The History Data Service – Using Technology to Enhance Access

by Cressida Chappell, Oscar Struijvé, Sheila Anderson *

The History Data Service (HDS) <http://hds.essex.ac.uk> is funded by the UK Joint Information Systems Committee (JISC) <http://www.jisc.ac.uk/> to collect, manage, and encourage re-use of digital resources which result from or support historical research and teaching. The HDS is located and integrated in the UK Data Archive <http://dawww.essex.ac.uk/> and is the Arts and Humanities Data Service (AHDS) <http://ahds.ac.uk/> service provider for the historical disciplines. The AHDS provides archival, training and other functions to the archaeology, history, performing arts, textual studies and visual arts communities, and consists of five subject-based service providers and a managing executive. The HDS collection covers a time period from the late tenth century to the mid twentieth century, and includes a wide range of historical data, which has been transcribed or compiled from original sources.

The HDS is committed to using technology to improve access to its collection through a programme of work that is essentially *needs* rather than *technology*-driven. Although this programme of work aims to make effective use of established and state of the art technologies and concepts for data and metadata storage, presentation and delivery, the real emphasis is on recognising and responding to end-users' needs. The HDS has an active and ongoing policy of consulting with actual and potential users. For example, in April 1998 the HDS held a workshop <http://hds.essex.ac.uk/reports/user_needs/final_report01.stm> to explore, assess and prioritise the needs of end-users in the historical community.

The central goal of this programme of work is to improve and increase access to historical data by making the location, identification, assessment and use of data easier. The objective is to get more people using and experimenting with historical data, and the HDS is aiming to gradually extend its user-base into the less computer-literate and non-computing segments of the historical community by acting as an information provider as well as a data provider. In a very broad sense, the HDS is seeking to improve the relationship between end-users and data and it is essential that this is carried out within existing resources.

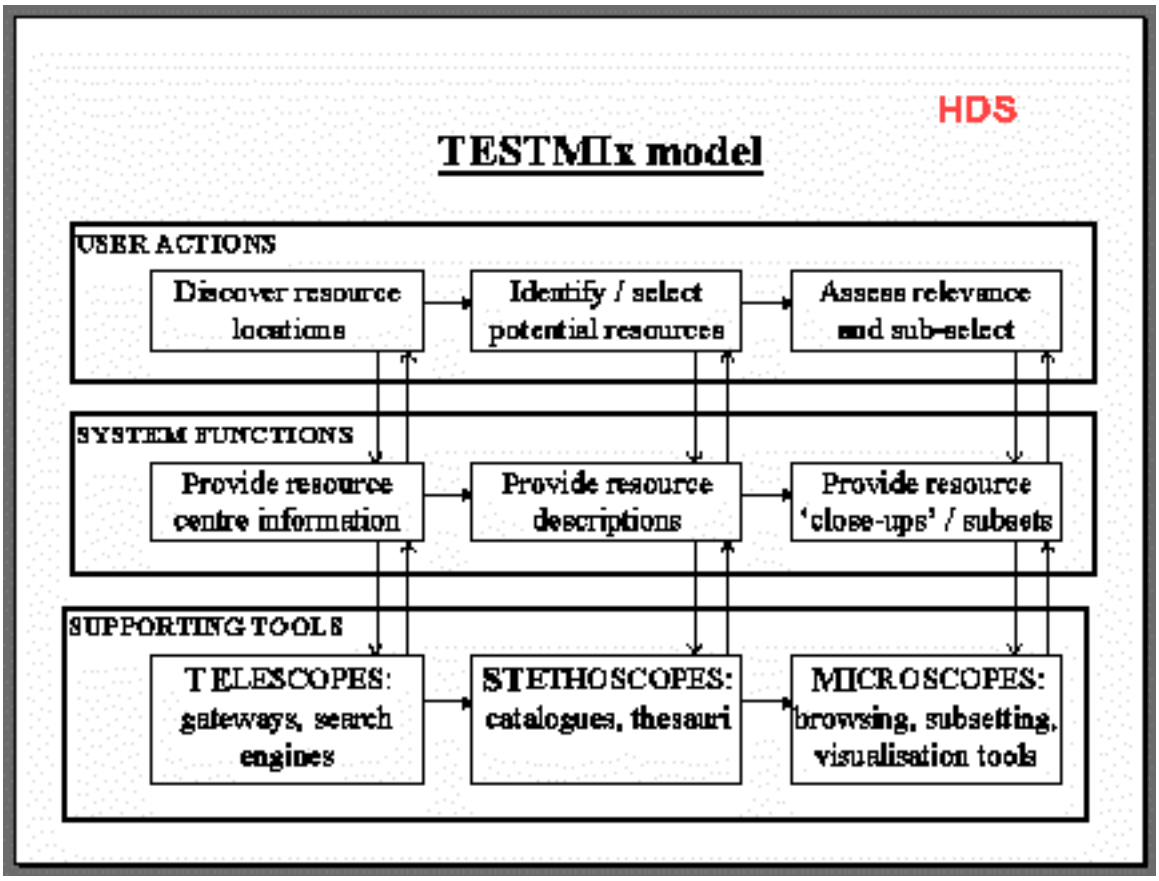
The HDS approach to using technology to enhance access is underpinned by a model called TESTMIx, short for telescope, stethoscope and microscope. The principle behind this model is that users want to locate, identify and assess data. The model maps user needs and actions to system functions, and it maps the system functions to the system components or tools required by users.

