



Sudamih Training Business Case

Supporting Data Management Infrastructure in the Humanities (Sudamih)

sudamih.oucs.ox.ac.uk

Author

Dr. James A. J. Wilson

Affiliation

Oxford University Computing Services

JISC

Table of Contents

Executive summary.....	2
Service Proposal.....	4
Background.....	5
Stakeholders.....	6
Strategic Alignment.....	6
Options Appraisal.....	7
Demand and Benefits Appraisal.....	9
Constraints and Dependencies.....	11
Risk Assessment.....	12
Cost Analysis.....	13
Timescales.....	17
Conclusions.....	17
Appendix A : Evidence of Demand and Effectiveness.....	19

EXECUTIVE SUMMARY

This document makes the case that the researcher training materials developed during the course of the JISC-funded ‘Supporting Data Management Infrastructure for the Humanities’ Project (Sudamih) should continue to be maintained and disseminated within the University of Oxford. The Sudamih Project ended on 31 March 2011, having produced and piloted four significant sets of training/learning outputs:

1. Two courses (‘Research Information Management : Organising Humanities Material’, and ‘Research Information Management : Tools for the Humanities’) with accompanying course-books, slides, and step-by-step practical exercises, intended for face-to-face delivery.
2. A suite of interlinked content written for the IT Learning Programme’s ‘Research Skills Toolkit’ website at the University of Oxford, consisting of data management advice and guides to relevant services and software.
3. Three slide packs for use in researcher induction sessions, entitled ‘An Introduction to Research Data Management in the Humanities’. The slide packs introduce new researchers to the data management tools and services available at the University. The three packs cover the same essential information but in different levels of detail, to fit different time allowances.
4. A data management factsheet designed to accompany the Humanities Division’s ‘Managing the D.Phil.’ course.

The training outputs were well received by researchers and project stakeholders. In FEC terms approximately £40,000 has been invested in developing and trialling the materials. They have been developed specifically for use in existing contexts within the University’s training infrastructure, although a ‘non-Oxford’ version of each of the outputs has also been created for use by the broader HE community (as mandated by the JISC).

With the expiration of project funding, the Sudamih Project proposes that responsibility for outputs 1, 2, and 3 be passed to the IT Learning Programme at Oxford University Computing Services, whilst the responsibility for output 4 should be passed to the Humanities Division.

Eight benefits have been identified as arising from the continued maintenance and dissemination of the materials developed by Sudamih:

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, thus enabling better 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 it is not easy to quantify all of these benefits financially, we have estimated where possible the impact that the training would need to have in order to cover the costs of continuing to provide the training. Results suggest that the outputs thus assessed would quickly and easily repay the costs incurred. A costs and benefits analysis is provided from pages 13 to 17. The proposal is aligned with the strategy of the University of Oxford as laid out in the Strategic Plan (see p. 6).

The total cost to the IT Learning Programme at Oxford University Computing Services of maintaining and delivering the training and learning materials described comes to £4,634 per annum. Given that this figure is relatively low, it is proposed that the Computing Services absorb the cost within its overall budget over the coming years.

SERVICE PROPOSAL

This proposal seeks commitment from the University of Oxford to implement and sustain training in research information management for researchers in the humanities.

During the Supporting Data Management Infrastructure for the Humanities (Sudamih) Project, a number of training materials were developed to help improve research data management practices in the humanities. These materials consist of:

1. Two courses ('Research Information Management : Organising Humanities Material', and 'Research Information Management : Tools for the Humanities') with accompanying course-books, slides, and step-by-step practical exercises, intended for face-to-face delivery.
2. A suite of interlinked content written for the IT Learning Programme's 'Research Skills Toolkit' website at the University of Oxford, consisting of data management advice and guides to relevant services and software.
3. Three slide packs for use in researcher induction sessions, entitled 'An Introduction to Research Data Management in the Humanities'. The slide packs introduce new researchers to the data management tools and services available at the University. The three packs cover the same essential information but in different levels of detail, to fit different time allowances.
4. A data management factsheet designed to accompany the Humanities Division's 'Managing the D.Phil.' course.

This proposal argues that these materials should be maintained, reused, and further developed to ensure that appropriate data and information management skills are embedded in future academic practice. Although all of the materials have been released for reuse under an open licence and may be freely reused and adapted by other individuals and institutions, this business case only considers the use and maintenance of the outputs within the University of Oxford. External reuse is considered to be out of scope of this proposal.

It is proposed that the IT Learning Programme (ITLP), based within Oxford University Computing Services (OUCS), is the section of the University best equipped to maintain (and potentially expand) these materials, although their effective dissemination to the Oxford research community may best be achieved by exploiting multiple channels besides the ITLP. The two face-to-face courses should be run on a termly basis unless demand drops to the extent that places on the courses are no longer being filled. The courses will need to be updated regularly and checked for currency each time they are run. The materials developed for online delivery will need to be fully checked and updated annually, as will the introductory slide packs (although these are not likely to require much time). The data management factsheet, unlike the other materials, should be the responsibility of the Humanities Division to maintain, given that it was written to accompany one of their regular courses.

