

iSkills: Data Sources for Research

Michaelmas Term

Feb 2025



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



John Pilbeam
Web/Digital Officer
john.pilbeam@sbs.ox.ac.uk

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
- Updated over time
- Intended for academic or commercial use?

Databases - some examples

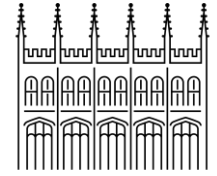
- Factiva: Business news sourced from Newspapers and business reports
 - Proquest Historical newspapers: 11 individual paper archives (1830 onwards) e.g. Times of India, Washington Post, Chinese newspapers, The Observer
 - Reaxys: database containing millions of chemical substances and reactions, with associated chemical and physical properties and literature references
 - Digital National Security Online: Declassified US policy documents
 - Roper iPoll: US Opinion Polls
 - Proquest Dissertations and Theses
- ... and Bibliographic Databases generally

Understanding the language (2)



A data archive / repository

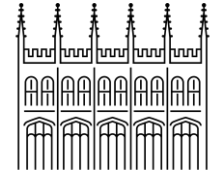
- Hosts a large collection of datasets or even databases
- Structured, indexed and documented
- Curated by an organisation; cared for and future-proofed
- Sometimes discipline based
- Sometimes based around a specific institution or organisation
- Content is ordered or downloaded
- Static and unchanging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



BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

Data archives - examples

- Governmental or Non-Governmental
- ORA: <https://ora.ox.ac.uk/about>
- UK Data Service: <http://ukdataservice.ac.uk/>
- UK Biobank: <https://www.ukbiobank.ac.uk/>
- ICPSR: <https://www.icpsr.umich.edu>
- Environmental Data Service (NERC):
<https://www.ukri.org/councils/nerc/facilities-and-resources/>
- British Oceanographic Ocean Data:
<https://www.le.ac.uk/emoha/catalogue.html>
- Research Data Repository Registry: <http://www.re3data.org/>



BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

UK Data Service

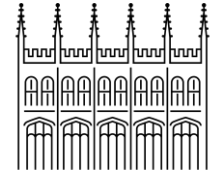
- International macro data from Eurostat, OECD, UN, World Bank, IMF, IEA.
- Survey and UK funded (e.g. ESRC) social science data
- Oral History and Qualitative data
- Advice on finding data deposits on the UK Data Service site <https://ukdataservice.ac.uk/find-data/>

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?



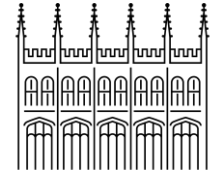
BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

Open Data portals - examples

Also called statistical data portals

- UNdata: <http://data.un.org/>
- World Bank data: <http://data.worldbank.org/>
- UK National Statistics: <http://www.statistics.gov.uk/>
- U.S. Census: <http://www.census.gov/>
- National Bureau of Statistics of China
<http://www.stats.gov.cn/english/>
- Hathi Trust Digital Library (repository of literary content from research libraries including content digitized via Google Books and the Internet Archive) <https://www.hathitrust.org/>





BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

Other data portals

Sometimes project specific...

- Oxford Poverty & Human Development Initiative:

<https://ophi.org.uk>

- The Oxford Roman Economy Database:

<http://www.romaneconomy.ox.ac.uk/>

- OII Research Program on AI & Work:

<https://www.oii.ox.ac.uk/research/projects/research-programme-on-ai-work/#news>

- Our World in Data:

<https://ourworldindata.org>

Browsing for data

- 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
 - Becomes a literature search objective

Focused searches for data

- 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 tool 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
 - Data and statistics for Social Sciences
<https://libguides.bodleian.ox.ac.uk/ssdata>
- Bodleian Data Service
 - Additional content on data management, analysis
 - <https://www.bodleian.ox.ac.uk/collections-and-resources/data>

Same resources but different ‘shop windows’

Subject and Research Guides

Subjects A-Z

Libraries

Getting Started

Specialist materials

For researchers

Reference Management

Workshops

Disabled readers

Contacts

Subjects A-F

- [African Studies](#)
- [Ancient Near Eastern Studies](#)
- [Anthropology](#)
- [Archaeology](#)
- [Art and Architecture](#)
- [Biochemistry](#)
- [Bioinformatics](#)
- [Biology](#)
- [Business](#)
- [Caribbean Studies](#)
- [Cartography](#)
- [Celtic](#)
- [Chemistry](#)
- [Chinese Studies](#)
- [Classics and Ancient History](#)
- [Commonwealth Studies](#)
- [Computer Science](#)
- [Continuing Education Subject Guides](#)
- [Criminology](#)
- [Development Studies](#)
- [Digital Humanities](#)
- [Dutch](#)
- [Earth Sciences](#)
- [ebooks](#)

Subjects G-O

- [Genetics](#)
- [Geography and the Environment](#)
- [German Language and Literature](#)
- [Greek](#)
- [Hebraica, Judaica, and Semitics](#)
- [History](#)
- [History of Science and Medicine](#)
- [Humanities](#)
- [Humanities \(Digital\)](#)
- [Human Sciences](#)
- [Internet Studies](#)
- [Italian Literature and Language](#)
- [Japanese Studies](#)
- [Korean Studies](#)
- [Latin American Studies](#)
- [Law](#)
- [Linguistics](#)
- [Management](#)
- [Maps](#)
- [Materials](#)
- [Mathematics](#)

