

Sudamih Training Recommendations

Supporting Data Management Infrastructure in the Humanities

Thursday 22nd July 2010

James A J Wilson

James.wilson@oucs.ox.ac.uk

JISC

SUPPORTING DATA MANAGEMENT
INFRASTRUCTURE IN THE HUMANITIES
(SUDAMIH)



UNIVERSITY OF
OXFORD

Supporting Data Management Infrastructure in the Humanities (Sudamih)

- Pilot project to develop infrastructure at the University of Oxford, but outputs will be re-usable in other institutions
- Funded by JISC as part of their 'Managing Research Data' Programme
- Project team:

Prof. Paul Jeffreys – PI

Dr. Michael Fraser – co-PI

Prof. Andrew Wilson – co-PI

Dr. Ian Archer – co-PI

Dr Andrew Fairweather-Tall – co-PI

Sally Rumsey – Library Services

Kathryn Dally – Research Services

John Ireland – Computing Services

Dr James A J Wilson – Project Manager

Dr Meriel Patrick – Project Analyst

Asif Akram – Software developer



SUPPORTING DATA MANAGEMENT
INFRASTRUCTURE IN THE HUMANITIES
(SUDAMIH)



UNIVERSITY OF
OXFORD

Sudamih challenges

- Understanding how scholars in the humanities manage the information they use in their research
 - Finding, storing, structuring, using, and re-using information
- Pilot data management training modules
 - How can we improve existing practices?
- Pilot ‘Database as a Service’ (DaaS) system
 - What advantages can we bring to researchers through this?
- Cost models for data management services
- Sudamih is a pilot project – we hope to expand the data management infrastructure developed to other disciplines

Data



- Broad interpretation
 - All of the information that goes into the research process, whether stored on electronic or physical media
 - May include content of books, journal articles, photographs, emails, hand-written notes, etc.

- Narrow interpretation
 - Information stored in digital format which may be operated on by computers



What's the point of data management?

- Ensures researchers can find the most appropriate information rapidly when required
- Helps maintain integrity of information
- Facilitate re-use of information, both by original author and others
- Ensures electronic information isn't lost and doesn't become inaccessible or obsolete
- Enables some types of data to be compared and analysed to produce new research insights
- Prevents embarrassment at hands of climate sceptics & others

What does ‘data management training’ mean to researchers?



- Not a familiar term!

“[dealing with] tabulated information, or numerical or statistical information”

“how to organize your bibliography, your notes, the resources you’re going to need”

“storing data, backing it up, ... versioning your files”

“encoding text according to TEI or other standards, and thinking about documenting your database”

“how to manage references”

“what sort of data are worth structuring in a database”

“the classification and categorization of all the materials I’ve got”

Effects of good data management

“I do believe that our research could be enhanced by having better ways of storing information, because the way I store my thoughts makes a difference to how I use them when progressing in my thinking. I can see that improving the way I store them might help the actual thinking – apart from saving time, it might be a bit more substantial than that: having a clear view of what I’ve already done, or how my different projects interconnect, might just be heuristic in a sense.”

[Lecturer in Philosophy]

Current Humanities data management practices

- Many different 'systems' for organising data, most of which have changed over time in response to new requirements
- Rarely time to go back and re-organize things
- Most scholars work with a mixture of paper and electronic materials and notes
- Electronic files tend to be placed in an hierarchical folder structure
- Variety of opinions regarding bibliographic software
- Little awareness of policies of funding bodies

Humanities Data - Storage

- Favourite storage medium – Laptop hard drive
- Favourite backing-up mechanism
 - External hard drive, every once in a while
- Frequently use more than one computer, with files transferred via memory stick
- Relatively little use of institutionally-provided storage
- Ignorant of, or confused by automatic back-up systems
- Don't overestimate researcher's awareness of centrally provided infrastructure

Humanities data – re-use and sharing

- Where a public web interface is not envisaged as an output from the outset, there are problems sharing data:
 - It's messy
 - It's employs personal, idiosyncratic standards
 - It's partial and specific
 - It's existence is not widely known
 - Needs to be milked for publications first
- However, humanities researchers are rarely opposed to sharing their data *in principle*.



Do humanities researchers need training?