BACKGROUND

Given the importance of managing one's research data properly, and the potential costs of failing to do so, it is surprising that it has taken the HE sector so long to implement structured training programmes. The work undertaken by the Sudamih Project turned out to be unexpectedly pioneering. Whereas at the start of the project we had envisaged creating a suite of training materials largely by recombining and customizing existing training materials from elsewhere, it became apparent during the course of the project that there was very little currently available in an appropriate form. We therefore needed to produce much of the content ourselves. As Sudamih was funded by the JISC it was required that the content created in the course of the project be open and freely available for the whole UK HE sector to adapt and reuse. The content was initially created specifically for Oxford, but alternative versions of all the training and learning materials have since been created with references to services restricted to students and staff at the University removed.

The need for improved research data management training has become a hot topic over the last few years due to the recognition of the possibilities of data-driven research,¹ the need to find ways of managing the 'data deluge',² the controversy over the climate change data held by the University of East Anglia,³ and a growing feeling that the data outcomes of publicly-funded research should be made available to the public that funded them. There has also been an increasing recognition that making data publicly available maximizes the value of that data, given that it reduces the chance of unwitting recreation and enables other researchers to reuse existing data in ways unimagined by its original creators.

On a more personal scale, better data management can simply help individual researchers improve the efficiency with which they use the information they have gathered. Less time spent searching for notes and sources means improved rates of publication. Better management of data and better connections between sources and notes can also mean better research, as the way researchers manage their information affects the way they recall and use it.

Data management training is just one aspect of the data management infrastructure that needs to be implemented in a university in order to maximize the value of that institution's research data, but it is an important aspect which underpins all other stages of the research data lifecycle.⁴ Furthermore, basic researcher training can be implemented before all the other aspects of data management infrastructure are fully established, enabling 'quick wins' in terms of better data management.

¹ See for example Hey, T., Tansley, S., Tolle, K. (eds.), 'The Fourth Paradigm : Data Intensive Scientific Discovery', 2009.

² See the JISC Briefing Paper, 'The Data Deluge: Preparing for the Explosion of Data', 2004: <http://www.jisc.ac.uk/media/documents/publications/datadelugebp.pdf>.

³ See for instance The Guardian's online section on: 'Hacked Climate Science Emails': <http://www.guardian.co.uk/environment/hacked-climate-science-emails>.

⁴ For a diagrammatic representation of the data lifecycle, see the Digital Curation Centre's model: <http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>.

The training materials developed by the Sudamih Project are aimed primarily at the personal information management aspect of data management infrastructure, although given that they already include substantial coverage of tools and techniques, it would be relatively straightforward to expand them in future to include instructions to make the most of the broader data management infrastructure being assembled at Oxford.

STAKEHOLDERS

The key stakeholders in the proposal are researchers in the humanities, although the institution itself also has a significant stake as it stands to benefit from more efficient research practices and improved publications deriving from that (potentially of data as well as articles).

Stakeholder	Nature of Impact
Early-stage research masters and doctoral students	<ul style="list-style-type: none"> • Awareness of University data management services • Awareness of appropriate tools for particular research cases • Awareness of different filing and retrieval methods
Senior Researchers & PIs	<p>As above, but also including:</p> <ul style="list-style-type: none"> • Better connections between research notes and sources improves use of evidence and quality of research • Less time spent searching for particular information results in faster writing-up of research papers • Better knowledge of technical approaches to research • Improved technical sections in research bids
University of Oxford	<ul style="list-style-type: none"> • Improved funding bids for external research money • More efficient research results in better outputs to enter for REF assessment • Data in format enabling easier University curation (this increasing data assets)

STRATEGIC ALIGNMENT

The 35-page University of Oxford Strategic Plan (2008/09 to 2012/23) refers to the importance of researcher training in several places, and emphasizes in particular the “increasing significance of doctoral training”.⁵ One of the principal learning and teaching strategies mentioned in the document (I(c)) is to “train postgraduate research students and postdoctoral researchers as academic apprentices”,⁶ and it is very much by understanding the information management practices of experienced researchers, and distilling best

⁵ University of Oxford Strategic Plan (2008/09 to 2012/23), §21.

