

1

Census Data Resources in the United Kingdom

Philip Rees, David Martin, and Paul Williamson

SYNOPSIS

This chapter provides the reader with an overview of the census data system, which ensures that the census is successfully administered, processed, corrected, turned into useful statistics, disseminated, and used. Census statistics—comprehensive in their coverage of a national population, the only data set to yield a rich picture of the population at small area scale—are described in terms of key features, dissemination funding and organization, the different types of output, how the Census Offices fulfil their guarantee of maintenance of individual confidentiality, the formats in which outputs are produced and the software that can be used to extract subsets and carry out analysis. The chapter then provides an outline of the actual census data resources associated with the different chapters and placed on the accompanying compact disk (CD) and concludes with an account of the logic of the structure of the whole book.

1.1 INTRODUCTION

1.1.1 Census Data Resources

The census is a device for counting people and recording their characteristics. Censuses are normally carried out once every decade for the whole population resident in a national territory. Although the census employs a fairly simple questionnaire, social science researchers are making increasingly sophisticated use of the comprehensive data on national, regional, local, and small area populations provided by the UK Censuses.

Census data resources are the raw census counts, the added value statistics, and the tools for accessing, manipulating, and visualizing the wealth of information provided. To produce the resources, a sequence of components is needed. The first consists of the people who complete census returns. The second component consists of the Census Offices that administer the data collection exercise, an extensive survey that delivers census forms to 24 to 25 million households in the United Kingdom. The Census Offices organize data entry, check and edit the input data, place the data into suitable databases, and produce many gigabytes of census output statistics and associated materials. The third component is made up of the user consortia that license use of the data from the

Census Offices and disseminate the data and added value products to user organizations. These organizations in turn distribute the data to individual users, who use the statistics for research, teaching, administrative, or commercial purposes. Individual users of the printed volumes and computer files may be researchers, teachers, students or pupils, local or central government officers, market analysts, or software developers. The academic community covering higher education institutions and research institutes makes up one customer sector that licenses most census outputs and distributes them to several thousand individual users.

Arrangements for disseminating census data from the 2001 Census will differ from procedures used with earlier censuses (1971, 1981, and 1991). In 2001, ONS announced the intention of providing free access via the Worldwide Web (referred to hereafter as the 'Web') to Key Statistics, Census Area Statistics (CAS), and Standard Tables (area statistics from the 2001 Census), in association with Neighbourhood Statistics, statistics derived from administrative sources that fill the gap between censuses in knowledge about small areas. The Cabinet Office announced in April 2001 the provision of a Web-based licensing facility for value-added use of official data, to be administered by Her Majesty's Stationery Office (HMSO) (see <http://www.hmso.gov.uk/>).

1.1.2 The Aims of the Book

Comprehensive reviews of the 1991 Census in the United Kingdom have been given in two previous books. The *1991 Census User's Guide* (Dale and Marsh, 1993) gave full details of census topics—taking, processing, and products. The *Census User's Handbook* (Openshaw, 1995a) described methods for extracting, processing, visualizing, and analysing census data, and contained a critique of what had been achieved (Openshaw, 1995b). The dissemination arrangements for UK 1991 Census data were described in detail in Chapter 2 of the Handbook (Rees, 1995a). This book builds on these accounts by detailing the additional data products and software that have been developed since original release of 1991 Census data. Although census data in the United Kingdom are produced only once in a decade, data sets evolve over time. For example, census data need to be reaggregated to new administrative policy or analysis regions; better ways of accessing the data for a wider set of users have been developed and new classifications of the data have been produced.

The principal aim of this book is to provide an account of these value-added products, created with the support of the Economic and Social Research Council (ESRC) and the Joint Information Systems Committee (JISC), that have been subsequently added to the basic census data to create a *Census Data System*. The book also provides information on plans for producing outputs from the 2001 Census, carried out on Sunday 29, April 2001. All of these new value-added products are to be found on computers connected to the Joint Academic Network (JANET), a subset of the Internet, established by JISC. They can be tracked down and used on-line via the Internet (see <http://census.ac.uk/>). However, users may be, at first, overwhelmed by the volume and variety of data provided via the ESRC/JISC Census Programme. This book provides a guide to the new derived data and value-added products associated with the census, held on different servers. It acts, therefore, as a guide to UK census data available in computer readable form. By providing links to the relevant Web pages and data locations, the intention is to future-proof the guide. Inevitably, data products and software will be enhanced over time and

the Web references (Unique Resource Locators or URLs) will change. The book editors will maintain a set of Web pages for the three years 2002, 2003, and 2004, in which the links referred to in the text will be kept up to date. The URL for this resource will be: <http://census.ac.uk/censusdatasystem/>.

Users outside the academic community may wonder what there is in this book for them. Many of the on-line resources consist of metadata (information about data sets) and basic statistics and they are made available in the public domain and are accessible without the need to register. These public domain resources have been gathered together and placed on a Compact Disk that accompanies the book, which users can use on their own personal computers. There is, of course, a great deal of detailed data for which potential users will need to obtain licences from the Census Offices. As indicated earlier, more and more census data will become available, free at the point of use, via the Web under new dissemination arrangements. There is likely to be a transition period during which 1991, 1981, and 1971 Census data sets are transferred to new licensing arrangements. New arrangements should be fully in place by the time 2001 Census outputs are released.

1.1.3 Chapter Organization

Section 1.2 describes the essential attributes of the census, and assesses its strengths and weaknesses as an information resource. Section 1.3 outlines how the licensing of data sets is funded and how dissemination is organized. Section 1.4 defines the fundamental data types used in census outputs and their internal structure. Section 1.5 briefly discusses the measures used by Census Offices to protect the confidentiality of the individual records embedded in the census outputs, measures that users need to be aware of when employing the data. Section 1.6 of the chapter gives details of output formats and the main software packages used to extract subsets of data from larger files. Section 1.7 gives guidance on the URLs that link to census data resources and on the contents of the CD. The final Section 1.8, outlines the organization of the book.

1.2 CENSUS DATA: ESSENTIAL ATTRIBUTES

The census is a simple questionnaire *survey* of the whole of the UK population that has been held every ten years since 1801, except in 1941, and with the addition of 1966 when the brave experiment of a mid-decade 'sample census' was attempted. The most recent census was held on 29 April, 2001. The census is administered separately in England and Wales, Scotland, and Northern Ireland, but the majority of the statistics published are common to all countries. The census covers a wide range of *topics* describing the characteristics of the population of the UK. Subjects covered include demography, households, families, housing, ethnicity, birthplace, migration, illness, economic status, occupation, industry, workplace, transport mode to work, cars, and language. Census data are available in computer format for a variety of geographies and spatial scales: (1) for *small areas* used in the collection of the census or in its most detailed output, (2) for *administrative areas* such as districts, counties, regions, and countries, (3) for *postal areas* such as unit postcodes, postal sectors, postal districts and postal areas, and (4) for *electoral areas* such as wards within local government areas, and Parliamentary and European constituencies.

