveness of the Data Audit Framework in the context of the Sudamih requirements gathering. July 2010. http://sudamih.oucs.ox.ac.uk/docs/Use%20of%20the%20DAF.pdf
Sudamih Benefits Case Study	Consideration and categorization of the benefits deriving from the training and DaaS components of the Sudamih Project. February 2011. http://sudamih.oucs.ox.ac.uk/docs/Sudamih_BenefitsCaseStudy_v2.0.pdf
Sudamih Research Data Management Training Business Case	Business case arguing that the training resources developed by the Sudamih Project should be maintained and disseminated by the University of Oxford's IT Learning Programme. Includes service costing and assessment of demand and benefits. http://sudamih.oucs.ox.ac.uk/docs/SudamihTrainingBusinessCase_v1.1.pdf
PUBLICATIONS:	
Wilson, James A. J., Fraser, Michael A., Martinez-Uribe, L., Jeffreys, P., Patrick, M., Akram, A., Mansoori, T., 'Developing Infrastructure for Research Data Management at the University of Oxford' Ariadne 65, October 2010.	Account of the approach taken by the University of Oxford to develop an infrastructure for supporting research data management, focusing on the tools and workflows developed during the Eidcsr and Sudamih Projects. Written September 2010, published in *Ariadne* 65, October 2010: http://www.ariadne.ac.uk/issue65/wilson-et-al/
Wilson, James A. J., Martinez-Uribe, L., Fraser, Michael A., Jeffreys, P., 'An Institutional Approach to Developing Research Data Management Infrastructure', International Journal of Digital Curation (forthcoming)	Account of research data infrastructure being developed at Oxford, concentrating on roles and rationale for approach. Written October-November 2010. Publication pending.
BIBLIOGRAPHIES AND LITERATURE REVIEWS:	
Data Management Bibliography	Annotated bibliography of literature relating to various aspects of data management, divided into sections on data sharing; curation and preservation; repositories;

	personal information management; and metadata and related issues. November 2010. http://sudamih.oucs.ox.ac.uk/docs/Data%20management%20bibliography.pdf
Personal Information Management - Literature Review	Critical summary of works relating to personal information management. November 2010. http://sudamih.oucs.ox.ac.uk/docs/Personal information management literature review.pdf
TRAINING MATERIALS:	
Slide packs: 'An Introduction to Research Data Management in the Humanities'	Three sets of PowerPoint slides introducing research data management principles and supporting services offered by University of Oxford. Intended to be used in induction sessions. The three packs cover similar information, but are designed to fit different time allocations. November 2010. http://sudamih.oucs.ox.ac.uk/docs/Data%20Management%20Intro%20-%20single%20slide.pptx http://sudamih.oucs.ox.ac.uk/docs/Data%20Management%20Intro%20-%20seven%20slides.pptx http://sudamih.oucs.ox.ac.uk/docs/Data%20Management%20Intro%20for%20Postdocs%20-%20eleven%20slides.pptx
'Managing Research Information' section of Research Skills Toolkit website	A suite of approximately twenty webpages and PDF documents providing information management advice and guidance. Includes sections on organising material, managing references, file synchronisation and versioning, keeping data safe, and managing structured data. Developed throughout project, and made live in February 2011. Original version integrated with Research Skills Toolkit (Oxford access only): https://weblearn.ox.ac.uk/access/content/group/e34f4cf9-1ecb-4244-a62b-ba3e96472790/SkTK_WebPages/index.html Content available for reuse (available via Jorum): http://resources.jorum.ac.uk/xmlui/handle/123456789/14724
Course: 'Research Information Management: Organising Humanities Material'	Three-hour face-to-face course with course-book, slides, and hands-on exercises. Covers information organisation strategies and linking notes with sources. January 2010. Version with references to Oxford context: http://sudamih.oucs.ox.ac.uk/docs/Generic%20Courses/Organising%20Humanities%20Material%20course%20book.docx (course book); http://sudamih.oucs.ox.ac.uk/docs/Generic%20Courses/Organising%20Humanities%20Material%20slides.ppt (accompanying slides);

	http://sudamih.oucs.ox.ac.uk/docs/Generic%20Courses/Organising%20Humanities%20Material%20exercise%20files.zip (exercise files) Non-localized version available via Jorum: http://resources.jorum.ac.uk/xmlui/handle/123456789/14725
Course: 'Research Information Management : Tools for the Humanities'	Three-hour face-to-face course with course-book, slides, and hands-on exercises. Introduction to various software tools that can help with aspects of research information management. February 2011. Version with references to restricted Oxford services: http://sudamih.oucs.ox.ac.uk/docs/Generic%20Courses/Tools%20for%20the%20Humanities%20course%20book.docx (course book); http://sudamih.oucs.ox.ac.uk/docs/Generic%20Courses/Tools%20for%20the%20Humanities%20slides.pptx (accompanying slides); http://sudamih.oucs.ox.ac.uk/docs/Generic%20Courses/Tools%20for%20the%20Humanities%20exercise%20files.zip (exercise files) Non-localized version available via Jorum: http://resources.jorum.ac.uk/xmlui/handle/123456789/14726
Research Data Management Factsheet	Two-page information sheet providing research information management advice and links to support services. Designed to accompany the 'Managing the D.Phil.' course offered by the Humanities Division at the University of Oxford. December 2010. http://sudamih.oucs.ox.ac.uk/docs/Research%20Data%20Management%20Factsheet.pdf
Leaflet: 'Managing Your Research Data at the University of Oxford'	Introductory leaflet designed in conjunction with the Research Services at the University of Oxford. A joint output with the Eidcsr Project. October 2010. http://sudamih.oucs.ox.ac.uk/docs/Research%20Data%20Management%20Leaflet%20v1%2025.pdf
Research Data Management Web-portal	Website designed in conjunction with the Research Services at the University of Oxford. A joint output with the Eidcsr Project. October 2010. http://www.admin.ox.ac.uk/rdm/
Humanities RDM website	A humanities-specific version of the general Research Data Management Web-portal, produced by the Humanities Division. March 2010. [URL not yet available]
SOFTWARE:	
Pilot 'Database as a Service' (DaaS) software	The DaaS as it stands enables users to create new databases, manage access permissions, import and export databases in common formats, add and edit

	data, and search for data. Please see ‘ Appendix B – Database as a Service Software’ for more information. http://daas.oucs.ox.ac.uk:8080/sudamihMySQL/home.seam (test site)
WORKSHOPS:	
‘Data Management Training for the Humanities – A Half-Day Workshop’	National workshop considering the state of data management training for researchers in the humanities. Slide presentations from the speakers are available from the workshop website. July 2010. http://sudamih.oucs.ox.ac.uk/training_workshop.xml
‘Research Databases in the Humanities – Where Next?’	National workshop looking at the ways in which humanities researchers build, maintain, and preserve databases, along with the processes currently in place to support such activities. Slide presentations from the speakers, plus a summary report of the workshop are available from the workshop website. January 2011. http://sudamih.oucs.ox.ac.uk/databases_workshop.xml
PRESENTATIONS:	
Sudamih slides for JISC launch meeting – Paul Jeffreys	Introductory slides outlining the Sudamih Project. http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20in%20%20slides%20for%20JISC%20meeting%20Nov%202009.ppt
'The Sudamih Project – Developing Data Infrastructure for the Humanities : Attitudes and Requirements' – James A. J. Wilson	A presentation given to the eContent Conference, May 2010, looking in particular at data management behaviour of humanities researchers and their self-perception of training needs. http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20slides%20for%20eConent%20conference%2011-05-2010.pdf
'The Sudamih Project – Researcher Requirements' – James A. J Wilson	A more concise version of the presentation about attitudes and requirements. Presented to the JISC Managing Research Data Programme Workshop, May 2010. http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20slides%20for%20JISC%20MRD%20workshop%20May%202010%20FINAL.pdf
'Data Management in the Humanities – the Sudamih Project' – James A. J. Wilson	Longer presentation with more background information about need for data management and information about the DaaS. Delivered to the ‘InterFace’ conference (bringing together humanities scholars with technical experts) at the University of Warwick, July 2010. http://sudamih.oucs.ox.ac.uk/docs/Sudamih_Interface2010.pdf
'Sudamih Project Training Recommendations’ – James A. J.	Outlining data management training needs of humanities researchers and the approach Sudamih will

Wilson	be taking to practically address those needs. Presentation at the 'Data Management Training for the Humanities' Workshop, July 2010. http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Training%20Workshop%20July%202010%20FINAL.pdf
'An Introduction to Research Data Management in the Humanities' – Meriel Patrick	Trial of the introductory managing research data slide pack at the Humanities Postdoctoral Induction Day, Oxford, October 2010.
'Managing Humanities Research Data' – Meriel Patrick	Presentation given as part of the Humanities Division's 'Introduction to the DPhil' workshop. October 2010.
'Research Data Management' – Meriel Patrick	Introduction to research data management for researchers attending the OUCS 'Research Skills Toolkit' face-to-face session for new researchers and postdocs. The topic 'Research Data Management' was chosen by researchers from a list of options that they could hear more about. December 2010.
'An Institutional Approach to Developing Research Data Management Infrastructure' – James A. J. Wilson	Summary of approach being taken to developing research data infrastructure at Oxford, with reference to both Eidcsr and Sudamih projects. Looking in particular at roles within the University. Delivered to the 6th International Digital Curation Conference in Chicago, December 2010. http://sudamih.oucs.ox.ac.uk/docs/OxfordInfrastructure_IDCC2010.pdf
English Studies Data Management Seminar – James A. J. Wilson	Talk given to graduate researchers in English advising on data management issues. January 2011.
'Database as a Service : a tool for researchers' – James A. J. Wilson	Introducing the DaaS to humanities researchers at the 'Research Databases in the Humanities – Where Next?' workshop. January 2011. http://sudamih.oucs.ox.ac.uk/docs/Databases_in_humanities_DaaS.pdf
Demonstration of the DaaS – James A. J. Wilson / Asif Akram	Demonstration of DaaS pilot software at the 2011 JISC Conference. March 2011.
'The Sudamih Project : Findings and Conclusions' – James A. J. Wilson	Short presentation outlining the challenges addressed by Sudamih, what we've done to address them, and what we've discovered in the process. March 2011. http://sudamih.oucs.ox.ac.uk/docs/JISCiMRD_March2011.pdf
INTANGIBLE OUTCOMES:	
Improved understanding of data management roles and responsibilities	By speaking to representatives from the Humanities Division, Bodleian Libraries, Research Services, and ITLP, as well as a large number of researchers, we have improved our knowledge of what different parts of the University regard their responsibilities as including and what they regard as outside of their remit or range of expertise.
Improved coordination of training	The various parts of the University that provide training

between institutional training providers	function largely independently from one another. Data management is, however, an issue that cuts across many different departments. During Sudamih we brought a number of different training providers together to discuss the issues involved and how each could play a part.
Improved contact and links between Computing Services and humanities researchers working on data projects	We have made a number of contacts which we hope to draw on again as the University's programme to improve its data management infrastructure progresses.