⁶ Ibid. §32.

relative claims of the ITLP and the Humanities Division as potential information management training providers:

ITLP:

<p><u>Strengths</u></p> <ul style="list-style-type: none"> • Good reputation • Alert to changes in technology • In-house expertise in data • Training materials already familiar due to being trialled in conjunction with ITLP • New materials already 'integrated' (in terms of delivery format, content, and branding) 	<p><u>Weaknesses</u></p> <ul style="list-style-type: none"> • Inferior disciplinary expertise (although OUCS is a centre for humanities computing)
<p><u>Opportunities</u></p> <ul style="list-style-type: none"> • Could recombine material with existing course content / cross-fertilize • Could expand to other disciplines or generalize relatively easily 	<p><u>Threats</u></p> <ul style="list-style-type: none"> • In the long term, changing practices in the humanities could render the current training obsolete • Funding reductions

Humanities Division:

<p><u>Strengths</u></p> <ul style="list-style-type: none"> • Trusted by humanities researchers • Good reputation 	<p><u>Weaknesses</u></p> <ul style="list-style-type: none"> • Currently employ a different style of training to that envisaged by Sudamih • Lack of technical expertise • Smaller training infrastructure (in terms of permanent staff)
<p><u>Opportunities</u></p> <ul style="list-style-type: none"> • Could expand with more real case studies 	<p><u>Threats</u></p> <ul style="list-style-type: none"> • Internal uncertainty whether such training should be run centrally or delegated to individual faculties – could result in reluctance to take responsibility • End of Roberts money • Funding reductions more generally

Weighing up both options it seems reasonable that the responsibility for maintaining and delivering the information and data management training materials should rest primarily with the ITLP. As the materials are already tailored toward the humanities (and include appropriate case studies) and changes to humanities research practices are likely to happen at a significantly slower pace than new technological developments and changes to software, the ITLP is better equipped to keep the content up to date. Furthermore, the fact that the face-to-face courses (which are the most expensive materials to develop and

deploy) are already formatted and adapted for ITLP use means that there would be significant additional development costs incurred if responsibility for their upkeep and implementation were now transferred to the Humanities Division.

Both the ITLP and the Humanities Division run induction courses for new researchers, as do many of the individual academic departments. For the reasons given above, the ITLP are best placed to maintain the introductory slide packs, but the intended purpose of them is to be used by anyone giving a research induction session, even if they have no personal expertise in the matter. The ITLP should therefore endeavour to ensure that they remain publicly available and that awareness of their existence and function is widespread. This may involve some sort of partnership between ITLP and the Humanities Division, possibly involving the Research Services as well, given their involvement in creating a research data management website for the University as a whole and an accompanying leaflet (produced in conjunction with Sudamih Project staff).

The factsheet, given that it is specifically intended to complement a Humanities Division course, should become the responsibility of the Humanities Division. It is a very small resource in comparison with the other newly-developed training materials.

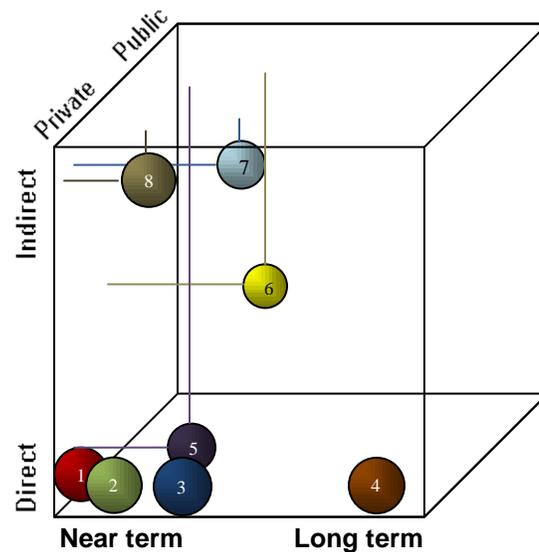
DEMAND AND BENEFITS APPRAISAL

The Options Appraisal section of this document (see above) concludes that the bulk of the training materials developed by the Sudamih Project should be maintained and used by the IT Learning Programme within the Computing Services, but does not consider whether there is a financial case for such maintenance and dissemination. The benefits arising from the maintenance and use of the materials need to be assessed to establish whether this can be justified.

A number of potential benefits arising from the training that Sudamih has been developing have been identified:

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, thus enabling better 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

These anticipated benefits can be framed and understood in terms of the three-dimensional benefits taxonomy developed by the JISC-funded ‘Keeping Research Data Safe 2’ Project (KRDS2).⁸ This classifies benefits in terms of: direct benefits, which save money; indirect benefits, which avoid potential costs; near-term benefits and long-term benefits; and lastly ‘private’ benefits, which accrue to particular agents, and ‘public’ benefits, which can bring various wider positive future outcomes.



