Data Sans Frontiers!
Removing the barriers to data use

Celia Russell – MIMAS
Luis Martinez – LSE Data Library
Introduction

In this talk we are going to identify the barriers that prevent people from using data in their social science research and teaching and then describe strategies to overcome these barriers.

In 2001, Robin Rice investigated barriers to the use of data in higher education.

ESDS International
Number of users and dataset release dates

Source: Athens user log March 2006

10/07/2006
Barriers to data use

- Problems locating data
- Lack of awareness of available data
- Lack of local user support
- Different service providers, multiple registration systems and interfaces
- Data not delivered via web
- The data users want is not the data available
- Lack of documentation
- Prohibitive data license costs
- Lack of appropriate data handling skills
- Lack of integrated user support service
Locating data

• Metadata from national data centres shared across institutions
  – National data centres to be ‘localised’ (Rice et al., 2001)

• Metadata standard for Social Sciences Data: Data Documentation Initiative (DDI)
  • Five sections
    – Document description
    – Study description
    – Files description
    – Data description
    – Other related materials
# Locating data 2

<table>
<thead>
<tr>
<th>DC Element</th>
<th>DDI Element</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>&lt;title&gt; 2.1.1.1</td>
<td>EUROBAROMETER 10 -- OCTOBER - NOVEMBER, 1973</td>
</tr>
<tr>
<td>Creator</td>
<td>&lt;AuthEntity&gt; 2.1.2.1</td>
<td>RABIER, JACQUES-RENE / INGLEHART, RONALD</td>
</tr>
<tr>
<td>Subject</td>
<td>&lt;keyword&gt; 2.2.1.1</td>
<td>Common Market, European Community, Europe, France, Belgium, Denmark, Ireland, Great Britain, Italy, Luxembourg.</td>
</tr>
<tr>
<td>Description</td>
<td>&lt;abstract&gt; 2.2.2</td>
<td>EUROBAROMETER 10 WAS CONDUCTED BY JACQUES-RENE RABIER, SPECIAL ADVISER TO THE COMMISSION OF THE EUROPEAN COMMUNITIES, AND BY RONALD INGLEHART OF THE UNIVERSITY OF MICHIGAN. THIS STUDY IS PART OF AN ONGOING PROGRAM OF PUBLIC OPINION RESEARCH.</td>
</tr>
<tr>
<td>Publisher</td>
<td>&lt;producer&gt; 2.1.3.1</td>
<td>Inter-university Consortium for Political and Social Research</td>
</tr>
<tr>
<td>Contributor</td>
<td>&lt;othcno&gt; 2.1.2.2</td>
<td>Other Identification/Acknowledgements – Data Collection</td>
</tr>
<tr>
<td>Date</td>
<td>&lt;prodDate&gt; 2.1.3.3</td>
<td>March 20, 2000</td>
</tr>
<tr>
<td>Type</td>
<td>&lt;dataKind&gt; 2.2.3.10</td>
<td>Survey data</td>
</tr>
<tr>
<td>Format</td>
<td>&lt;fileType&gt; 3.1.6</td>
<td>Type of File</td>
</tr>
<tr>
<td>Identifier</td>
<td>&lt;IDNo&gt; 2.1.1.5</td>
<td>7728</td>
</tr>
<tr>
<td>Source</td>
<td>&lt;sources&gt; 2.3.1.8</td>
<td>Sources – Used for Data Collection</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td>&lt;othrSticky&gt; 2.5</td>
<td>The Eurobarometer survey series, launched by the Commission of the European Community in 1974 and continuing today</td>
</tr>
<tr>
<td>Coverage</td>
<td>&lt;timePnt&gt; 2.2.3.1</td>
<td>October 1978-November 1978</td>
</tr>
<tr>
<td></td>
<td>&lt;collDate&gt; 2.2.3.2</td>
<td>Nov 1978</td>
</tr>
<tr>
<td></td>
<td>&lt;nation&gt; 2.2.3.3</td>
<td>France, Belgium, Netherlands, Germany, Italy, Luxembourg, Denmark, Ireland, Great Britain</td>
</tr>
<tr>
<td></td>
<td>&lt;geogCover&gt; 2.2.3.4</td>
<td>Nine countries forming the European Community in 1978: France, Germany, Great Britain, Italy, the Netherlands, Belgium, Denmark, Ireland and Luxembourg.</td>
</tr>
<tr>
<td>Rights</td>
<td>&lt;copyright&gt; 2.1.3.2</td>
<td>Copyright the University of Michigan, 1980</td>
</tr>
</tbody>
</table>

