Geography Data Unit (UKBORDERS)

Biannual Report to the ESRC Census Advisory Committee

August 2006 – July 2007

EDINA

UKBORDERS

Contact Details: James S Reid,
Data Library, Main Library,
George Square,
Edinburgh,EH8 9LJ.

Executive Summary

- **Continued successful service operation**
- **Completion of full system port to new ESRC machines**
- **Discussions with and input into Census Portal**
- **Census Portal integration work completed and launched**
- **Ongoing programme engagement and service promotion activities**

1 Introduction

This report contains a high level review of the activities of the Geography Data Unit (UKBORDERS) during the first year of the new Programme period 2006-2011 for the period August 2006 to July 2007.

Aims

The core aims of the UKBORDERS project are:

- to provide those in UK higher education involved in teaching & research with cost-effective access to a resource of digitised boundaries for thematic mapping and GIS operations; and
- to provide those in UK higher education involved in teaching & research with cost-effective access to geographical lookup directories; and
- to provide training, support and assistance to that community in the exploitation of those resources.

Objectives

The key objectives for the service for the period 2006-2011 are:

- to maintain the data facility as a robust on-line system serving up data in a variety of formats
- to explore and develop new methods and channels to assist users in exploiting the census outputs
- to support users via on-line help, individual assistance and via workshops
- to promote the ease of use of Census materials in research and teaching
- to interoperate with other Census Units to assist in the delivery of census products to the community via the new Census Portal

Start and end dates of the project

The UKBORDERS project commenced on 1 August 2006 and will run until 31 July 2011.

Staffing

The GDU team comprises Dr David Medyckyj-Scott as Unit Director. Day-to-day responsibility for the service and ongoing development is the responsibility of James Reid. Technical development, GIS analytical support and helpdesk support are performed by a team comprised (in part) of Duncan Clarkson and James Crone, Stuart MacDonald, Tony Mathys and Tim Riley . The generic EDINA helpdesk and promotion infrastructure is also utilised for support and promotion.

2 Overview of Activities

The key activity during this period has been geared towards maintaining and operating the existing service whilst in parallel porting the existing UKBORDERS service off JISC funded infrastructure to a dedicated ESRC one. Another key focus of activity has been in assisting in the establishment of the Census Portal. Work has been completed, integrating style and functional changes as, required by the Census Portal, to the newly ported UKBORDERS service. Preliminary work on a metadata service has commenced.

3 Progress

3.1 Continued Operation and Completion of Infrastructure Port

The existing UKBORDERS service has continued to be supported and maintained and necessary changes implemented where these could not await the roll out of the new service (see below). One such change was the implementation of the new Special Conditions and authentication/authorisation checks necessitated by the changes to the licensing and Terms of Use for the postcode directory resources - the agreement of a more liberal Terms of Use between ONS and ESRC required that all users were required to agree to a new set of Special Conditions for this resource. In order not to adversely impact upon the Programme timing for these changes, both the pre-existing UKBORDERS service and the newly ported version were upgraded to support the Special Conditions changes.

Routine data updates (e.g. new versions of the National Statistics Postcode Directory) have been ongoing as have low level maintenance tasks (e.g. security patches to OS or database). These have availed of the 'at-risk' period and not inconvenienced users. Under the new infrastructure (see below), outstanding application and data/interface bugs are formally maintained in a separate Bug Tracking software repository. These are routinely cleared and any new issues, added, prioritised and serviced as a matter of course.