Why are census data important? Census data describe the state of the *whole* nation, area by area. No other data set provides such comprehensive *spatial* coverage. The data are extremely relevant for *policy analysis*. They are used by government in the allocation of billions of pounds of *public expenditure*. The data are very valuable *commercially*. They are essential ingredients in marketing analysis and retail modelling. Finally, census data are important teaching resources, as students need to be trained in census use and analysis to meet the demands in public and private organizations.

It is helpful to provide a few simple definitions. A census is a complete counting of a population at a point in time. National states have long regarded such counts as providing essential information for such purposes as congressional apportionment (USA), resource allocation from central to local government (UK), planning of welfare schemes, pensions, school provision by public bodies, and the planning of delivery systems by private retail, finance, and information organizations (most developed countries).

Users of census data need to be aware of their strengths and weaknesses. Population censuses held by national statistical offices have the following *strengths*: (1) they are geographically comprehensive; (2) they represent the 'gold standard' of data collection; (3) they provide data for many geographical scales; (4) they provide objective attributes for the population; and (5) they have the confidence of the people. However, they have the following *weaknesses*: (1) they increasingly suffer from underenumeration (see Chapter 12); (2) there is debate about how to estimate and locate the missing population; (3) the data are only collected at periodic intervals; (4) the range of characteristics gathered is limited; (5) respondents make many mistakes when they fill in the census questionnaire and (6) output is modified to protect respondent confidentiality, with 'raw' microdata never being released. Work has been carried out and is ongoing, both in the Census Offices and in user communities, to overcome these deficiencies.

The UK Censuses use rather simple household questionnaires. The 1991 Census of Great Britain asked only 19 individual and 5 household questions. However, the questions were asked of 22 million households and an additional 0.8 million individuals living in communal establishments (Central Statistical Office, 1995). The 2001 Census in England, for example, asked 36 individual and 10 household questions. Because a census questionnaire is completed by the householder (and not by an interviewer), it must be simple, clear, and easy to complete.

From these simple beginnings, an enormous volume and staggering variety of information can be produced. How many cross-tabulations might be produced from the 24 questions of the 1991 Census? The number of possible cross-tabulations is the sum of all combinations of 24 variables taken 2 at a time, 3 at a time, and so on up to 24 at a time, which is 16 777 215. However, these tables can be designed in a variety of ways as each variable can be classified in numerous ways. There are, therefore, millions of possible tables that can be generated containing millions of cell counts. Most of these tables will never sensibly be requested but there are still a very large number that make sense. There is no problem in providing the software necessary to construct such tables. The Census Offices have licensed from Space-Time Pty, a suite of fast cross-tabulation programs called SUPERSTAR, SUPERCROSS, and SUPERTABLE.

So, for any census there is enormous scope for analysis and for the production of outputs, which we examine in Sections 1.4, 1.5, and 1.6. First, we discuss how census statistics are provided to the individual user.

1.3 CENSUS DATA: FUNDING AND DISSEMINATION

We now address the issue of how all the census outputs from the 1991 Census (and earlier censuses) can be accessed by the researcher in the United Kingdom with information on arrangements for the 2001 Census outputs. Two aspects are discussed: funding and agreements, and then access and support. It is perfectly possible for individual researchers or organizations to arrange the purchase or licensing of census output data directly with the appropriate UK Census Office. However, there is considerable value in making collective arrangements. The UK Census Offices have agreements covering data purchase and use with Central Government departments, with Local Governments, the Health Service, and marketing analysis firms acting as Census Agencies (value-added re-sellers of Census data with their own products). Arrangements made for supply to and distribution within the Higher Education community of census information in place for the 1991 to 2001 period are described in the following section. However, for the 2001 to 2011 period, the arrangements are likely to be much simplified. We try to indicate how they might change, although at the time of writing there is much to be negotiated.

1.3.1 Funding and Agreements for Outputs from the 1991 and Earlier Censuses

In connection with the 1981 and 1991 Censuses, two academic bodies, ESRC and JISC, have collaborated to make agreements with the UK Census Offices and fund collective purchase, for academic users, research, and teaching purposes. ESRC funds research in the social sciences in UK universities and prefers to avoid the expense of each researcher bidding separately for support to buy census data. JISC funds central computer provision for individual universities, larger servers for collective use, and JANET. JISC aims to provide the information to populate the IT systems funded for research and teaching use. In total, about £1.5 million was spent by ESRC and JISC in acquiring the 1991 Census of Population outputs for academic use.

The census in the United Kingdom is administered by three separate offices, which have recently undergone reorganization. Figure 1.1 gives full contact details for the Census Offices. The 1991 Census in England and Wales was carried out by the Office for Population Censuses and Surveys (OPCS). This merged in April 1996 with the Central Statistical Office to form the Office for National Statistics (ONS). The 1991 Census in Scotland was carried out by the General Register Office Scotland (GROS) based in Edinburgh. The Census Office Northern Ireland, which administered the 1991 Census in Northern Ireland, has since been included in 1996 within the Northern Ireland Statistics and Research Agency (NISRA). Further bodies involved in the negotiations were the ED (enumeration districts)-Line consortium and Ordnance Survey Northern Ireland (OSNI), which provide access to digital boundary data for census areas.

The key advantages for the Census Offices in negotiating collective agreements with the academic community are reduced administration and guaranteed purchase of all data sets for the whole country. The ESRC/JISC constitute the Census Offices' largest individual customer. Other advantages include the high degree of expertise available in the academic community to help check census outputs.

Licensing agreements were agreed separately for each of the separate data sets (discussed in Section 1.4). The general model for the data agreements is shown in Figure 1.2. This general model has been applied to some data sets (e.g., SARs, DBD), but others are

England and Wales Census Data

People:	Census Customer Services
Address:	Census Marketing, Office for National Statistics, Segensworth Road, Titchfield, Fareham, Hampshire PO15 5RR
URL:	http://www.statistics.gov.uk/
Email:	census.customerservices@ons.gov.uk
Tel:	01329 813800
Fax:	01329 813532

England and Wales Census Area Boundaries (1991)

People:	The ED-line Consortium
Address:	Data Management and Analysis Group, Greater London Authority, 81 Black Prince Road, London SE1 7SZ
URL:	http://www.london-research.gov.uk/dshome.htm
Email:	dsinfo@london.gov.uk
Tel:	020 7 787 5500
Fax:	020 7 787 5606

Scotland Census Data and Boundaries

People:	Census Customer Services
Address:	General Register Office for Scotland, Ladywell House, Ladywell Road, Edinburgh EH12 7TF
URL:	http://www.open.gov.uk/gros/groshome.htm or http://wood.ccta.gov.uk/grosweb/grosweb.nsf
Email:	customer@gro-scotland.gov.uk
Tel:	0131 314 4254, 0131 334 0380
Fax:	0131 314 4696

Northern Ireland Census Data

People:	Customer Services
Address:	Northern Ireland Statistics and Research Agency, McAuley House, 2-14 Castle Street, Belfast BT1 1SA
URL:	http://www.nisra.gov.uk/census/
Email:	Censusnisra@dfpni.gov.uk
Tel:	028 90 348160
Fax:	028 90 348161

Northern Ireland Census Area Boundaries:

People:	Customer Services
Address:	Ordnance Survey Northern Ireland, Colby House, Stranmillis Court, Belfast BT9 5BJ
URL:	http://www.osni.gov.uk/
Email:	osni@nics.gov.uk
Tel:	028 90 255755
Fax:	028 90 255700

(Note: Contact details may change over time but up-to-date references can be easily found using a search engine such as Google and Web sites www.open.gov.uk or www.ukonline.gov.uk.)