The top level of this model deals with user actions and needs. From left to right we have the need to discover information about data centres, the need to identify and select suitable data, and lastly the need to make detailed assessments of the relevance of data, along with the need to explore data in detail and make sub-selections. The next layer maps user needs and actions to system functions, so we have the provision of information about data centres, the provision of searchable data descriptions, and lastly the provision of data 'close-ups' and subsets. The next layer maps the system functions to the system components or supporting tools that are required, so we have gateways and search engines likened to telescopes; catalogues and thesauri likened to stethoscopes; and browsing, subsetting and visualisation tools likened to microscopes.

The HDS uses TESTMIx as a framework to structure and prioritise development work; to clarify relationships between services and systems; to identify scope for improvement and collaboration; and as a focus point for technical and service activities. It stimulates a coherent approach, and provides simple metaphors for communication. Within this framework the HDS is implementing a multilevelled strategy to improve and increase access to data.

Increasing the number of metadata access points

The first level involves increasing the number of metadata access points and the HDS has two different approaches. One approach uses conventional online catalogues and operational examples include the UK Data Archive's information retrieval system, BIRON <http://biron.essex.ac.uk/cgi-bin/biron/> and the CESSDA IDC (Council of European Social Science Data Archives Integrated Data Catalogue) <http://dastar.essex.ac.uk/Cessda/IDC/> Information about the HDS collection is



also being made available through the prototype AHDS Integrated Access Gateway <http://prospero.ahds.ac.uk:8080/ahds_live> which is based upon the Dublin Core and the Z39.50 network applications protocol, and which acts as a virtual union catalogue for the collections of the five subject-based service providers. In the coming months information about the HDS collection will also be accessible via the Cheshire Information Retrieval System – an SGML-based system which utilises the Data Documentation Initiative (DDI) Codebook DTD.

The other approach uses a tree-based structure as an alternative way of accessing information about the HDS collection. This will allow users to adopt a 'drill down' approach to locating data in addition to the more sophisticated search options offered by online catalogues. Users will be able to drill down to metadata via three dimensions. A time dimension by centuries, a geographic dimension by countries and administrative subdivisions within the UK, and a subject categories dimension.

Providing online access to additional information
The second level involves providing online access to additional information and will allow users to access to types of information that are not generally found in catalogue records. In particular we are interested in providing users with access to online documentation with the option to preview a sample of data. We believe that this will make it much easier for users to make detailed assessments of the suitability of data and that it is a more efficient way of supplying users with information.

These services will initially apply to areas of the HDS collection where there is critical mass of related materials, because there is a greater potential to create additional documentation. We are intending to apply this approach to a collection of early twentieth century surveys, a collection of European state finance data, and a collection of electoral poll book data. These services will be freely available to all users and will not require registration.

Providing a data and documentation ftp service
The third level involves providing a data and documentation ftp service and will give registered users online access to the vast majority of the HDS collection. We envisage that users who register with the HDS will be able to select and download data when they require it in suitable easy-to-use formats such as tab or comma delimited ASCII. The main exceptions will be difficult to use data and the minority of HDS data which has more

restrictive access conditions.

Developing online browsing, sub-setting, combining and downloading facilities

The fourth level involves the development of online browsing, sub-setting, combining and downloading facilities for major collections of value-added data and will allow registered users to explore fully documented data collections online. The HDS has developed this service for a large collection of nineteenth and twentieth century statistics, the Great Britain Historical Database (GBHD), and work is now being carried out on developing a similar service for a large collection of individual-level non-anonymised British historical census data which includes the 1881 Census for England and Wales digitised by the Genealogical Society of Utah and the UK Federation of Family History Societies.

The GBHD has been assembled by Humphrey Southall at Queen Mary Westfield College, London and incorporates demographic statistics, marriage statistics, mortality statistics, employment statistics, trade union statistics, government unemployment statistics, poor law statistics and small debt statistics. The GBHD Online system allows users to sub-set this data by geographical area, at present either standard regions and/or counties. The system also allows users to specify the tables to be searched for relevant data and to specify the variables to be included in the result. Finally the resulting subset and customised documentation can be viewed and browsed online or downloaded to the user's workstation by FTP. Data is formatted as fixed width ASCII with headers, which can be imported into a variety of software packages for further manipulation and analysis. For more information about GBHD Online (including details about registering as a user) please see the GBHD Online webpages at <http://hds.essex.ac.uk/gbh.stm>

The HDS is confident that this multilevelled strategy will encourage and enhance use of and experimentation with the HDS collection. We believe that this programme of work will widen and improve access to historical data, and that it will contribute to the creation of an infrastructure which will enable historians and others to explore the full potential offered by digital resources.

*Paper presented at the 1998 IASSIST/CSS Conference, Yale University, New Haven, USA, May 19-22, 1998, by C.Chappell, O. Struijvé, S. Anderson, History Data Service, UK Data Archive, University of Essex