Subjects P-Z

- [Patents](#)
- [Pathology](#)
- [Pharmacology](#)
- [Philosophy](#)
- [Physics](#)
- [Physiology](#)
- [Plant Sciences](#)
- [Plant Taxonomy](#)
- [Politics and International Relations](#)
- [Portuguese and Brazilian Language and Literature](#)
- [Printed Ephemera](#)
- [Psychology](#)
- [Refugee and Forced Migration Studies](#)
- [Russian and Slavonic Language and Literature](#)
- [Science](#)
- [Slavonic, Central and Eastern European Studies](#)
- [Social Policy](#)
- [Social Science Data and Statistics](#)
- [Social Sciences](#)
- [Socio-Legal Studies](#)

Contact us

Subject Specific Enquiries

If you have a subject enquiry please contact your [subject librarian](#).

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:


Bodleian Live Chat





Library staff are online to help. [Chat here](#)

Report a problem with LibGuides

[Bodleian Libraries](#) / [Oxford LibGuides](#)

/ [Research support, tools and tips for accessing materials and information about our special collections](#)

/ [Data and Statistics for Social Sciences](#) / [Home](#)

Data and Statistics for Social Sciences: Home

[Home](#)
[Macrodata](#)
[Economics & Finance](#)
[Politics & IR](#)
[Society](#)
[Media, IT and Communication](#)
[Public Opinion](#)
[Digital archives](#)
[Data analysis tools & training](#)
[Citation & sharing](#)
[Managing research data](#)

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.

How to find

Macrodata

- [International Data](#)
- [Country level data](#)

Economics & Finance

Politics & IR

Society

- [Education](#)
- [Energy & Environment](#)
- [Health & Social care](#)
- [Science & Technology](#)

Media, IT & Communication

Public Opinion: Polls & Surveys

Digital archives

- [Data Archives &](#)

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

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 .

News

UK Data Service introductory training series: Autumn 2021

Subject and Data Librarian



John Southall

Contact:

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



Subjects:

[Economics, Economics - General](#)

Caveats

- Search engines collate available resources and present them in a superficially unified way
- But how you access, view and use them can vary
 - Be clear what is being made available
 - Be clear in what format

Some common examples

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

Some other caveats

- Restrictions on how you use it
 - For analysis only - cannot be quoted
 - Use only with suppliers tools (e.g TDM)
 - 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

Restricted data

- Information collected at the individual i.e. micro-level
- Very sensitive and confidential
- May include direct personal identifiers
- Or require special licensing or access approval
 - E.g. economic data, population data, literature still in copyright, large corpora via API
- Only available pending strict application approval
 - From e.g. Eurostat, Statistics Canada, UKDS Secure lab, ICPSR, Hathi Trust Research Centre
 - Open data portals will not supply it
- Support making applications via Bodleian Data Librarian

Data

Explore data



Database

Access the complete database



Statistical themes

Get background information on the data



Stats finder A - Z

Search alphabetical list of topics



Experimental statistics

Discover new sources and methods



Data visualisations

Explore data in an easy way



Education corner

Find materials for teachers and students

Advanced tools & services



Web services / API



Bulk data download



Geographical
information (GISCO)



Microdata

HathiTrust Research Center

Overview	
Member List	
Services & Programs	▲
Accessible Text Request Service	
Copyright Review Program	
Emergency Temporary Access Service	
HathiTrust Research Center	
Metadata Quality Improvement Program	
Shared Print Program	
U.S. Federal Documents Program	
Resources for Librarians	▼
Contribute Content	
Get Involved	▼

The HathiTrust Research Center focuses on the combined strengths of the collection to spark curiosity: what can we learn from so many books? [HTRC](#) can help you and your users find the right questions and help answer them using text and data mining. Find easy entry points to digital humanities tools, as the entire collection is open for text and data mining explorations. Members may also access copyrighted titles for research in more advanced computing environments. Whether for classroom use, individual research projects, or ongoing library instruction, HTRC tools illuminate the landscape of human knowledge.

Jump to Section

Text and Data Mining Tools ▼

Select

Benefits of the HathiTrust Research Center

Support Digital Humanities at All Levels

- Novice digital humanities researchers can dig into their questions using simple tools without any programming experience.
- Advanced researchers can access robust datasets and apply for powerful computing environments to support demanding workloads. Advanced researchers can access robust datasets and apply for access to secure computing environments to support advanced text analysis.
- Help your library build functional knowledge in digital humanities methods and tools with hands-on workshops and training.
- Research support is provided to HathiTrust members through a periodic competitive awards program called [Advanced Collaborative Support](#).

Explore the Collection at Scale

- Investigate 18+ million volumes and trillions of words with powerful, behind-the-scenes computation.
- Subjects range from Archeology to Zoology with titles in more than 400 languages.
- Discover new lines of research or new findings to answer long-standing questions.

Documenting data

- Keep a track/log of resources consulted
- Good research practice
- Data **AND** contextual documentation
- Make use of citation tools
- Key element of Reproducible Research, transparency and Open Science

Thank you



- Follow up this session...
- See Subject Guides:
<https://libguides.bodleian.ox.ac.uk/home/subjects>
- Data Research Support
john.southall@bodleian.ox.ac.uk
john.pilbeam@sbs.ox.ac.uk