3.3 How did you go about achieving your outputs / outcomes?

The initial objectives of the Sudamih Project, as defined in the project proposal, were to:

- Develop institutional services for data management, curation, and long-term preservation for selected humanities research activities, but with a view to ensuring the deliverables can then be expanded to other activities within the humanities and beyond;
- Address database support needs within the humanities, including support for specific data types but also sustainability and service costing models;
- Understand the training and support requirements for humanities researchers and developing a data management skills course and other support activities, in collaboration with the DCC but tailored to the requirements of the humanities;
- Investigate the roles and responsibilities of service providers in Oxford for supporting humanities researchers in the management and curation of research data; developing a deeper understanding of research workflows and how they may interface with institutional services.

These broad objectives did not substantially change during the project, and all have been addressed in the work undertaken.

The principal institutional services developed during the course of the project were the data management training and the pilot 'Database as a Service' software. Sudamih focused specifically upon the needs of researchers in the humanities, and this is reflected in the outputs. The training materials have been written to reflect the kinds of working practices that we identified as common amongst humanities researchers, and the case studies used as examples reflect the kinds of tasks and situations that are likely to be familiar across the disciplinary boundaries within the Humanities Division at Oxford. Implementation of DaaS functionality also reflects the concerns and preferences we heard during interviews with humanities scholars during the requirements-gathering phase of the project. We anticipate that the key DaaS priorities of humanities researchers are unlikely to differ too much from the priorities of researchers in the other academic divisions, although more care will need to be taken when extending the coverage of the training materials (particularly the courses), as demonstrating an awareness of real research practices and the problems faced by researchers helped to build confidence and trust.

The Sudamih Project funding proposal envisaged three distinct phases of work: the analysis or requirements-gathering phase; the pre-implementation phase; and finally the implementation phase. Whilst it is difficult to clearly define, even retrospectively, exactly where one phase ended and another began, and the priorities established in the pre-implementation phase were wont to change as they came up against the practicalities of piloting real services in the unforgiving timetable of the Oxford academic year, this three-

stage framework is still useful for defining the narrative of events. Perhaps a brief preliminary 'initialization' stage and a final 'carry-forward' stage should, however, be added: the former during which the project tools were set up; and the latter in which we needed to focus more urgently on how the outputs we were implementing could and should be continued in a sustainable manner.

3.3.1 Project Initialization Phase

The project initialization involved producing a detail project plan and setting up dissemination channels such as the project website and blog, along with a public-facing 'events diary' which we used to indicate relevant conferences and workshops. We also needed to recruit an analyst and a software developer, which delayed the commencement of the requirements-gathering stage proper until towards the end of January 2010. The initialization phase also involved the establishment of the Sudamih Project Steering Group, consisting of senior stakeholders who could direct the project and ensure that its outcomes met the needs of the identified stakeholder groups.[7] The Steering Group met on three occasions during the project, whilst an Oxford-based Project Working Group met monthly to discuss progress. Both groups included active researchers in the humanities who could ensure that the project was not losing sight of its ultimate objectives.

3.3.2 Requirements Gathering Phase

The principal aim of the requirements-gathering phase was to understand the existing research data management landscape in the humanities disciplines, in order that we could develop relevant tools and training which would actually improve practices and meet the needs of researchers. Although the 'Scoping Digital Repository Services for Research Data Management' project from 2008 had identified the need to provide training and improved database tools, it had not gone in to the level of detail needed to determine precise requirements and priorities.

Between February and May 2010, the Sudamih Project conducted a total of thirty-two semi-structured interviews, twenty-nine with researchers and an additional three with support staff at the University. We selected our interviewees initially on the basis of recommendations by the Sudamih Project's Principal Investigators, and then to a lesser extent on following 'friend of a friend' recommendations. This ensured that we got to speak to a number of senior researchers with quite different approaches to data. We identified the majority of our interviewees, however, via the World Wide Web, looking for people with diverse research interests and experiences across the spectrum of the Oxford Humanities Division. We did *not* attempt to interview an entirely representative sample of the Humanities Division. Instead, we contacted a disproportionate number of researchers whose work relied heavily on databases and highly structured information, in order to ensure that we could get 'expert' opinions as to how we should approach the development of the DaaS, and also to see if their attitudes to data management training differed from researchers less heavily involved in data collection and analysis.

The first part of each interview focused on the nature of the research in which the interviewee was currently engaged. During this phase we sought to learn about the working practices of the researcher and in particular their procedures relating to data. We then moved on to ask interviewees whether they could see possible applications for the DaaS in their research, and, if so, what features such a service would need to offer to make it attractive. The final part of each interview consisted of questions concerning data management training: what the interviewee thought was actually meant by this; whether they were aware of training in this area or had even undertaken such training themselves; whether they felt there was any real need for data management training for humanities

[7] See Appendix A.

researchers; and, if so, what such training should focus on and how it should be imparted for maximum effect.

Each interview was recorded as an mp3 file and then summarized by the project analyst to highlight the interviewees' most relevant or interesting responses. We did not produce full transcriptions of the interviews.

Where researchers had been involved in creating data assets, we used the Digital Curation Centre's 'Data Audit Framework' (DAF) methodology to follow up the interviews and obtain additional information. We asked seven of the researchers whom we had interviewed and who had developed a database resource to provide further details about the development and maintenance of those resources. The questionnaire answers suggest that researchers realize that their data often has long-term value, but also that they are inclined to undervalue the time investment that they had put into creating the data and the potential reusability of the data by others.[8] We also helped assess the DCC's AIDA benchmarking tools,[9] although we found that the structure of the University of Oxford, plus the time and resources that would be required to test AIDA properly, meant that we could not undertake a full trial.

3.3.3 Pre-implementation Phase

The pre-implementation phase of the Sudamih Project consisted of the analysis of the researcher interviews and, stemming from that, the prioritization of various aspects of training and DaaS functionality. The resulting Researcher Requirements Report was approved by the Sudamih Project Working Group and the Steering Group and published online in July 2010.[10] We also staged the first of the two Sudamih workshops in July. Entitled 'Data Management Training for the Humanities', this attracted participants from across the UK and was in part intended to help the Sudamih Project ensure that we were aware of training programmes under development elsewhere with which we could partner or share content ahead of our own implementation phase. Although the workshop did not reveal the existence of much additional pre-existing training that we were not already aware of, it was a useful opportunity to compare our findings and plans with those of the Incremental Project, and also to learn more about Vitae, especially their increasing focus on research data management. As a result of the workshop, Vitae's representative Ross English was invited to join the Project Steering Group.[11]

The DCC was the only group represented at the workshop which had already developed significant data management training content. We had invited the DCC to run a customized version of their 'Digital Curation 101 Lite: How to Manage Research Data' course at Oxford in June 2010. This was adapted to be particularly relevant to researchers, and proved both popular and well-received. Some attendees indicated that they felt the course could be improved by including more worked examples or case studies and more emphasis on the practicalities of managing and curating data, advice that we took into account when developing our own course materials.

The training plan developed during the pre-implementation stage of Sudamih (and subsequently revised at intervals thereafter) identified the following five grouped aspects of training as being the most important with regards to improving research data management practices in the humanities:

[8] See 'Use of the Data Audit Framework within the Sudamih Project' report. http://sudamih.oucs.ox.ac.uk/docs/Use%20of%20the%20DAF.pdf
[9] AIDA Project, http://aida.jiscinvolve.org/wp/.
[10] Sudamih Researcher Requirements Report v.1.0, http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Researcher%20Requirements%20Report.pdf
[11] The workshop is summarized in a blog post available at http://blogs.oucs.ox.ac.uk/sudamih/2010/08/02/data-management-training-in-the-humanities-lessons-from-the-sudamih-workshop/.

1. *Introduction to data management.* What is data management? What tools and services does Oxford provide to help do it?
2. *Tools to help manage research data*. Considering real-life research challenges and problems faced – which tools/methods are best for solving them? Spreadsheets; databases; XML; bibliographic software; other 'research' software; how to structure and query data.
3. *Organizing and linking research information for later retrieval*. Including information on organizing paper-based as well as all forms of electronic sources (notes, journal articles, books, references, images, numerical data, multimedia). Includes: versioning; file & folder structures; classification; linking sources and themes; making things searchable.
4. *Technical aspects of funding bids*. How to plan and write about the technical aspects of research bids; examples of successful bids; technical advisory; IPR; long-term data curation.
5. *Database design for humanities research data*. Associating people, places, things, and events; distinguishing entities with shared names; GIS data; uncertain dates; incomplete data; non-Roman alphabets; recording sources.

The activities planned to address these training needs focussed mostly on the first three aspects, partly due to the findings of the analysis, partly due to resource constraints, and partly due to the fact that aspects four and five were already addressed to some degree via existing infrastructure within the University. The *Infodev* team within Oxford University Computing Services already offers advice to researchers planning humanities computing projects and has considerable experience in this field. It was therefore felt that promoting this existing service would be more cost-effective than creating new materials which would furthermore lack the relevance to specific cases that the dedicated consultancy service can offer.

Feature priorities for the DaaS implementation were also discussed, and ranked according to whether they were core, medium, low, or simply not worth taking forward given the trade-off between the time they would take to implement against demand (again bearing in mind the relatively limited resources at the project's disposal). The DaaS priorities were significantly refined and re-ordered as the project progressed.

The final aspect of the pre-implementation phase worth mentioning is the drawing together of a DaaS user-testing team, consisting of four of the humanities researchers we had interviewed who were either working on, or planning, some sort of database, and a fifth (technical) member who had previously worked on creating database software for the Ashmolean Museum.

3.3.4 Implementation and 'Carry-Forward' Phase

The implementation phase itself commenced between August and October 2010, so as to start getting training materials out in time for the induction sessions that are usually run at the beginning of the new academic year (which at Oxford officially began on the 10th October in 2010). The training events at which Sudamih presented are listed amongst the project outputs above (3.2).

The second of the Sudamih workshops was staged on the 21st January 2011. Entitled 'Research Databases in the Humanities – Where Next?', the workshop brought together almost sixty delegates from around the UK (and one from the Netherlands) to exchange ideas and good practice relating to the development and support of humanities research databases. It also provided Sudamih with an opportunity to introduce the DaaS to a broad

sample of its intended user community and get a better sense of whether they were really likely to use it, and for what.[12]

The final two months of the Sudamih Project were primarily taken up with presenting and assessing project outputs, and considering the steps required to take them forward. The project developed two three-hour face-to-face courses: 'Research Information Management : Organising Humanities Material'; and 'Research Information Management : Tools for the Humanities'. The series course title 'Research Information Management' was chosen so that courses relating to disciplines other than the humanities could potentially be added in future. The 'Tools for the Humanities' course addresses aspect two of the training plan (above), whereas the 'Organising Humanities Material' session addresses aspect three. Both of the courses were debuted in the 'Isis' course room at Oxford University Computing Services. Besides the face-to-face courses, the data management Web content for the Research Skills Toolkit was made live in February 2011, and feedback sought.