Benefits one, two, and three are direct, near-term, and mostly private in character. Benefit four is direct, long-term, and private, although potentially also with public benefits if the improved comprehensibility of the information enables new research in the future. Benefit five, the improved awareness and use of software tools is interesting: it is of direct, near-term, private benefit to the researcher using the software, but the improved uptake of the best research tools also benefits the providers of such tools by increasing their user community and therefore their commercial strength, better enabling the producers of the software tools to improve their products and bring them to the attention of others who might benefit from them. Benefit six, awareness of central infrastructure services, not only benefits researchers but is also likely to improve the cost effectiveness of the central services being provided, due to economies of scale. When researchers interact with central service providers it tends to increase the knowledge and understanding of both parties, which can lead to improved and better-focussed services in the future, or research that makes better use of the expertise of university support staff. Benefits seven and eight – the reduced risk of data loss and better version control – are indirect benefits, as they help avoid the costs of having to recreate or ‘untangle’ data, which in some instances can be considerable.

All of the identified benefits of the data management training apply to the researchers using the material, but several are likely also to bring consequent financial benefits to the

⁸ Beagrie, N., Lavoie, B., Woollard M., ‘Keeping Research Data Safe 2’, 2010, p. 55.
<http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>.

University, such as: greater external research income from funding bodies due to improved funding bids; greater government research funding as more efficient research results in better outputs to enter for REF assessment (including data outputs); increased data assets due to data being assembled by researchers in formats enabling easier institutional preservation and curation.

Whilst these benefits are not straightforward to quantify, we have undertaken a number of activities to try to measure researcher demand for data and information management training, and have surveyed responses to the new training materials where possible within the constraints of the time and resources available to the Sudamih Project. These are detailed in Appendix A.

CONSTRAINTS AND DEPENDENCIES

The principal constraints affecting the ITLP's ability to take over the maintenance and dissemination of the information management learning and training materials developed by the Sudamih Project are staff-related. The current ITLP staff lack specialist expertise in humanities data management issues, so money will either need to be invested in re-skilling existing staff or hiring external experts.

Normally, when the ITLP staff at the University of Oxford recognize the need to learn a new technology for teaching purposes, the member of staff nominated to 'take ownership' of the skill will spend a considerable length of time (about 8-10 weeks) learning the topic and attending courses themselves to adequately understand it. Given that data management training is still in its infancy, with little in the way of existing 'train-the-trainer' material yet available, and no research information management courses that cater specifically for humanities researchers (apart from those they would be expected to teach), it would be very difficult for core ITLP staff to acquire the necessary expertise. Furthermore, none of the members of the teaching staff at the ITLP has undertaken original academic research themselves, so 'speaking from experience' would also be difficult.

Maintaining and teaching the courses produced by Sudamih will, therefore, necessitate the commissioning of external experts. This is already fairly routine practice, so little effort would be needed to create new processes, and the costs of using external experts are already understood, but finding and recruiting appropriately skilled tutors could prove difficult, and experts are not necessarily cheap.

RISK ASSESSMENT

RISK	Probability	Severity	Score (PxS)	MITIGATING ACTIONS
Software described in courses and on Research Skills Toolkit becomes obsolete or is withdrawn from market. This risk is particularly acute when dealing with online services, which can vanish overnight, potentially taking data with them.	5	4	20	Ensure courses and RSTk content is updated regularly, preferably termly. Make the risks of software obsolescence and service withdrawal abundantly clear, include strategies for preventing data loss and explain data formats and exporting.
Lack of available staff with suitable data management knowledge to update and teach the training materials	5	3	15	Start quickly and leave plenty time to find and commission an appropriate external teacher and allow them time to assess relevant software and modify the course to suit their own preferences and expertise.
Staff lack expertise to respond to follow-up enquiries	4	3	12	Given the lack of in-house expertise, it is likely that staff will not be able to adequately respond to specific software enquiries or questions demanding specialist knowledge. This weakness should be acknowledged during courses so as not to cause disappointment later. The general Computing Services help-desk may be able to help in some instances.
Demand for courses outstrips availability of places	4	2	8	Direct those wishing to register for the face-to-face courses to the RSTk content and online materials.
Course attendance is poor	1	5	5	Initial piloting of courses suggests this is unlikely in the near-term, but if this starts to be an issue then it may be worth marketing more vigorously or arranging 'closed sessions' organized with particular departments of divisions, who are well placed to reach researchers who would otherwise ignore ITLP promotions.
Cessation of Research Skills Toolkit Service (if the RSTk as a whole proves unaffordable, the information management material within it is also lost)	2	2	4	Keep content and presentation of RSTk separate to enable content to be easily removed and re-presented in different contexts.

Uptake of RSTk research information management material drops significantly / costs of maintaining content prove too high to justify low level of use.	1	4	4	Firstly ensure awareness, including working with Humanities Division and departments to promote material; otherwise pull training materials, but informing organizations such as the Digital Curation Centre and Research Information Network to see if they can provide an alternative home.
Different departments and divisions duplicate material or inconsistently update it	3	1	3	Due to the cross-departmental relevance of the material, ITLP should adopt a coordinating role and inform relevant departments and divisions as content is updated, whilst integrating their feedback.