- Yes

Do humanities researchers *want* training?

- Most would appreciate training

“many colleagues have no idea, no conceptualization of that sort of structure of files, which means they lose things”

“I’m sure there are useful tricks one could learn, and ways of saving quite a bit of time in terms of not losing references and finding things more quickly”

“Data management is what we do, really, as researchers, and being trained to be a good researcher is always a good thing”

- Graduates least convinced, but,

“As you acquire more and more data, managing it becomes more of a challenge: if you have a rather esoteric filing system, any problems with that are likely to become more obvious as you add more to it.”

- Some scepticism regarding ‘generic’ data management training

And what training do they want?

- ‘Broad’ training

“Training in ways to organize material would be useful – computer file structures, organizing paper notes, that sort of thing”

“Case studies and examples of what people have done in the past [to organize all their information]”

“Finding out how to connect pictures to searchable notes would be really useful.”

- Technical training

“A review of different software packages – an overview which covers their advantages and disadvantages and shows what they might be used for”

“It might be useful to learn about specific bits of technology, such as scanner pens”

“It would be helpful to teach people data modelling, and how to deal with complex data. They need to understand data structures, and to know how to look at their own data”

Training for Graduates

- Suggestions for Graduates included:
 - Think in advance about how you need to structure your information
 - Good backing-up practices
 - recording your sources and what you've read
 - Bibliographic software
 - Versioning
 - Basic technological knowledge – Excel, database software, XML
- “People often start off by creating massive Word or Excel files, then two years down the line they have to change everything as they realize that the best way to manage the material is in a database”
- Catch graduates early(ish)

The sales pitch

- ‘Data management training’ may lack appeal given the other pressures on researchers’ time

“If a training course titled ‘Data Management’ were offered, most humanities students would consider it irrelevant to them”

“People would probably think it sounded useful, but might not actually do it”

Solutions?

“Any training in this field would need to make it very explicit what it was going to cover and what it would help people to do.”

“training would be more attractive to graduate students if it happened in their faculty.”

- Several people suggested that it would be worth making this sort of training ‘compulsory’

DCC 101 Lite at Oxford

- Marketed to “researchers and those who support researchers and want to learn more about how to develop sound data management and curation plans”
- Places filled within 48 hours
- Well-received, although participants wanted more
 - Worked examples or case studies
 - Emphasis on practicalities of managing and curating data
- Attendees thought institutional training should cover
 - Developing a data management plan; the curation life-cycle model; what data to keep and what to discard; integration with other information skills training; specific decision-making sessions

The format of delivery

- Classroom training
 - Gives opportunity to ask questions
 - Easier to adapt to circumstances
 - Social element
 - Hard to find time
- Online training
 - Can be done at a convenient time
 - You can refer back to it later
 - May be cheaper to run
 - Scales better
 - Easy to put off
- No clear preference – maybe classes with online support?
- Several senior researchers wanted a one-to-one advice service, particularly so that they could talk through the technical aspects of a project, which are likely to vary greatly

What does Oxford already offer?

- ‘Data management’ is not recognized as a field in its own right
- Humanities Divisional training
 - Introduction to the D.Phil; academic career skills; career management; managing the doctorate
- Humanities Research Facilitators
 - Advice on funding applications
- Individual faculty training
 - Varies greatly, but no real coverage of data management
- Computing Services
 - Courses on software packages, database design, and text encoding
 - Project advisory service

What we are proposing (first thoughts!)

1. Introduction to data management and the services provided by the University
 - To integrate with existing introductory courses
 2. Which tools to use for particular research problems
 - Face-to-face with online factsheets
 3. Organising and linking research notes and sources
 - Face-to-face
 4. Preparing technical funding bids
 - Via improved consultancy service
 5. Database design for humanities research data
 - Via consultancy service and possible face-to-face
 6. Tools and practices for collaborative research
 - Face-to-face with online factsheets
- Plus training to use the DaaS

Thanks!

Any Questions?

Contact me at james.wilson@oucs.ox.ac.uk

JISC

SUPPORTING DATA MANAGEMENT
INFRASTRUCTURE IN THE HUMANITIES
(SUDAMIH)



UNIVERSITY OF
OXFORD