

# "LIBRARY SUPPORT FOR YOUR MPHIL THESIS: USING BODLEIAN COLLECTIONS FOR RESEARCH AND DATA MINING"

JOHN SOUTHALL

BODLEIAN DATA LIBRARIAN

& ECONOMICS SUBJECT CONSULTANT

JUNE 2026

# COLLECTIONS FOR RESEARCH

- Bodleian Subscriptions give access to **authoritative** sources for;
  - Text based journal and book content
  - Text based newspaper and media content
  - Bibliographic or Systematic Review platforms for text
  - Statistical data from national statistical agencies and archives
  - Statistical data from business intelligence and analysis vendors
- Authoritative in what sense?
  - Citable
  - Reproducible
  - Cleared / licenced for academic use

# OPTIONS FOR SEARCHING & USE

- SOLO – main search engine for all Bodleian collections
- Database A-Z – dedicated to database platforms
- Research guides of various sorts
  - Economics libguide - <https://libguides.bodleian.ox.ac.uk/economics>
  - Business Libguide - <https://libguides.bodleian.ox.ac.uk/business>
- **Curated** suggestions of journal, statistics and grey literature content
  - Abstracts
  - Full text

SEARCH CRITERIA    BARCODE SEARCH    ^

Search Profile: Search Everything ▾

Search filters

Any field ▾ contains ▾ Enter a search term

AND ▾ Any field ▾ contains ▾ Enter a search term

+ ADD A NEW LINE    ↻ CLEAR

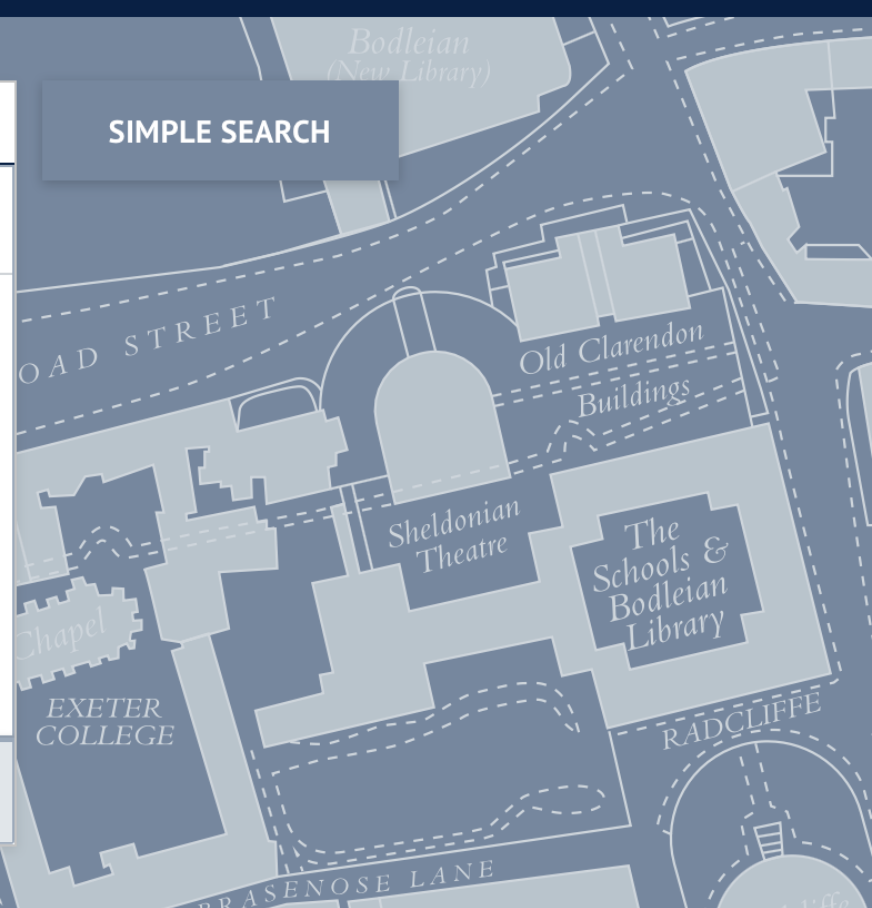
Material Type  
All items ▾

Language  
Any language ▾

Start Date:  
Day ▾ Month ▾ Start Year

End Date:  
Day ▾ Month ▾ End Year

→ Any field contains Enter a search term AND Any field contains Enter a search term    🔍 SEARCH



### Announcements

✔ An interim solution providing access to electronic Legal Deposit content collected before October 2023 is available on Bodleian Libraries reading room PCs via the SOLO Browser Electronic

### Getting Help

If you have any questions or need help making the most of the extensive electronic resources available to you then please:

### Useful links

- > [Bodleian Libraries Home Page](#)
- > [Subject & Research Guides](#)
- > [Find A Library](#)
- > [Journal Search | BrowZine](#)

## Sort & Filter Results

Include results without full-text access

Sort by Relevance ▾

Show only: ^

Available Online

Open Access

Peer-Reviewed Articles 

Physical Resources

Library ▾

Resource Type ▾

Author/Creator ▾

Publication Date ▾

Journal Title ▾

Subject ▾

Genre ▾

Language ▾

Collection ▾

Publisher ▾

New Records ▾

Data Source ▾

SELECT ALL 0 selected PAGE 1 1-10 of 6,376 Results ▾



1



BOOK

**Microeconomics**

Dobson, Stephen. | Miller, Ellen M. | Maddala, G. S.  
c1995 | London : McGraw-Hill Book Co | xx, 378 p. : ill. ; 25 cm

[Find & Request](#) >



2



MULTIPLE VERSIONS

**Microeconomics.**

Estrin, Saul.; Laidler, David E. W.;Dietrich, Michael, 1953-;Estrin, Saul.