As reported in previous progress reports, the pre-existing UKBORDERS service (in operation over the 2001-2006 period), had been developed and delivered on JISC funded hardware infrastructure. Growing demands from other JISC services sharing this infrastructure was starting to have detrimental effects upon the reliability and performance of the UKBORDERS service and a completely new infrastructure was increasingly urgently required. The new funding period (2006-2011) provided the UKBORDERS team with the opportunity and resources to develop this and a new UKBORDERS service (ostensibly a like-for-like port of the existing service but in reality a complete rebuild) was launched in April 2007. The most salient aspects of this new infrastructure are:

- An ESRC funded infrastructure (development and deployment platforms) dedicated to delivery of Census resources
- A largely Open Source software stack (Postgres+Geoserver+Mapserver) replacing expensive proprietary solutions (Oracle/LSL Gothic)
- An improved service delivery infrastructure utilising industry sanctioned practices (e.g. Service Oriented Architecture approaches, separation of development and live servers, scalable open standards based component technologies)

- 'In waiting' datasets that were acquired during the redevelopment stage of the new service have been added and release under the new service, including:
    - The 1981 Enumeration districts (from Nigel Walford, Kingston University, developed under an ESRC small grant application). The full 1981 census hierarchy was rebuilt from this and released at the same time to remove any consistencies between the new and old 1981 data.
    - Updates and additions to the historical Census data (English and Welsh data 1911-1971) and the addition of data for Ireland (developed as part of ESRC Research Methods work conducted by Ian Gregory, Queens University of Belfast).
    - N.Ireland grid square data (1Km and 10om variants)
    - Output Area Classification data for UK

Whilst end users will witness little immediate changes (and for service continuity purposes this is a major advantage), the upshot of this new architecture has been:

- A more robust, dependable service delivery platform – users should see speed gains in request processing and context map delivery is now much faster than it had been previously
- Easier maintenance and deployment – the planned development of new interfaces onto e.g. the postcode directories will as a consequence of the new infrastructure, be easier to develop, test and deploy

3.2 Census Portal activities

The GDU has actively supported the establishment of the Census Portal by:

- Providing extensive metadata to the CP team for comment and integration into the Portal;
- Providing metadata mappings (AGMAP/ISO 19115 to DDI) and advice;
- Providing a Dublin Core OAI-PMH interface for remote harvesting of the UKBORDERS datasets;
- Refinement of the CP Style Elements Guide and adoption of same across the UKBORDERS services;
- Iterative development of CP Information Sheets (version included for information in Appendix A).

3.3 Related developments

EDINA have been engaged in other projects and development activities that may have consequences for the UKBORDERS service in the medium to longer term. Amongst these initiatives was the exploration of 'grid enabling' datasets (via the JISC Grid Enabling EDINA Services – GEESE project) and the investigation of appropriate security mechanisms (e.g. Shibboleth and Grid integration) for facilitating data interoperability under the JISC e-Infrastructure programme (JISC Circular 02/06). These (ongoing) activities are particularly germane to the e-Social Science community who may wish to use the digital boundary data within Virtual Research Environments and as components of workflow engines (see also below). Findings from these activities will be fed back to the social science constituency and inform any subsequently funded activities under the Innovation Strand of the 2006-2011 Census Programme.

Given the relevance of Open Geospatial Consortia (OGC) standards and specifications both for the UKBORDERS service as well as to the wider UK geospatial community, EDINA have continued their involvement with the OGC. Under the auspices of the JISC funded SEE-GEO project, EDINA, in association with the National eScience Centre (NeSC), the National Centre for eSocial Science (NCeSS) via Leeds University, and MIMAS have been participating in the Open Geospatial Consortiums (OGC) Geolinking Interoperability Experiment (IE). Within the OGC process, the purpose of IEs is to progress candidate specifications through the standards specification programme by testing their robustness in a test environment comprising several OGC members.

The Geolinking IE is exercising the Geolinked Data Access Service (GDAS) and the GeoLinking Service (GLS) candidate specifications which endeavour to provide a standards based method of integrating attribute data and geometry data e.g. census statistics with census output geographies. OGC members from Canada, Spain and the UK are creating a number of GDAS, GLS clients and servers. Led by EDINA, our implementation involves a MIMAS GDAS on top of the census statistics, an OGC Web Feature Service (WFS) on top of UKBORDERS boundary data, and a GLS server implementation housed inside OGSA-DAI. The Use Case and GLS client is being provided by the NCeSS MoSeS node at the University of Leeds, this will provide a visual demonstration of selected MIMAS statistics being geolinked to UKBORDERS boundary data.

