Sudamih

• Supporting Data Management Infrastructure for the Humanities
• Part of a programme of data management infrastructure development at the University of Oxford
  – Scoping Digital Repository Services for Research Data Management (internally funded)
  – EIDCSR (Embedding Institutional Data Curation Services in Research) (JISC funded)
• Intention to extend disciplinary coverage and create integrated research data management processes
Database as a Service

Research Services
Divisional Research Services
Ethics Committee
Researchers
Computing Services

Project Planning
Project Set up
Data Creation
Documentation
Local Storage & Retrieval

Researchers
Academic departments

Ingest

Institutional Storage
Rediscovery Mechanism
Retrieval Mechanism

Computing Services
Libraries

Training

Research Services; Divisional Research Services; academic divisions; academic faculties; Computing Services; Libraries.
Data Management Training

Research Services
Divisional Research Services
Ethics Committee
Researchers
Computing Services

Researchers
Academic departments
Computing Services

Researchers
Academic departments

Local Storage & Retrieval
Documentation

Ingest

Institutional Storage
Rediscovery Mechanism
Retrieval Mechanism

Libraries
Computing Services

Ingest

Research Services; Divisional Research Services; academic divisions; academic faculties; Computing Services; Libraries.
Information/data Management Practices

• Little consideration given to data management
• Diverse and idiosyncratic data management practices
  – Lone scholar tradition contributes to this
• Low awareness of central services
• Little or no prior training
Database practices vary greatly in the humanities

But the process often looks like this...

Is this really right?

This data looks interesting. I wonder where it came from?
Databases in the humanities – current practices

- Database practices vary greatly in the humanities
- Or this...

Nice data – I can use this!
Problems identified

- Lack of technological awareness
- Poor backing-up practices
- Collaboration difficult
- Difficult to re-discover and re-use data
- Risk of technical obsolescence
- Money and long-term sustainability
- Technical expertise required
- What happens to data when its creators move?
Using the DaaS

I can access and cite good research data

I can find what data other people have been gathering

We can quickly & easily add & edit data and open it to the public
Future possibilities

• Full productionization and scalability
• Registry service for descriptive metadata
• Extended functionality beyond the humanities
• Incorporation of data storage models other than relational databases
  – e.g. XML-based, document-based, and CouchDB
• Deployment as a cloud-based software service
  – Involves advance monitoring and management tools
  – Capable of running on other institutional virtual infrastructures
• End-user documentation and support
Demand for training

Is there a need for data management training?
- Yes
- Possibly
- No

What should this training cover?
- Databases
- Organising files
- Software tools
- Bibliographic software
- Versioning
- Backing up
- Media files
- Structuring data
- Dealing with paper
- Linking notes and sources
- Websites
- Tracking sources
- File naming
- Organising notes
- Case studies / best practice
- Collaboration
- Recording everything
- Ideas tools
- OTHER
Approach to Training

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. **Organising and linking research information for later retrieval.** Including information on organising 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.
Training/support Materials Developed
Response to Courses

Have you or will you change any aspects of your own information management practices as a result of the course?

- 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)
- After the course I started to work as a 'horizontal organizer' and I feel it works much better for me. Also now I plan ahead how to get these materials organized and traceable, rather than working it out afterwards as I did before.
- I'm getting more and more organized and happy now! It changed not only my ways of working, but also my mood and my attitude

Yes, significantly
Yes, one or two aspects
Considering it
Unlikely
No
Costs & Benefits of Training

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
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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Resource</th>
<th>Costs per term (to run both courses once)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update course content</td>
<td>25 hours at £13.77</td>
<td>£344.25</td>
</tr>
<tr>
<td>OUCS Training Room hire &amp; set-up</td>
<td>25-person capacity room for 6 hours. Not charged to ITLP-led courses.</td>
<td>£0 (would otherwise cost c. £180)</td>
</tr>
<tr>
<td>Printing of course booklets</td>
<td>25 booklets, covered by course price.</td>
<td>£0 (would otherwise cost c. £150)</td>
</tr>
<tr>
<td>Cost of teacher (preparation)</td>
<td>6 hours at £13.77</td>
<td>£83.62</td>
</tr>
<tr>
<td>Cost of teacher (delivery)</td>
<td>6 hours at £27.54</td>
<td>£165.24</td>
</tr>
<tr>
<td>Cost of assistant (preparation &amp; support)</td>
<td>12 hours at £13.77</td>
<td>£165.24</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>£758.35</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(£2,275 annually)</td>
</tr>
</tbody>
</table>
## Costs & Benefits of Training

