LSE Research Online

Article (refereed)

[Luis Martinez](#) and Stuart MacDonald

Supporting local data users in the UK academic community

Supporting Local Data Users in the UK Academic Community

Stuart Macdonald and Luis Martinez describe the current situation of local data support in the UK looking back to where it all started and ahead with the new developments in this area.

Introduction

This paper will report on existing local data support infrastructures within the UK tertiary education community. It will discuss briefly 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. The paper will outline examples of local data support services, their organizational role and areas of expertise in addition to the origins of the Data Information Specialist Committee UK, DISC-UK. The paper will conclude with an exploration of future developments which may impact on the data professional.

Background
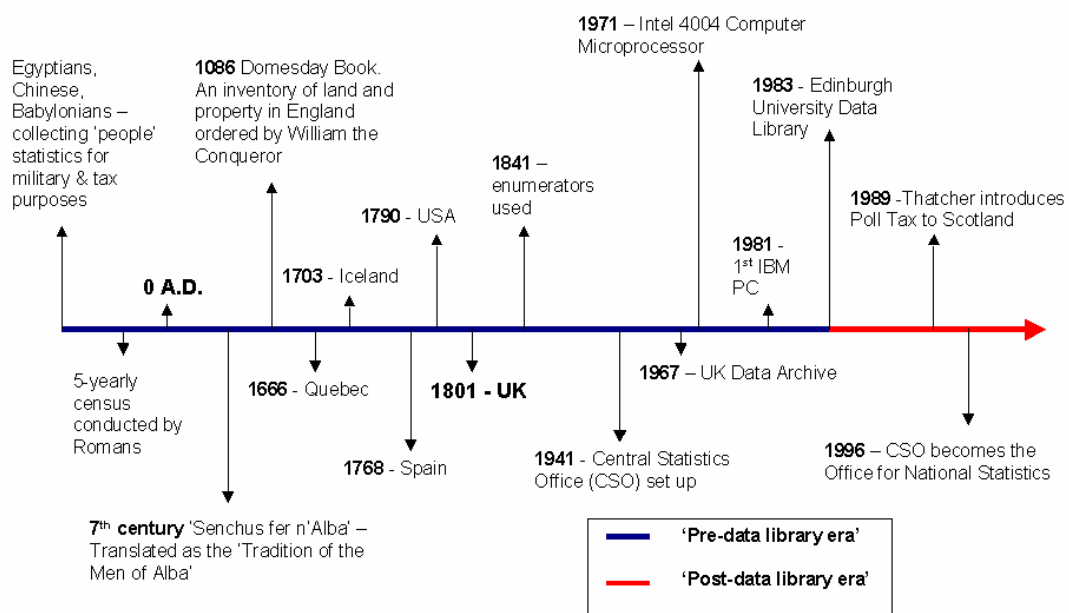
imagefile01.jpg - Figure 1: Historical data events in timeline

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). In 1086 the *Domesday Book* was commissioned by William the Conqueror for administration purposes. 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 1830s but 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 [1]). 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 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 is the microprocessor and PC in conjunction with advances in telecommunications and web technologies. 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)

> '*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*'[2]

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 Agencies and national data centres

Currently the United Kingdom has several government statistical agencies and national data centres which deal with the data collection, storage and dissemination. The number of requests for resources from these agencies and data centres are significant thus local data support staff need to have a good understanding of what is available in addition to data access conditions, format and delivery method.

Government agencies provide statistical and registration services, the Office of National Statistics (ONS) [3] is responsible for such activities in England in addition to conducting the decennial census for England and Wales. The General Register Office for Scotland [4] and the Scottish Executive are charged with similar roles in Scotland. The Northern Ireland Statistics and Research Agency (NISRA) [5] and the Statistical Directorate of the National Assembly of Wales have similar functions within the other two UK territories.

National data centres are distributed in nature providing a range of cross-disciplinary data services. They offer the UK tertiary education and research community network access to a library of data, information and research resources, in the majority of cases services are available free of charge for academic use.

