Data Sources for Research: Discovery, Access and Use

John Southall
Bodleian Data Librarian

Michaelmas Term
November 2021
Sources of Data at Oxford

We’re considering data as a secondary resource
  ◦ Already gathered
  ◦ Available for re-use
  ◦ Quantitative or qualitative
  ◦ Not data you're collecting yourself
  ◦ Some disciplines see this as novel others as routine

Described as databases, datasets or archives
Understanding the language (1)

Database or dataset
- A collection of information; numerical, textual etc.
- Structured, e.g. Excel spreadsheets, Word etc.
- Indexed
- Derived from a particular research project, source or programme
- Often downloaded or accessed
- May be updated
- Intended for academic use
Databases – some examples

• Factiva: Business news sourced from Newspapers and business reports
• Proquest Historical newspapers
• Amadeus: Pan-European company information derived from annual reports
• Proquest Dissertations and Theses
• … and Bibliographic Databases generally
Understanding the language (2)

A data archive

- Hosts a large collection of datasets or even databases
- Structured, indexed and documented
- Curated by an organisation; cared for and future-proofed
- Sometimes discipline based
- Sometimes based around a specific institution or organisation
- Content is ordered or downloaded
- Static and unchanging
Data archives - examples

Also known as repositories

- ORA and ORA-Data: http://ox.libguides.com/home/researchers
- UK Data Service: http://ukdataservice.ac.uk/
- ICPSR: https://www.icpsr.umich.edu
- East Midland Oral History Archive: https://www.le.ac.uk/emoha/catalogue.html
- Research Data Repository Registry: http://www.re3data.org/
UK Data Service

• International macro data from Eurostat, OECD, UN, World Bank, IMF, IEA.
• Survey and UK funded social science data
• Oral History and Qualitative data
• Advice on finding data deposits on the UK Data Service site [https://ukdataservice.ac.uk/find-data/](https://ukdataservice.ac.uk/find-data/)

Can browse and read documentation before ordering data
Understanding the language (3)

- Open Data portals
  - Recent innovation
  - Usually concentrate on statistical data
  - Make data 'Open' publicly available via web
  - May seem to duplicate what’s in archives and databases

- Question – do they differ?
Open Data portals - examples

Also called statistical data portals

- UNdata: http://data.un.org/
- World Bank data: http://data.worldbank.org/
- U.S. Census: http://www.census.gov/
Other data portals

May be project specific…


• The Oxford Roman Economy Database: http://www.romaneconomy.ox.ac.uk/

• Charles Booth’s London: https://booth.lse.ac.uk/
Exercises

Take the time to look at some examples

- OECD iLibrary
- UK Data Service
Browse/Search

◦ The sheer amount being made available creates problems of discovery
◦ Interested in seeing what’s available?
  ◦ Browse for them

◦ Increasingly datasets are cited
  ◦ e.g. Iterate (political terrorism database), British Household Panel Survey, Bloomberg Professional

◦ Look for them using public/university searches
Search/Browse

- Search engines include data alongside books, journal articles etc.
  - SOLO, Databases A-Z
- Encouraged by publishers like Proquest
  - Creating bibliographic databases
- Often available via subscription
  - Asked for payment? Enter via SOLO etc with single sign-on
What to use?

- **SOLO**
  - Integrates with searches for traditional resources

- **Databases A-Z**
  - Tighter focus on e-resources

- **Subject Guides (LibGuides)**
  - Curated by your subject librarian

- **Bodleian Data Library**
  - Additional content on data management, analysis
  - [https://www2.bodleian.ox.ac.uk/data](https://www2.bodleian.ox.ac.uk/data)

Same resources but different ‘display’ windows
Data and Statistics for Social Sciences: Home

Purpose of this guide

This guide is intended primarily for researchers and students studying Social Sciences at the University of Oxford. It provides information on where to find general and subject-specific data and statistics, how to manage data, and what tools are available for data analysis. The guide introduces resources available only to the Bodleian card holders as well as freely available to everyone.

