

# The Local Data Support Landscape in the UK

## Introduction

This paper will report on existing data support infrastructures within the UK tertiary education community. The paper will then discuss early methods and traditions of data collection within UK territories. In addition it will focus on the current UK data landscape with particular reference to specialized national data centres which provide access to large-scale government surveys, macro socio-economic data, population censuses and spatial data. It will conclude by outlining examples of local data support services with particular reference to the 'accidental' nature of the data librarian, their organizational role and areas of expertise in addition to future developments.

## Data in the UK

The 'tradition' of data collection within the United Kingdom can be traced back to the 7th century 'Senchus fer n'Alba' in Gaelic Scotland (translated as tradition/census of the men of Alba). The original document is now lost but was translated in the 10th century and is regarded as the earliest native tax assessment known in Britain. It is part genealogical record, part inventory of the territories of the descendants of Eochiad Muin and consists of a list of the numbers of men that the various families of the Scoto-Irish kingdom of Dal Riata (centred on modern-day Argyll) could provide for their navy.

In 1086 the Domesday Book (an inventory of land use and property in England) was commissioned by William the Conqueror for administration purposes.

The 17th and 18th centuries saw a number of countries conduct national censuses e.g. 1666 Quebec province, 1703 Iceland, 1749 Sweden and 1790 USA however it was not until 1801 that the first comprehensive UK Census was conducted, partly to ascertain the number of men able to fight in Napoleonic wars, to be subsequently carried out on a decennial basis.

The need for a Central Statistics Office was recognised as early as the 1830's with the introduction of Tables of Revenue (a type of statistical year book instituted by George Richardson Porter from the Board of Trade). A decision to create a CSO was agreed in 1880 but never

by *Stuart Macdonald & Luis Martinez*<sup>1</sup>

came to fruition and, indeed, it wasn't until 1941, with the aim of ensuring coherent statistical information, that the Central Statistics Office was founded by Winston Churchill.

Following the advent of mainframe computing the SSRC Data Bank was established at the University of Essex in 1967 (later to become the UK Data

Archive<sup>2</sup>). 1971 saw the Social Statistics Laboratory at Strathclyde University<sup>3</sup> being set up in addition to the appearance of the first microprocessor. In 1981 the first IBM desktop PC was appearing in the midst of the developing Internet. The first data library based in a UK tertiary education institution was set up at Edinburgh University in 1983. In 1992 the World Wide Web was released by CERN (the European Organization for Nuclear Research) and in 1996 the CSO merged with the Office for Population, Censuses and Surveys to become the Office for National Statistics.

From the abridged account above it is evident that the collection, organisation and analysis of records about people is nothing new. What is new, in relative terms, is the microprocessor and PC in addition to advances in telecommunications and web technologies. And as Robin Rice suggested in the recent article 'the Internet and democratisation of access to data' (as part of an online discussion for ESRC Social Sciences Online – Past, Present and Future<sup>4</sup>)

“Data collection has arguably been changed more by computers than analysis itself, which has been dominated for decades by a few well-known statistical, and qualitative, analysis packages”

Analysis of large research datasets at the desktop requires a different set of skills to those of data discovery. These skills include tools to make the data usable in addition to a familiarity with the construct of the dataset. Thus it is as a result of this march of technology that there have emerged data professionals who not only have the necessary data discovery skills but also provide access to, support and train those wishing to use research and statistical data.

## Government Statistical Services

The Office for National Statistics (ONS) is the government department that provides UK statistical and registration services in addition to planning and conducting the decennial census for England and Wales. It is also responsible for producing a wide range of key economic and social statistics which are used by policy makers across government to create evidence-based policies and monitor performance against them. Publications include:

**Labour Force Survey** - quarterly sample survey of households living at private addresses in Great Britain. Its purpose is to provide information on the UK labour market that can then be used to develop, manage, evaluate and report on labour market policies.

**Family Expenditure Survey** - continuous survey of household expenditure and income that has been in existence since 1957. Annual samples of around 10,000 households (about 1 in 2000 of all United Kingdom households) are selected each year.