The UK Data Archive (UKDA), based at the University of Essex acts a repository for the largest collection of digital data in the social sciences and humanities in the UK. Its remit includes data acquisition, preservation, dissemination and promotion of social scientific data.

EDINA [6] is hosted by Edinburgh University Data Library. Services include abstract and index bibliographic databases such as BIOSIS; spatial data services such as Digimap and UKBORDERS; multimedia services such as Education Media Online (EMOL) and the Education Image Gallery (EIG).

Based at Manchester Computing at the University of Manchester, MIMAS [7] provides spatial data services, census data via the Census Dissemination Unit and international data banks (via the Economic and Social Data Service, ESDS [9]).

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.

The Arts and Humanities Data Service (AHDS) [8] is a data centre which aids the discovery, creation and preservation of digital resources in and for research, teaching and learning in the arts and humanities.

United Kingdom local data support services

'Institutions provide support for data services in a variety of ways, these being 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.' [10]

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 services. Below there are examples of four different local data support services in the UK.

Edinburgh University Data Library

Edinburgh University Data Library (EUDL)[11] was established in 1983 and as such was the first such service in the UK.

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 [12] 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 [13] 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. A wide range of aggregated databases providing worldwide socio-economic indicators from IGOs such as IMF, OECD or EUROSTAT. Some geographic information systems, GIS, covering 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, it 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 [14] 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 5 people, therefore the data manager has the support of two systems professionals, and a part time information professional 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 [15].

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.

It was perceived however that there was required much closer collaboration in order to deal with day-to-day data issues within the UK academic community.

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 [16], meeting formally for the first time in February 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
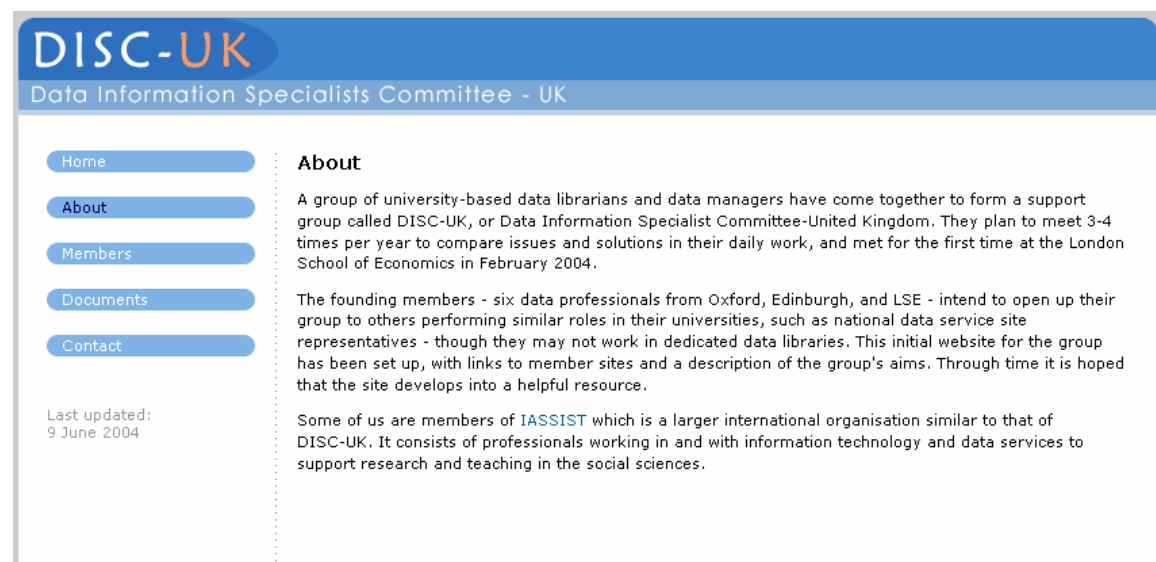
imagefile02.jpg – Figure 2: Screenshot of the DISC-UK Web site

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.

A much better response has been obtained when individuals have been approached. Universities such as Warwick, Glasgow, Birkbeck, Southampton and others have been contacted and links with those institutions have been established for future collaboration. An interesting fact deriving from the contacts 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 possibility of running 'Train the Trainers' events. The arrangement of regular meetings with those 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.