Some risks associated with maintaining and disseminating the training materials developed by the Sudamih Project should simply be considered as part of the regular costs of service. The pace of technological change, for instance, will rapidly render the training materials obsolete without time spent every year re-researching and updating the content.

Courses should be monitored via feedback surveys and feedback passed to teachers to help improve matters. The Research Skills Toolkit should have an annual review procedure which includes the targeted surveying of users and potential users.

COST ANALYSIS

It is estimated that in FEC (Full Economic Cost) 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 ongoing costs are likely to be as follows for each of the five main types of training output listed at the beginning of this document:

1. Two courses (with accompanying course-books and slides) intended for face-to-face delivery

The two related courses: ‘Research Information Management : Organising Humanities Material’, and ‘Research Information Management : Tools for the Humanities’ are major outputs of the Sudamih Project, and the feedback received from their pilot runs has been very positive.⁹ There are a number of different costs involved in maintaining and staging these three-hour face-to-face courses with ‘hands-on’ exercises. Running and updating the courses will require the commission of an external expert, who are generally paid at £27.54 per hour for their teaching time, and £13.77 for preparation time. External

⁹ See Appendix A for details of the demand for the courses and their effectiveness.

commissioned worker do not incur FEC costs.

Activity	Resource	Costs per term (to run both courses once)
Identifying and instructing an appropriate (internal) teacher (one per year)	1 day of core staff work per year at grade 7 (FEC)	£132 (£396 / 3)
Update course content	25 hours at £13.77	£344.25
OUCS Training Room hire & set-up	25-person capacity room for 6 hours. Not charged to ITLP-led courses	£0 (would otherwise cost c. £180)
Printing of course booklets	25 booklets, covered by course price	£0 (would otherwise cost c. £150)
Cost of teacher (preparation)	6 hours at £13.77	£83.62
Cost of teacher (delivery)	6 hours at £27.54	£165.24
Cost of assistant (preparation & support)	12 hours at £13.77	£165.24
TOTAL		£890.35 (£2,671.05 annually)

The ITLP estimate that it takes approximately 50 hours to undertake a substantial annual update of an existing course. This cost is reduced if the updates are conducted more regularly. Here it is assumed that termly (three times per year) updates to the two three-hour courses will add up to approximately 75 hours annually. We also assume that the ITLP will need to spend 1 day per year of cores staff time locating an appropriately-qualified teacher within the University and explaining the work required of them.

If the two courses were both staged once per term, the annual cost would come to approximately £2,670, with course attendees each being charged £6 to attend. This is the usual fee charged by ITLP, primarily to cover the printing costs of the course book.

2. A suite of texts written for the ITLP's 'Research Skills Toolkit' covering data management tips and tools

The content written for the Research Skills Toolkit forms a significant output of the Sudamih Project. These materials constitute approximately 20% of RSTk webpages as the service currently stands. The ITLP are proposing to employ an intern to perform an annual update of the RSTk, although they believe that due to the specialist nature of some of the content of the Managing Research Information section input from an external expert may also be required. The external expert is in practice likely to be the person recruited to teach the courses above, and as such will not incur additional recruitment costs.

Activity	Resource	Annual cost (at 2011 prices)
20% of the costs of recruiting and instructing an appropriate intern (one intern will review the entire RSTk)	20% of 2 days core staff work per year at grade 7 (FEC)	£158
Verifying content and identifying material needing updating	2 weeks work by an intern (at £11.89 per hour)	£892
Updating material	1 week of external expertise (at £13.77 per hour)	£516
Hosting and server maintenance for Managing Research Information material	No additional funding required*	£0
TOTAL		£1,566

The Managing Research Information section of the RSTk does not incur any hosting charges in addition to those for the service as a whole.

3. Slide packs for use in induction sessions, introducing new researchers to the data management tools and services available at the University

The slide packs mostly refer to broad principles of data management with reference to key University services. As such, the amount of updating required every year is likely to be light. If they are to be useful they will need to be promoted to divisions and departments ahead of the start of each academic year (i.e. in September). The three slide packs will also need to be stored somewhere accessible, such as on a server hosted by ITLP or the Humanities Division. As the slide packs do not require external expertise to update this can be performed by in-house staff, although doing so does incur FEC.

Activity	Resource	Annual cost (at 2011 prices)
Update slide packs	0.5 days (FTE) g.7.8	£198
Promotion to key stakeholders (by email)	0.5 days (FTE) g.7.8	£198
Web hosting	Negligible	£1
TOTAL		£397

4. A data management factsheet designed to accompany the Humanities Division's 'Managing the D.Phil.' course.

The factsheet to accompany the 'Managing the D.Phil.' course run by the Humanities Division is only a short document covering a couple of sides of A4 paper. Although it will need to be checked and updated when the course is run, this is unlikely to take more than an hour, so costs are negligible.