This work is, at time of writing, still at pre-reporting stage. Whilst aimed initially at e-Science users, the results may inform innovative service integration activities between the Data Centres at a later date. A range of OGC compliant interfaces onto the UKBORDERS data have been made available to the project to assist in the investigations.

3.4 ONS Data and Licensing Issues

EDINA had been offered data by the ONS which we would have liked to incorporate into the UKBORDERS service (see Appendix B for a listing). However, from discussions with ONS it is clear that there is an issue of residual Ordnance Survey (OS) IPR that precludes an immediate release of the data through UKBORDERS. Extensive discussions with OS and ONS have been ongoing but from email correspondence in late July 2007 it as become clear that the OS will only permit redistribution of the data under EDINAs Digimap License which rules out redistribution via UKBORDERS. Furthermore, OS policy is apparently to only permit the organisation which originally derived the data, to act as sole redistributors. EDINA had earlier sought advice and comment from the Census Advisory Committee on this issue, specifically on the extent to which the existing ESRC license covers these datasets, however, the OS position in the interim has solidified and would seem to offer little room for further negotiation. A clearer policy on how and under what conditions data derived from the ESRC census holdings can be distributed remains required and we believe that this is a policy issue for the ESRC Data Strategy to resolve.

EDINA still wish to make the data available to the academic community irrespective of distribution channel and are still in discussions with ONS as to how best to service legitimate academic access without infringing OS. This will likely see EDINA Digimap and UKBORDERS act as gatekeepers for requests which will be channelled to ONS who (in line with OS policy) can distribute OS derived data sets.

4 Internal Communication

As part of the normal operation of the unit various internal (to the census programme) meetings and discussions were held:

- All hands Census Programme Meeting at Windsor 23-24 October 2006
- Final delivery of the QA'd postcode directory data to ONS
- Discussions with CRS and CDU with respect to the Special conditions, especially in relation to the new Postcode directories;
- Email exchanges with CRS technical team on implementation updates re: Special Conditions and changes to Athens client;
- Meeting with Census Portal team to discuss requirements and timetable;
- Teleconferences with CRS on CP issues – e.g. metadata, style guide;
- 2011 Census Data Distributor Workshop, London March 20th 2007
- Email exchanges with ONS geography on redistribution of OS derived data sets
- Meetings and email exchanges with OS to clarify data redistribution issues pertaining to derived data

5 External Communication

UKBORDERS, and the Census Access Programme (CAP) more generally, benefit from the promotional activities that EDINA as a JISC National Datacentre engage in. The following summarises the conferences, workshops and other promotional activities over the reporting period at which UKBORDERS and the census programme has been promoted.

- EDINA Annual Report (Oct 2006)
  [http://edina.ac.uk/about/annrep/0506/]
- Community Report (Feb 2007)
  [http://edina.ac.uk/about/annrep/communityreport07.pdf]
- New Quick reference Guide (Oct 2006)
  [http://edina.ac.uk/ukborders/docs/UKBORDERS_Quick_Reference_Guide.pdf]
- RGS / IBG conference - 30th August 2006 - London
- eScience Workshop, Lancaster, 6-7th September
- Internet Librarian - 10th October 2006 – London
- Online Information - 30th November 2006 – London
- *Digital Gazetteer Research and Practice,* December 7-9, 2006
  Upham Hotel, Santa Barbara, California
  [http://www.ncgia.ucsb.edu/projects/nga/ncgia.html]
- JISC Conference - 13th March 2007 – Birmingham
- GISRUK 2007, Maynooth 11-13 April
- EDINA Geoforum, Leeds 15 June
- HE Academy - 3rd annual conference Harrogate 3-4 July