Social Science Data in the Bodleian

- **Bodleian Libraries**
  Provide access to an extensive collection of databases via

- **Databases A-Z**
  A browsable list of databases available at Oxford

- **Nuffield College Library**
  The largest social science library outside of the Bodleian Libraries, and holds sociology collections relating to the interests of College members since 1937.

  ![Bodleian Data Library](image)

  **Bodleian Data Library**
  BDL provides advice and guidance on both finding data and statistics to support your research and managing your own research data.

  *more...*

  Contact the **Bodleian Social Science Library** if you wish to ask a question or seek more help.

Subject and Data Librarian

![John Southall](image)

**John Southall**

**Email Me**

**Contact:**
Bodleian Social Science Library
Manor Road, Oxford, OX1 3UQ
+44 (0)1865 2-85936

**Subjects:**
Economics, Economics - General
Data Archiving and Sharing
Find out about securely storing and sharing your research data.

More...

News: UKDS: Online seminar: Data sources for research students - Released 8th September by UKDS

Data Analysis
A selection of tools for making best use of the data you have collected for your research.

Data Obligations
What you need to know about using data at Oxford. Your responsibilities and how to comply with data use terms and conditions.

Research Data Management
Find out more about effectively organising data as part of your research.
Any questions?

◦ Points of clarification?
◦ Comments?
Optional exercise

- For ten minutes try some searches
- Identify three resources relevant to you
- Tell us why these resources are relevant to you, and say how you found them
- Are they public or subscription based
Caveats & annoyances

◦ Search engines collate available resources and present them in a unified way

◦ But how you access, view and use them can vary
  ◦ Be clear what is being made available
  ◦ Be clear in what format
  ◦ Check for usage restrictions
  ◦ Time delays
Common examples

- Registration suggested or required
  - Subscription database environment
  - Common with external data archives
  - Providing authentication
- Restrictions
  - Sensitivity dealt with by making access harder
  - Archived data discoverable but not accessible
More caveats

- Restrictions on how you use it
  - For analysis only - cannot be quoted
  - Cannot be shared
  - Supplied tabulations cannot be altered
- Make sure you are aware of these!
- May need to analyse data within their browser/interface
- Data often available from multiple sources but with differing documentation and support

On 20 October, World Statistics Day is being celebrated for the second time. This event highlights the importance of official statistics in our society. Official statistics help decision makers develop informed policies that impact millions of people. Improved data sources, sound statistical... more

Key figures

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Government gross debt</td>
<td></td>
<td></td>
<td></td>
<td>86.8%</td>
</tr>
<tr>
<td>Government deficit &amp; surplus</td>
<td>-3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government expenditure</td>
<td>48.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP growth rate</td>
<td></td>
<td></td>
<td></td>
<td>1.4%</td>
</tr>
</tbody>
</table>

General government gross debt

Percentage of gross domestic product (GDP)
Restricted/Microdata

- Very sensitive or includes personal identifiers
  - E.g. economic data, population data
- Only available pending application approval
  - From e.g. Eurostat, Statistics Canada, UKDS Secure lab, ICPSR
  - Open data portals will not supply it
- Support making applications via Bodleian Data Library
More Examples

Restrictions on how you use it

- For analysis only cannot be quoted
- Cannot be shared
- Supplied tabulations cannot be altered

Make sure you are aware of these!

May need to analyse data within their browser

25
Documenting Data

- Good research practice
- Data **AND** contextual documentation
- Make use of citation tools
- An aid in interpreting/preserving your own data
- Key element of Reproducible Research, transparency and Open Science
Thank you

- Follow up this session with your own subject librarian (or any other subject specialist)
- See Subject Guides: https://libguides.bodleian.ox.ac.uk/home/subjects

- Bodleian Data Librarian: John Southall
  john.southall@bodleian.ox.ac.uk