Illustration of cost savings

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 get a sense of the impact that each course 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 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. Average, 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. Average's research time is spent writing up his research in the form of journal articles and other publications. Dr. Average 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 (at 2011 prices). This alone covers the ongoing costs of both the training courses in that term.

Now consider that Dr. Average 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.

We could also approach the issue of savings from the perspective of benefit 7 – the reduced risk of data loss (combined with elements of benefit 6: 'Better awareness and uptake of central infrastructure services intended to help researchers'; and benefit 8: 'Improved version control'). The introductory and induction materials created by Sudamih both remind researchers of the existence of the centrally-provided backing-up service offered by the University, whilst the Research Skills Toolkit content and the face-to-face courses go into greater detail about the importance of versioning, synchronizing, and tools which can help with these activities.

Roughly two-thirds (61%) of the researchers who attended the data management courses could recall that they had 'lost' research material (data, notes, files) due to files becoming corrupt, being overwritten, or simply going missing. The other third could not recall having lost material when asked, but none could say with certainty that they had

never lost any of their material. Those who reported losing information said that what they had lost amounted to between a few hours work and a month or so. None of the researchers had suffered any catastrophic loss. Based on the information received (and leaving aside those researchers who were uncertain), we calculated that approximately 1.15% of research information gathered is lost per year due to poor management and/or the failure to use the highest-reliability backing-up mechanisms. For Dr. Average (assuming the same conditions as above)¹⁰ this rate of loss would cost the institution £479 per year. If the training reduces these information losses by a half, the costs of running the courses he attended would be repaid within four years.

TIMESCALES

The online content developed by the Sudamih Project is already incorporated into the ITLP's Research Skills Toolkit, and can to a large degree be updated alongside the rest of the service in its standard schedule. The research data management training courses, on the other hand, cannot be implemented immediately due to the need to source new, appropriately skilled trainers and to give those trainers time to adapt the course content to suit their preferences and teaching style. It is possible that the members of the Sudamih Project Team who authored the original content may be available to teach the courses again, but the nature of project staff means that this cannot be guaranteed.

Despite the recognized possibility of having to leave some time to recruit a new teaching, it is important that the gap before the course is taught again is not too great. The new academic year after the long summer vacation would seem to be the ideal time to begin running the courses again. Failing to embed the training outputs of the Sudamih Project into the University's training infrastructure in time for next academic year will lead to greater costs in the future, as the currency of the materials will necessarily degenerate as new tools and technologies become available. The cost of updating the courses will be lower if the existing materials can be updated termly from the summer rather than if the materials need to be re-researched and re-written two or more years down the line. Furthermore, given that the content created by the Sudamih Project is open for any university to pick up and adapt, there is also the risk that Oxford will begin to lag behind its competitors, and consequently researchers at other universities make improvements to their research efficiency that comparatively disadvantages the University which created the materials in the first place.

CONCLUSIONS

The total cost to the IT Learning Programme at Oxford University Computing Services of maintaining and delivering the training and learning materials described comes to £4,634 per annum. The cost savings generated by improvements to researchers' skills and awareness are likely to amount to significantly more than this each year, and may

¹⁰ i.e. employing Dr. Average costs the institution under FEC terms around £111,000 per year, 50% of his time is spent on research, and 25% of that time is spent writing up results – leaving 75% of his research time (or 37.5% of his total time) to gather the information which he subsequently loses.

reasonably be expected to continue to accrue in subsequent years beyond that in which the training took place or the learning materials were consulted.

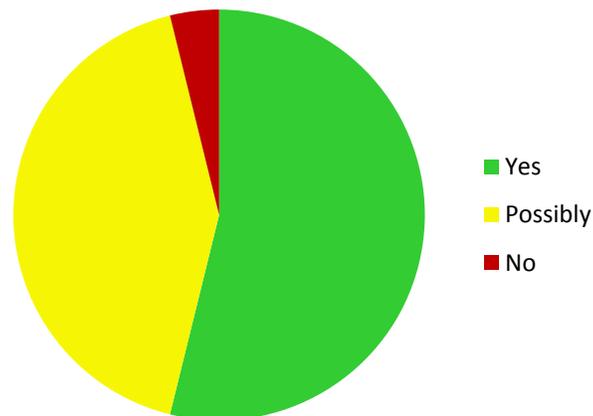
Given that the figure of £4,634 per annum is relatively low, and certainly not substantial enough to justify the recruitment of additional core staff (who in any case would require specialist knowledge that few training staff yet have), it is proposed that this money is provided centrally from the Computing Services budget (option *a* as described in the Options Appraisal).¹¹

¹¹ See p.7.