**General Household Survey** - continuous national survey of people living in private households. The main aim of the survey is to collect data on a range of core topics, covering household, family and individual information.

**Health Survey for England** - annual surveys about the health of people living in England. It began in 1991 and has been carried out annually since then. A number of core questions are included every year but each year's survey also has a particular focus on a disease or condition or population group.

The General Register Office for Scotland (GROS)<sup>5</sup> is charged with a similar role in Scotland. GROS is the department of the devolved Scottish Administration responsible for the registration of births, marriages, deaths, divorces, and adoptions in Scotland as well as conducting and publishing the output from the Scottish Census.

In addition, the Scottish Executive<sup>6</sup> provides relevant and reliable statistical information, analysis and advice that meet the needs of government, business and the people of Scotland. They are also responsible for commissioning a number of surveys specific to Scotland, these include:

Scottish Household Survey

Scottish Crime Survey

Scottish Health Survey

Scottish House Condition Survey

The Northern Ireland Statistics and Research Agency (NISRA) and the Statistical Directorate of the National Assembly of Wales have similar functions within the other

UK territories.

## UK Data Centres for tertiary education

There are a number of UK National Data Centres offering the UK tertiary education and research community networked access to a library of data, information and research resources. In the majority of cases services are available free of charge to members of UK tertiary education institutions for academic use, although institutional subscription and end-user registration are required for most services.

The UK Data Archive (UKDA), based at the University of Essex acts as a repository for the largest collection of digital data in the social sciences and humanities in the UK including the aforementioned economic and socio-demographic survey data, amongst many others. Its remit includes data acquisition, preservation, dissemination and promotion of social scientific data. It also houses a major collection of computerised historical material and is responsible for the Census Registration Service which facilitates access to census data resources for UK higher and further education.

The UKDA is a lead partner in the Economic and Social Data Service (ESDS)<sup>7</sup> which was launched in January 2003. ESDS is a distributed service, based on the collaboration between four key centres of expertise, UKDA, MIMAS, the Institute for Social and Economic Research (ISER), and the Cathie Marsh Centre for Census and Survey Research (CCSR). It acts as a national data service providing access and support for an extensive collection of quantitative and qualitative datasets for the research, learning and teaching communities. It is organised in five sections :

- ESDS Government – dedicated to the government surveys such as the General Household Survey.
- ESDS Longitudinal – supporting the use of the UK longitudinal collection, datasets like British Household Panel Survey or British Cohort Study.
- ESDS International - provides access to a range of international macro-economic datasets from organizations like the UN, Eurostat, OECD and IMF.
- ESDS Qualidata – specialist service providing access and support to qualitative datasets. Data includes interviews, focus groups, personal documents and photographs.
- ESDS Access and preservation – focuses on the data acquisition, processing, preservation and dissemination.

Hosted by Edinburgh University Data Library, EDINA<sup>8</sup>

was launched as a JISC-funded National Data Centre in January 1996. Services include abstract and index bibliographic databases such as BIOSIS, Cab Abstracts; spatial data services such as Digimap (which provides the UK tertiary education community with access to a range of Ordnance Survey mapping products), UKBORDERS, EuroGlobalMap and EDINA agcensus; multimedia services such as Education Media Online (EMOL) and the Education Image Gallery (EIG). EDINA are also partners in Learning and Teaching projects (JORUM and NLN) in addition to spatial development projects such as eMapScholar and the resource discovery tool Go-Geo!, and infrastructure projects such as Shibboleth Development and Support Services.

MIMAS<sup>9</sup> (based at Manchester Computing at the University of Manchester) is a JISC and ESRC-funded national data centre, and like EDINA provides UK higher education, further education and research community with networked access to key data and information resources to support teaching, learning and research across a wide range of disciplines. Services include bibliographic discovery tools such as COPAC, zetoc, JSTOR and the ISI Web of Knowledge; learning materials such as NLN; spatial data services such as LandMap which provides access to UK Satellite Data; Census data via the Census Dissemination Unit; International Databanks (via ESDS International).

