Data Sources for Research: Discovery, Access and Use

John Southall
Bodleian Data Librarian

Cathy Scutt
Bodleian Education Librarian

Trinity Term
(May 2021)
Research 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
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 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.

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 – does it differ?
Open data portals - examples

Also 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

Project specific?


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

- Charles Booth’s London: https://booth.lse.ac.uk/
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
  ◦ Search for them using public / university search engines
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 LibGuides**
  - Curated by your subject librarian
- **Bodleian Data Library**
  - Additional content on data management, analysis

Same resources but different shop windows
This is just a selection of the datasets related to the topic, while the Bodleian Data Library provides more extensive list of resources.

Internet WorldStats

Global Internet Usage Based on Time of Day

Mobile Coverage Explorer

Mobile Coverage Explorer
Mobile Coverage Explorer is a representation of the area covered by mobile networks around the world. It is supplied as polygon data which has been created from submissions made directly to Collins Bartholomew or the GSMA from network operators who provide roaming detail for inclusion in the GSMA’s Roaming and Coverage Maps web application. This data is supplemented with coverage created from OpenCellID, the world’s largest open database of cell towers.

Online news resources

BBC Monitoring International Reports
Full text from 1979-present available via LexisLibrary. Provides access to broadcasts from 120 countries, and emphasises information from counties for which other news sources are not readily available.

FBIS - Foreign Broadcast Information Service Daily Reports 1975-1996

UK Data Service

Information and communication
Information and communication technologies are reshaping the world, transforming the way in which we communicate, govern, work, manage crises, do business and spend our free time. These changes are, in turn, driving new policy development to address the societal impacts of digital technologies. The UK Data Service holds data on a wide range of ICT topics, such as mobile communication, telework, social-media, radio, television, internet use, mass media, infrastructure, security and trust. This key research area helps us to understand how the fast evolving world of digital technologies impacts people’s lives and communities, changes the way we interact, affects our cultural experiences and could shape our future society and economy.

Go to Discover to search and browse data collections, series, case studies, and ESRC outputs on Media, communication and language.

WPT database

World Association of Newspapers and News Publishers

World Press Trends (WPT) Database
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.
Optional exercise

Questions?

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 may be available from multiple sources but with differing documentation and support
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
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.

24
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 library specialist)
- See LibGuides: https://libguides.bodleian.ox.ac.uk/home/subjects

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