Chapter 2
Access to Census Interaction Data

Adam Dennett
University of Leeds, UK

John Stillwell
University of Leeds, UK

Oliver Duke-Williams
University of Leeds, UK

ABSTRACT

This chapter is concerned with how users gain access to census interaction data. The authors outline a brief history of electronic access to interaction data sources and identify a number of issues and problems which led to the development of the Web-based Interface to Census Interaction Data (WICID). After presenting a number of practical and technical prerequisites for WICID, the authors explain in detail the architecture underpinning the system and the importance of the metadata framework for both initial successful implementation and ongoing maintenance and flexibility. Much of this chapter is devoted to explaining the basic query building and data extraction processes from a user perspective and further guidance relating to some of WICID’s less basic but no less useful features, is provided.

INTRODUCTION

We live in times when gaining access to electronic secondary data is increasingly quick and easy. In the UK, there are a range of avenues for those seeking access to go down, at the end of which a vast quantity of data are available. A researcher wishing to gain access to the latest Census Key Statistics, for example, needs only log on to the ONS neighbourhood statistics website (http://neighbourhood.statistics.gov.uk/), select the desired area(s) and select whichever variables are required before simply clicking a button which will deliver the selected data over the internet to his/her desktop. Access to these data are unrestricted and freely available to anyone with an internet-enabled computer. It is not just standard census data that are available with no more effort than a few clicks of a button. The UK Data Archive (http://www.data-archive.ac.uk/) provides access to a colossal range of data for academic users (geographically referenced or otherwise) from an increasingly wide range of social
surveys and other sources. Advances in computing technology, coupled with the proliferation of digital data storage and developments in network capacity and speed mean that data, which, not more than twenty years ago were difficult to gain access to let alone process with ease, are now so freely available that many users take their availability for granted.

The development of technologies that have brought us to the current zenith of rapid access to numerous sources of secondary data has been gradual. Particular technological developments in data access have been driven and shaped by the demands of users, the nature of the datasets themselves and the concerns of the data custodians relating both to who accesses the data and exactly how much detail users are able to access. For example, as outlined in the Neighbourhood Statistics Programme evaluation report (ONS, 2006), the development of the current online system came about as the result of the poor availability of small area key statistics that could be used in local government policy and planning. Data that were available were from disparate sources and often only accessible after payment to custodians was made. With the desire to increase the availability of data as the main driver, the particular technological developments which have taken place to help provide access to these data have been guided by a better understanding of the particular requirements of end-users as the system developed. Interaction data can be viewed as a particular sub-set of statistical information with unique characteristics; characteristics which have resulted in different approaches to access. Accessibility and ease of access drove the development of the Neighbourhood Statistics programme; it could be argued that due to the unique nature of interaction data and its more specialist utility and resultant smaller user-base, the route to a more universal access has been very different to that of neighbourhood and other statistics required more readily by governmental decision makers. One of the problems in the past with using origin-destination data has been that the available software has been difficult to use, and this can certainly be attributed to the smaller user-base for the data. This is especially true in comparison to access software for the area statistics. To provide context for the latest access developments for interaction data, a brief review of software packages that were used to access origin-destination data prior to the Web-based Interface to Census Interaction Data (WICID – the system developed for use by the academic community in conjunction with the outputs from the 2001, 1991 and 1981 Censuses) will be given.

THE ACCESS PROBLEM

There now exist some excellent web-based interfaces to data sets from the Census of Population and other sources. The neighbourhood statistics website already mentioned provides access to the most recent census data counts. Another example is CASWEB (http://casweb.mimas.ac.uk/), providing access to UK Census Area Statistics and related information to academic users from the most recent 2001 Census, but also the 1991, 1981 and 1971 Censuses in the form of counts of persons and households for various geographical units. CASWEB has been developed by the Census Dissemination Unit (CDU) which forms part of the Manchester University-based MIMAS service and its development, like that of WICID, has been funded as part of the Census Programme by the Economic and Social Research Council (ESRC) and the Joint Information Systems Committee (JISC). Another example is NomisWeb (http://www.nomisweb.co.uk/), giving online access to the most up-to-date and detailed labour market data produced from official sources and run by the University of Durham. Both CASWEB and NomisWeb are similar in that they are primarily based on stock variables; that is, counts or values relating to specific geographical areas which