Figure 1.1 UK Census Offices and mapping agencies: contact information.



Figure 1.2 The general model of data agreements, 1991 Census.

simpler (e.g., SMS, SAS/LBS) or have different arrangements (LS). In total, ten agreements cover the census data sets from the 1981 and 1991 Census. Users had to fill in ten registration forms and institutions had to agree to four contracts in order to gain access to all the information. The process has been streamlined to some extent by making the individual registration forms accessible via the Web and a Web Site linking to all registration systems has been set up (see <http://census.ac.uk/>).

1.3.2 Funding and Agreements for Outputs from the 2001 Census

ESRC and JISC have set aside resources for a 2001 to 2006 Census Programme that will fund the purchase of census data sets and their support. The 2001 CAS, which the Census Offices intend to make freely available on the Web, still needs to be commissioned under the Census Acts and funded. This will be done through the *Census Access Project*, set up by ONS with a Treasury 'Invest and Save' grant in partnership with organizations representing the major census user communities. These organizations are ESRC/JISC representing the Higher and Further Education and Research Community, the Department of the Environment, Transport, and the Regions (until reorganization in 2001) representing central government departments, the Department of Health representing the National

Health Service and the Local Government Association representing member authorities. A partner representing the private sector may be added. Partners contribute a share of project costs and receive free access to the area statistics data, along with any other user. The major data sets to be produced from the 2001 Census, such as the Origin-Destination Statistics (ODS) (see Chapter 18) and the Samples of Anonymized Records (see Chapter 14), will need to be purchased by user organizations and licences negotiated.

Where census data are placed on Web sites for universal access, as ONS intend for the area statistics, then a user will need to obtain a click-use licence from HMSO, for the use of data in publications and other products. The click-use licence may follow the general model set up by HMSO or may be a tailored licence, the contents of which will be negotiated.

1.3.3 Access and Support

The Census Offices have, to date, supplied data to and interacted with a handful of people, appointed and funded by ESRC/JISC in order to disseminate census data to all academic users (the Data Units). In the 1981 Census round, the Regional Computing Centres (funded by JISC) were the data distributors along with the ESRC supported Data Archive. In the 1991 and 2001 Census rounds, national data distributors were appointed for the data sets as set out in Table 1.1, which provides full contact details for the Data Units. The *Census Dissemination Unit* (CDU), part of the Manchester Information and Associated Services (MIMAS) service at the University of Manchester, has principal responsibility for the area statistics, and has also looked after the interaction statistics and some DBD. Digital boundary data for census areas are supported by the *UKBORDERS service* within the University of Edinburgh's EDINA Data Centre. The *Census Microdata Unit* has responsibility for the Samples of Anonymized Records for the 1991 and 2001 Censuses. The *Centre for Longitudinal Studies* looks after project users who access the complex

Table 1.1 Arrangements for the dissemination of UK Census data in higher and further education.

Data unit	Data sets and functions supported
Census Dissemination Unit (CDU) (1992–2006)	
Keith Cole (Director), Justin Hayes MIMAS, Manchester Computing, University of Manchester, Oxford Road, Manchester M13 9PL web: http://Census.ac.uk/cdu/ email: info@mimas.ac.uk tel: 0161 275 6109, fax 0161 275 6040	Key Statistics UK, 2001 Small Area Statistics GB, 1981, 1991, NI 1991 Census Area Statistics UK, 2001 Local Base Statistics GB, 1991 Standard Tables UK, 2001 Special Migration Statistics GB, 1981, 1991 Special Workplace Statistics GB, 1981, 1991 Lookup Tables 1991, 2001

Table 1.1 (*continued*)

Data unit	Data sets and functions supported
Census Interaction Data Service (CIDS) (2001–2006)	
John Stillwell (Director), Paul Boyle, Oliver Duke-Williams, Zhiqiang Feng School of Geography, University of Leeds, Leeds LS2 9JT web: http://www.geog.leeds.ac.uk/ email: j.stillwell@geog.leeds.ac.uk , o.duke-williams@geog.leeds.ac.uk tel 0113 233 3315, fax 0113 233 3308	Special Migration Statistics GB, 1981, 1991 Special Workplace Statistics GB, 1981, 1991 Origin-Destination Statistics UK, 2001
UKBORDERS Service (1996–2006)	
David Medyckyj-Scott (Director), Alistair Towers EDINA, Edinburgh University Data Library, George Square, Edinburgh EH8 9LJ web: http://edina.ac.uk/ukborders/ email: ukborders@ed.ac.uk , edina@ed.ac.uk tel: 0131 650 3302, fax: 0131 650 3308	Digital Boundary Data (England & Wales) 1981, 1991 Digital Boundary Data (Scotland) 1981, 1991 Digital Boundary Data (Northern Ireland) 1991 Lookup Tables 1991, 2001
Census Microdata Unit (1992–2006)	
Angela Dale (Director), Ruth Durrell Cathie Marsh Centre for Census and Survey Research, Faculty of Economic and Social Studies, University of Manchester, Dover Street Manchester M13 9PL web: http://les.man.ac.uk/ccsr/cmu/ email: ccsr@man.ac.uk tel: 0161 275 4721, fax: 0161 275 4722	Individual Sample of Anonymized Records GB, 1991 Household Sample of Anonymized Records GB, 1991 Individual Sample of Anonymized Records NI, 1991 Household Sample of Anonymized Records NI, 1991 Individual Sample of Anonymized Records UK, 2001 Household Sample of Anonymized Records UK, 2001 Small Area Microdata UK, 2001
Longitudinal Study Support Programme (1991–2001)	
Heather Joshi (Director), Kevin Lynch The Centre for Longitudinal Studies, Institute of Education, 20 Bedford Way, London WC1 H 0AL web: http://www.cls.ioe.ac.uk/Ls/ lshomepage.htm email: hj@cls.ioe.ac.uk tel: 0171 612 6875, fax 0171 612 6880	Longitudinal Study, EW 1971-1981-1991 Longitudinal Study, EW 1971-1981-1991-2001

(continued overleaf)

Table 1.1 (*continued*)