The Arts and Humanities Data Service (AHDS)<sup>10</sup> is a JISC and AHRB-funded data centre which aids the discovery, creation and preservation of digital resources in and for research, teaching and learning in the arts and humanities. It currently consists of 5 distributed hubs: the Archaeology Data Service; AHDS History; AHDS Visual Arts; AHDS Literature, Languages and Linguistics; AHDS Performing Arts. The Hubs collect, preserve, catalogue, and distribute digital resources which are relevant to their subject areas, facilitate good practice in their creation and use, and offer some user services

### **United Kingdom local data support services**

Academic Institutions provide support for data services in a variety of ways<sup>11</sup>. This is reflected in the diversity of organisational representatives for ESDS, for example. Some work in a data library, a university library, a university computing centre, a central research office or an academic department.

However, the data support offered by data libraries goes beyond supporting the national data services. Data librarians/managers deal with the management and implementation of such services. Among their multiple tasks, data libraries:

- act as data repositories, developing and preserving local collections

- serve as reference services helping researchers to identify appropriate resources and troubleshooting
- educate users to access and handle data resources through teaching and learning activities.

Although there are common activities, the level of support and areas of expertise varies amongst data support services. Below there are examples of four different local data support services in the UK.

### **Edinburgh University Data Library**

Edinburgh University Data Library (EUDL)<sup>12</sup> was the first such service in the UK, started in 1983.

It was set up as a small group with a Sociology lecturer as part time manager, with 1.5 staff (one programmer and one computing assistant). Currently two qualified librarians provide the service, with administrative and technical support from EDINA.

The current collection covers large scale government surveys, macroeconomic and financial time series, population and agricultural census data and geospatial resources.

EUDL specialises in data for Scotland, and GIS resources due to its relationship with EDINA. The Data Library staff actively participates in local training activities in addition to providing a consultancy service, helping with the extraction, merging, matching and customization of data; for time-consuming jobs a fee is charged.

### **University of Oxford Data Library**

The University of Oxford Data Library<sup>13</sup> started in 1988. Three people formed the Computing and Research Support Unit, with one statistician, one computer/statistical software specialist and one data manager. At present it consists of one data manager, with no dedicated IT support and is part of the Nuffield College library .

The current collection comprises survey micro-datasets from UK and elsewhere, including the large government continuous surveys, and many ad hoc, repeated cross-section and panel/cohort academic surveys. Subsets of General Household Survey and Labour Force Survey variables combined over time have been compiled, and are widely used by researchers.

For seventeen years now the University of Oxford Data Library has supported researchers in using quantitative datasets; some of the key support functions have been:

- Data searching and conversion
- Negotiations and management of contracts with data providers

- Storage, protection and access arrangements
- Questionnaire design, data collection and management methods, knowledge of question wording, coding structures and methods, data cleaning.

### **London School of Economics Data Library**

The LSE Data Library<sup>14</sup> was launched in 1997 to support LSE researchers in the task of locating quantitative data. It provides an advisory service to PhD students, contract researchers and academics, helping them to locate and access datasets.

Its collection includes a microdata archive covering large scale government surveys, longitudinal data collections and international opinion polls as well as a wide range of aggregated databases providing worldwide socio-economic indicators from IGOs such as IMF, OECD or EUROSTAT. It also includes spatial data such as UK boundaries and EU and World administrative regions. Lastly, financial databases are increasingly becoming an important part of the collection, providing company accounts, indexes and bond data, exchange and interest rates, etc.

The Data Librarian offers direct support for users through the weekly data surgery (one-to-one advice) and Information Literacy courses, helping to locate, access, and format data.

A data laboratory, Datalab, is being implemented that will provide gigabit connectivity between PCs in a computer classroom and a dedicated server. The Datalab will be used for using datasets for teaching as a first stage, and will be hosting all microdata and managing access and metadata.

A high level advisory group for guidance on academic priorities formed by one academic from each department is also in place. The group will identify academic priorities across LSE departments and will shape strategic planning for the future.

### **London School of Economics RLab Data Services**

The London School of Economics RLab<sup>15</sup> Data Services started in 1999 providing data support to LSE's research laboratory, a unique institution bringing together leading research centres in economics, finance, industrial relations, social policy and demography.

