

E-social science is here

"e-science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it"

Dr John Taylor, Director General of Research Councils

The e-social science vision is of a globally-connected scholarly community made up of virtual co-laboratories aimed at promoting the highest quality scientific research. The realisation of this vision rests on the development of an IT infrastructure that will enable flexible, secure, co-ordinated resource sharing among dynamic collections of individuals, institutions and resources. This includes computational systems, data storage and specialised experimental facilities. It can be seen as a natural and potentially massive extension of existing web services. The next generation of infrastructure is known as the Grid.



The Grid points to great opportunities for social science by: broadening access to existing data repositories; developing new forms of analysis software; providing increased remote access to computational facilities; new forms of middleware to tie together social science facilities; new visualisation technologies; and exploring the broadening of information available via GIS facilities.

ESRC has invested in e-science by establishing a national centre for e-social science — the NCeSS. This centre will have a distributed structure comprising: a co-ordinating hub based at the University of Manchester and supported by the UKDA; and a set of research-based Nodes distributed across the UK. The Hub has been granted an initial three-year budget of £1.5 million from 1 April 2004. It draws on the expertise of the UKDA, Manchester Computing and staff in economics, sociology, geography, and computer science at the University of Manchester.

The role of the Hub is to act as the central resource base for e-social science issues and activities in the UK, integrating them with ESRC research methods initiatives and the existing e-science core programme. The Hub will provide a one-stop shop for awareness-raising, expertise, training, technical infrastructure, data resources, computer facilities and user support for e-social

science research. The Nodes will each pursue a designated part of a wide-ranging research agenda, on a set of key themes. The task of the Nodes is to develop Grid technologies and apply them to substantive social science research problems. The Nodes will be commissioned in the second half of 2004 and begin work in April 2005.

A series of demonstrator projects and consultancies have been commissioned by ESRC on e-social science. Details of these initiatives are available from the NCeSS web site: www.ncess.org. Other aspects of ESRC's e-science strategy can be found on the ESRC web site at: www.esrc.ac.uk/esrccontent/researchfunding/esciencecentre.asp

ESDS Qualidata Online Conference

On 15 January ESDS Qualidata held a workshop, *Online Access to Qualitative Data: Opportunities and Challenges*. The conference focused on issues surrounding the provision of online access, and technical strategies for publishing and exchanging qualitative data.

Louise Corti and Libby Bishop of ESDS Qualidata set the scene for the day by providing the ESRC and JISC vision for access to qualitative data, emphasising why an XML model for preserving, sharing and accessing data online is essential. They introduced the drivers and principles of the Qualidata Online project, currently under development, and drew attention to the behind-the-scenes work required to load or publish the data into the system. Extensive work is required to prepare textual data (primarily from interview transcripts) for loading into a web-based exploratory system, and any tools that can help automate this process would provide significant advantages. Currently many of the most vital and data-contextualising researcher products that are generated during the data management and analysis stages using a CAQDAS package are lost. There exists no way to export, for example, coded thematic data or annotations, in a non-proprietary and exchangeable data format. Finally, existing metadata standards for social science data were highlighted.

Content providers in search of web-based technical solutions was chaired by Professor Nigel Fielding, University of Surrey, and covered three online resource projects in progress. Firstly, Graham Gibbs, Department of Behavioural Sciences at the University of Huddersfield, spoke on *Establishing an accessible data corpus for professional users using the Victoria Climbié inquiry transcript: merging user needs and research dimensions*. The data corpus will use

XML-based standards for archiving qualitative data, and will incorporate thematic elements and analytic features suggested by an investigation of user needs. Michael Kandiah, Centre for Contemporary British History (CCBH), University of London, then offered an insight into *Witness seminars online*. The full transcripts of witness seminars, together with introductory papers and chronologies, are published onto the web as pdf files. How to make these potentially very rich materials available to researchers beyond the narrow historical community was discussed. Finally, Rob Perks and Jonathan Robinson from the National Sound Archive, British Library, presented *The way we speak: web access to sound recordings at the British Library*. This project has been exploring ways of providing remote online user access to its two million recordings: both to the catalogue data and to the actual recordings.