Meanwhile, the DaaS was, after some delays, moved to a live server where it could be publicly accessed whilst feature implementation work continued unabated. The delays here were due to unforeseen issues relating to the provision of long-term support to the technical platform (JBoss) that had been chosen to host the DaaS. The benefits of both strands of work were considered alongside the costs of their continuing development and support.[13] Finally, this report was completed and submitted for approval by the JISC.

The initial Project Plan changed relatively little over the course of the project, although the 'evaluation' work package was jettisoned due to time pressure and uncertainty regarding how the project stood to benefit as a result of it. The aim of the evaluation work package had been to produce an independent, formative evaluation of the project approximately half-way through and make recommendations which could be implemented to improve matters. This had worked well during the Eidcsr project, giving a renewed sense of focus to what needed to be covered during the latter stages. In Sudamih, however, the Project Working Group and Steering Group both took a very active guiding role, meeting regularly to help evaluate progress, reconsider priorities, and deal with potential issues as they arose. As a result of this regular intervention there were no obvious issues to commission an external evaluator to evaluate. It was agreed at the second Steering Group meeting that the evaluation work package should be dropped provided that suitable feedback mechanisms were embedded in the day-to-day work.

Dissemination activities were undertaken during the whole course of the project to ensure that stakeholders were aware of developments and could bring to our attention any issues that they felt we should be aware of. A list of dissemination activities is provided in section 3.2.

3.4 What did you learn?

Besides the general experience acquired during the course of running a project such as Sudamih, there were two periods in particular dedicated to analysing and reflecting upon the information generated: the first after the researcher requirements gathering phase, and the second towards the end of the project when we needed to consider the value of what had been created. Sections 3.4.1, 3.4.2, and 3.4.3 consider the lessons learnt from the first phase (and are predominantly based on the Researchers Requirements Report),[14] whereas sections 3.4.4 and 3.4.5 consider the latter. Some of the general lessons learnt from running

[12] A summary report of the second Sudamih workshop is available here: http://blogs.oucs.ox.ac.uk/sudamih/2010/08/02/data-management-training-in-the-humanities-lessons-from-the-sudamih-workshop/.
[13] See sections 3.4.4 and 3.4.5.
[14] Sudamih Researcher Requirements Report v.1.0, http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Researcher%20Requirements%20Report.pdf

the project are distilled in section 3.4.6 for the edification of anyone running similar projects in future.

3.4.1 Research data characteristics and practices in the humanities

One of the initial aims of the Sudamih Project was to better understand the processes by which researchers in the humanities disciplines create, store, and recall the data they use in their research. Whilst the Project Manager and Analyst both had research backgrounds in the humanities themselves, the differences between the various disciplines and between individual researchers meant that it could (and indeed would) have proved presumptuous to carry out no wider research. Given the lack of any prior training in data management issues and the fact that data management is apparently not a popular topic of discussion at social gatherings, most humanities researchers seem to have evolved their own organizational practices quite independently of one another, resulting in a very broad range of practices and little comprehension of what works best. The lack of training, combined with the traditionally independent ('lone scholar') nature of much humanities research probably also contributes to a frequent lack of awareness of many centrally-provided services.

Key findings from the requirements-gathering interviews include:

- Humanities research is hugely diverse, and makes use of an enormous range of types of sources.
- The intellectual value of humanities datasets tends not to depreciate over time – a database of Roman cities is potentially of as much use to researchers in fifty years time as it is today, provided it is not rendered obsolescent through technological change.
- Humanities scholarship often aggregates to a 'life's work' body of research, with any given researcher often wishing to go back to old notes, sources, or datasets in order to find new information.
- There appears to be a growing trend towards structuring data within the humanities. This is partly because technological advances make more sophisticated data projects possible, and perhaps partly because of the changing priorities of funding bodies.
- Methods of organizing data also vary considerably, as does the extent to which researchers succeed in creating and maintaining a well-functioning system.
- Good information management is time consuming, and academics often find themselves with insufficient time to keep on top of it.
- Most researchers are willing in principle to share their data with others, but in practice do not regularly do so, for a variety of reasons. In the humanities, issues surrounding the incompleteness of the original data, or the layer of interpretation often required to render it consistent, can lead to reluctance to share, as researchers worry that their 'processed' data may be misinterpreted by others.
- Data storage is generally on personally-owned machines and backing-up is generally also to personal devices on an *ad hoc* basis. Knowledge of centrally-provided services is limited and they are seldom used.
- There is a significant amount of confusion over the ownership of research data. This is exacerbated by complex situations in which multiple people or organizations may have different claims on the same resource.

3.4.2 Training Requirements

Most of the researchers we spoke to had not previously encountered the concept of 'research data management training' before agreeing to be interviewed by the project, so we had to offer an explanation of what we considered such training might include. Throughout the project we adopted a broad definition of what constituted 'data', regarding it as encompassing not just structured information on computers, but the whole range of materials

that researchers must assemble and analyse in order to produce their research outputs. This could include, therefore, printed books, articles, references, photographs, handwritten notes, and electronic files in any format, in addition to datasets as they might traditionally be understood. Despite this breadth, more than 75% of the researchers we interviewed reported that they had never received any data management training, and most of the rest had not really received much more than an introduction to bibliographic software. The interviewees were, however, keen on the idea of data management training. Asked if there was a need for this sort of training, most thought that there was:

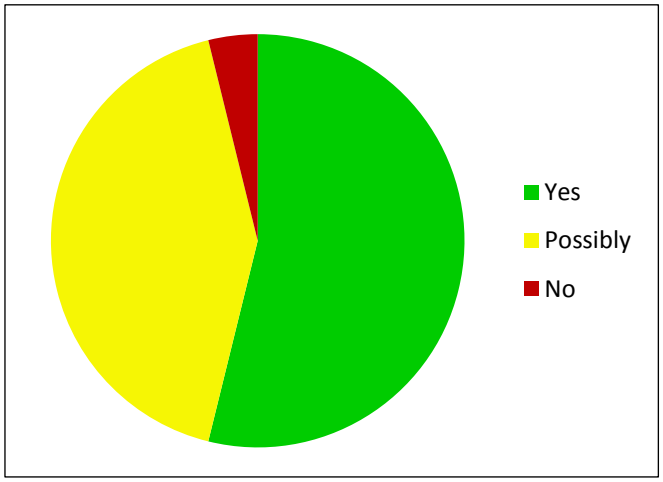


Figure 1. Humanities researchers' responses when asked if there was a need for data management training

When it came to asking the researchers which aspects of data management training they thought were most important to learn, their responses were varied:

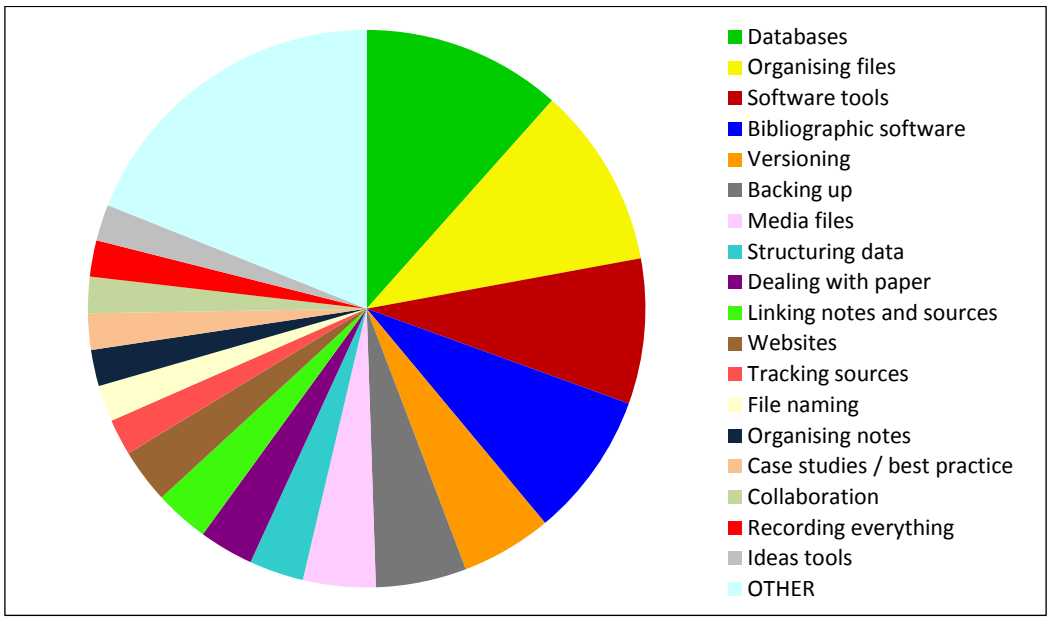


Figure 2. What researchers thought it would be useful for data management training to cover (either for themselves or new graduate research students)[15]

The training materials created by the Sudamih Project cover the most requested aspects of data management training.

[15] As responses we free and largely un-guided by the interviewers it is not always easy to group them retrospectively. Although this is what we have tried to do for clarity in this diagram, this categorization necessarily hides nuances between responses.

Noteworthy conclusions arising from what researchers said in the interviews included:

- Training needs to be based upon actual research problems commonly faced, not promoted as generic skills training.
- Training should on the whole be provided via face-to-face courses, with supplementary online content.
- Graduate students should receive data management training early in their research careers, but ideally not before they have already had the opportunity to assess and start to gather the kinds of sources that their research will demand.
- Aspects of data management training should be integrated into existing training where possible.
- If possible, information management issues should be included in compulsory training.
- If possible, courses should be customizable to allow for the needs of particular faculties. It should be ensured that default examples are broadly applicable across disciplines.
- Training should be offered in both 'broad' data (or information) management skills and also in 'narrow' (technical) data management skills:
  - 'Broad' data management would include: organizing your files in such a way that you can retrieve information quickly and easily; backing up; versioning; managing email; linking notes to content; keeping track of your sources.
  - 'Narrow' data management would include: which type of software is most suited to particular requirements; structuring data in relational databases; querying and retrieving information; long-term curation – data formats, obsolescence and migration issues; using the DaaS.
- An advisory service should be on hand to offer one-to-one technical advice about data management (particularly in the narrow sense) to Principal Investigators wishing to apply for project funding. [As mentioned earlier, such a service does in fact already exist at Oxford, although awareness of it was limited.]

