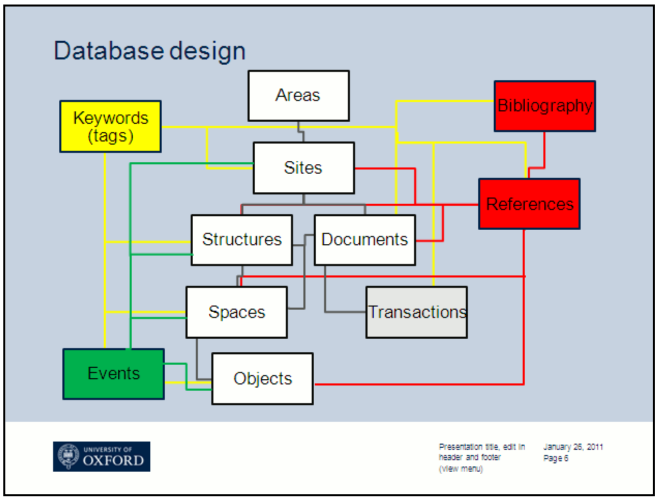
Workshop Report: Databases in the Humanities – Where Next?

The second Sudamih workshop, entitled 'Databases in the Humanities – Where Next?' took place on the afternoon of Friday 21st January in Oxford, attracting 62 delegates from a variety of universities and other organisations across the UK (including one from SURFnet in the Netherlands). The aim of the workshop was to consider the ways in which humanities researchers build, maintain, and preserve databases, along with the processes currently in place to support such activities, and to provide a forum in which ideas could be exchanged and new approaches to humanities data illustrated. Part of the motivation from the Sudamih Project's perspective was to introduce the Database as a Service (DaaS) system to a wider group of potential users and better understand the work that they were already engaged with to assess how the DaaS might fit in to this research landscape.

Professor Paul Jeffreys, Director of IT at the University of Oxford, introduced the workshop and explained its context within the broader programme to develop research data management infrastructure for the University. I then gave the first of the eight 20-minute presentations and questions sessions with an overview of the DaaS – why we believed it would be useful for researchers and the functionality we were in the process of implementing. It was evident from the questions that followed that there was a lot of interest in the potential of the DaaS to improve the sustainability of databases beyond their initial period of funding, an aspect that we have yet to consider in depth but clearly need to address. It is reasonable to expect that a scalable, generic, centrally-hosted system such as the DaaS should be able to offer considerable economies of scale and this is something that we shall try to better quantify over the next two months.

Dr. Miko Flohr then took to the stage to introduce one of the research projects that we have been working with as an exemplar: The Roman Economy Project (REP). One of the common elements across a number of the presentations was the problems associated with combining existing databases into a new and more useful online database, and this is the process that the REP is currently going through. Dr. Flohr spoke about the difficulties of cleaning up data (especially when that data has simply been placed in text fields and in various different formats) and the fairly elaborate table structure that the REP is adopting, with linking tables for keywords, references, and bibliographic information connected to a number of 'original' data tables.



Database design

Keywords (tags)
Areas
Bibliography
Sites
References
Structures
Documents
Spaces
Transactions
Events
Objects

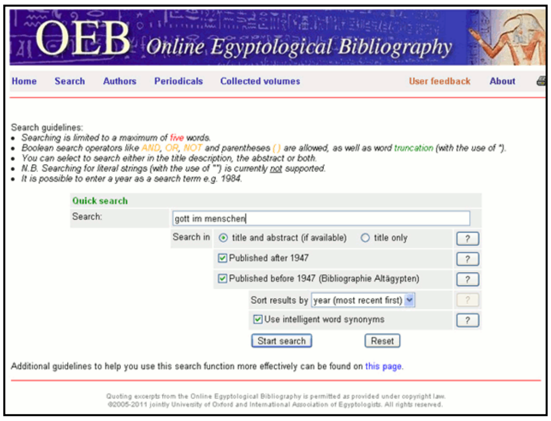
UNIVERSITY OF OXFORD

Presentation title, edit in header and footer (view menu)