- AHRC ICT Methods Network Workshop on Space and Time: Methods of Geospatial Computing for Mapping the Past, e-Science Institute, Edinburgh, 23rd-24th July
- Data Citation in "Poverty, wealth and place in Britain, 1968 to 2005" - Daniel Dorling, Jan Rigby, Ben Wheeler, Dimitris Ballas, Bethan Thomas, Eldin Fahmy, David Gordon and Ruth Lupton

6 Future Developments

The early phase work concentrating on service migration and integration work with the Census Portal, has been a necessary precursor to the more user visible changes that are planned and that will form the focus for the coming year. These developments include:

- The addition of a new Metadata search service. This will exploit the full IS0 19115 / AGMAP [http://www.gogeo.ac.uk/geoPortal10/Help/Guidelines1.html] profile metadata records created for the UKBORDERS datasets and will provide users with more informative descriptive details about how datasets where sourced and their technical specifications.
- The development of new interfaces for the Postcode Directories to enable users to perform fine grained data subsetting for download.
- User commissioned case studies
- Training activities (provisionally confirmed with MIMAS for November/December 2007).
- Ongoing contributions to the development of the Census Portal.
- Ongoing programme engagement and outreach activities.

Annex A – (Summary) STATISTICS

Usage - downloads

The average monthly number of downloads for this period was over 2000 per month, which is significantly higher than the previous reporting period (1900), indicating a continuing growth in service usage (seasonal fluctuations notwithstanding).