3.4.3 DaaS Requirements

In general, the interviewees who showed most interest in the DaaS were those who were about to embark on a database project. With one or two exceptions, researchers involved with existing mature online databases demonstrated little inclination to move from their present systems. There was, however, considerable interest in the DaaS as a potential means of ensuring the continued availability of Web resources which currently lacked stable, long-term hosting arrangements, and as a tool which would enable datasets which are not presently publicly accessible to be made available online.

The response to the proposed DaaS was generally positive, and while such a service will naturally not be of use to all researchers a substantial proportion could see some potential application for their own work. The following conclusions were drawn from the interviews:

- The DaaS should be trialled both with researchers with existing databases, to ensure it offers the functionality they expect, and also with researchers still in the planning stages, to ensure that it is appropriately intuitive and can meet their expectations.
- It should be simple for researchers to import existing databases into the DaaS, otherwise they are unlikely to move to the new system despite the other advantages it might offer.
- If the DaaS is going to be broadly useful to humanities researchers, it needs to meet the following key user requirements:
  - Intuitive and easy-to-use interface
  - Flexible searching and querying
  - Able to deal with a range of data types

  - Records may be linked to external sources
  - Support for diacritics and text in non-Roman alphabets
  - Can be edited by multiple people
  - Must be easy to edit online via a Web interface
  - Query outputs presented in Web browser
  - Data can be downloaded and worked on with desktop applications
  - Data integrity can be preserved
  - System is stable and secure
  - There is good user support and training
- The University must clarify the ownership rights of researchers to any data they place in the DaaS. Researchers would be unlikely to use the services if they felt their ownership of that data would be compromised.

3.4.4 Costs and benefits of training developed

The project identified eight benefits arising from the continued maintenance and dissemination of the training and learning materials developed by the Sudamih Project:

1. Time saved by researchers by locating and retrieving relevant research notes and information more rapidly
2. Improved quality of research by locating better, more relevant research information than would otherwise be the case
3. Improved quality of research by linking materials in such a way as to highlight connections and trigger new ideas
4. Improved comprehensibility of research information and data after long time periods, facilitating reuse
5. Better awareness and use of software tools to assist research management
6. Better awareness and uptake of central infrastructure services intended to help researchers, including technical help and assistance with funding bids
7. Reduced risk of data loss
8. Improved version control

Whilst these benefits are not straightforward to quantify, the project has surveyed responses to the new training materials where possible to assess the level of impact they would need to justify the costs of continued provision.

3.4.4.1 Data Management Training Demand and Impact

The first point to make is that the early indications of high demand for data management training seemed to be borne out by the speed with which places on the two face-to-face courses filled. These three-hour courses were both staged in the ‘Isis’ training room at the Computing Services, which has a capacity of 25 people at computer terminals. Both courses were fully booked and had waiting lists of an additional 10 people within two days of being announced via the standard ITLP course-list and email advertising forthcoming courses.

The second thing to note regarding the face-to-face courses is that they were effective, at least insofar as effectiveness can be measured by altering behaviour. Asked ‘have you/will you change any aspects of your own information management as a result of the course?’, 23% of respondents said that they had made or would make ‘significant’ changes, 69% said that they had or would change at least one or two aspects of their current practice, and 8% that they were considering changes. None of the respondents reported that they were not even considering changes after attending the courses.

When asked about particular changes that course respondents were going to make as a result of the courses, there were some interesting replies. A number of attendees mentioned specific software packages there were using or going to use, and several mentioned that they would try tagging files and notes with key words rather than simply storing everything in

the default hierarchical manner that Windows and Mac operating systems encourage. Other comments included: 'I will look at IT tools from a different perspective (how they will move my project forward, rather than how I fit my project to the IT tools)'; 'I am going to make a database for each new project as an aid to writing from now on. I see it as making a sort of visual "first draft"'; and 'I'm getting more and more organized and happy now! It changed not only my ways of working, but also my mood and attitude'.

86% of those who attended the second face-to-face course (Research Information Management: Tools for the Humanities) and completed the feedback form specifically mentioned software tools which they were going to try as a result, which certainly suggests that benefit 5 of the above list (better awareness and use of software tools to assist research management) is met by the training materials.

In an attempt to measure benefit 1 (time saved by researchers by locating and retrieving relevant research notes and information more rapidly) we asked course attendees to estimate how much of their time spent writing up their research outputs is actually spent looking for notes/files/data that they know they already have and wish to refer to. The average was 18%, although in some instances it was substantially more, especially amongst those who had already spent many years engaged in research (and presumably therefore had more material to sift through). This would indicate that there is at least considerable scope to save time (and improve research efficiency) by offering training that over the long term could improve information management practices.

We undertook a similar approach to address benefits 6, 7 and 8 – the related benefits aimed at reducing the risk of data loss and improving version control. 61% of those course attendees who completed the feedback survey indicated that they had at least on occasions lost information or data that they had wished to refer back to. Most respondents did not attempt to quantify precisely how many days' work they had lost, and none had suffered any catastrophic loss, but two respondents estimated that they had lost at least a month's work during their five or six years of research. By reminding researchers about, or introducing them to, centrally provided back-up and security services and tools and methods for file synchronization and version control, it can reasonably be expected that these figures can be reduced.

3.4.4.2 Online Provision of Data Management Learning Materials

The Sudamih Project researcher requirements gathering phase identified different learning preferences amongst humanities researchers, and it was felt to be important that training content was available both online, to suit researchers who like to access information as and when the opportunity (or requirement) arises, as well as via face-to-face training courses, which better suit researchers who like to have the opportunity to ask questions and interact whilst learning. We therefore developed both the two face-to-face courses mentioned above (alongside presentational materials that could be used by others) and a complementary suite of online material.

The information management training content developed by the Sudamih Project to form part of the University of Oxford's online 'Research Skills Toolkit' went live on 7th February, 2011. It is hosted on the section of the website named 'Managing Information' although this also contains some content developed elsewhere and some of the materials developed by Sudamih are additionally linked to from other sections of the website. The 'Managing Information' part of the Toolkit was publicized via an email sent on 18th February to approximately 740 individuals who had attended or 'nearly attended' Toolkit events over about 3 years. It was also promoted via a blog post on the ITLP blog and an RSS newsfeed on the ITLP home page. Given the short time that the content has been available and the relatively small publicity push in a university the size of Oxford, the low numbers of page hits so far is understandable, although we anticipate that usage will grow over the coming months. The Research Skills Toolkit is currently only available to members of the University

of Oxford, although the training materials developed during the Sudamih Project have now additionally been made publicly available for reuse via Jorum.[16] As of the 21st March, the different parts of the Research Skills Toolkit site had attracted the following number of visits:

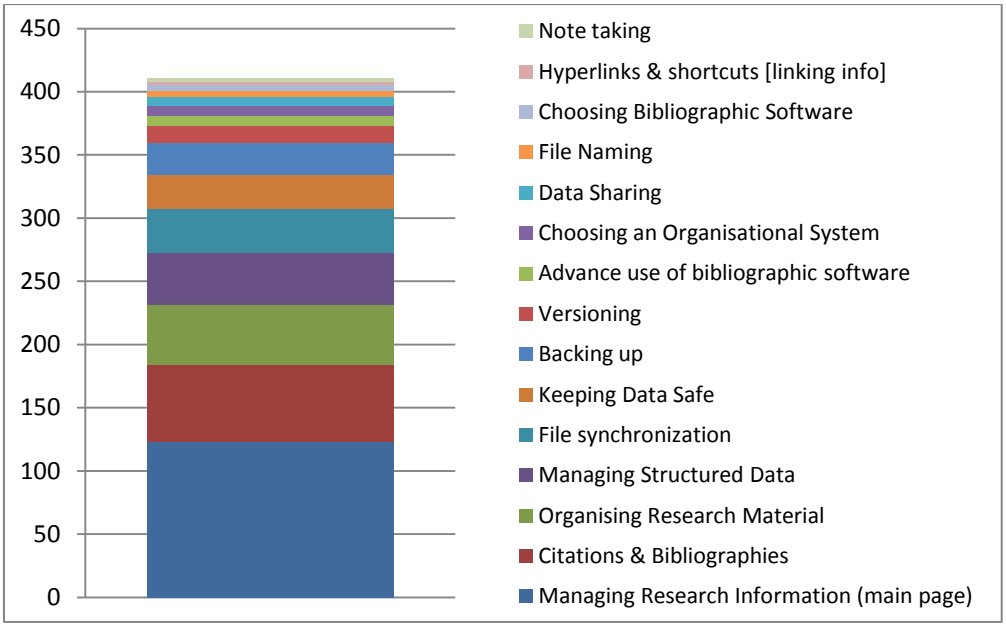


Figure 3. Visits to webpages and PDF articles relating to data management in the Research Skills Toolkit

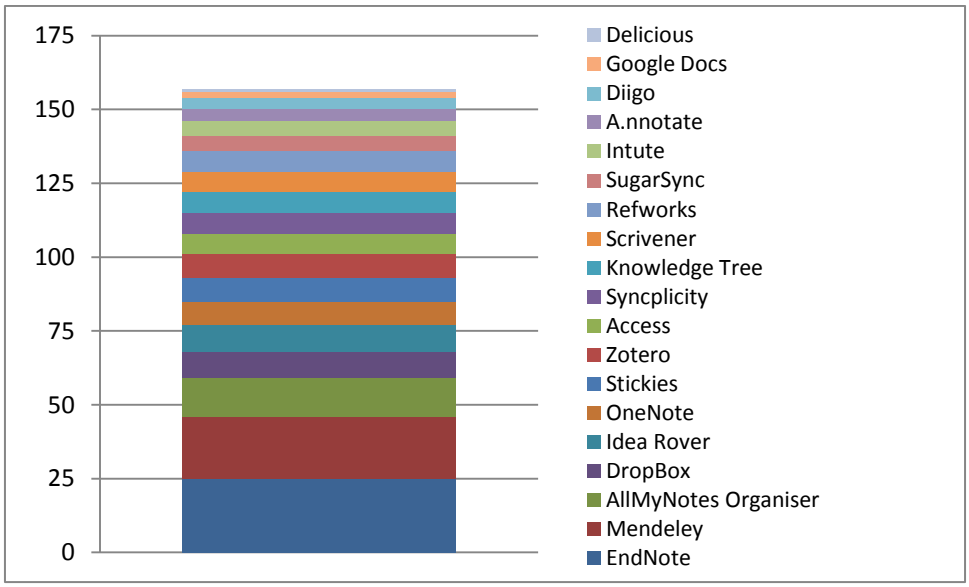


Figure 4. Visits to webpages relating to particular data management tools and software included in the Research Skills Toolkit

Although Web statistics do not make it possible to establish the precise impact of a given webpage on its readers' research practices, when coupled with the feedback received from attendees to the face-to-face courses one can begin to see that if the online content is anything like as effective at changing research information-handling behaviour as the

[16] Research Skills Toolkit content available via Jorum, http://resources.jorum.ac.uk/xmlui/handle/123456789/14724.

physical courses are proving to be, the potential impact (due to the wider accessibility) is significant.