Data unit	Data sets and functions supported
Census Registration Unit (2001–2006)	
Kevin Shurer (Director) Data Archive, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ web: http://www.data-archive.ac.uk/ email: archive@essex.ac.uk tel: 01206 872 009, fax: 01206 872 003	Making arrangements for authentication, registration and licensing of users of data sets. Promoting use of the whole suite of census information.
ESRC/JISC Census Programme Coordinator (1992–2002)	
Phil Rees School of Geography, University of Leeds, Leeds LS2 9JT web: http://www.geog.leeds.ac.uk/ <i>staff/p.rees/</i> email: p.rees@geog.leeds.ac.uk tel 0113 233 3341, fax 0113 233 3308	Negotiations with data suppliers, licensing arrangements, programme development and coordination, steering group organization (ESRC/JISC Census Advisory Committee), consultation with users, organizing workshops, preparing annual reports on the programme, disseminating knowledge about the programme

Note: GB: Great Britain, NI: Northern Ireland, UK: United Kingdom, EW: England and Wales, SC: Scotland; Contact details may change over time but up-to-date references can be easily found using a search engine or Web sites <http://www.census.ac.uk/> or <http://census.ac.uk/>

census and vital events longitudinal data, linking 1971, 1981, and 1991 Censuses and which will be extended to the 2001 Census by 2004.

The Data Units perform the following functions: (1) They register users and institutions for access to the data, though this function will be centralized in the 2001 to 2006 period; (2) They receive data sets from the Census Offices, mount them on servers, load them into appropriate software, and carry out quality checks; (3) They liaise with the Census Offices if problems are detected with the data and receive new versions; (4) They create new derived data from that supplied from the Census Offices; (5) They provide comprehensive user on-line documentation for the data sets and interface software; (6) They provide courses and seminars for users on the nature of census data set, on how to access the statistics, and on how to carry out analysis; (7) They field inquiries from users; (8) They carry out some data processing for users who lack the resources and skills to do the processing themselves, or where the resulting products can be captured for general use; (9) They receive from users the results of analysis, particularly derived data, which can be mounted for wider access; and (10) Finally, they carry out surveys of user activity and gather information about publications.

A couple of examples illustrate the value of the support that ESRC/JISC Census Programme Data Units provide. The UK Censuses have not, to date, asked a question about income. The Census Microdata Unit has been able to add to the SARs (Samples of

Anonymized Records) earning scores derived by imputing to SAR individuals the earnings of individuals matching characteristics in the New Earnings Survey and in the Labour Force Survey (Drinkwater and Leslie, 1996). In the 2001 Census, similar procedures may be applied more generally, in the absence of a question on individual income. A second example is the creation of SAS for postal districts and postal areas for general use by the CDU from postal sector SAS supplied by the Census Offices.

1.4 CENSUS DATA: TYPES AND STRUCTURES

Figure 1.3 represents the five forms in which census outputs are currently produced in the UK. These outputs are produced principally on computer media, except for statistics for larger areas, which are also published in printed volumes. The diagram provides a picture of the data type on the left-hand side and a representation of the data structure on the right-hand side. There are five data types: census microdata, CAS, census area boundaries, census lookup tables, and census interaction data, which are discussed in turn.

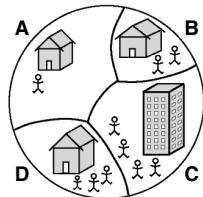
1.4.1 Census Microdata

The census collects completed questionnaires from the nation's households and its communal establishments (colleges, prisons, military bases, long-stay hospitals) and enters them into a database. From this database are produced the outputs released to researchers. Microdata constitute the data type closest in format to the master database and consist of household and individual records.

There are two microdata sets associated with the 1991 Census: the Longitudinal Study and the Samples of Anonymized Records. The *LS* is a linked sample from the 1971, 1981, and 1991 Censuses together with their associated demographic events (births, deaths, marriages, divorces), which will be linked to the 2001 Census by 2004. The sample is just over 1%, selected on the basis of four birthdays during the year. It is widely used for research on mortality because deaths can be linked to prior census characteristics, and it is also of great value for looking at social transitions of more than 10-, 20-, or 30-year intervals. The raw data are not released *per se* because it would be easy to recognize individuals in the information. Instead, researchers request tables or statistical analyses that are vetted before release. Chapter 17 describes the new information that has been added to the *LS* since 1991 and outlines the plans for adding the 2001 Census link.

The second microdata set, the *Samples of Anonymized Records (SARs)*, was extracted from the 1991 Census in both Great Britain and Northern Ireland, with modified coding of sensitive variables and without any identifying names and addresses. The microdata are released as two public use samples, a 2% sample containing individuals (including those in communal establishments rarely covered in household surveys) and a separate 1% sample of households and their members. Researchers can access these data on-line or can transfer them to their desktop/notebook computers. Note that the geography coding, regions, in the household SAR is much coarser than in the individual SAR, districts or grouped districts, because of the greater risks of disclosure of linked individuals in a household—their combined characteristics are much more likely to be unique within a subnational population. Chapter 15 outlines the SARs data sets and discusses the derived variables that have been added since 1991 and proposals for SARs from the 2001 Census.

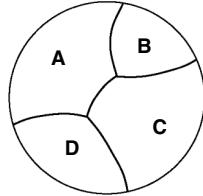
Census Microdata



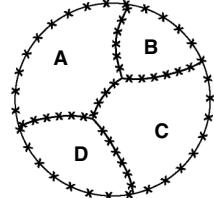
(Individual entity memberships or attributes)

Household No.	Person No.	Area	Household Size	Individual variable		
				Young	Middle Aged	Old
1	1	A	1	0	1	0
2	1	B	2	0	0	1
2	2	B	2	0	0	1
3	1	C	4	0	1	0
3	2	C	4	0	1	0
3	3	C	4	1	0	0
3	4	C	4	1	0	0
4	1	D	3	0	1	0
4	2	D	3	0	1	0
4	3	D	3	1	0	0

Census Area Statistics (counts in classes)

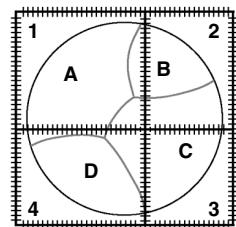


Area	Average Hsld Size	Young	Middle Aged	Old
A	1	0	1	0
B	2	0	0	2
C	4	2	2	0
D	3	1	2	0

Census Area Boundaries (x, y coordinates)

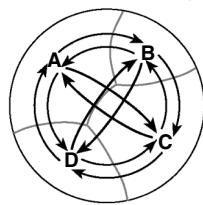
Area	Polygon coordinate pairs
A	(1, 3) (2, 6) (4, 7) (4, 4) (3, 3) (1, 3)
B	(4, 7) (6, 6) (7, 4) (4, 4) (4, 7)
C	(4, 4) (7, 4) (6, 2) (4, 1) (3, 3) (4, 4)
D	(3, 3) (4, 1) (2, 2) (1, 3) (3, 3)

Census Look-up Tables (zone indexes and proportions)



Area	New Zone	Weight
A	1	0.9
A	4	0.1
B	1	0.2
B	2	0.8
C	1	0.1
C	2	0.2
C	3	0.5
C	4	0.2
D	4	1

Census Interaction Data (counts of individuals)



Origin	Destination			
	A	B	C	D
A	1	1	1	0
B	0	1	0	0
C	0	0	1	0
D	0	1	0	1

Figure 1.3 Data types and structures for census outputs.