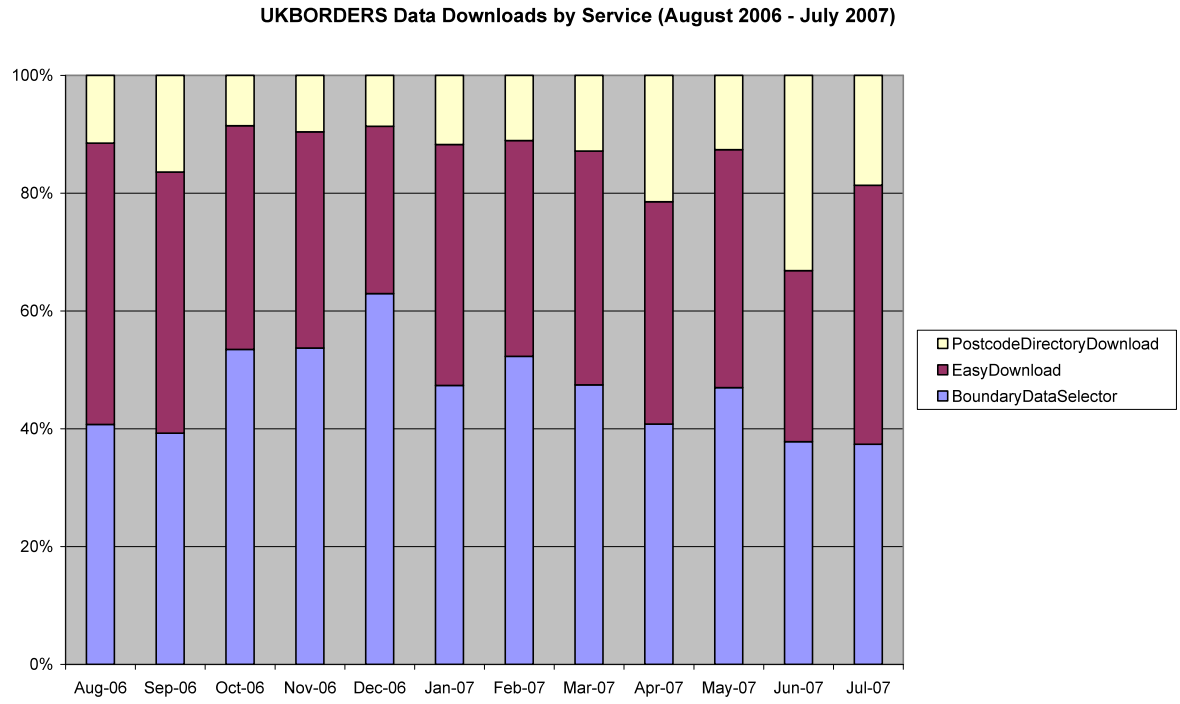
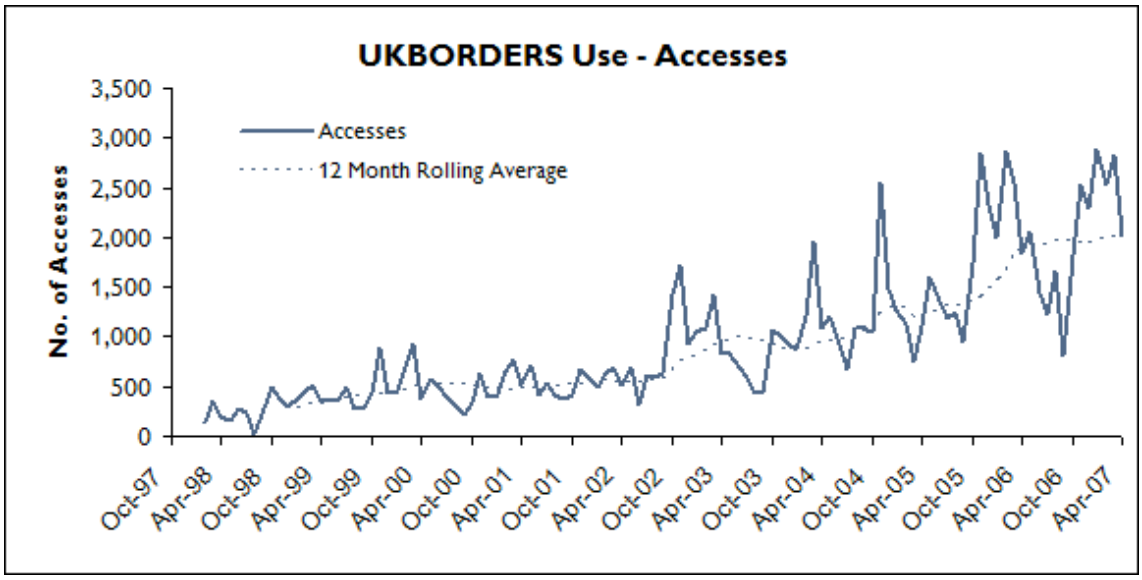


Fig. 1 Download Statistics August 2006 - July 2007.

Number of Accesses - For details on number of accesses to the Census Programme and other DSUs, please see http://www.mu.jisc.ac.uk/servicedata/census/trend/index.html. The relevant UKBORDERS data is shown below.


UKBORDERS Use - Accesses
No. of Accesses
3,500
3,000
2,500
2,000
1,500
1,000
500
0
Accesses
12 Month Rolling Average
Oct-97
Apr-98
Oct-98
Apr-99
Oct-99
Apr-00
Oct-00
Apr-01
Oct-01
Apr-02
Oct-02
Apr-03
Oct-03
Apr-04
Oct-04
Apr-05
Oct-05
Apr-06
Oct-06
Apr-07

Service	*Aug-06*	*Sep-06*	*Oct-06*	*Nov-06*	*Dec-06*	*Jan-07*	*Feb-07*	*Mar-07*	*Apr-07*	*May-07*	*June-07*	*July-07*	*Grand Total*
BoundaryDataSelector	748	354	1040	1361	1438	1375	1354	1389	807	806	430	445	11547
EasyDownload	877	400	739	930	649	1187	948	1162	746	693	330	524	9185
PostcodeDirectoryDownload	211	148	166	243	197	341	286	376	424	216	377	222	3207
Grand Total	1836	902	1945	2534	2284	2903	2588	2927	2017	1715	1137	1191	23939

Table 1. BoundaryDataSelector, EasyDownload and PostcodeDirectoryDownload numbers of downloads by month, August 2006 – July 2007.

