Data Sources for Research: Discovery, Access and Use

John Southall
Bodleian Data Librarian
Subject Consultant for Economics and Sociology

Catherine Scutt
Subject Consultant for Education

Hilary Term 2020
Research data at Oxford

– We’re considering data as a secondary resource
  • i.e. not data you are collecting yourself
  • Already gathered data available for re-use
  • Some disciplines see this as novel others as routine
  • Quantitative or qualitative

– Described as databases, datasets or archives
Understanding the language

– A database or dataset
  – A collection of information; numerical, textual etc.
  – Structured – excel spreadsheets, Word et al
  – Indexed
  – Derived from a particular research project, source or programme
  – Often downloaded or accessed
  – May be updated
  – Intended for academic use
Be aware of options

Apparently Wikipedia doesn't count as a literature review

Use the best tools
Databases

- 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

A data archive

- 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
- National archives – Institutional repositories
Data Archives and repositories

- ORA and ORA-Data: [http://ox.libguides.com/home/researchers](http://ox.libguides.com/home/researchers)
- UK Data Service: [http://ukdataservice.ac.uk/](http://ukdataservice.ac.uk/)
- ICPSR: [https://www.icpsr.umich.edu](https://www.icpsr.umich.edu)
- CESSDA (European Social Science Data Archives): [http://cessda.net/](http://cessda.net/)
- Research Data Repository Registry: [http://www.re3data.org/](http://www.re3data.org/)
Welcome to the UK Data Service

Your resource for quality social research data

A unified portal to access to data from ESS, Census Programme, Secure Data Service and others.

International macro data from Eurostat, OECD, UN, World Bank, IMF, IEA.

Survey and UK funded social science data.
Advice on finding data deposits on the UK Data Service site.

Can browse and read documentation before ordering data.
Open data portals

• Recent innovation
• Usually concentrate on statistics
• Making data publicly available via web browsers
• May seem to duplicate what’s in archives and databases
• Q – does it differ?
Statistical data portals

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

- Project specific?
- The Oxford Roman Economy Database: http://www.romaneconomy.ox.ac.uk/
- Charles Booth’s London: https://booth.lse.ac.uk/
Exercises

– 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 and 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
  • Additional requests to sign in?
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
<table>
<thead>
<tr>
<th>Subjects A-F</th>
<th>Subjects G-O</th>
<th>Subjects P-Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>African Studies</td>
<td>Genetics</td>
<td>Patients</td>
</tr>
<tr>
<td>Ancient Near East</td>
<td>Geography and the Environment</td>
<td>Pathology</td>
</tr>
<tr>
<td>Anthropology</td>
<td>German Language and Literature</td>
<td>Pharmacology</td>
</tr>
<tr>
<td>Archaeology</td>
<td>Greek (Modern)</td>
<td>Philosophy</td>
</tr>
<tr>
<td>Art and Architecture</td>
<td>Hebraica, Judaica, and Semitics</td>
<td>Physics</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>History</td>
<td>Physiology</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>History of Science and Medicine</td>
<td>Plant Sciences</td>
</tr>
<tr>
<td>Biology</td>
<td>Internet Studies</td>
<td>Plant Taxonomy</td>
</tr>
<tr>
<td>Biomedical Sciences</td>
<td>Italian Literature and Language</td>
<td>Politics and International Relations</td>
</tr>
<tr>
<td>Business</td>
<td>Japan</td>
<td>Portuguese and Brazilian Language</td>
</tr>
<tr>
<td>Cartography</td>
<td>Korea</td>
<td>Literature</td>
</tr>
<tr>
<td>Celtic</td>
<td>Latin American Studies</td>
<td>Printed Ephemera</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Law</td>
<td>Psychology</td>
</tr>
<tr>
<td>China and Inner Asia</td>
<td>Linguistics</td>
<td>Public Policy</td>
</tr>
<tr>
<td>Classics and Ancient History</td>
<td>Management</td>
<td>Refugee Studies</td>
</tr>
<tr>
<td>Commonwealth Studies</td>
<td>Maps</td>
<td>Russian and Slavonic Language and</td>
</tr>
<tr>
<td>Computer Science</td>
<td>Materials</td>
<td>Literature</td>
</tr>
<tr>
<td>Continuing Education Subject Guides</td>
<td>Maps</td>
<td>Science</td>
</tr>
<tr>
<td>Criminology</td>
<td>Mathematics</td>
<td>Slavonic, Central and Eastern</td>
</tr>
<tr>
<td>Development Studies</td>
<td>Medicine</td>
<td>European Studies</td>
</tr>
<tr>
<td>Dutch</td>
<td>Middle East &amp; Islamic Studies</td>
<td>Social Policy</td>
</tr>
</tbody>
</table>

**Contact us**

**General Enquiries**
Email: reader.services@bodleian.ox.ac.uk
Telephone: 01865 277162
Post: Main Enquiry Desk, Bodleian Library,
Broad Street, Oxford OX1 3BG

**Chat to a Librarian**
Live Help is offline. Please use our Ask A Librarian service.
Live Help is usually staffed during normal working hours (Monday-Friday, 9am-5pm).

**Text a Librarian**
You can also text message the "chat to a librarian" service on 02030 953910 (Mon-Fri 9:00-17:00)

**Subject Specific Enquiries**
If you have a subject enquiry please contact your subject librarian.

**Report a problem with LibGuides**
Please contact users@bodleian.ox.ac.uk

Follows us on...
Pew Research Center

**Internet, Science & Tech**

The Pew Research Center's Internet Project offers access to raw data sets from their research which are made available as single compressed archive files (.zip file).

All uses of this data should reference the Pew Research Center as the source of the data and acknowledge that the Pew Research bears no responsibility for interpretations presented or conclusions reached based on analysis of the data. Pew Research is interested in learning about other ways that scholars use our data. If you publish something based on our data, please let us know by sending us an email. Individuals will need to fill out a brief registration before downloading data. Questions concerning the data sets may be directed to the Pew Research Center.

- A Majority of Teens Have Experienced Some Form of Cyberbullying
- How Teens and Parents Navigate Screen Time and Device Distractions
- Most Americans Accept Genetic Engineering of Animals That Benefits Human Health, but Many Oppose Other Uses
- August 16, 2018

**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 impinges 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**

Vibrid Press Trends (WPT) Database

A dynamic continuously updated database includes the summary tables, also allowing users to produce their own, customized reports, which can be exported to PDF or to Excel format.


Available in PDF format via Weblearn.

**European Audiovisual Observatory**

**Yearbook**

The Yearbook provides a wide range of statistical data on the four main industry branches of the audiovisual sector in Europe: the film industry; television; video/DVD; on-demand audiovisual services. The Yearbook provides three sub-sets of data, delivered as it becomes available. Detailed tables and figures; country profiles; key trends reports. The figures are available in Excel tables and graphs, and the text is available in PDF version.

Username/password are required for this resource: see Weblearn for details.
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.
Exercise

– Introduction complete
  • Questions?

– For ten minutes try some searches
  • Identify three resources relevant to your discipline – why?
  • 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
– Are they seen as ‘valid data’ in your discipline
– How you access, view and use them can vary
  • Be clear what is being made available
  • Be clear in what format
  • Are there usage restrictions
– Extra hurdles
  • 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
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
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
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

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