Technical and software solutions: XML addressed metadata and technical solutions for facilitating the sharing, presentation and exchange of digital qualitative data via the web. Lou Burnard, Oxford University Computing Services, provided an illuminating paper, *Text Encoding for Interchange: Myths and Realities*. He presented an orthodoxy on the way digital resources are created and used in academia, and contrasted this with the myth that the process of digitization creates a culturally-neutral and objective version of a cultural artefact superior to everything that has preceded it.

The final session of the afternoon took up the gauntlet of introducing e-science to social scientists — a new area beset by complex technical jargon that is currently largely unfamiliar to the regular social scientist. Nigel Fielding introduced the findings from a recent ESRC scoping study he had undertaken on potential applications of e-social science to qualitative methods and research. The social science of Grid technologies — or e-social science — offers new resources that can support new modes of research enquiry. In his talk he highlighted some of the most pertinent applications of relevance to online access to qualitative data — development of research instruments and other tools for online data collection; tools for content analysis of large natural language corpora; extending CAQDAS to analysis of visual data, and applying AI routines for more systematic qualitative data analysis; applying XML to support CAQDAS-type tools for web use; specifications for real-time distributed manipulation, annotation and analysis of streamed data; and the potential of the Access Grid to work collaboratively. Finally, Michael D. Fischer, Reader in Anthropological Sciences and Director of the Centre for Social Anthropology and Computing at the University of Kent at Canterbury spoke on *From Patches to Patchwork: Grid*

applications and qualitative research. His current e-science demonstrator project is exploring if and how the Grid might be used to advance qualitative research — for example, by utilising new technologies to address key issues such as stable access to document segments, managing layered documents, new methods for collaborative work, reuse of information and methods, and prospects for integrating the products of qualitative research with other information streams in society.

The day was well attended by researchers and technical specialists from a variety of backgrounds, including overseas data archives and social science projects. The applications of generic standards and tools are essential for enabling a new culture of data sharing, and in many ways represents the ground work to grid-enable qualitative data for the e-science applications the ESRC is envisioning. Unfortunately, very few of the CAQDAS programs are using XML, but both opportunities and pressure arising from e-science and the more web-aware collaborative user communities may entice them to consider implementing a data and metadata model that could facilitate data exchange.

ESDS Qualidata are planning talks with teams at Sheffield, Oxford and Edinburgh about tools for natural language processing and XML mark-up strategies. A first draft of a specific DTD for describing and marking up the content of qualitative data, that builds upon the DDI and TEI, will be presented at the International Association for Social Science Information Services and Technology (IASSIST) conference, which will be held at the University of Wisconsin, Madison in May 2004.

The abstracts and slides from the presentations are available at:
www.esds.ac.uk/news/esdspastevents.asp

Enhancements to Nesstar

ESDS is dedicated to providing easy online access to our most popular data holdings. The cornerstone of this endeavour in the improved Nesstar data browsing service. The UKDA Nesstar Catalogue has been renamed ESDS Nesstar Light Catalogue to reflect the UKDA's role as service provider for the ESDS.

In April 2003, Nesstar Version 2 software included improvements to the look and feel, and ease-of-use of the interface. This version also improved the overall stability of the software and the tools used to publish the data. A major change was that applets were removed so that all data exploration now occurs in a centralised frameset with no pop-ups and easier navigation. Also, instead of being presented with a list of datasets, data are now organised into a series of easy-to-browse folders and sub-folders grouped by data series, UK geographic focus and teaching datasets.

There are now 93 studies in Nesstar including 38 large-scale ESDS Government and two ESDS Longitudinal datasets. All new deposits from the following data series are now routinely prepared for and published to Nesstar: Quarterly Labour Force Survey (including Household Datasets); ONS Omnibus Survey; General Household Survey; British Social Attitudes; Scottish Social Attitudes; Scottish Household Survey; British Election Study (cross-

“...potentially very rich materials available to researchers...”

“...applications of e-social science to qualitative methods and research...”

section data only); Northern Ireland Life and Times/Young Life and Times; Wales Life and Times; and teaching datasets. The teaching datasets currently include British Crime Survey and Labour Force Survey data.