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
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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Resource</th>
<th>Costs per term (to run both courses once)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update course content</td>
<td>25 hours at £13.77</td>
<td>£344.25</td>
</tr>
<tr>
<td>OUCS Training Room hire &amp; set-up</td>
<td>25-person capacity room for 6 hours. Not charged to ITLP-led courses.</td>
<td>£0 (would otherwise cost c. £180)</td>
</tr>
<tr>
<td>Printing of course booklets</td>
<td>25 booklets, covered by course price.</td>
<td>£0 (would otherwise cost c. £150)</td>
</tr>
<tr>
<td>Cost of teacher (preparation)</td>
<td>6 hours at £13.77</td>
<td>£83.62</td>
</tr>
<tr>
<td>Cost of teacher (delivery)</td>
<td>6 hours at £27.54</td>
<td>£165.24</td>
</tr>
<tr>
<td>Cost of assistant (preparation &amp; support)</td>
<td>12 hours at £13.77</td>
<td>£165.24</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>£758.35</strong> (£2,275 annually)</td>
</tr>
</tbody>
</table>

I spend about 25% of my time writing up research publications

25%
### Costs & Benefits of Training

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
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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Resource</th>
<th>Costs per term (to run both courses once)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update course content</td>
<td>25 hours at £13.77</td>
<td>£344.25</td>
</tr>
<tr>
<td>OUCS Training Room hire &amp; set-up</td>
<td>25-person capacity room for 6 hours. Not charged to ITLP-led courses.</td>
<td>£0 (would otherwise cost c. £180)</td>
</tr>
<tr>
<td>Printing of course booklets</td>
<td>25 booklets, covered by course price.</td>
<td>£0 (would otherwise cost c. £150)</td>
</tr>
<tr>
<td>Cost of teacher (preparation)</td>
<td>6 hours at £13.77</td>
<td>£83.62</td>
</tr>
<tr>
<td>Cost of teacher (delivery)</td>
<td>6 hours at £27.54</td>
<td>£165.24</td>
</tr>
<tr>
<td>Cost of assistant (preparation &amp; support)</td>
<td>12 hours at £13.77</td>
<td>£165.24</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>£758.35</strong> (£2,275 annually)**</td>
</tr>
</tbody>
</table>

About 18% of that time is spent looking for notes and sources that I know I already have... somewhere

25% X 18% = 4.5%
Costs & Benefits of Training

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
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

<table>
<thead>
<tr>
<th>Activity</th>
<th>Resource</th>
<th>Costs per term (to run both courses once)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update course content</td>
<td>25 hours at £13.77</td>
<td>£344.25</td>
</tr>
<tr>
<td>OUCS Training Room hire &amp; set-up</td>
<td>25-person capacity room for 6 hours. Not charged to ITLP-led courses.</td>
<td>£0 (would otherwise cost c. £180)</td>
</tr>
<tr>
<td>Printing of course booklets</td>
<td>25 booklets, covered by course price.</td>
<td>£0 (would otherwise cost c. £150)</td>
</tr>
<tr>
<td>Cost of teacher (preparation)</td>
<td>6 hours at £13.77</td>
<td>£83.62</td>
</tr>
<tr>
<td>Cost of teacher (delivery)</td>
<td>6 hours at £27.54</td>
<td>£165.24</td>
</tr>
<tr>
<td>Cost of assistant (preparation &amp; support)</td>
<td>12 hours at £13.77</td>
<td>£165.24</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>£758.35</strong> (£2,275 annually)</td>
</tr>
</tbody>
</table>

I’m paid roughly £44,000 each year, but in terms of FEC, I cost £110,000

\[
25\% \times 18\% = 4.5\%
\]

Of £110,000 = £4,950 per year
Costs & Benefits of Training

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
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

