Using NESSTAR Light to Access Major UK Datasets Online

**Karen Dennison and
Louise Corti
University of Essex**

kdenn@essex.ac.uk

corti@essex.ac.uk

NESSTAR is a set of internet-based tools to provide an enhanced online service for data providers and data users that gives easier and quicker access to data and information about that data. NESSTAR has been funded by the European Commission to the sum of £2 million under the 4th and 5th Framework IT Programmes.

Many government departments, research institutes, companies, and others collect social and economic data. Many of these datasets are available for re-use, particularly for research and education, typically via data archives and libraries. Currently, NESSTAR users can access this type of data from a number of European data archives, including the 3 main partners in the project: UK Data Archive, Norwegian Social Science Data Services and the Danish Data Archive. The number and type of data and data providers is continually expanding and NESSTAR is also being used by other data communities as an Intranet service. This article focuses upon the use of NESSTAR in serving the needs of users of data archives.

A common scenario for a user of a data archive is that s/he locates a dataset of possible interest through an archive's online catalogue through which s/he can also read about the survey design and contents and possibly view the codebook and associated user guides. However, at this stage, the user does not usually have immediate access to the data or to any frequencies or percentages. S/he may then need to contact the archive to ascertain its suitability, order the data and then wait to receive it, for example on a CD. NESSTAR radically facilitates the data access process, allowing users immediate access to frequencies and percentages and allowing registered users to perform exploratory online analyses (at no cost) and then download the data for offline analysis immediately.

Online, NESSTAR users can:

- search for datasets
- browse dataset documentation (study description, codebook, user guide)
- analyse data (tabulations, regressions, scatterplots, descriptive statistics)
- visualise data (produce graphs, bar charts, pie charts)
- download data for offline analysis (including subsets) in a number of formats (including SPSS, SAS and STATA)
- save outputs and bookmark searches and analyses

Getting Started - Access to Data

Browsing the dataset documentation is normally unrestricted. The data owners often restrict use of the data. To apply to access such data, users will need to authenticate themselves. For example, to access UK restricted data, users will have to register for a UK Data Archive account at http://www.data-archive.ac.uk. This will provide them with a username and password that they will be prompted to enter when attempting to access restricted data. An access control unit will then check the access conditions against the profile of the user and allow or deny access immediately.

http://www.nesstar.org

Searching for Datasets

There are three different ways to search for data:

Simple Search
A freetext search is carried out within the text associated with a dataset, i.e. the catalogue and data description.

Field Search
Using the field search you can narrow your search to one or more fields. This is similar to the way searches are done in libraries.

Advanced Search
The advanced search is for specialist users. Like the field search you can target your search to special fields, and a combination of fields. This gives more flexibility than the field search and you can search directly on the variables and question text.

At the UK Data Archive, survey data now available online include selected years from the major data series, some of which are produced by the UK Office for National Statistics:

- General Household Survey
- Quarterly Labour Force Survey
- Health Survey for England
- British Election Surveys (also available to unregistered users)
- British Social Attitudes Surveys
- ONS Omnibus Survey

together with some other individual popular UK based datasets such as the Scottish Social Attitudes Survey. This online service will be extended to other data collections within the UKDA on a continuous basis, with more datasets being mounted on the system over the coming months.

Searching at the UK Data Archive can be bypassed by linking straight in to the NESSTAR catalogue records from pages on our web site dedicated to helping users find datasets from our major series (http://www.data-archive.ac.uk/findingData/majorStudies.asp).

Browsing dataset documentation

From a list of search results, users can choose to 'Browse' dataset documentation.

The datasets in NESSTAR are documented according to the Document Type Definition (DTD) for the Data Documentation Initiative (DDI - see the DDI web site at http://www.icpsr.umich.edu/DDI/), a standard for

documenting survey data sets. The information is structured into major fields such as:

Study Description, including for example the title, abstract and keywords;

Data Files Description providing information about each data file;

Variable Description providing information about each variable (including frequencies and percentages) and variable groups in the dataset;