1.4.2 Census Area Statistics (CAS)

The second type of census data consists of counts for areas organized in tables. Figure 1.3 contains a simple example of a table that counts the people in four areas according to their age. Extraction software (e.g., CASWEB, described in Chapter 8) is designed to reconstitute the cell counts into tables with relevant row and column headings. The content and layout of the tables are agreed in advance between the UK Census Offices and the user communities. The proposals for the 2001 Census have been the subject of consultation during 1999 to 2001 and should be finalized by the end of the year (see Chapter 20 for an account of planned outputs). The area data constitute a fixed set of tables, which are said to be safe from disclosure risk. In many countries the same kind of statistics are produced (for example, Census Tract data in the United States and Canada), although the tables are usually less complex. In other countries, reliance is placed on a bespoke tabulation service for users. The problem with bespoke tabulation is that the national Census Office has to vet each table for disclosure risk in the light of previous requests. This has to be done in order to avoid the ‘differencing problem’ of subtraction of successive tables, which yield new tables subject to disclosure risk. Publication of a set of fixed tables for areas constitutes a solution to this differencing problem.

Three kinds of area data (tables) were produced from the 1991 Census round in the United Kingdom: (1) National Topic Statistics, (2) LBS, and (3) SAS.

National statistics consist of detailed tables published in Topic Reports (for example, on migration, ethnicity, and country of birth) for the countries making up the UK (England, Wales, Scotland, Northern Ireland). Topic Reports also contain data not available elsewhere (for example, the full 34-group ethnic classification for all districts in Great Britain in Table A of OPCS/GROS, 1993a). These tables cover demographic, housing, occupation, employment, ethnicity, birthplace, migration, car availability, journey to work, and health. From the 2001 Census, a similar set of topic reports are likely to be published, but with fewer tables and more commentary and analysis, with accompanying CDs holding further detailed tables in a variety of computer formats.

The second kind of area data are the Local Base Statistics (LBS) that consist of circa 20,000 counts arranged in 99 tables. LBS tables are published in computer and printed form for UK countries, regions, counties/Scottish regions, local government districts, and in computer readable form for wards (England and Wales) or postal sectors (Scotland). The areas for which LBS are available are set out in Table 1.2. There are slight differences between countries in the areas for which LBS are provided, and none were produced for Northern Ireland. The LBS tables for the smallest areas (wards in England and Wales, and pseudo-sectors in Scotland) differ from those for districts and larger areas in the disclosure control measures applied. The interior cells of most tables in the ward/sector level LBS have been randomly perturbed by the addition of 0, +1, or -1 in secret proportions in two rounds, though Rees and Duke-Williams (1994) have suggested only single blurring be applied.

The equivalent data set to be produced from the 2001 Census will be called the *Standard Tables*. These will be simpler and more numerous than the LBS tables in the 1991 Census, avoiding the concatenation of many cross-tabulations into one table. Chapter 20 provides more detailed information and references to planning documents.

The Small Area Statistics (SAS) constitute the third type of CAS. The SAS consist of about 9,000 counts arranged in 86 tables. The tables are ‘junior’ versions of those in the LBS with fewer variable categories. Table 1.3 shows that SAS were produced for a

Table 1.2 LBS from the 1991 UK Census: geographical coverage and number of spatial units.

Geographical coverage	Spatial scale	Number of units	Blurring applied
England	Wards	8985	Yes
	Districts	366	No
	Counties	46	No
	Regions	8	No
	Country	1	No
	District health authorities	186	No
	Family health service areas	90	No
	Regional health authorities	14	No
Wales	Wards	945	Yes
	Districts	37	No
	Counties	8	No
	Country	1	No
	District health authorities	9	No
	Family health service areas	8	No
Scotland	Pseudo-sectors	1003	Yes
	Districts	56	No
	Scottish regions	9	No
	Island areas	3	No
	Health board areas	15	No

Source: Denham (1993: 58, Figure 3.1)

Note: Blurring is applied to cell counts by random addition of 0, +1, -1.

very wide range of areas in each UK constituent country. For areas below district scale, the cell counts are blurred in the way described above. The SAS were produced in two tranches: primary and secondary.

Primary SAS were produced first for the main census and administrative areas, followed at varying intervals by the SAS for secondary areas. The secondary area SAS met the needs of smaller groups of users such as political scientists (SAS for Parliamentary and European constituencies) and business users (postal sector SAS in England and Wales). The equivalent data set to be produced from the 2001 Census will be termed the *CAS*. These will contain fewer counts than the 1991 SAS, in simpler tables avoiding concatenation (see Chapter 20 for more details).

1.4.3 Census Boundary Data

In order to map the statistics for geographic areas, using cartographic or geographical information system (GIS) software, researchers require digital files that give the coordinates of the area boundaries. This is the third type of data depicted in Figure 1.3. The data

Table 1.3 SAS from the 1991 UK Census: geographical coverage and number of spatial units.

Geographical coverage		Spatial scale	Number of units	Blurring
England	Primary	Enumeration districts	106 866	Yes
		Wards	8985	Yes
		Districts	366	No
		Counties	46	No
		Regions	8	No
		Country	1	No
	Secondary	Civil parishes	10 262	Yes
		Parliamentary constituencies	524	Yes
		European constituencies	38	Yes
Wales	Primary	Enumeration districts	6330	Yes
		Wards	945	Yes
		Districts	37	No
		Counties	8	No
		Country	1	No
	Secondary	Communities	866	Yes
		Parliamentary constituencies	66	Yes
		European constituencies	4	Yes
England & Wales	Secondary	Postal sectors	7892	Yes
		Urban areas	1859	Yes
Scotland	Primary	Output areas	38 255	Yes
		Pseudo-sectors	1003	Yes
		Districts	56	No
		Scottish regions	9	No
		Island areas	3	No
		Wards	1158	Yes
	Secondary	Regional electoral divisions	524	Yes
		Civil parishes	916	Yes
		Inhabited islands	144	Yes
		Parliamentary constituencies	72	Yes
		European constituencies	8	Yes
		Localities	534	Yes
Northern Ireland	Primary	Enumeration districts	3729	Yes
		Wards	566	Yes
		Districts	26	No
	Secondary	Postal sectors	400	Yes
		Health board areas	4	No
		Education board areas	5	No

Source: Denham (1993: 58, Figure 3.1) and MIMAS Web pages (<http://census.ac.uk/>)

consist of geographic coordinates (grid references or latitude, longitude) representing the boundaries. The boundary data matrix consists of rows of codes and labels for each area together with the grid coordinates describing the polygon or set of polygons that make up the area's boundaries. The exact format in which the data are organized will depend on the software being used. The UK digital boundary data are created and supplied by a variety of organizations. The academic community (through the ESRC/JISC purchase) has acquired 1991 Census boundary data for Scottish Census areas from GROS, and for English and Welsh areas from the ED-line Consortium. This consortium consists of MVA Systematica, the London Research Centre (now part of the Greater London Authority), Ordnance Survey Great Britain, and the Data Consultancy. Boundary data for Northern Ireland are supplied by OSNI direct to users under an annual licence.