Dataset
England CAS Wards 2001
England Output Areas 2001
England Districts 2001
England Parishes 1851
England Super Output Areas (middle) 2001
England Super Output Areas (lower) 2001
England Standard Table Wards 2001
England Counties 2001
England Unitary authorities CAS Wards 2001
England Unitary Authorities 2001
England Wards 2001
England Registration Districts 1911
England Unitary Authority Output Areas 2001
England Urban Areas 2001
England CAS Postcode Sectors 2001
England CAS Parishes 2001
Scotland CAS Postcode Sectors 2001
England Government Office Regions 2001
England Parishes 1911
Scotland Output Areas 2001

Table 2. BoundaryDataSelector Top 20 dataset downloads over the period.

Helpdesk Calls

UKBORDERS utilises the EDINA helpdesk infrastructure. The majority of calls come through as emails, a few enquiries are by letter, and the remainder are by telephone. Table 3 summarises the type of queries received by the EDINA helpdesk regarding UKBORDERS. The majority of queries are logged and have, as minimum, an initial response within 2hrs. All queries are ultimately resolved and logged as closed by the HelpDesk tracking system.

Short Description	**Number of Enquiries**
Login or Access Problem	6
Registration	6
Database Contents	83
Interface Use	28
Error Reports	15
Terms	3
Request to View	8
Forward Use	22
Other	5
Total	**176**

Table 3 Summary of calls received by helpdesk August 2006 – July 2007.

APPENDIX A
Draft UKBORDERS Information Sheet supplied to Census Portal

Structure of theme/data information at portal level

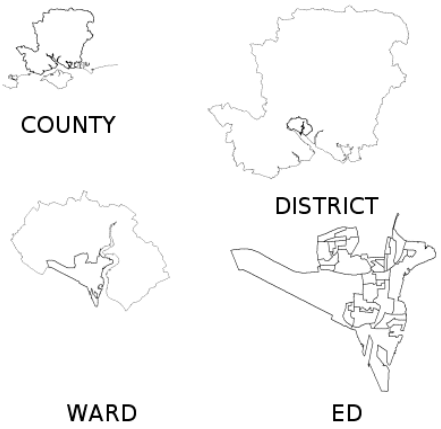
What is [theme/data]?

[Brief description of the theme/data]

Census Area Statistics provide counts of people or households for geographical areas broken down by socio-demographic characteristics such as age, gender or employment. Digital Boundary Datasets (DBDs) are a digital representation of the underlying geography of the census.

The geography of the census consists of a hierarchical subdivision of UK local government areas of various types down to sub-authority areas such as wards to lower levels created specifically for census purposes such as Enumeration Districts or Output Areas.

The geography of the 1991 census for England consisted of a 4 level hierarchy. Enumeration Districts (ED) at the lowest level nest within wards, districts and counties. A similar system was used in 1981 and most recently for the 2001 census.



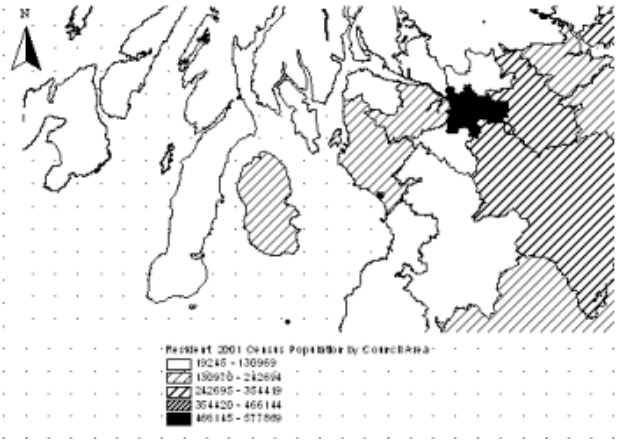
The coordinates which make up these census geographies are available as Digital Boundary Datasets.