The centerpiece of the collection is an electronic library housing approximately 150GB of data. The data is mainly social survey data, with some financial, geographical and medical data. Both macro and micro datasets are held in the library, data from individual countries throughout the world, as well as a wealth of international sources from the US, Europe, India, China.

Data Support is part of the RLAB IT Service, there is a

team of 6 people, as well as a full-time data assistant. The data manager has the support of two systems professionals, and two information professionals to help with the web site.

RLAB's data manager, Tanvi Desai, is now involved in the ESRC Review of International Data Sources and Needs. The aim of the project is to gain an understanding of the opportunities and the obstacles presented by international data resources, enabling us to recommend strategies for improving international research and collaboration.<sup>16</sup>

### **Data Information Specialist Committee, DISC-UK**

UK Data librarians have always been represented by the International Association of Social Science Information Service & Technology [IASSIST]. This association also represents international data archives, statistical agencies, government departments and non-profit organizations. However, a need existed for much closer collaboration to deal with data issues within and relevant to the UK.

In October 2002, a mailing list "Digging for Data" was set up as a forum to help data librarians, national data centre site representatives' and any other academic staff, support staff, statistical consultants or students to locate and use quantitative data. It was an informal initiative amongst UK data librarians represented by Oxford University, the University of Edinburgh and the London School of Economics.

In September 2003 the data libraries from these Universities formed a support group called DISC-UK, or Data Information Specialist Committee-UK<sup>17</sup>, holding their first meeting in April 2004. The group meets several times per year to compare issues and solutions arising from their daily work and its aims are the following:

- Foster understanding between data users and providers
- Raise awareness of the value of data support in Universities
- Share information and resources among local data support staff

The founding members intend to open up their group to others performing similar roles in their universities, though they may not work in dedicated data libraries. ESDS site representatives were emailed a simple questionnaire to find out the level of data support offered at their institutions. Although this elicited only one response from a site representative who had little to do with data support in his institution, it is an area which DISC-UK will pursue further.



Fig.1 Data Information Specialists Committee (DISC-UK) website

A much better response has been obtained when individuals have been approached. Universities such as Glasgow, Birbeck College, Southampton and others have been contacted and links with those institutions have been established for future collaboration. An interesting fact to emerge from such contact is that in most cases people doing data support in those institutions are subject librarians dealing with other electronic resources such as bibliographic databases.

The website for the group has been set up, with links to member sites and a description of the group's aims. Through time it is hoped that the site develops into a helpful resource hosting online training materials and links to relevant articles.

Channels of communication between DISC-UK and the national data centres are already in place. ESDS workshops

have been organised in each of the member institutions and several improvements suggested to UK Data Archive's administration web interfaces.

The next step for the group is to plan and coordinate the necessary resources to act as a broker between member institutions and the national data centres in addition to investigating the viability of running 'Train the Trainers' events. The arrangement of regular meetings with such centres will benefit both, establishing means to provide feedback and develop common strategies for the promotion of the data hosted at the national data centres.

**Accidental Data Librarian**

So where are the data librarians coming from? Is there a common educational background? What are the skills required to be a data librarian?

Data librarianship is still an underdeveloped profession in the United Kingdom. The UK example shows the variety of

backgrounds for its data librarian/manager professionals.

Jane Roberts, University of Oxford Data Library, comes from a Social Science background with a Politics degree, and worked as a Social Policy researcher, from questionnaire design and data collection through to analysis. Jane also provided a support service for data management and SPSS.

Tanvi Desai, LSE RLAB Data Service, comes from a Science background with a degree in Mathematics with European studies. She worked as a research assistant doing data analysis on the key government social surveys. Subsequently she managed a number of primary data collection projects.

Luis Martinez, LSE Data Library, comes also from a Science background with a degree in Mathematics specialised in Statistics and Computer Science with very little data experience.

Robin Rice is Data Librarian to the University of Edinburgh. She has a Masters degree in Library and Information Studies from the University of Wisconsin-Madison, where she worked as a data librarian through the 1990s.