3.4.4.3 Data Management Training Business Case

It was always the intention of the Sudamih Project that the training and learning materials developed, if well received and regarded as beneficial, would be sustained beyond the project itself. As with any other such initiative, however, taking the training from project to resourced service required the development of a business case. Recognizing that this would be a common feature of the Research Data Management Infrastructure projects funded by the MRD Programme, JISC employed Charles Beagrie Ltd, as part of a broader support and synthesis role, to create a comprehensive business case template, which the Sudamih Project used to frame the case for maintaining and continuing to disseminate the Sudamih training outputs.[17]

The Sudamih Training Business Case proposal seeks a commitment from the University of Oxford that the University will implement and sustain training in research information management for researchers in the humanities. The Business Case argues that the small (two-side) Research Data Management Factsheet should be the responsibility of the Humanities Division, as it was created to accompany one of the Division's taught courses and is unlikely to need to be updated regularly. The bulk of the training materials, however, should be the responsibility of the IT Learning Programme (ITLP) at the Computing Services, as they are best equipped with the technological expertise to update the training (it was felt that the technologies referred to in the training were likely to change far more rapidly than the disciplinary practices). A further reason for the ITLP to assume responsibility was that the Sudamih material was initially trialled using their formats and processes, so it they will not need to make additional investments in adapting the materials.

It is estimated that in FEC terms, approximately £40k has been invested in creating the data management training materials developed by the Sudamih Project. This includes the costs of researching the material and trialling it. The annual costs to the ITLP of maintaining the major learning and training outputs of Sudamih are as follows:

Training Material	Activity	Annual cost (at 2011 prices)
Two 'Research Information Management' face-to-face courses	Commissioning external expert to update and deliver both courses each term (3 terms per year at Oxford)	**£2,671 (£890 per term)**
Data management methods and tools resources in the online Research Skills Toolkit	Annual website maintenance by intern, plus one week of external expertise to update content	**£1,566**
Three data management slide-packs intended for researcher induction sessions	Internal staff time updating content and promoting to divisions/faculties	**£397**
TOTAL		**£4,634**

A full breakdown of these costs can be found in the Sudamih Training Business Case.[18]

It is extremely difficult to accurately estimate the costs savings likely to be derived from individual training courses, but given that we have a good idea of the costs involved in running the courses we can at least get a sense of the impact that the materials would need to have in order to make a return on investment.

Consider only the first of the identified benefits accruing from the research information management training courses: time saved by researchers locating and retrieving relevant

[17] Sudamih Training Business Case. http://sudamih.oucs.ox.ac.uk/docs/SudamihTrainingBusinessCase_v1.1.pdf
[18] Sudamih Training Business Case, pp.13-15.

research notes and information more rapidly. Whilst individual estimates varied quite significantly, the researchers who attended the courses estimated that they spend on average about 18% of their writing-up time looking for notes, files, or data that they know they already have and wish to refer to. Those who had been undertaking research for longer tended to spend a greater percentage of their time seeking such materials, but for our purposes here the average figure of 18% will suffice. Now imagine one particular researcher, Dr. Example, who is on a salary of £40,200 and gets to spend half of his working year undertaking actual research (as opposed to teaching and administrative duties). 25% of Dr. Example's research time is spent writing up his research in the form of journal articles and other publications. Dr. Example attends both the 'Research Information Management : Organising Humanities Material' and 'Research Information Management : Tools for the Humanities' courses run by the ITLP. As a result of what he learns on these courses, the proportion of his writing-up time he spends unproductively hunting for material he knows he already has is reduced from 18% to 16% for the next ten years of his career he spends at Oxford. This saves, when the full economic costs of his work are taken into account, a total of £998. This alone covers the ongoing costs of both the training courses in that term.

Now consider that Dr. Example is only one of up to 25 researchers who may have attended both of the courses in a given term. Over 90% of the initial course attendees who completed the feedback forms indicated that they would change their information management process as a result of the courses. Furthermore, it is quite conceivable that attendees may be able to shave more than two percentage points off the proportion of time they spend hunting for the information they have gathered. The possible savings arising purely from this one area of potential benefit could be significant.

3.4.5 Costs and benefits of DaaS

The anticipated benefits identified as arising from the use of the DaaS are as follows:

1. Improved sharing and reuse of data
2. Improved data security (storage; access & identity management)
3. Less duplication of data
4. Faster preparation of structured data
5. Greater technical support efficiency
6. Economies of scale due to centralized hosting
7. Greater data consistency between projects facilitating the repurposing of data and mash-ups
8. Data becomes more reliably citable due to use of DOIs
9. Strengthened research grant applications
10. Greater awareness of existing data provides inspiration for new research

As with the training outputs of the Sudamih Project, it is not possible to quantify many of the benefits arising from the uptake of the DaaS within the timescale of the project itself. Given that the DaaS is only just taking shape in its pilot form as the project draws to its conclusion, we cannot yet fairly compare its performance against more established database management software used by researchers. We have, however, tried to measure future demand, and we have also asked humanities researchers who work with structured data to give us their assessment of the intended service.

3.4.5.1 DaaS Demand and Impact

Our initial findings suggest that demand for the DaaS amongst humanities researchers conducting data-based research is very high. The second Sudamih workshop, 'Databases in the Humanities : Where Next?', attracted almost sixty delegates, most of whom were humanities researchers either already working with structured data or planning to do so. The

workshop gave us the opportunity to introduce our plans for the DaaS and explain the rationale behind its development.

After the workshop we sent feedback forms to the attendees with some 'bonus questions' relating specifically to the DaaS. We asked the delegates the question, 'would you consider using a system such as the DaaS (once it is fully operational) to develop/host a database?' Excluding those who indicated that the question was not relevant to them, 56% answered 'yes' whilst 44% gave a more cautious 'possibly'. None of the 27 respondents said that they would not consider using the DaaS.

We also asked the delegates how they envisaged actually using the DaaS. The responses were very varied, and in some instances suggested uses that we had not fully considered ourselves. Responses included:

- For migrating existing databases
- For a crowd-sourced database
- Perhaps for trialling datasets/schemas, or advising others to trial there, depending on what functionality is available.
- Shared data input
- To expand upon projects I have started/planned out using Access databases
- For relatively small-scale projects of my own or of others I work with or advise. For a large-scale, long-running project, it would probably require more personalization
- Using it to store and backup my database, accessing it remotely myself and allow others to do the same
- I would use it for my own project on Medieval Cambridge Families, which has been left dormant for several years due to lack of funding – I couldn't have paid technical people to develop a DB for my research

Ten of the respondents mentioned specific existing or planned databases which they would consider using the DaaS to support.

A number of the workshop delegates, both in the feedback and at the event itself, stressed that the DaaS, if it could be offered free at the point of use, would be a godsend for the sustainability of their data. It is a very common problem that there are no resources available to maintain digital outputs after the end of a project, yet Web-hosting costs money. Although obviously the DaaS itself has maintenance and hosting costs, there was a clear hope that the economies of scale it might afford, combined with a University-supported charging model, could open up the possibility of small datasets finding a long-term home without requiring annual fees to paid by the researchers themselves.

Some responses were from the perspective of those involved in institutional support services, who also saw a role for the DaaS:

- InfoDev often does bespoke database-backed websites for clients where these are university research projects, units/departments, or sometimes external clients. If there was a clear and consistent costing model that we could just pass on to them, this would be useful and make their hosting needs much simpler
- I would like to give it to the researcher community (SURFnet) and the research fellows of the NIAS (www.nias.knaw.nl)
- I would like to see the library (for which I work) and computing services departments in my institution working together to provide a system in which academics can deposit, preserve and make available their data.
- Would require user consultation at Bristol to determine whether a local or externally hosted service would be most appropriate, possibly trialling both, but the project certainly addresses a gap in the current curation and sharing of research data.

Whilst the response to the DaaS was overwhelmingly positive, we also tried to capture the concerns of potential users, to better understand the issues that might put people off using the new system. Again, there was quite a variety of responses, several focusing on possible technical limitations, data security, and usability:

- The feature that you will probably not have which I would *very* much like is for it also to be able to handle large deeply and arbitrarily nested mixed content
- XML files as input, and for these to be addressable by XPath (or even better XQuery)
- [Our] current database is both very complex and finely tuned to its multiple tasks; especially reports; unclear whether this could be transferred without major amount of work.
- Limited functionality in terms of types of analysis available
- If the data would be unsecured
- It might take some convincing to get colleagues to invest the time and effort necessary for a local installation of the DaaS, so it needs to be something that works right out of the box for us to be able to demonstrate its potential value.

Other concerns related to administration, economics, and sustainability:

- Both of these datasets are not from funded projects, so cost would be an issue
- Lack of a very simple straightforward costing model for internal members of the university might put me off recommending its use
- Being from another university, I might be concerned about support and sustainability
- Sustainability and cost of any remotely hosted service versus a local installation versus "free" services like Google's
- The cost, and I would prefer to keep the database within my institution if possible

3.4.5.2 DaaS Business Case

Besides the feedback from the research data management workshop, the Sudamih Project has also received useful and valuable advice from the Oxford Roman Economy Project (OXREP) which can be used here as a case study.[19] The OXREP is an ongoing project at the University of Oxford that is seeking to build a comprehensive (and quite complex) database of sites of economic activity in the Roman World. It is being assembled from a number of existing databases, with data being cleaned, standardized, and added to the new database. The project involves original research and it is intended that the OXREP database will continue to grow and be added to in the future.

Dr. Miko Flohr, the Assistant Director of the OXREP, estimated that if the database work they had conducted during 2010 had used a complete version of the DaaS rather than the Access database which they were in practice working with, they would have saved approximately 21% on staff time, primarily due to simplifications to database structuring and through controlling and standardizing data contributions. He indicated that for other projects the DaaS would be likely to save money by reducing expertise requirements.

The savings made by moving the OXREP to a centrally-hosted Virtual Machine we estimated to be even greater. The IT Officer of the Classics Faculty, which currently hosts the OXREP database and Web front-end on a departmental VM, is planning to move it to a centrally-hosted and maintained VM. The cost savings of so doing, when the staff time needed to look after the VM is taken into account, amount to approximately 37%. Given the economies of scale offered by a centrally-supported VI, this saving may be expected to increase further as the infrastructure is enlarged.

Although obviously every humanities database project will be different, the OXREP hopefully illustrates the kind of savings which could be achieved by projects by switching to a

[19] http://oxrep.classics.ox.ac.uk/nw/index.php

centrally-hosted DaaS service in the future. Furthermore, this short case study only attempts to quantify benefits 4 (faster preparation of structured data) and 6 (economies of scale due to centralized hosting) from the benefits list at the start of this section. The other benefits are harder to assess at this stage.