1.4.4 Lookup Tables

The fourth type of data associated with the 1991 Census consists of lookup tables (LUTs), which can be used for the conversion of census data from an official source geography to a user target geography. Figure 1.3 shows that the entity in this data structure is the intersection of the source areas for which area statistics have been supplied (areas A, B, C, D) and target geography (zones 1, 2, 3, 4). The simplest lookup table is simply an index that records which other area the basic unit belongs to. However, often the census geography and the alternative geography overlap, creating intersections. These intersections form the rows of the example data matrix shown in Figure 1.3. The third column of the lookup table holds the weights for converting area statistics from the standard census geography to an alternative. In the absence of other information, the weight might be the area of the overlap. In the case of the ED/postcode directory produced by OPCS/ONS, the weights are the number of households in the intersection resident at the 1991 Census. Such a lookup table enables researchers to convert data from census areas to postal areas at a small scale. One recommendation emerging from the current consultation process between Census Offices and the academic community is that LUTs be developed more systematically and accurately in response to user demands for census data for non-standard geographies, and in order to update the standard geographies as they change between censuses. Chapter 5 gives a systematic account of the derived lookup tables generated by census users and made available for general use. A facility for generating user lookup tables from the All Fields Postcode Directory (AFPD) has been developed through an ESRC project (Yu and Simpson, 1999; Simpson and Yu, 2000 and Simpson 2001) and is available via the MIMAS service (<http://census.ac.uk>).

1.4.5 Interaction Data

The fifth data type shown on Figure 1.3 consists of interaction data, which involve the flow of people from origins to destinations. Two kinds of flow are captured by the 1991 Census in the United Kingdom: migrant flows and journey to work flows. The number of persons in each flow can be counted in a variety of categories, which are organized into tables (see Flowerdew and Green, 1993 and Rees and Duke-Williams, 1995a for details). Interaction data are most naturally represented as origin–destination tables as shown in Figure 1.3. Each cell in the table refers to an origin (the row) and a destination (column). These data will be known as Origin–Destination Statistics in the 2001 Census.

The UK Census Offices created software systems for producing migration and journey to work statistics called the Special Migration Statistics (SMS) and the Special Workplace Statistics (SWS), respectively. Users were able to submit their own definitions of origins and destinations, and flow statistics were produced. The tables and counts made available depended on spatial scale in the case of the SMS, with very few counts being provided at ward scale. The ESRC/JISC Census Programme purchased both SMS and SWS data sets for wards/(England and Wales) and sectors (Scotland) and an SMS set for districts. The Census Offices had major difficulties in producing these complex statistics and the academic community had problems in producing suitable software to handle 10 933 by 10 933 flow matrices. ESRC/JISC are funding a new Centre for Interaction Data to provide support for use of these statistics (Table 1.1). Chapters 18 and 19 discuss what has been achieved with the 1991 Census, SMS and SWS, and what is proposed for the 2001 Census.

1.5 CENSUS DATA: MEASURES TO PROTECT CONFIDENTIALITY

The UK Census Offices pledge to their respondents that their individual data will not be disclosed to any person except Census Office staff involved in the initial processing of the census. This pledge is a requirement of the Census Act of 1920 amended by the Census (Confidentiality) Act 1991 (Marsh, 1993a). Names and addresses are recorded on the census form, but will remain there unused for 100 years. Only the unit postcode part of the address is captured in digital form together with the collection area code.

The 2001 Census is being processed initially by scanning the census returns. The scanned images will be used only for checking purposes. Software will convert the ticks and characters, except for names and addresses, into digital data for editing, imputation (where responses are missing) and output processing. All records will be assigned a unique address reference incorporating the grid coordinates of a point within the property parcel. These address references will be made available to the Census Offices by the mapping agencies in England and Wales (Ordnance Survey GB to ONS) and in Northern Ireland (OSNI to NISRA). GROS will be able to generate their own address references from their own in-house GIS system.

Two strategies are adopted to protect data from the risk of disclosure (Marsh *et al.*, 1994). The first is to keep the data within the Census Offices in a secure computer environment and to release outputs, such as tables, which have been carefully vetted for unique and sensitive classifications. The second strategy makes sure that the data is safe by applying one or more of a number of protection devices and then releasing the data to the users.

The census master database is classified as too risky to release *in toto* and so remains on Census Office computers, protected by both physical security and by the absence of physical connection to computer networks. From the master database are prepared the standard outputs of published tables and published data sets. Approved users are allowed to request tables and analyses from the LS. A service is offered to users for the production of bespoke, special tables, but these must be vetted in detail before release.

Census data are made safe in a variety of ways and then released for use by researchers or the public. Table 1.4 gives details of the nine protection devices used for the different data sets. The final column refers to record modification, which is likely to be used in the 2001 Census instead of random perturbation or suppression, as it makes for outputs that are more consistent.

Table 1.4 Protection devices used with census data sets.

Dataset	Vetting	Tables	Broad coding	Samp-ling	Thresholds	Random changes	Suppression	Licence agreement	Record modification
Longitudinal data									
LS 1971–1981–1991–2001	✓		✓	✓			✓		
Microdata									
SARs 1991		✓	✓	✓	✓		✓	✓	✓
SARs 2001		✓	✓	✓	✓		✓	✓	✓
Special tabulations 1991		✓	✓	✓			✓		
Area statistics									
SAS 1991 for EDs, wards	✓	✓	✓	✓	✓		✓		
SAS 1991 for districts	✓	✓	✓	✓	✓		✓		
CAS 2001	✓	✓	✓	✓	✓		✓		
LBS 1991 for wards, sectors	✓	✓	✓	✓	✓		✓		
LBS for districts	✓	✓	✓	✓	✓		✓		
ST 2001	✓	✓	✓	✓	✓		✓		
Interaction data									
SMS 1991, wards/sectors	✓	✓	✓	✓	✓		✓		
SMS 1991 districts	✓	✓	✓	✓	✓		✓		
SWS 1991	✓	✓	✓	✓	✓		✓		
ODS 2001	✓	✓	✓	✓	✓		✓		

1.6 CENSUS DATA: OUTPUT FORMATS AND SOFTWARE

The forms in which census data are delivered to the user are under rapid evolution. For most of the history of the census the main output format has been in books of printed tables, still an important resource for research, and purchased by libraries. Typeset output is, however, expensive to produce and does not have a great market beyond research institution libraries. Therefore, Census Offices in the 1960s and 1970s offered computer printouts. In the 1980s, with the development of desktop computers, computer printouts were largely superseded by files of digital counts that were the numbers in the tables, supplied initially on magnetic tape (reels or cassettes), but increasingly on easier-to-handle media such as diskettes, DAT tapes, and CD-ROMs.