January 26, 2011
Page 6

Dr. Jacob Dahl emphasised the importance of making data available to other researchers, as without that availability important sections of history can be omitted from the historical narrative. Speaking particularly about the huge database of cuneiform inscriptions he has been working on, Dr. Dahl spoke of the benefits of retaining the 'slightly archaic' software that the project has been using, in order to allow the data to be managed by researchers themselves. In the past, the technical people employed by the project have had a habit of being poached by commercial companies, leaving the researchers themselves in an awkward situation. The transliterations of the texts are also stored in a simple self-devised standard which can be converted into XML. The use of simple formats is in part so the data can be taken back out of the database if needed, for instance in the event of a cessation of funding. At the conclusion of the presentation, Jacob reiterated that the most important achievement of the project was that people were actually using the data, linking to it, citing it, and getting their students to use it too.

In his presentation 'A Small Field and its bibliographies', Professor John Baines spoke of the challenges faced in the process of integrating pre-existing Egyptology bibliographic databases into the new Online Egyptological Bibliography (OEB). Costs were an important issue for this project. Several of the source bibliographies being added to the OEB had fallen foul of funding cuts and could only continue to be made available via such an enterprise. An automated data capture system was also being used to minimize costs, meaning that only one full-time employee was needed to edit the new system. The OEB was itself using a subscription-based model to ensure its long-term sustainability.



Search guidelines:

- Searching is limited to a maximum of five words.
- Boolean search operators like AND, OR, NOT and parentheses ( ) are allowed, as well as word truncation (with the use of *).
- You can select to search either in the title description, the abstract or both.
- N.B. Searching for literal strings (with the use of "") is currently not supported.
- It is possible to enter a year as a search term e.g. 1984.

Quick search

Search: gott im menschen

Search in ● title and abstract (if available) ○ title only

☑ Published after 1947

☑ Published before 1947 (Bibliographie Altägypten)

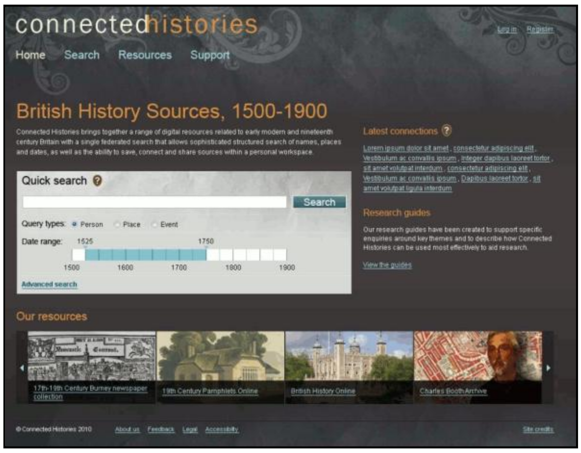
Sort results by year (most recent first)

☑ Use intelligent word synonyms

Additional guidelines to help you use this search function more effectively can be found on this page.

Quoting excerpts from the Online Egyptological Bibliography is permitted as provided under copyright law.
©2005-2011 jointly University of Oxford and International Association of Egyptologists. All rights reserved.

After the coffee break Dr. Jonathan Blaney from the Institute of Historical Research explained the processes behind the Connected Histories project. Again, the integration of various existing databases was the central challenge, although here the solution was being sought not so much via close integration of database structures but via a looser approach involving the semantic tagging (sometimes by hand, sometimes by machine) of disparate sources to enable structured searching. The presentation included a warning against reliance on some of the standard Web 2.0 software available at present due to the long-term unreliability of such services being sustained.



British History Sources, 1500-1900

Connected Histories brings together a range of digital resources related to early modern and nineteenth century Britain with a single federated search that allows sophisticated structured search of names, places and dates, as well as the ability to save, connect and share sources within a personal workspace.