What can [theme/data] tell you?

[In what way does it contribute to our understanding?]

Census Area Statistics contain a pointer to the geographical census areas which they relate to. By linking Census Area Statistics with the corresponding Digital Boundary Datasets for a specific census year, the census attributes can be visualised as a map. Mapping census datasets in this way allows for an exploration of the characteristics of census datasets geographically and may provide additional demographic, socio-economic and cultural insights into the census data.

A researcher may be interested in exploring patterns of housing tenure recorded in the census such as the proportion of people that live in local authority housing. By linking the census statistics to DBD`s of county boundaries or enumeration districts within a specific region and producing a shaded choropleth map of the numerical values held in the census dataset the researcher could see how housing in one region differs from another and whether there are any interesting patterns in the geographical distribution of census variables.



Using the census statistics and boundaries in a Geographical Information System (GIS) allows for spatial analysis of the census data and it's combination with other non-census geographically referenced datasets.

Digital Boundary Datasets could be used for:

Map production for research articles.
Data synthesis and development of residential neighbourhoods.
Geostatistical analysis of demographic or employment change.
Small area analysis and deprivation studies.
Health care research – incidence mapping and analysis.
Historical demographic research.

How can [theme/data] be used?

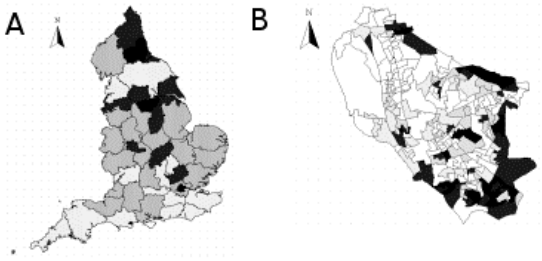
[How to approach the use of the data. Limitations. Specific issues. What manipulation is available – e.g. DSU access/software]

The geography of the decennial census is not fixed. For the same local area the output geography used in the 1971, 1981, 1991 or 2001 census may be quite different.

1991 Census Enumeration District Boundaries.

2001 Census Output Area Boundaries for the same location.

Different research questions may require mapping of the same census statistic at different scales and in different locations.



Households renting from local authority 1991 census variable mapped using county (A) and enumeration district (B) DBDs from UKBORDERS.

The UKBORDERS data support unit provides facilities by which researchers can access a full collection of Digital Boundary Datasets and supporting datasets including geographic look up tables. These datasets are available either pre-packaged or through a dynamic user driven interface permitting user-defined selection of boundaries.

The Easy Download facility provides access to the most regularly requested boundaries as ready-to-use national datasets.

The Boundary Data Selector facility allows users to select boundaries for the area required, for the census year required, and in the data format required. This flexibility allows the researcher to download census output areas for several counties or for a specific ward or district. During the boundary selection process, the chosen boundaries can be previewed over a topographic back-drop map before finally being extracted in one of several data formats for use with different GIS and mapping packages.

What resources are available?

[Outline of resources which can be accessed through the programme, with specific reference to the DSU(s) holding and providing the resources. Special restrictions on access/use.]

Many boundary types are available for England, Wales, Scotland (1840 to 2003) and Northern Ireland (2001 data only) including:

Administrative Boundaries (eg. Districts, Unitary Authorities, health boundaries etc)

Census Boundaries (2001,1991,1981 and 1971 Census boundaries for use with Census statistics)

Electoral Boundaries (e.g Wards, Parlimentary Constituencies)

Environmental Boundaries (e.g. National Parks, Urban Footprints)

Postal Boundaries (Postcode related boundaries)

Historical Boundaries (pre 1971 Census and Administrative boundaries from 1840 onwards)

Other Boundaries (e.g. Synthetic Neighbourhood Localities)

Important supporting datasets include Geographic lookup tables. These include the latest versions of the National Statistics Postcode Directory (NSPD) from National Statistics. The NSPD provides details of the locations of current and historic postcodes along with details of other geographic areas in which the postcode is located. This dataset provides a valuable means by which events or occurrences (such as disease, crimes, customer residence etc) can be allocated from a postcode to another area such as an electoral ward or health area.