APPENDIX A : EVIDENCE OF DEMAND AND EFFECTIVENESS

The Sudamih Project began by conducting a set of interviews with researchers in the humanities to gauge demand for data/information management training, and understand the particular aspects of such training that researchers felt would be most useful. Our sample of 28 researchers was not entirely random or representative as we deliberately selected several researchers who were actively engaged in research involving structured data with planned digital outputs. The results are interesting nevertheless:

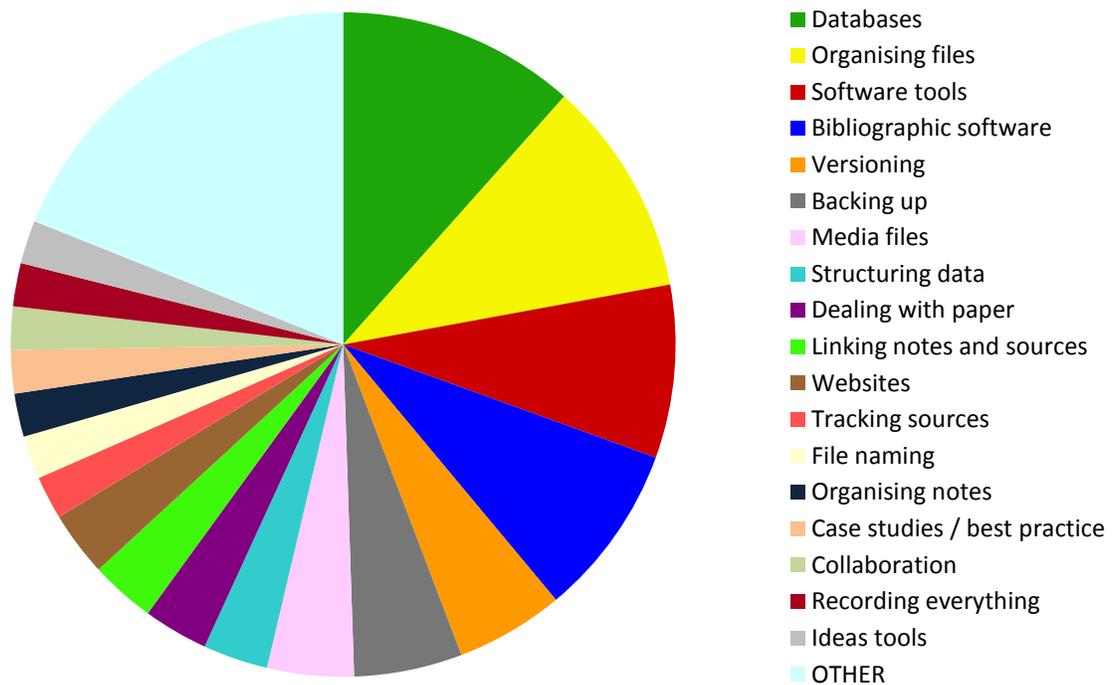
Humanities researchers' responses when asked if there was a need for data management training:



We also asked the interviewees whether they themselves had ever received any data management training. More than 75% reported that they had never received any such training, and most of the rest had not really received data management training in the sense that we understood it to mean, suggesting that the training materials developed by the project were essentially establishing a new market, but one in which there was already significant demand.

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

What researchers thought it would be useful for data management training to cover (either for themselves or new graduate research students):



N.B. we tried to leave it fairly open what we considered 'data management' to include, although we did define 'data' as not just including structured or numerical data: we explained that we intended it to cover all the information sources and notes that were gathered during the course of research for subsequent analysis and reference when creating research outputs.

As responses were 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.

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

The initial evidence of demand for training was confirmed by the two pilot training courses run at the Computing Services on the 31st January and the 28th February 2011. These three-hour courses were both staged in the 'Isis' training room at the Computing Services, which has a capacity of 25 people at computing terminals. Both courses were fully booked and had waiting lists of an additional 10 people within two days of being announced.

The effectiveness of the courses can be gauged by the number of attendees reporting that they would change the way they worked as a result. 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 would make 'significant' changes, 69% said that they 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.

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 they were using or going to use, and several mentioned that they would try tagging files and notes with keywords 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)” 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”. Improved happiness and well-being was not one of the initially-identified benefits of the training, but maybe could be considered as an indirect, long-term, private benefit.

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.