Quick search

Query types: ● Person ○ Place ○ Event

Date range: 1525 1750

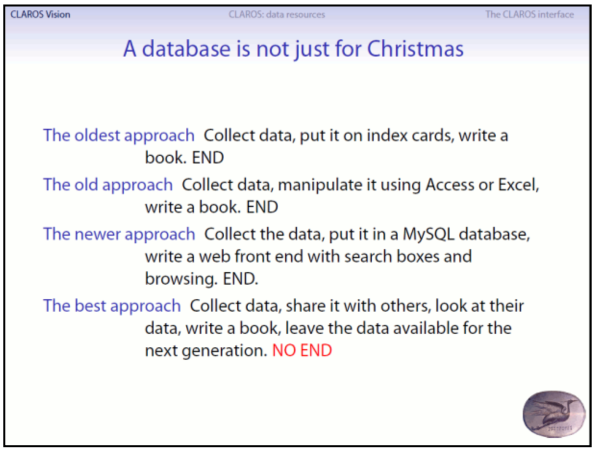
Latest connections

Lorem ipsum dolor sit amet, consectetur adipiscing elit, Vestibulum ac convallis ipsum, Integer dapibus laoreet tortor, sit amet volutpat interdum, consectetur adipiscing elit, Vestibulum ac convallis ipsum, Dapibus laoreet tortor, sit amet volutpat ligula interdum

Research guides

Our research guides have been created to support specific enquiries around key themes and to describe how Connected Histories can be used most effectively to aid research.

Dr. Hilde De Weerdt then demonstrated how the data she had been gathering during her research on 12th and 13th century imperial China could be used to analyse social and political networks. Via a series of visual representations of the data, including outputs on maps to illustrate geographic connections, and network diagrams indicating personal correspondence and citations gathered from notebooks of the era, Dr. De Weerdt showed how technologies could be applied to the data to challenge or substantiate various expectations. In this case the underlying texts had been worked up using TEI XML to enable the analysis being undertaken.



CLAROS Vision CLAROS: data resources The CLAROS interface

A database is not just for Christmas

The oldest approach Collect data, put it on index cards, write a book. END

The old approach Collect data, manipulate it using Access or Excel, write a book. END

The newer approach Collect the data, put it in a MySQL database, write a web front end with search boxes and browsing. END.

The best approach Collect data, share it with others, look at their data, write a book, leave the data available for the next generation. NO END

Next up, Sebastian Rahtz introduced the CLAROS Project to the workshop. He explained that the aim of CLAROS, unlike the other projects we had heard from so far, was not to produce websites but rather to produce linked open data and form a 'dataweb' of classical art and archaeology. The idea is that CLAROS does not seek to change resources already available in existing databases but rather makes them mappable to assist resource discovery. This is achieved via the use of RDF triplets to describe the various relationships between objects and by encouraging the use of CIDOC CRM as a common ontology. Responses to the papers in the workshop suggested that RDF is a relatively new metadata model for researchers in the humanities and its applications are still being explored. It was explained in this paper that whilst the performance of RDF databases had in the past been regarded as a barrier to adoption, many of these issues had now been addressed, and the field was developing rapidly.

The final presentation was given by Dr. Claire Warwick, who addressed the problem of data usability and maintenance when funding runs out. The importance of the website came to the fore again as Dr. Warwick emphasised the importance of maintaining the interface to the data as well as the data itself. Research suggests that appearances matter a lot when researchers use data resources, and if the web interface looks 'dodgy' or out of date, people tend to be reluctant to use it regardless of the quality of the underlying data. Money spent developing such websites is therefore wasted if they are not being used. On a positive note, the new Research Excellence Framework was, Dr. Warwick thought, likely to act as a driver for researchers and universities to improve their website maintenance, as digital objects can now be assessed in the same manner as journal articles. Tips for database developers included: get user feedback early in the development process; ensure the web front-end can be detached from the data back-end; document the data; and ensure the documentation remains attached to the data it describes.

The slide presentations from the workshop are available from the workshop website (http://sudamih.oucs.ox.ac.uk/databases_workshop.xml)

James A. J. Wilson

11th February, 2011