Table 1. Dublin Core and DDI associations
Awareness of available data

- Data Centres and Libraries coordinated promotion to raise awareness
- Training courses, workshops and conferences
  - Links to local data support staff / teaching staff
- Other channels of communication
  - Mailing list
  - RSS feeds
• Libraries need “someone” knowledgeable about data resources and support
  – Understanding of local user needs
  – Data sources
    • National Data Archives (ESDS, ICPSR…)
    • Government Statistical Offices (ONS)
    • Commercial datasets
    • Free data
  – Other support areas in the institution
    • Statistical software support/ Research methodology/ Survey design
  – Support groups and associations (DISC-UK, IASSIST)
IASSIST/DISC-UK

• International Association of Social Science Information and Service Technology (IASSIST)
  – Professionals working in an with IT and data services
  – Blog/Mailing list
  – Annual Conference

• Data Information Specialist Committee UK (DISC-UK)
  – Group of university based data librarians and data managers
    • Edinburgh/Oxford/LSE and LSE RLab
  – Raise awareness of value of local data support
  – Foster understanding between data users and providers
  – Share info and resources among members
Accessible local support

- Locating and raising awareness
- Trouble shooting
- Hands-on training
- One to one help
- Easing tedious registration
- Streamline ordering
- Manage institutional subscriptions
• Conducting data reference interview
  – Active listening and prompting for info
  – Determine level of enquiry
  – Other aspects
    • Unit of analysis (country, region, household)
    • Key variables needed (wages, debt, education, health, etc)
    • Time span (weekly/monthly/yearly)
    • Cross-sectional or longitudinal
Reducing the number of interfaces, service providers and registrations systems

Authentication methods

• Global usernames and passwords (Athens)
• Shibboleth in UK, Holland Germany, France
Reducing the number of interfaces, service providers and registrations systems 2

- Data aggregators, national data services and data libraries have a role in making data available on web
- This means that data previously available as ftp files, CD-Roms or on mainframes can be accessed through web interfaces
- This substantially increases data use
- Use of common web delivery software eg Nesstar or beyond 20/20
- Introduction of online registrations with click agreements
Grid technologies

- originally developed in the context of large-scale physical science projects
- are more efficient than web technologies in handling data
- provide single sign-on meaning users can cross search datasets and perform cross analyses of multiple datasets from different providers
- Grant easier access to high performance computing facilities
- Run on existing internet hardware
Current web architecture
Grid Architecture

Diagram shows a network of connections between datasets and HPCs with a single sign-on point for users.
Providing access to the right data

- Need to identify future potential demands for data
- Liaison with research councils, research community and government to identify research strategies and priorities
- Coordinated data acquisition strategies
- User consultation eg surveys, literature reviews, use elsewhere
- Provision of up to date data
- Data wishes pages
  - Authoritative source
  - widely cited
  - consistent data domains,
  - long time series
  - high quality data
  - up to date
ESDS International host 30 databases from many different agencies. For each database we provide a consistent and comprehensive descriptions of the database content and documentation. It’s the only place on the web that pulls together this information in a consistent way and has proved to be an immensely usable resource, used by people from around 100 countries.
Reducing data license costs

• Publically funded data should be publically available
• Ideally free to end user, either through institutional subscriptions or consortia approach
• Consortia approaches mean greater equality of access between universities, encouraging collaborative research
UK consortium

- In the UK, provision of data for UK social sciences is comparatively centralised
- This enables us to negotiate nation-wide agreements with international governmental agencies like the World Bank and the IMF
- The national licences each cost around the equivalent of 15 institutional licences but allow all UK universities to access the data for free
- Before the introduction of this licensing agreement, around 10 UK universities used the data. Now virtually around 150 universities access the databanks.
- We can help you create a similar data service in your country through iGoStats http://www.igostats.org/
Improving statistical capacities, data handling skills

- Teaching the teachers
- Teaching datasets and learning objects with identifiable learning goals (course level, pedagogical aims eg data analysis, substantive)
- Commissioned, expert content
- Role of reusable leaning object depositories – jorum
- Use of storytelling and examples
Summary and conclusions

• Metadata for all to make data easier to find
• Closer collaboration between local support and national data centres
• Partnerships for teaching and learning
• Sufficient resources allocated to data support at the institutional level
• Data on the web and free at point of use
• Adoption of new technologies for data access and data management