At present, the DaaS is a pilot service. Although the functionality is in place for users to import and export databases, structure, edit, and search databases, the system is not yet robust enough to be used for more than test purposes, nor is it intuitive or user-friendly enough for users to be able to carry out regular activities without step-by-step instructions. We will, therefore, need to develop the DaaS further to bring it to the standard expected of a supported, documented, production-ready service for research.

As well as improving the usability of the DaaS, we shall be conducting a more thorough return on investment analysis and creating a full business plan during the follow-on Virtual Infrastructure with Database as a Service (VIDaaS) Project.

3.4.6 Lessons learnt in running the project

- Language

Researchers do not understand the terminology used by data librarians. Care must be taken to avoid technical jargon and use unambiguous but straightforward terminology when talking about data management.

Humanities researchers often do not think of themselves as using or generating 'data', unless they are actively involved in constructing relational databases. One must either make it clear that when one refers to 'data' this encompasses all of the materials used in the process of reaching conclusions about research questions, potentially including notes, books, journal articles, and non-electronic material, or refer to 'research information' instead. 'Data' is off-putting and should be avoided when referring to training, unless that training focuses specifically on the use of databases or other very obviously consistently structured information.

- Quantification of research costs can be difficult

Our attempts to apply the DAF methodology to data assets created by researchers was complicated by the fact that the researchers themselves were not cost-aware and had no real sense of the value of their work in financial terms. As reported in the Sudamih DAF evaluation, 'in addition to the difficulties in estimating the financial value of researcher time, none of the respondents in this sample made any reference to costs to the institution (for server space, library access, and so forth), despite costs borne by the university or department being specifically mentioned in the question. This suggests that asking researchers alone may not always give a complete picture of the costs of creating data resources.'[20] When we came to assessing the costs and benefits of the services we were creating we found it helpful to ask researchers to estimate the time taken spent performing quite specific and clearly-defined tasks and basing our costings on this. It was also important to ask precisely who else had a role in the research tasks undertaken, as researchers tended not to think about the non-academic support they received.

- Measuring the benefits of data management can be difficult

Whilst the potential benefits of implementing research data management tools or training might be easy to enumerate, they can be hard to quantify. Illustrative case studies or the measurement of the minimum impact required to bring a return on investment may have to suffice as adequate alternatives to a thoroughly-costed business plan in such circumstances.

[20] Patrick, M., 'Use of the Data Audit Framework within the Sudamih Project', 2010, pp.10-11. http://sudamih.oucs.ox.ac.uk/docs/Use%20of%20the%20DAF.pdf

- You may need to compromise when embedding data management training into existing training

If you intend to integrate data management training into existing training programmes (which is advisable in order to reach researchers who might otherwise not attend such training) be prepared to compromise regarding how that training is delivered. Our research suggested, for instance, that handing out leaflets to new researchers during induction sessions is not the most effective means of communication, as such hand-outs tend to get lost or ignored amongst all the other materials that new researchers are bombarded with during their first couple of weeks, but if that's all that the trainer is prepared to countenance, it's better than nothing.

- When developing software, ensure that the technologies you are using can be supported by the institution

Sudamih encountered delays to the implementation and user testing of the DaaS due to the fact that the University's systems support team were unfamiliar with the technical platform the project had developed the software on. Whereas the software developer had used JBoss (which has good libraries for supporting databases), existing IT services were mostly based upon an Apache Tomcat platform. This mismatch caused problems when transferring the service to live servers. Clearer communication between teams earlier in the project would have flagged up this potential problem and may have made the process quicker.

- Manage expectations

The DaaS has generated an unexpected degree of excitement amongst some researchers, and we have been concerned that they are failing to appreciate that all the promised features cannot be implemented at once. It takes time to develop any new software to an acceptable service level, and this needs to be clearly communicated.

3.5 Immediate Impact

The Sudamih Project Plan identified seven important stakeholders: the University of Oxford (and in particular the Humanities Division and the service units with responsibility for supporting research); the JISC; the UK Research Data Service Scoping Project; the Digital Curation Centre (DCC); other research data management projects; HE institutions with research interests in the humanities; and the Research Information Network (RIN).

Arguably the most obvious and important impact of the project within Oxford has been on the researchers who have attended the research information training courses, read the materials on the Research Skills Toolkit, or been present at the introductory talks we have given at various induction days. We know from the feedback to the courses that over 90% of attendees either have made changes to the way they work or will try different approaches in future, and it is reasonable to hope that these changes will improve the efficiency of their research and result in better data management.

Events such as the workshops and even the initial round of interviews with researchers have helped raise the profile of data management as an important aspect of research which should be treated as such. Simply talking about the subject and exchanging experiences regarding methods and tools which work and problems overcome has had the visible effect of getting researchers to think about how their information and data management affects the process of producing research outputs.

As far as JISC are concerned, the findings of the Sudamih Project should help shape the future of the Managing Research Data Programme, and related programmes. At the start of the project there was surprisingly little training material to take and adapt, whereas now there is a suite of assessed material that can be reused and improved by others. Combined with the training outputs of other MRD projects such as Incremental, JISC should be in a

good position to synthesize findings and make recommendations to others. Sudamih has also participated in the benefits analysis and business case activities as requested by JISC. Again, this participation should provide JISC with a range of comparable case studies, the findings of which can be synthesized and distilled into good practice for future projects.

The University of Oxford has for the last few years had a formal liaison role with the UKRDS scoping project to plan Pathfinder services and submit a business plan and proposal to HEFCE. The DaaS is one such Pathfinder service which the UKRDS project has been interested in. Sudamih has helped gauge support for the service and develop the core functionality. Building on this scoping work, as intended, we have now put forward a proposal for HEFCE resources (via a JISC/UMF call) to develop the DaaS into a national service.

We have involved the DCC with the Sudamih Project from the beginning, contributing to DCC events such as the International Data Curation Conference, and helping them assess their tools and training. We provided feedback on the DCC 101 Lite course they kindly agreed to run at Oxford, and also on the DAF and AIDA methodologies. We know from an AIDA presentation at a JISC workshop in November 2010 that our feedback has been taken on board and helped inform development.

During the lifespan of Sudamih we have maintained contact with the other JISC MRD projects, not just at JISC workshops but also via project-organized workshops and emails. The project has exchanged findings and ideas in particular with Admiral and Incremental.

Both of the workshops staged by the Sudamih Project have attracted wide participation from universities besides Oxford, and this has enabled us to impart our findings at a national level, as well as using the experiences and concerns of those at other universities to inform our own decisions. The interest generated by the DaaS in particular has been encouraging, although its full impact on research practices will not be apparent until further development work has been completed.

Whilst the project did not in the end work closely with the RIN, we did engage with Vitae – the UK's national organization supporting the personal, professional, and career development of doctoral researchers and research staff. Sudamih only became aware of Vitae after the project had commenced, but invited a representative from Vitae to speak at the 'Data Management Training for the Humanities Workshop' and subsequently asked him to join the Project Steering Group. Whereas data/information management from the researchers' perspective had not been a prominent feature of the Vitae strategy before 2010, in their new Researcher Development Framework (published September 2010)[21] it assumes a more prominent role. By engaging closely with Vitae, Sudamih has contributed to this recognition of the importance of research information management.

Finally, Sudamih has had a major impact in extending and improving data management expertise at the University of Oxford, and helping define the roles and responsibilities of the various academic support services. The University has a long-term commitment to implementing a data management infrastructure that addresses all parts of the research data life-cycle, and Sudamih has added components to this whilst improving our sense of how various aspects of infrastructure need to interface with one another.

3.6 Future Impact

Certain aspects of the Sudamih Project are unlikely to have a significant immediate effect, but will have long-term impact. We know that the researcher training implemented by the project has already had an effect on user behaviour, for instance, but the real benefits derived from the training are likely to accrue over the longer term, as researchers save time

[21] Researcher Development Framework, http://www.vitae.ac.uk/CMS/files/upload/Vitae-RDF-Sept-2010.doc.291181.download.

which would have been spent hunting for sources, or see new connections which spark new ideas as a result of better linking and organizing their data.

The future impact of the DaaS will not really be felt until the current pilot service can be developed to a point at which it can be used as the primary host for research data, functioning at a production level with guarantees of stability and security. We already know that the Humanities Division are intending to recommend the DaaS to future research projects applying for funding and that they feel that indicating the use of the DaaS will strengthen a bid by addressing the need for long-term hosting with economies of scale over departmental or *ad hoc* hosting. The Infodev research support team at Oxford have also stated that they will consider recommending the DaaS to researchers at the University. At present the Infodev team often construct bespoke database-backed websites for clients, but have indicated that if there were a clear and consistent costing model that could be passed on this would make hosting much simpler. As already covered, the interest in the DaaS amongst individual humanities researchers is considerable.[22]

4 Conclusions

4.1 General Conclusions

The Sudamih Project was successful in meeting its initial aims and objectives, and responses to the significant outputs have been reassuringly positive. It transpired that there was a real gap in training provision when it came to research data management. Whilst the DCC already provided good training for those directly involved with data curation, this was not really targeted at 'normal' researchers whose primary concern is producing well-received research outputs and who have no formal responsibility to look after data. Sudamih recognized at its inception, partly based upon the findings of projects such as Paradigm[23] and FutureArch[24] that had looked at the management and curation of private archives, that early intervention in the research process is required to assist researchers manage their own data with a view to future curation. The need for such training was borne out by our study of current practices and recognized within the research community itself (at least in the humanities). Our findings also confirmed the 'life's work' nature of much humanities research, and the corresponding need for data to be managed and kept accessible over very long time periods.

The project has also confirmed the need for better database support within the humanities. In part, this relates to the training needs. It appears that there is growing interest in data-focused research in the humanities, but many who have ideas for specific projects are unsure about what technologies are available to help them or which would be most appropriate to their needs. In some cases this lack of knowledge is putting researchers off undertaking such projects. Where researchers had opted to work with databases they often had little idea of best practice, and idiosyncratic solutions were found to common problems. This tendency was exacerbated by a lack of awareness of centrally-provided services which could help. The proposed DaaS was widely seen as offering advantages over traditional database management systems which were not well suited to collaborative projects, easy dissemination of data, and, most importantly, came with various sustainability issues to which there were no easy solutions.

4.2 Conclusions relevant to the wider community

It is highly likely that humanities researchers at institutions other than Oxford face many of the same issues. The library and information management communities are becoming

[22] See p. 22.
[23] Paradigm, http://www.paradigm.ac.uk/.
[24] FutureArch, http://www.bodleian.ox.ac.uk/beam/projects/futurearch.

increasingly adept at dealing with the challenges of digital data preservation and curation, and several HE institutions are starting to get to grips with data repositories and suchlike. The same is also true for particular academic disciplines, such as astronomy and crystallography, where there are clear and obvious benefits of sharing data outputs. In many fields, however, the researchers who originate research data have little sense of how best they can manage it, and receive little support to do so. Given that data curation is often described in terms of a life-cycle, with conceptualization and creation as its birth,[25] it is perhaps surprising that so little effort has been made to help the researchers involved in these earliest stages of data management. After all, good early management of research data leads not only to more efficient use of that data, but also helps the custodians of that data in the later stages of the life-cycle.