ESDS Nesstar Light Catalogue:
nesstar.esds.ac.uk/nesstarlight



BHPS Nesstar sampler file available, second sampler planned

Via the online Nesstar software tool, registered ESDS users can now analyse, visualise, subset and download selected data from the British Household Panel Survey (BHPS). Frequencies can also be run on the data by users who are not registered with ESDS.

The BHPS sampler is the first file with a longitudinal component to be added to the ESDS Nesstar Light Catalogue. This unique resource was created to provide potential users of this large and complex dataset with a taste of how the BHPS might be used. As with other Nesstar datasets, the sampler is in an accessible rectangular format and is therefore ideal both for exploratory research and as a tool for teaching on research methods courses.

The sampler contains wave one respondents, and follows them for eleven waves, selecting broadly the same variables at each wave. Substantively, the dataset is in the area of work, family and health, but it can also be used to examine social mobility and income mobility. The dataset contains a number of background variables indicating parental social class, and some opinion variables relating to the work and family theme.

A second BHPS sampler file is currently being constructed and is due to be made available via the ESDS Nesstar Light Catalogue in the coming weeks. This file will be based around the theme of social and political attitudes.

Latest collections available for researchers

The following brief descriptions highlight a selection of the newly-released datasets:

SN 4786 British Crime Survey, 2001

SN 4787 British Crime Survey, 2001-2002

The British Crime Survey (BCS) is one of the largest social surveys conducted in Britain. It is primarily a 'victimisation' survey, in which respondents are asked about the experiences of property crimes of the household and personal crimes which they have experienced. From 2001, the series became annual instead of biennial, and from

the 2001-2002 wave the survey moved to the financial year cycle used by other large government surveys.

SN 4767 Northern Ireland Life and Times Survey, 2002

The Northern Ireland Life and Times Survey (NILT) series began in 1998, and was the successor to the previous Northern Ireland Social Attitudes series. The aims of the NILT series are to: monitor public attitudes towards social policy and political issues in Northern Ireland; provide a time series on attitudes to key social policy areas; facilitate academic social policy analysis; and provide a freely-available resource on public attitudes for the wider community of users in Northern Ireland.

SN 4733 Constructing Children's Welfare : a Comparative Study of Professional Practice, 2000-2001

This research sought to compare and contrast how prevailing models of childhood are selectively drawn upon and interpreted by Guardians *ad litem* (GALs) and Family Court Welfare Officers (FCWOs) in their everyday working lives. The main topics covered include childhood; child welfare; family life; legal system; and working practices.

SN 4805 Quarterly Labour Force Survey, September - November, 2003

The Labour Force Survey is a unique source of articulated information using international definitions of employment and unemployment and economic inactivity, together with a wide range of related topics such as occupation, training, hours of work and personal characteristics of household members aged 16 and over. The questionnaire comprises core questions included in every survey, together with non-core questions which vary from quarter to quarter. The first part contains questions on information about household, family structure, housing and demographic details. The second part includes questions on economic activity, education, and health.

SN 4783 ONS Omnibus Survey, October 2002

The Omnibus is a regular, multi-purpose survey which in recent years has gone into the field in eight months of the year. Each month's questionnaire consists of two elements: core questions, covering demographic information, and non-core questions that vary from month to month. The non-core questions for this month were tobacco consumption; telephones; internet access; human rights; walking and cycling.

New datasets and new editions are continuously added to the UKDA catalogue. The online catalogue may be searched via: a free-text or fielded search, including searching by title, study number, time period, data creator etc.; browsing by subject; a thesaurus-aided search; or by browsing a list of new data releases.

For a full list of recently-released datasets see:
www.data-archive.ac.uk/home/newData.asp

Descriptions of all datasets can be found from the Search Catalogue web page:
www.data-archive.ac.uk/search/searchStart.asp

"...providing easy online access to our most popular data..."

"...examine social mobility and income mobility..."