Future Developments

Web and telecommunications developments and the culture of educational technology serve as the backdrop for the future of local data support within the UK academic community. The following issues (as relevant as they are at time of publication) have to be considered on the strength of their relationship to the abovementioned factors:

- Rationalisation of large scale surveys
  For example, the General Household Survey's transition to a longitudinal survey and eventual integration into the Continuous Population Survey (along with the Labour Force Survey, Expenditure and Food Survey and the National Statistics Omnibus Survey). As has been reported earlier analysis of large research datasets at the desktop requires a different set of skills to those of data collection and discovery. By rationalising large scale government surveys and the Population Census this could pave the way for a new generation of uniform online analysis tools.

- Increased use of European survey data.
  Recently commissioned European surveys such as the European Social Survey, and the EU Survey on Income and Living Conditions (EU-SILC – preceded by the European Household Panel Survey) will become valuable resources as each wave adds value to the time-series element of analysis. In an increasingly global environment analysis at this level may take on greater relevance. Other European data sources (including macro-economic and financial data) may in turn see greater demand depending upon the UK's role within the European Community.

- Consolidation of library support.
  The UK model has been national data centres providing centralised access to and support for data resources. With the advent of search engines such as Google in addition to budgetary constraints academic institutions are reviewing 'Reference Desk' services. This has ramifications for the 'data reference interview' for example identifying those users whose data needs are not met by the above methods in addition to a whole range of user support issues including training and outreach.

- Advances in 'portable' educational technology.
  These include delivery of digital course and lecture notes to students via mobile phone, wireless and WAP technologies. In addition to opening up endless possibilities and consequent re-thinking of delivery of teaching content there is potential for data professionals to take on a more pro-active role. Will there be palmtop data analysis?

- Rising expectations of an increasingly IT literate student population.
  By the end of the decade a new generation of users totally familiar with Bluetooth, WAP, MP3s, and pixels will be participating in tertiary education. In order to use and benefit from technological advances teachers and practitioners (including data professionals) have to be at ease in this environment. This may well prove to be a bigger challenge than first anticipated.

- Centralisation versus institutionalisation.
  Centrally-funded (JISC, ESRC AHRC) data centres brokering deals with large data suppliers to supply data to the academic community means no more individually brokered subscriptions for a whole range of spatial, financial, macro-economic and survey data. This trend towards centralisation is being countered by open access publishing and institutional repositories as each institution seeks to gain greater control over their own 'assets'.

- Diversification: words, numbers, pictures, sounds.
  Diversification into support for data from other disciplines such as physics, astronomy, meteorology, genetics. It can be argued that the disciplines mentioned have been revolutionised by new media and digital technology. The 'Statistics, Librarianship and Computing' elements which comprise the current role of local data professionals (and refer predominantly to the Social Science domain) may need to be expanded to include branches of mathematics such as algebra! Another area of diverse activity is in the Humanities where there is increasing use of digital resources such as images, video, and spoken word.

- Data curation.
  Scientists and researchers across the UK generate increasingly vast amounts of digital data, with further investment in digitisation and purchase of digital content and information. The scientific record and the documentary heritage created in digital form are at risk, by technology obsolescence and by the fragility of digital media. The role of the data professional could be to assist universities rise to the challenge of 'curating' their own scholarly assets including perhaps the actual or derived datasets upon which published research papers are based.

- Professional recognition.
  The Introduction of data librarianship as a module or component of Library and Information Science qualifications. Alternatively potential certification of such a profession via professional organisations such as IASSIST.

- Grid technology and e-Social Science.
  The Grid may well prove to be the first major advance with regard to data analysis and more importantly data integration since the establishment of a few well-known statistical and analysis packages in the late sixties.
  The Seamless Access to Multiple Datasets (SAMD) project [17] is one example of integrating and streamlining the access to data repositories and analysis engines in a social science context.