The organization of supply of data from a few sources to users at many locations is a problem faced by many manufacturing, wholesale, retail, and financial organizations. Two models of dissemination coexist: the packet and the on-line network. The former was invented in 1840 by Roland Hill as the prepaid post at uniform rates, and the latter by Alexander Graham Bell in 1876 as the telephone. There is a strong case to be made for each of these two models. The convenience of packet data dissemination has vastly improved with the development of CD and associated technologies (DVD), while the convenience of on-line access to information resources has improved considerably with the adoption of the Web standard for interfaces on the internet.

The model of Web access to central resources is likely to become more important over time: the user can keep up-to-date with new versions of data products and can learn about and use new derived data sets deposited at central data units. However, the speed of access to the Web via normal telephone lines and modems, although improving, is currently slow and relatively expensive for most users, which is why there still remains a market for packet delivery via CD or DVD. In planning a dissemination strategy for the 2001 Census, the Census Offices and user communities will need to consider carefully the balance between these two dissemination strategies.

The Census Offices delivered 1991 data to the academic community very largely as ASCII (plain text) files in fixed format or comma-delimited form, together with documentation on its meaning, usually in paper form. Essentially, the files were just streams of numbers. The only exception was the LS, which could only be released in the form of vetted tables. Three kinds of software were needed to make sense of the numbers: (1) extraction software, (2) analysis software, and (3) display software.

Extraction software allows the user to make choices about filtering a large data set, selecting the table or tables and formatting the outputs. Filtering might involve choosing a set of geographic areas or types of household; selecting tables might involve choosing the row variables and column variables, and how each was organized into categories; and formatting might involve selecting a data structure, which could be used with a spreadsheet. Extraction software is specific to the census data sets from which the extractions are being done, whereas analysis and display software can handle a wide variety of data input by the user.

To read the files of digital counts or records required specialist software written in higher level programming code. This was beyond the skills of most researchers, so there was a demand for software for extracting the data required from the large digital file and presenting it in an easy-to-understand format. At its simplest, this software simply marries the digital counts in a table with the metadata describing the data—that is, the row labels and the column headings. Local authorities in Great Britain commissioned a University

team of researchers to develop software to access the Small Area Statistics from the 1981 Census (Rhind, 1983). This software, SASPAC, was further developed to handle the 1991 Census (Davies, 1995) for UNIX, DOS, and Windows environments. Licences for SASPAC were purchased by the academic community for use at individual universities or via the MIMAS service at Manchester Computing. An alternative but similar product called C91 was developed for PC use by Powys County Council, which has also been used within the academic community. More recently, a new system for Web-based access to the SAS/LBS, CASWEB, has been developed. These issues are discussed more fully in Chapter 8.

These extraction packages provide rather limited analysis capabilities, but they do contain facilities for exporting extracted data in formats compatible with widely used spreadsheet, statistics, and mapping/GIS software. Integrated extraction and graphical presentation software has been prepared by CD-ROM designers such as Chadwyck-Healey. An academic team (Unwin and Dykes, 1996) has developed a software system called CDV (Census Data Visualizer) with JISC funding to facilitate exploratory visual analysis of integrated SAS-DBD data sets (see Chapter 7). A similar package, DESCARTES, developed by the German national GIS centre, GMD, has recently become available for use via the Web.

Table 1.5 sets out the main software packages made available by the ESRC/JISC Census Data Units. The LS is held in database package MODEL 204, not widely used and for which the ONS had to write specific interface software, FLEXTRACT, to help both in-house and LS support programme staff to produce tabulations and prepare database extracts for use with statistical packages. The SARs are held in a variety of statistical

Table 1.5 Interface and analysis software for census data sets.

Dataset	Database data format	Extraction software	Analysis/display software (examples)
LS	MODEL 204	FLEXTRACT	SPSS, SAS, MLN, GLIM
SARs	ASCII	SPSS, SAS, QUANVERT, SIR, QUICKTAB, <i>USAR</i> , NDSTAT	SPSS, SAS, GLIM, NDSTAT
KS, SAS, CAS, LBS, ST	ASCII, DATABASE	SASPAC, <i>CASWEB</i>	SPSS, SAS, GLIM, CDV, DESCARTES
SMS, SWS, ODS	ASCII, DATABASE	QUANVERT, <i>SMSTAB</i> , <i>WICID</i>	SPSS, SIM, GLIM
DBD	ASCII, ARC-INFO	<i>UKBORDERS</i> , <i>DBD91</i> , <i>CASWEB</i>	ARC-INFO, MAP-INFO, CDV, DESCARTES

Note: ASCII = American Standard Code for Information Interchange; *SMSTAB* = italics indicates academic community software.

package system file formats that cover most user needs. Users can also download the files in ASCII format along with full data dictionaries and glossaries for use with other packages. A project funded by the ESRC/JISC Census Programme has produced a very fast interactive tabulation package called USAR (Turton and Openshaw, 1995) for use on computers running the UNIX and DOS operating systems (described in Chapter 15).

Bespoke software written by projects funded by the ESRC/JISC Census Programme are important for providing access to the more complex interaction data sets and DBD. SMSTAB was originally written by Duke-Williams (see Rees and Duke-Williams, 1995a) as a program to read and check the SMS data set for wards/sectors, but proved useful as a general extraction tool. Proprietary software, QUANVERT, is also used to access the interaction statistics. SMSTAB code was adapted very quickly by the Census Dissemination Unit to access the interaction (journey to work) statistics in the SWS. An ESRC/JISC Census Programme project has developed a more user-friendly Web Interface to Census Interaction Data (WICID).

A project at the Data library of the University of Edinburgh has developed an interactive software interface called UKBORDERS, described in Chapter 6, to access the digital boundary data for Great Britain held as polygon descriptions of the smallest Census units (enumeration districts, EDs, and output areas, OAs) in ARC/INFO format. The user is able to create subsets of the data at small or larger scales. An equivalent interface called DBD91 carries out similar functions for England and Wales ED boundaries within counties (see <http://census.ac.uk>). Users carrying out interaction or DBD extractions need considerable support because the data sets are both large and complex. Chapter 6 describes the UKBORDERS interface.

1.7 WEB RESOURCES AND THE COMPACT DISK

All these data and software products are housed in computer files on various university servers. The information about the files is freely accessible via the Web, while most of the data are available on-line to registered users on university servers or government computers. It is likely that much of the 2001 Census outputs will become freely available via the Web. However, as with many other Web resources, it takes a great deal of time and knowledge to find these resources. The chapters in this book provide Unique Resource Locators (URLs), which are the addresses that users need to track down the Census Data System. URLs may change over time, so the editors of the book have established a set of Web pages. This can be found at <http://census.ac.uk/censusdatasystem/>, which will be maintained over the print life of the book. For those readers who do not have an internet connection but have a suitable disk drive on their personal computer, the editors have assembled the contents of the Web resources referenced in the book that are in the public domain (free to users) on a CD. The website and CD are referred to collectively in the following pages as the *CDS Resources*.