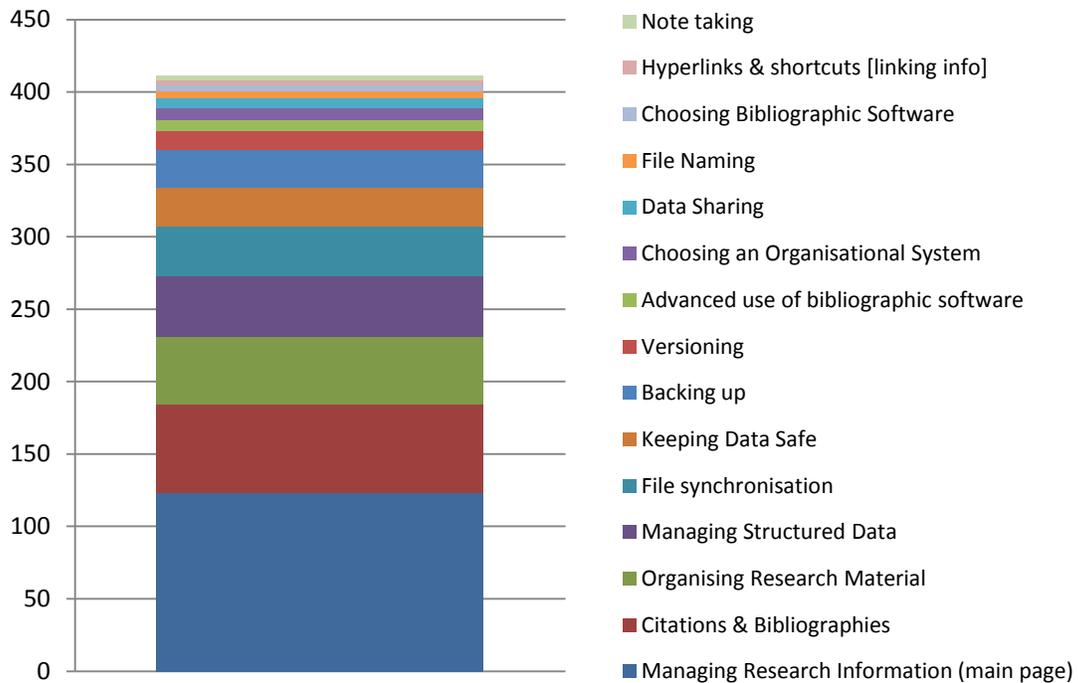
In an attempt to measure benefit 1 (time saved by researchers by locating and retrieving relevant research notes and information more rapidly) we asked course attendees to estimate how much of their time spent writing up their research outputs is actually spent looking for notes/files/data that they 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 occasion 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 or introducing researchers 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.

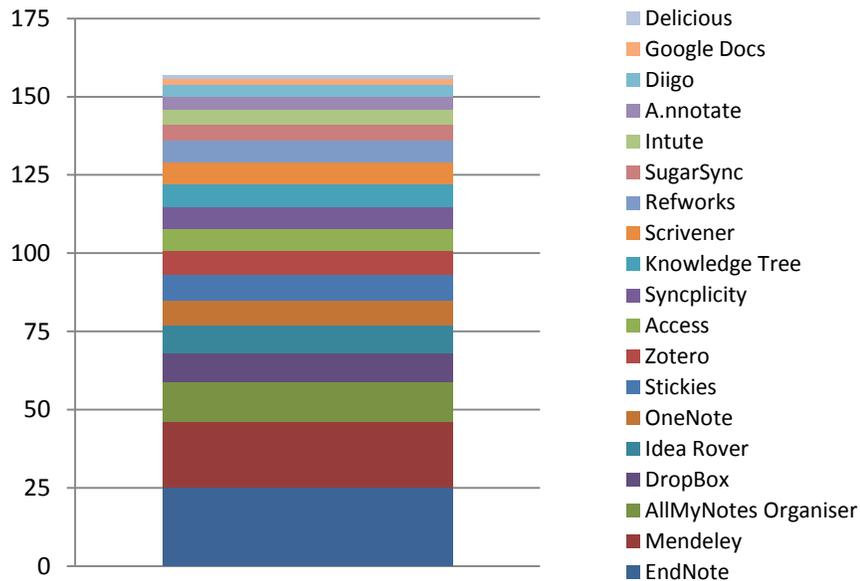
The Sudamih Project researcher requirements gathering phases identified different learning preferences amongst humanities researchers, and it was felt to be important that training content was available both online, to suit researchers who like to access information as and when an opportunity arises, and via face-to-face training courses, which better suit researchers who like to have the opportunity to ask questions and interact whilst learning.

The information management training content developed by the Sudamih Project to form part of the University of Oxford’s ‘Research Skills Toolkit’ went live on the 7/2/2011. It is hosted on the section of the website named ‘Managing Research 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 Research Information’ part of the Toolkit was publicized via an email sent on the 18/2/2011 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 will be made publicly available in due course. As of the 21st March, the different parts of the site have attracted the following number of visits:

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



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