AHDS History News

PCC Wills, 1384-1858

AHDS History has recently acquired a new study from The National Archives (TNA), *Prerogative Court of Canterbury (PCC) and Related Probate Jurisdiction: Will Registers*. Covering all wills in this series proven between 1384 and 1858, the database includes detailed biographical information on over one million people. The key data covers names, occupation, place of residence and date of probate. There is also a direct link to the TNA's Documents Online system which can be used to download an image of the original will. (The TNA make a charge for this service.) The PCC wills metadata represents an important historical resource, not only as a finding aid for surviving PCC wills, but also as a means for prosopographical research, investigating testators' behaviour by location and social group over time. The data also has the

potential to be linked with other sources facilitating investigation into a range of research questions.

TNA's Documents Online system:
www.documentsonline.pro.gov.uk

Online Historical Population Reports Project

This JISC-funded project will be looking to set up a panel of users, to test the developing system and provide feedback on issues surrounding the development of the web site and the selection of material to be made machine-readable. Given that, within the lifetime of the project, all tables in the published census and Registrar-General's Reports for the period 1801-1931 cannot be made available, suggestions of tables for digitisation would be welcomed.

Email: info@history.ahds.ac.uk.

Forthcoming Events

- 23 March 2004:** JISC Conference 2004, International Convention Centre, Birmingham
- 25 March 2004:** Social Science Online - The Internet for Economics, Business, Management and Accountancy, University of Bristol
- 26 March 2004:** CASS seminar: Survey Research Resources and Secondary Datasets, University of Southampton
- 30 March 2004:** Exploring Survey Non-response in Great Britain: the 2001 Census-linked Study of Survey Non-response, Royal Statistical Society, London
- 31 March 2004:** National Child Development Study (NCDS) and 1970 British Cohort Study (BCS70): a workshop for potential users, Institute of Education, London
- 2 April 2004:** Social Science Data Archives: creating, depositing and using data, Wolfson Suite, Main Library, University of Edinburgh
- 14 April 2004:** Introduction to the ESDS International Databanks, Manchester Computing, University of Manchester
- 15 April 2004:** Introduction to the Analysis of Time Series using Stata, Manchester Computing, University of Manchester
- 23 April 2004:** Social Science Data Archives: creating, depositing and using data, University of Oxford
- 12 May 2004:** Social Science Online - The Internet for Politics, London School of Economics and Political Science
- 25-28 May 2004:** IASSIST 2004, Madison, Wisconsin, USA
- 21-25 June 2004:** ESRC Social Science Week, various locations
- 23 June 2004:** ESDS Government Workshop: Analysing hierarchical survey data, University of Manchester
- 23 June 2004:** ESDS Government Workshop: Using government surveys for research, University of Manchester
- 30 June 2004:** MIMAS Open Forum 2004, University of Manchester
- 1-3 July 2004:** ESRC Research Methods Festival, St Catherine's College, Oxford

Further details of these and other events may be found on the ESDS web site.

Editorial group enquiries
Email: publicity@esds.ac.uk
Tel: +44 (0)1206 872319/873159
Fax: +44 (0)1206 872003

General ESDS enquiries
Email: help@esds.ac.uk
Tel: +44 (0)1206 872143
Fax: +44 (0)1206 872003

Acquisitions
Email: acquisitions@esds.ac.uk
Tel: +44 (0)1206 872572/872974
Fax: +44 (0)1206 872003

ESDS Government
Email: govsurveys@esds.ac.uk
Tel: +44 (0)161 275 1980
Fax: +44 (0)161 275 4722

ESDS International
Email: international@esds.ac.uk
Tel: +44 (0)161 275 6109
Fax: +44 (0)161 275 6071

ESDS Longitudinal
Email: longitudinal@esds.ac.uk
Tel: +44 (0)1206 872143
Fax: +44 (0)1206 872003

ESDS Qualidata
Email: qualidata@esds.ac.uk
Tel: +44 (0)1206 873058
Fax: +44 (0)1206 872003

AHDS History
Email: info@history.ahds.ac.uk
Tel: +44 (0)1206 872326
Fax: +44 (0)1206 872003

Economic and Social Data Service
UK Data Archive
University of Essex
Wivenhoe Park
Colchester
Essex CO4 3SQ, UK