Special Condtions.

Access to DBD datasets and lookup tables requires acceptance of additional special conditions. Principally these restrict use of the digital boundary datasets and associated data to teaching and research.

Further reading

[Brief bibliography covering both general introduction to the data/theme and significant specific issues]

The UK Census Geography:

Rees, Martin and Williamson (2002). The Census Data System, Wiley.

Office for National Statistics Beginner's Guide to UK Geography
http://www.statistics.gov.uk/geography/beginners_guide.asp

General Register Office for Scotland Geography
http://www.gro-scotland.gov.uk/statistics/geography/index.html

Handling spatial data and Geographical Information Systems.

Walford, N (2002). Geographical Data: Characteristics and Sources, Wiley.

Martin, D (1996) Geographic Information Systems: Socioeconomic Applications, Routledge.

Monmonier, M (1996) How to Lie with Maps, The University of Chicago Press

Longley, Goodchild, Maguire and Rhind (2001), Geographic Information Systems and Science. Wiley.

APPENDIX B
Data (potentially) Offered by ONS to UKBORDERS. Note that some of these data are already available through UKBORDERS.

Administrative
Civil Parishes/Communities (England and Wales)
Electoral Wards/Divisions (England and Wales)
Electoral Wards (Scotland)
*Electoral Wards (Northern Ireland)
Local Authority Districts (England and Wales)
Council Areas (Scotland)
*District Council Areas (Northern Ireland)
Shire Counties (England)
Government Office Regions (England)
Country (England and Wales)
Country (Scotland)
*Country (Northern Ireland)
Statistical
Statistical Wards (England and Wales) refer to guidance
Output Areas (England and Wales)
Output Areas (Scotland)
*Output Area (Northern Ireland)
SOA Lower Layer (England and Wales)
SOA Middle Layer (England and Wales)
Data zones (Scotland)
Intermediate Geography (Scotland)
*SOA Lower Layer (Northern Ireland)
Census Area Statistics Wards (England and Wales)
Census Area Statistics Wards (Scotland)
*Census Area Statistics Wards (Northern Ireland)
Standard Table Wards (England and Wales)
**Standard Table Wards (Scotland)
Electoral
European Electoral Regions (United Kingdom)
Westminster Parliamentary Constituencies (Great Britain)
*Westminster Parliamentary Constituencies (N Ireland)
National Assembly for Wales Constituencies (Wales)
National Assembly for Wales Electoral Regions (Wales)
Scottish Parliamentary Constituencies (Scotland)
Scottish Parliamentary Regions (Scotland)
Health
Strategic Health Authorities (England)
Primary Care Organisations (England)
Local Health Boards (Wales)
**Health Boards (Scotland)
IT Clusters (England)
Other
National Assembly Regional Committee areas (Wales)
National Assembly Economic Regions (Wales)
Urban Areas (England and Wales)
Urban Area Sub-Divisions (England and Wales)
Local Education Authority areas (England and Wales)
*Education and Library Board areas (Northern Ireland)
Local Learning and Skills Council areas (England)

Education and Learning Wales regions (Wales)
**Local Enterprise Company areas (Scotland)
National Parks (Great Britain)
Registration Districts (England and Wales)
Other (Cont.)
NUTS 1 (England and Wales)
NUTS 2 (England and Wales)
NUTS 3 (England and Wales)
LAU1 (England and Wales)
LAU2 (England and Wales)

*Internal customers only
**Currently unavailable