### Activity

<table>
<thead>
<tr>
<th>Activity</th>
<th>Resource</th>
<th>Costs per term (to run both courses once)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update course content</td>
<td>25 hours at £13.77</td>
<td>£344.25</td>
</tr>
<tr>
<td>OUCS Training Room hire &amp; set-up</td>
<td>25-person capacity room for 6 hours. Not charged to ITLP-led courses.</td>
<td>£0 (would otherwise cost c. £180)</td>
</tr>
<tr>
<td>Printing of course booklets</td>
<td>25 booklets, covered by course price.</td>
<td>£0 (would otherwise cost c. £150)</td>
</tr>
<tr>
<td>Cost of teacher (preparation)</td>
<td>6 hours at £13.77</td>
<td>£83.62</td>
</tr>
<tr>
<td>Cost of teacher (delivery)</td>
<td>6 hours at £27.54</td>
<td>£165.24</td>
</tr>
<tr>
<td>Cost of assistant (preparation &amp; support)</td>
<td>12 hours at £13.77</td>
<td>£165.24</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>£758.35</strong> (£2,275 annually)</td>
</tr>
</tbody>
</table>

If I attend the two training courses and, as a result, I only spend 16% of my time hunting notes & sources

- **25% X 18% = 4.5%**
- Of **£110,000 = £4,950 per year**
- 2% of £4,950 = £99
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
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

### Activity

<table>
<thead>
<tr>
<th>Activity</th>
<th>Resource</th>
<th>Costs per term (to run both courses once)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update course content</td>
<td>25 hours at £13.77</td>
<td>£344.25</td>
</tr>
<tr>
<td>OUCS Training Room hire &amp; set-up</td>
<td>25-person capacity room for 6 hours. Not charged to ITLP-led courses.</td>
<td>£0 (would otherwise cost c. £180)</td>
</tr>
<tr>
<td>Printing of course booklets</td>
<td>25 booklets, covered by course price.</td>
<td>£0 (would otherwise cost c. £150)</td>
</tr>
<tr>
<td>Cost of teacher (preparation)</td>
<td>6 hours at £13.77</td>
<td>£83.62</td>
</tr>
<tr>
<td>Cost of teacher (delivery)</td>
<td>6 hours at £27.54</td>
<td>£165.24</td>
</tr>
<tr>
<td>Cost of assistant (preparation &amp; support)</td>
<td>12 hours at £13.77</td>
<td>£165.24</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>£758.35</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>(£2,275 annually)</strong></td>
</tr>
</tbody>
</table>

The costs of running the courses is recouped over the next 8 years of my employment

\[
25\% \times 18\% = 4.5\% \\
\text{Of } £110,000 = £4,950 \text{ per year} \\
2\% \text{ of } £4,950 = £99 \times 8 \text{ years} = £792
\]
Costs & Benefits of Training

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
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

### Activity vs. Resource vs. Costs per Term

<table>
<thead>
<tr>
<th>Activity</th>
<th>Resource</th>
<th>Costs per term (to run both courses once)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update course content</td>
<td>25 hours at £13.77</td>
<td>£344.25</td>
</tr>
<tr>
<td>OUCS Training Room hire &amp; set-up</td>
<td>25-person capacity room for 6 hours. Not charged to ITLP-led courses.</td>
<td>£0 (would otherwise cost c. £180)</td>
</tr>
<tr>
<td>Printing of course booklets</td>
<td>25 booklets, covered by course price.</td>
<td>£0 (would otherwise cost c. £150)</td>
</tr>
<tr>
<td>Cost of teacher (preparation)</td>
<td>6 hours at £13.77</td>
<td>£83.62</td>
</tr>
<tr>
<td>Cost of teacher (delivery)</td>
<td>6 hours at £27.54</td>
<td>£165.24</td>
</tr>
<tr>
<td>Cost of assistant (preparation &amp; support)</td>
<td>12 hours at £13.77</td>
<td>£165.24</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>£758.35 (£2,275 annually)</td>
</tr>
</tbody>
</table>

And there’s only one of me!

25% X 18%
= 4.5%
Of £110,000
= £4,950 per year
2% of £4,950
= £99 x 8 years = £792
Lessons learnt

• Mind your language
• Communication and engagement is important. Need buy-in at multiple levels
• Researchers are key, but they probably don’t think all that much about re-usability, long-term curation, etc.
• Understand the rhythms of the academic year
• Use existing infrastructure where possible
• You may need to compromise when embedding data management training into existing training
Thanks!