1.8 ORGANIZATION OF THE BOOK

The book is divided into six parts. Part I discusses the *Geography of the Census and Lookup Tables*. In the 1991 Census the small areas used for publishing data from the census coincided with the EDs used to collect the census in England, Wales, and Northern Ireland. In Scotland, the use of OAs separate from collection areas was introduced. Chapter 2,

The Debate about Census Geography, describes the discussions that took place during a two-year period, 1995 to 1997, on the question of how small areas for the 2001 Census should be designed. Chapter 3 describes the solution to be adopted in England and Wales, *Output Areas for 2001* built up from unit postcodes, which adapts and improves on the Scottish model for 1991. In Chapter 4, the authors describe the software *Tools for designing geographical zones* for a variety of purposes. Part I of the book concludes with Chapter 5, *Lookup tables and new area statistics for the 1991 Census*, which gives an account of the lookup tables available to convert 1991 Census data from the geography current at the time of the census to new geographies.

Part II continues the geography theme by reviewing *Boundary Data and Visualization*. Chapter 6 describes *Handling and accessing boundary data of the census*, while Chapter 7 introduces innovative software systems for *Visualizing census data* on maps, graphs, and cartograms.

Part III takes the reader through the most widely used census data set: the *Area statistics*. Until recently, the range of software for extracting data items from the area statistics was dominated by one software package, SASPAC, originally developed for local government use. In Chapter 8, *Disseminating census area statistics over the Web*, a new interface is described, which makes possible Web-based access to area statistics. This is the way in which many data products from the 2001 Census will be delivered. The next four chapters describe products, based on CAS, developed by the Census Offices or researchers. Chapter 9 gives an account of *Deprivation indicators* based wholly or partly on census data. Chapter 10 describes how data for irregular reporting areas has been converted to a regular grid and expressed as *Census population surfaces*. Chapter 11, on *ONS classifications and GB profiles*, shows how an enormous amount of census data for areas can be summarized into a set of intuitively meaningful *census typologies for researchers*, at a variety of spatial scales. Chapter 12 tackles the problem of *Dealing with the undercount* that mars all censuses and shows how revised estimates of population statistics after the census were prepared for general use. Chapter 13 summarizes the research carried out since the 1991 Census on approaches to *Updating the census*, which aim to overcome the decay over time in statistical relevance of census data.

Part IV of the book provides reviews of census *Microdata*. Chapter 14 describes the enhancements that have been effected to the *Microdata data from the census: samples of anonymized records*. Chapter 15 provides a guide to the software for *On-line tabulation for the SARs*. Chapter 16 covers the main features of the post-1991 *ONS Longitudinal Study*. Chapter 17 reviews methods that have been employed to generate *Synthetic microdata* needed in a variety of research applications.

Part V tackles *Interaction Data* (origin-destination flows). Chapter 18 summarizes the variety of machine-readable census *Migration data* now accessible, together with an account of the derived data that fill crucial gaps in the district flows created by data suppression, and which corrects for the undercount. The chapter also discusses how the improved migration information to be generated from the 2001 Census. Chapter 19 describes the *Workplace data*, its structure, derived additions, and methods of analysis.

Part VI, the concluding part of the book, looks forward to the *outputs from the 2001 Census*. Chapter 20 discusses the new questions for the 2001 Census and comments on the research opportunities in prospect. Chapter 21 reviews the careful plans being put in place to overcome anticipated underenumeration to produce *The one number census*. Chapter 22 outlines *An outputs strategy* for the 2001 Census, as currently planned. Users

will need guidance about how to find particular topics and variables in the census outputs, and Chapter 23 on *Metadata*, describes ways in which this guidance will be provided. The last two chapters introduce the topic of disclosure control measures, designed to provide protection of confidentiality for census data with minimum loss of information to the potential user. The final Chapter 24, introduces new methods for *Testing user-requested geographies*, the first brick in a plan to provide much more information from the 2001 Census to users on request. Readers of this sixth section should make use of the book's website to trace developments concerning the 2001 Census data.

This chapter describes how the extensive outputs of the 1991 Census of Population in the United Kingdom have been made available to researchers in the country's universities. This has been possible because of a productive partnership between government (the Census Offices), higher education (ESRC/JISC and the Universities involved in the Census Programme), business (software suppliers, digitizing companies) and local government. The academic community has performed an important quality control function, has developed critical pieces of extraction software, and has filled in critical lacunae. Its most important role, however, has been to use the data to provide reports on the state of nation and to draw socioeconomic patterns, trends, and problems to the attention of policy makers. The ESRC/JISC Census Programme has provided part of the information infrastructure needed for the academic community to play this role. Many lessons have been learnt in the process of building this information infrastructure that has informed the consultative process leading up to the UK Census in 2001 and the dissemination strategy for the new census outputs.

Box 1.1 Key web links for Chapter 1.

<p>http://www.hmso.gov.uk/</p> <p>http://census.ac.uk/</p> <p>http://census.ac.uk/censusdatasystem/</p> <p>http://census.ac.uk/cdu/</p> <p>http://www.geog.leeds.ac.uk/</p> <p>http://edina.ac.uk/ukborders/</p> <p>http://les.man.ac.uk/cscr/cmu/</p> <p>http://www.cls.ioe.ac.uk/Ls/lshomepage.html</p> <p>http://www.data-archive.ac.uk/</p> <p>http://www.geog.leeds.ac.uk/staff/p.rees/</p> <p>http://www.statistics.gov.uk/</p> <p>http://www.london-research.gov.uk/</p>	<p>A Web-based licensing facility for value-added use of official data, administered by Her Majesty's Stationery Office (HMSO)</p> <p>Web page giving access to census data resources supported by the ESRC/JISC Census Programme for academic use</p> <p>The book Web pages in which the links referenced in the text will be kept up-to-date</p> <p>CDU home page</p> <p>CIDS, home department web page</p> <p>UKBORDERS home page</p> <p>Census Microdata Unit home page</p> <p>Longitudinal Study Support Programme home page</p> <p>Census Registration Unit home department web page</p> <p>ESRC/JISC Census Programme Coordinator staff page</p> <p>National Statistics and ONS web site</p> <p>London Research Centre (now Greater London Authority) home page</p>
---	--

(continued overleaf)

<i>http://www.open.gov.uk/gros/ http://www.nisra.gov.uk/ census/start.html http://www.ordsvy.gov.uk/ http://www.osni.gov.uk/ http://www.open.gov.uk http://www.ukonline.gov.uk</i>	General Register Office Scotland home page Northern Ireland Statistics and Research Agency Census web page Ordnance Survey Great Britain Web site Ordnance Survey Northern Ireland web site UK Government web site UK Government web site
---	--