4.3 Conclusions relevant to JISC

The experience of the Sudamih Project suggests that JISC is right to target resources at the 'upstream' stages of the data management lifecycle, as this is likely to ensure that a much larger volume of important research data makes it down to the point at which institutional or other long-term curation services can bring their expertise to bear and facilitate the preservation and wider reuse of that data. Although the audience for research data management training is large, dispersed, and generally disinclined to seek assistance without prompting, there do appear to be a number of training requirements common to researchers across disciplines. The general principles of file organization and awareness of the kinds of technologies that can help data management fall into this category. With good coordination across UK HE a relatively small body of generalized but high-quality training materials could potentially improve practices across a broad cross-section of researchers. Without also addressing specific disciplinary needs, however, issues of data quality are likely to remain. Therefore some degree of subject-based customization and institutional localization is required to maximize training benefits. When it comes to understanding how best to structure data in databases a more customized, even bespoke, approach to training is useful; when addressing issues such as backing-up and the availability of support services, it obviously helps to focus on specific local provision.

The JISC may wish to consider investing in a central repository of reusable generalized training materials with sections for discipline-specific content, whilst issuing advice to institutions regarding the kinds of training that should be provided locally. Jorum is of course the most obvious place to store general training materials, although it may need to adopt a slightly different structure than at present to enable the straightforward discovery and adaption of such content. The current HE subject headings do not include an obvious home for data management training resources, and while the area may be deemed too narrow to merit its own subject heading, a general 'Study and research skills' category would be a useful addition, for both data management resources and many others. It would also perhaps be helpful to include categories for resources which are relevant to a broad disciplinary area (e.g. humanities, social sciences, or sciences): some headings of this kind already exist for FE, but not for HE.

With regard to the development of tools to assist researchers to manage their data better, there is evident demand for tools that address real research problems faced by researchers for which generic or commercially-provided software is not well suited. There is little point simply recreating existing software, but academic research, at least in the humanities, is a niche market that is seldom well-addressed by commercial developers but which offers scope for greater efficiency, hence the demand for the DaaS. Whilst the development of such tools for a single institution may not be cost effective, most tools are likely to be of interest to disciplinary communities beyond the single institution.

[25] DCC Data Curation Lifecycle Model, http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf.

The long life-span of humanities data poses particular problems. Unlike much scientific data, the value of humanities data tends not to depreciate over time, whilst the integrity of academic research depends upon the continued availability of sources remaining as those sources were when cited. As such, some way of guaranteeing the long-term preservation, access, and integrity of research data (and at low cost) is important. Researchers need to learn to conceptualize data as a long-term resource that will be used by others, in a similar way to how they presently conceptualize publications as long-term resources which will be read and interpreted many times by researchers working on potentially quite different research questions. The HE sector, likewise, needs to be able to offer and support technology which will encourage and make it possible for researchers to treat data in this manner.

5 Recommendations

General recommendations

1. It is helpful for an institution to have a single location or point of contact which researchers can be referred to when dealing with data management issues, even if this location actually just redirects people to more specialized sources of advice or expertise. The Research Data Management website (http://www.admin.ox.ac.uk/rdm/) now serves this function at Oxford.
2. Different academic departments and institutional service providers should work together to understand who should be responsible for implementing, and sustaining, various aspects of data management training.
3. 'Data management' or even 'information management' are broad terms that can mean very different things to different people, even within a single discipline. When offering data management training it needs to be immediately obvious what the training covers and who it is intended to benefit.
4. Encouraging good data management in the humanities will, over the longer term, require a change of mindset amongst researchers. At present, the value of data assets, both in terms of costs of creation and potential for reuse, is not fully appreciated. Researchers need to learn to see data outputs more as publications, and institutions need to provide infrastructure that will assist this change in mentality.
5. Institutions that are serious about winning research funding should in the future have a specialist technical advisory service which researchers can consult for assistance with the technical aspect of bids.
6. Universities should clearly disseminate information about central services that support data management to ensure researchers are aware of them.
7. Researchers need help to discover the most appropriate software tools for specific research challenges.
8. Researchers should be trained in organizational principles and strategies to enable them to better manage their information and sources.
9. Universities should clarify the intellectual property rights that researchers have with regard to their structured data outputs, and in particular their rights when depositing data in a repository or service such as the DaaS. If researchers do not feel that they will retain 'ownership' of their data, even when moving between institutions, this is likely to reduce their willingness to use the central services provided and, as a consequence, this will reduce many of the benefits that such central services are intended to bring about.
10. Researchers value the security of their data very highly and any tools or repository intended to help manage research data will require high levels of security.

Recommendations for JISC

1. A repository of data management training and learning materials should be established, where training providers at an institutional, divisional, and departmental

level can easily discover appropriate materials. Jorum may provide the basis of such a repository. The existence and significance of this repository will need to be actively communicated to training providers within institutions.

2. Given the common need for institutions to provide technical advisory services to assist researchers with funding bids, JISC should consider the supporting role that the DCC can play. Besides providing technical bid templates, some sort of registry of tools that can help address particular challenges at particular stages of the research data life-cycle might be useful, as many researchers lack awareness of even basic data management tools and methods at present. Such a registry may need to indicate which tools are likely to be of most interest to which broad academic areas in order to be accepted and used by researchers. The DCC is likely to be better placed to develop and maintain such a registry than individual universities.
3. There may be a role for JISC in clarifying legal matters relating to data IPR. Researchers need to be reassured that they will not 'lose control' of their data in the process of making it reusable.
4. JISC should consider investing in research into cost models for supporting very-long-term data sustainability, such as is required in particular by researchers in the humanities.
5. Due to the long-term nature of the impact of the Sudamih outputs, the benefits deriving from them cannot be easily assessed. JISC should consider investing in a long-term impact evaluation framework that can be coordinated at a programme rather than project level. This would overcome the problem faced when project teams are dispersed or move on to new projects and are therefore unable to revisit previous projects and seek evidence of their impact several years later.
6. JISC should consider undertaking a longitudinal study of data management awareness amongst researchers to help measure progress at a national level.

6 Implications for the future

- Given the existence of an established community of IT experts employed to support research in HE, plus the apparent widespread need for tools such as the DaaS, there is a realistic chance that an active open source development community can be built around the DaaS, drawing upon experience at various UK Universities. To facilitate this, the software source code and documentation will need to be placed in an appropriate open-source software repository with standard supporting tools (tools for handling bugs, feature requests, community discussion and coordination, etc.).
- As our survey of envisaged uses of the DaaS suggested, project outputs are not necessarily employed in the way that their creators initially imagined, and the real benefits to users may be different from those assumed. Enabling communication between the future user community of the DaaS and the development community may see the software taken in new directions and adapted to different purposes, whether this be for institutional support services, 'big data' applications, or something else entirely.
- Further requirements gathering needs to be done to establish precisely which aspects of data management can be addressed via general cross-disciplinary training and learning materials, and which require subject-specific or localized training resources.
- There is a need in the humanities (and possibly other disciplines) for very-long-term data sustainability solutions and cost models designed to deal with effectively permanent storage and access. A service such as DaaS may provide a good tool for serving both 'active' high-access or in-development databases as well as dormant or low-use databases, possibly using different service-level agreements and a cost model that can cross-subsidize one from the other to produce a workable long-term solution.

7 Appendix A – Composition of Steering Group and Project Working Group

The Sudamih Project was led by Oxford University Computing Services (OUCS) and reported internally to the University's Research Committee Information Management Sub-committee (RIMSC). This Sub-committee ensures coordination, communication and collaboration between activities addressing information from research. In addition to this, the project was ably supported by a Steering Group comprised of representatives of the key stakeholder groups, both from within Oxford and from the national community. The Steering Group met on three occasions, to provide strategic direction to the project and ensure that activities were appropriately prioritized. We are grateful to Professor David Shepherd, Dean of Humanities and Social Sciences at Keele University and formerly the Director of Sheffield's Humanities Research Institute, for chairing the Steering Group.

Steering Group Membership	
Prof. David Shepherd	Chair
Prof. Paul W. Jeffreys	Principal Investigator
Dr. Michael Fraser	Co-Investigator
Dr. Ian Archer	Co-Investigator
Prof. Andrew Wilson	Co-Investigator
Dr. Andrew Fairweather-Tall	Assistant Registrar (Research)
Dr. Simon Hodson	JISC Programme Manager
Prof. David Robey	Oxford e-Research Centre
Sally Rumsey	Bodleian Libraries
Kathryn Dally	Research Services
Joy Davidson	DCC Representative
Dr. Ross English	Vitae Representative
Dr. James A. J. Wilson (in attendance)	Project Manager

The Project Working Group, comprising members of the core project team, humanities academics, Research Services Office and Bodleian Libraries representatives, met each month to monitor progress and ensure that project outputs were delivered according to plan.