Stuart Macdonald, also from Edinburgh University Data Library has a biosciences background with a degree in Biochemistry in addition to a postgraduate qualification in Information Studies. He has worked for the last six years in the world of data.

“Future of data library development will depend critically upon the combined efforts and expertise of three sorts of practitioner. These are the reference librarian, the applied statistician and the software engineer.”<sup>18</sup>

As Peter Burnhill suggested in his paper “Towards the Development of Data Libraries in the UK”, the skills required to deal with data management can be summarised as an amalgamation of three core areas :

- Computing; to manage the systems that will store the data and metadata and will provide access to it. To be able to deal with data formatting, merging and customization. To be aware of the latest technologies which will make a significant impact in dissemination and curation of data.
- Statistics; to interpret the data and help users to take decisions on the appropriate data analysis. To support conducting new surveys, helping with questionnaire design, sampling methods and coding structures.
- Librarianship; required to catalogue and organise

resources. To be able to structure the information in such a way that users can make the most out of what is available. To conduct training session to raise awareness and educate data users.

Armed with such skills data specialists educate themselves through the day to day exposure to the world of data, specialising according to the needs of their institutions, which in turn have strengths and weakness in the disciplines that they cover.

### Conclusion

To conclude, new web and telecommunication technologies coupled to value-added metadata has resulted in advances in data discovery and dissemination. However, as we move towards the use of powerful grid technologies with the next generation of internet, and institutional repositories there may well be a shift towards a more devolved data services infrastructure. The volumes of data will increase exponentially, bringing with it the need for appropriate data curation activities. Access to data will be better controlled although possibly more and more restricted. Thus it is evident that the 21st Century data professional will need to evolve as fast as these technological events take place in addition to being actively involved in the dialogue protecting access to data.

### Footnotes

<sup>1</sup>Paper presented at the IASSIST Conference, Edinburgh, May 2005, by Stuart Macdonald (Edinburgh University Data Library) and Luis Martinez (London School of Economics Data Library). Contact: Stuart.Macdonald@ed.ac.uk and L.Martinez@lse.ac.uk.

<sup>2</sup> [www.data-archive.ac.uk](http://www.data-archive.ac.uk)

<sup>3</sup> [www.strath.ac.uk/Departments/SocialStats/](http://www.strath.ac.uk/Departments/SocialStats/)

<sup>4</sup> ‘The Internet and democratisation of access to data’ Robin Rice (EUDL), blog for Social Science Week on SOSIG, [www.sosig.ac.uk/socsciweek/](http://www.sosig.ac.uk/socsciweek/)

<sup>5</sup> [www.gro-scotland.gov.uk/](http://www.gro-scotland.gov.uk/)

<sup>6</sup> [www.scotland.gov.uk/Home](http://www.scotland.gov.uk/Home)

<sup>7</sup> [www.esds.ac.uk](http://www.esds.ac.uk)

<sup>8</sup> <http://edina/>

<sup>9</sup> [www.mimas.ac.uk](http://www.mimas.ac.uk)

<sup>10</sup> <http://ahds.ac.uk/>

<sup>11</sup> “Providing local data support for academic data libraries.” Robin Rice (EUDL), The Data Archive Bulletin:

8-11. [Available: <http://datalib.ed.ac.uk/discuk/docs/rice2000.pdf>]

<sup>12</sup> <http://datalib.ed.ac.uk/>

<sup>13</sup> <http://www.nuff.ox.ac.uk/projects/datalibrary/>

<sup>14</sup> <http://www.lse.ac.uk/library/datlib/>

<sup>15</sup> <http://rlab.lse.ac.uk/data/>

<sup>16</sup> ESRC Review of International Data Resources and Needs: <http://rlab.lse.ac.uk/ESRCData/default.asp>

<sup>17</sup> <http://datalib.ed.ac.uk/discuk/>

<sup>18</sup> Peter Burnhill, 1985. Towards the development of data libraries in the UK. [online] Unpublished paper. University of Edinburgh Data Library. Available at : <http://datalib.ed.ac.uk/discuk/docs/devdatlibuk.pdf>