- The Semantic Web.
  Tim Berners-Lee, inventor of the WWW, has been promoting for a few years now what is meant to be an extension of the current web:

  "Most of the Web's content today is designed for humans to read, not for computer programs to manipulate meaningfully. The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users." [18]

  The Semantic Web in conjunction with Web services and Grid technologies will bring new standards which can be used to represent information and to exchange knowledge. Data discovery will also be highly affected and it will become a much easier practice.

- Learning and teaching.
  There exists the potential for the involvement of data professionals in the creation, promotion and use of learning and teaching datasets including tools for using data in the classroom. An example being the X4L SDiT project [19] which aims to increase the use of real data sources held within the JISC Information Environment within the classroom

Other notable developments from a data support perspective could include:
- next generation Virtual Learning Environments,
- text and data mining technologies,
- open access publishing and digital repositories (such as the The University of California's eScholarship Repository - http://repositories.cdlib.org/escholarship/)
- Blogs,and Wikis,
- Other internet communication applications such as Podcasting

Conclusion

The tradition of local data support in the UK can be traced back to the first Data Library set up 1983. In more recent times data support professionals have come together to form a group (DISC-UK) in order to occupy the perceived gap between national data centres and users at their respective institutions. Currently there is a national commitment towards providing distributed access to digital resources and materials produced in academic environments, however at present this does not suitably cover data. A centralised approach for data repositories and disseminators dominates the data landscape in the UK with centrally-funded data centres dealing with the acquisition, preservation, support and access issues relating to data. In this context, DISC-UK members hope to play an increasingly important role establishing communication bridges between data users and data centres.

This will be set against the backdrop of the continuous and multifarious advances of technology which continue to shape data practices, with current and future data professionals having to adapt, evolve and embrace the diversity of data-related activities.

References

1. UK Data Archive site http://www.data-archive.ac.uk
2. Rice, R, 'The Internet and democratisation of access to data', blog for Social Science Week SOSIG, July 2005, www.sosig.ac.uk/socsciweek/
3. Office of National Statistics site : http://www.statistics.gov.uk
4. General Registry for Scotland site : www.gro-scotland.gov.uk/
5. Scottish Executive site : www.scotland.gov.uk/Home
6. EDINA site : http://edina/
7. MIMAS site : http://www.mimas.ac.uk
8. Arts and History Data Service site : http://ahds.ac.uk
9. Economic and Social Data Service site : www.esds.ac.uk
10. Rice, R "Providing local data support for academic data libraries", The Data Archive Bulletin 8-11, http://datalib.ed.ac.uk/discuk/docs/rice2000.pdf
11. Edinburgh University Data Library site : http://datalib.ed.ac.uk

12. Oxford University Data Library site :
    http://www.nuff.ox.ac.uk/projects/datalibrary/
13. London School of Economics Data Library site :
    http://www.lse.ac.uk/library/datlib/
14. LSE RLab Data Service site : http://rlab.lse.ac.uk/data/
15. ESRC Review of International Data Resources and Needs site :
    http://rlab.lse.ac.uk/ESRCData/default.asp
16. Data Information Specialist Committee United Kingdom, DISC-UK site :
    http://datalib.ed.ac.uk/discuk/
17. The Seamless Access to Multiple Datasets (SAMD) project site :
    http://www.sve.man.ac.uk/Research/AtoZ/SAMD
18. Tim Berners-Lee, James Hendler and Ora Lassila, "The Semantic Web" Scientific
    American, May 2001.
19. X4L SDiT project site :http://x4l.data-archive.ac.uk

Author Details

Stuart Macdonald
Assistant Data Librarian
Edinburgh University Data Library (EUDL)
Email: Stuart.Macdonald@ed.ac.uk
Web site: http://datalib.ed.ac.uk/
Luis Martinez
Data Librarian
London School of Economics Data Library
Email : L.Martinez@lse.ac.uk
Web site : http://lse.ac.uk/library/datlib/

_____

Keywords: please list in order of importance the keywords associated with the content of your article (usually up to 8). These will not be made visible in the text of your article but may be added to the embedded metadata:

1. *Data Library*

2. *Local Data Support*

3. *Survey data*

4. *Quantitative datasets*

5. *GIS*

6. *Data management*

7. *Numeric data*

8. *Statistical data*