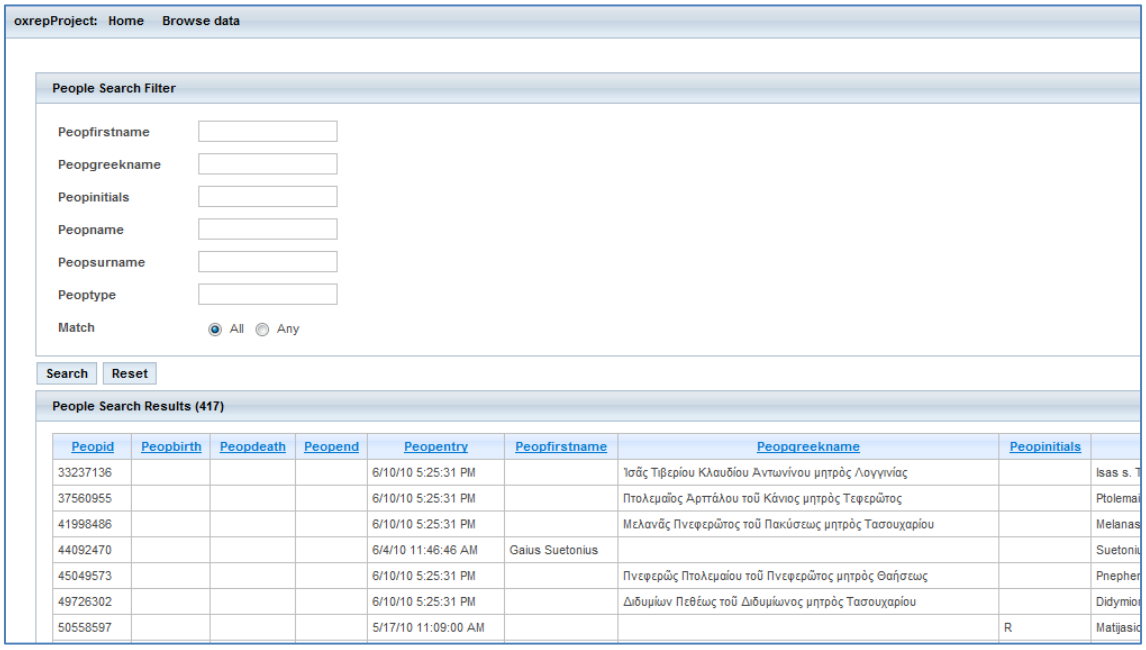
Project Working Group Membership	
Prof. Paul W. Jeffreys	Principal Investigator
Dr. Michael Fraser	Co-Investigator
Dr. Ian Archer	Co-Investigator
Prof. Andrew Wilson	Co-Investigator
Dr. James A. J. Wilson	Project Manager
Dr. Andrew Fairweather-Tall	Assistant Registrar (Research)
Dr. Miko Flohr	Faculty of Classics
Erin Cooper (to July 2010)	Bodleian Libraries
Sally Rumsey (from September 2010)	Bodleian Libraries
Kathryn Dally	Research Services
John Ireland	Computing Services
Asif Akram	Systems Developer
Dr. Meriel Patrick	Analyst

8 Appendix B – Database as a Service Software

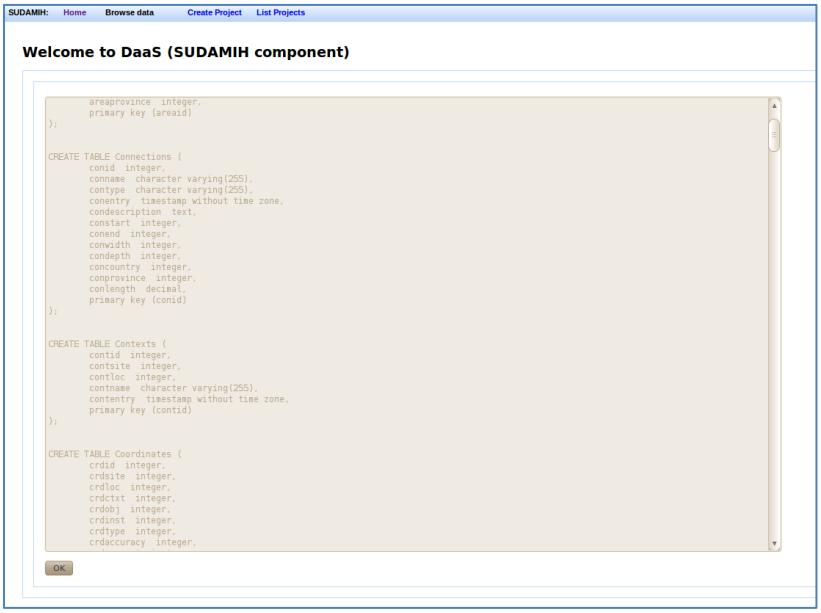
The pilot DaaS has been developed as a number of semi-independent components that can pass information between each other as required. The components are:

1. The 'core' DaaS database management system, which handles database administration and provides an interface through which users can browse and edit relational databases.
2. A conversion utility that can convert existing databases in Access format or saved as comma-separated values into PostgreSQL
3. A graphical SQL-designer utility that enables the user to create or modify database structures via a simple Web interface. All possible types of relationship between tables are available. Databases structured using this tool are not limited to use in the DaaS but can be used by other relational database management systems as well.
4. A graphical form-builder utility that enables the user to drag and drop buttons, text fields, multiple choice menus, and other standard form components via a straightforward Web interface. This potentially has uses beyond the DaaS.
5. An advanced SQL query-builder to enable users to construct sophisticated search queries without needing to be an SQL expert. This component is currently incomplete. but will be completed during the early stages of the VIDaaS follow-up project.

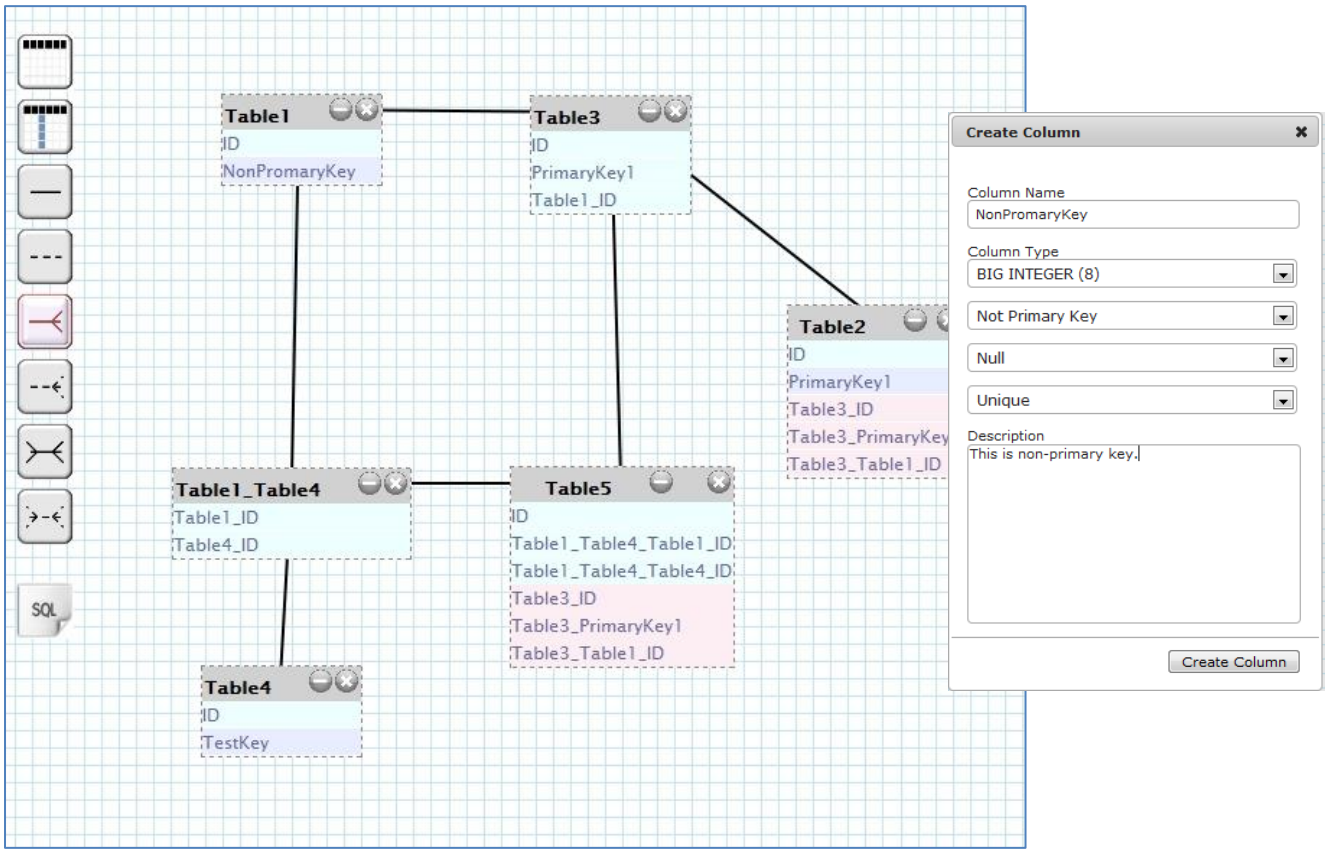
These components are illustrated in the screenshots below:



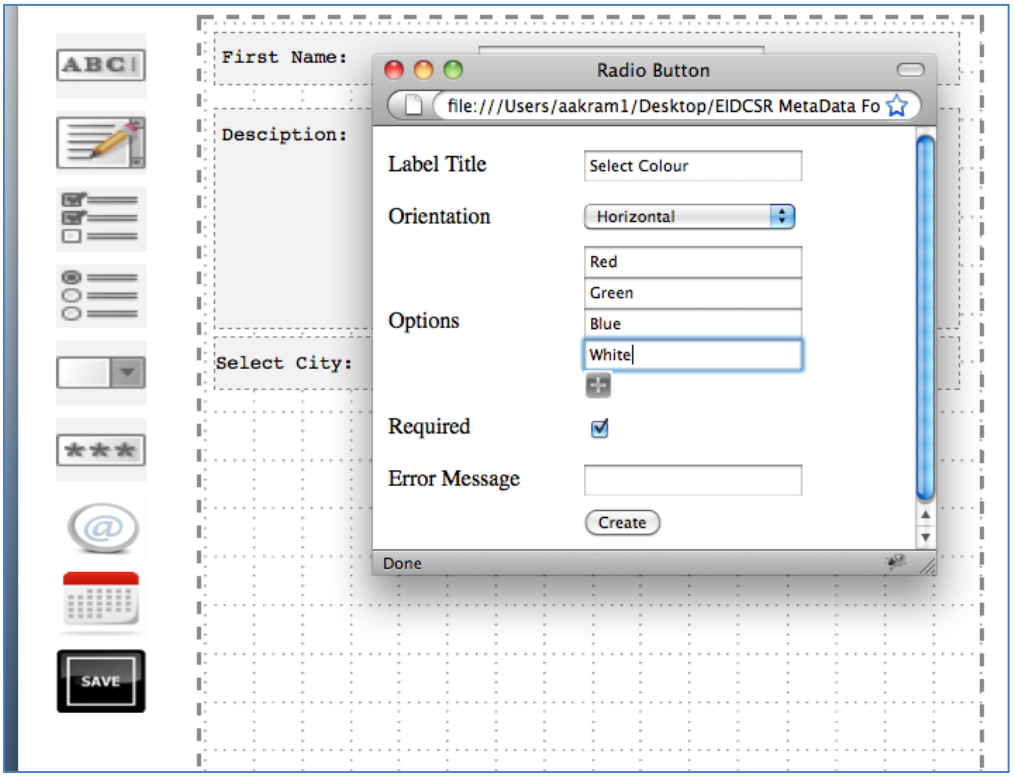
Component 1 – DaaS Database browse / simple search interface



Component 2 – Access Database converter (test screenshot – user does not see code generated)



Component 3 - Graphical SQL-designer for structuring relational databases



Component 4 – Form-builder utility

Work on taking the DaaS from a pilot to a production-standard service will be undertaken during the JISC-funded Virtual Infrastructure with Database as a Service (VIDaaS) Project.[26]

[26] VIDaaS Project, http://vidaas.oucs.ox.ac.uk/.