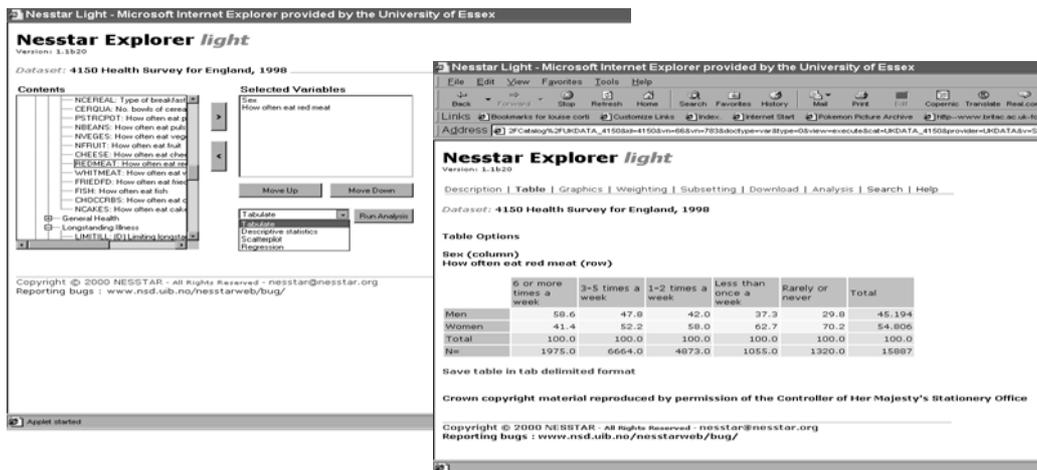Other Documentation providing links to User Guides or references to articles.

Analysing data

Users can choose to 'Explore' the data. This is an ideal way for researchers to ascertain whether a dataset contains the type of data they are looking for, allowing them to explore the contents before proceeding to obtain the data for offline analysis. This is particularly useful for those users (non-academic in the case of the UK Data Archive) who may be charged at the point of download.

Users can choose to produce frequency counts, crosstabulations (with up to 4 variables) descriptive statistics, scatterplots and regressions (with up to nine independent variables). Depending upon the type of analysis chosen, there are various options for further manipulation of the data. For example, for tabulation, you can choose between percentages and raw numbers and select missing categories to include in the analysis.

Users can also choose to subset by case before conducting an analysis. As an example, a user might choose to include only females, unemployed females or unemployed females aged between 20 and 30 years in the analysis. Analyses can also be weighted using predefined weighting variables. Weights allow the user to produce estimates at a national level and / or to compensate for over or under-representation of certain respondents in the survey.

NESSTAR is able to handle hierarchical datasets and merge files on the fly to produce the results of analyses using variables from files at different hierarchical levels. For example, a dataset may contain one file holding data about each individual in the household and another file containing summary information about each household. A user would be able to use variables from both files in an online analysis.

© UKDA, July 2001

Visualising data

Having conducted an analysis, a simple click allows the user to produce a number of graphical representations, depending on the type of analysis that has been carried out. For example, with tabulations you can produce simple and stacked bar charts and pie charts.
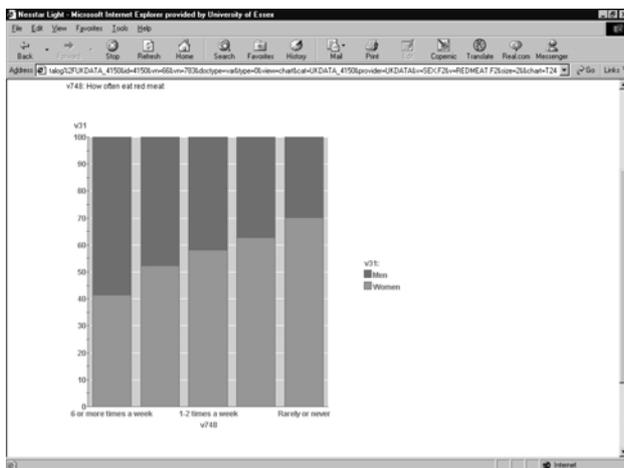
© UKDA, July 2001

Downloading data

The data can be downloaded in the following formats:

- SPSS system file
- SPSS portable file
- NSDstat
- Statistica
- Stata 6
- Data interchange format (DIF)
- Dbase 3
- SAS

Users can also download subsets. This means they can select the variables and / or cases they wish to include in the downloaded file.

Saving outputs and bookmarks
Tables can be saved in tab delimited format. Tables, graphs and documentation can also be saved as html.

Users can bookmark searches and analyses to return to in a later session by saving the URLs of the appropriate web pages. These URLs can then be sent to colleagues and also embedded in online reports. This means that another user could go directly to a table or graph that had been produced.

> **More about NESSTAR light...**To find out more about NESSTAR light at the UK Data Archive, see http://www.data-archive.ac.uk/usingData/nesstar.asp. You can read more about the R&D aspects of NESSTAR at: http://www.nesstar.org.   We also have a User Support team who can answer any specific enquiries at archive-userservices@essex.ac.uk.