 4 versions of this record exist. [See all versions](#) >

3



MULTIPLE VERSIONS

**Microeconomics**

Eaton, Buford Curtis, 1943-; Eaton, Diane F.;Allen, Douglas W. (Douglas Ward), 1960-;Devlin, Rose Anne, 1956-;Gallini, Nancy Theresa.

 3 versions of this record exist. [See all versions](#) >

4



MULTIPLE VERSIONS

**Microeconomics**

Krugman, Paul R.; Wells, Robin, 1959-;Cunningham, Rosemary Thomas.;Kelly, Elizabeth Sawyer.

 10 versions of this record exist. [See all versions](#) >

5



MULTIPLE VERSIONS

**Microeconomics**

Parkin, Michael, 1939-



JOURNAL

### The Academy of Management review.

Academy of Management

1976- | Mississippi State, Miss. : Academy of Management | Vol. 1, no. 1 (Jan. 1976)-

PEER REVIEWED

[Online Access >](#)

[TOP](#)

[EXPORT OPTIONS](#)

[VIEW ONLINE](#)

[DETAILS](#)

[LINKS](#)

[BROWSE RELAT...](#)

[TAGS](#)

#### Export Options



QR



FAVOURITES/ORLO



EXCEL



CITATION



PRINT



PERMALINK



EMAIL



BIBTEX



ENDNOTE



MENDELEY



REFWORKS



RIS

#### View Online

##### Full text availability

[ABI/INFORM Global](#)

Available from 01/01/1976 volume: 1 until 31/10/1985 volume: 10 issue: 4.

[ABI/INFORM Global](#)

Available from 01/01/1987 volume: 12 issue: 1 until 31/10/2001 volume: 26 issue: 4.

[EBSCOhost Business Source Ultimate](#)

Available from 01/01/1976.

[JSTOR Business Collection I](#)

Available from 01/01/1976 volume: 1 issue: 1 until 31/10/2018 volume: 43 issue: 4.

#### Details

<b>Title</b>	The Academy of Management review.
<b>Uniform Title</b>	Academy of Management review (Online)
<b>Contributor</b>	Academy of Management
<b>Publisher Details</b>	Mississippi State, Miss. : Academy of Management





MyEBSCO  
Sign in

## Search articles, books, journals & more

Searching: **Business Source Ultimate**

Basic search

All fields

AND

All fields

AND

All fields

+ Add row | - Delete row

Clear all

Search

Filters Search options **Publications** Subjects

Search for publications in Business Source Ultimate -- Publications

Alphabetical  Match any words  Subject and description

← Back

Add to search

# Typical 'browser based' platform

### My dashboard ^

Projects

Saved

Recent activity

Holds & checkouts

Alerts

### Research tools

New search

Help

Publications authority

Subjects authority

# Academy of Management. The Academy of Management Review

Briarcliff Manor: Academy of Management



Scholarly Journal



[Advanced Search](#)

Peer reviewed.

**Full text coverage** Jan 1987 (Vol. 12, no. 1) - Oct 2001 (Vol. 26, no. 4)

**Citation/Abstract coverage** JAN. 1976 (Vol. 1, no. 1) - present

[More publication information](#) ▾

## Journal Citation Indicator

The citation impact of this journal compared to the average journal (1.0) in the same category.

2024 2023 2022 2021 2020 2019 2018

**3.83 3.43 3.38 2.66 3.08 3.32 3.95**

JCR CATEGORY (2 of 2)

BUSINESS

MANAGEMENT

JCI RANK

1/317

2/421

JCI QUARTILE

Q1

Q1

Source: Journal Citation Reports

[? What's this?](#)

## Choose an issue to view

2026 ▾



Apr 2026; Vol. 51 (2) ▾

[Show issue contents](#)

Issue contents (latest available): Apr 2026; Vol. 51 (2)

[← Previous issue](#)

Select 1-12





[How to R.E.S.P.O.N.D.: A Framework for Thoughtful Revisions and Scholarly Dialogue](#)

1

Rogers, Kristie; Shropshire, Christine; Bolino, Mark. **Academy of Management. The Academy of Management Review; Briarcliff Manor** Vol. 51, Iss. 2, (Apr 2026): 217.



Citation/Abstract

[Abstract/Details](#)

[FIND IT @ OXFORD](#)

[2025 Presidential Address: Activism, Impact, and Science: Management Scholarship in a Challenging Era](#)



# Academy of Management. The Academy of Management Review

Briarcliff Manor, Academy of Management



Scholarly Journal

Advanced Search

Peer reviewed.

**Full text coverage** Jan 1987 (Vol. 12, no. 1) - Oct 2001 (Vol. 26, no. 4)

**Citation/Abstract coverage** JAN. 1976 (Vol. 1, no. 1) - present

[More publication information](#)

## Journal Citation Indicator

The citation impact of this journal compared to the average journal (1.0) in the same category.

2024 2023 2022 2021 2020 2019 2018

**3.83 3.43 3.38 2.66 3.08 3.32 3.95**

JCR CATEGORY (2 of 2)	JCI RANK	JCI QUARTILE
BUSINESS	1/317	Q1
MANAGEMENT	2/421	Q1

Source: Journal Citation Reports

[? What's](#)

## Choose an issue to view

Show issue contents

Issue contents (latest available): Apr 2026; Vol. 51 (2)

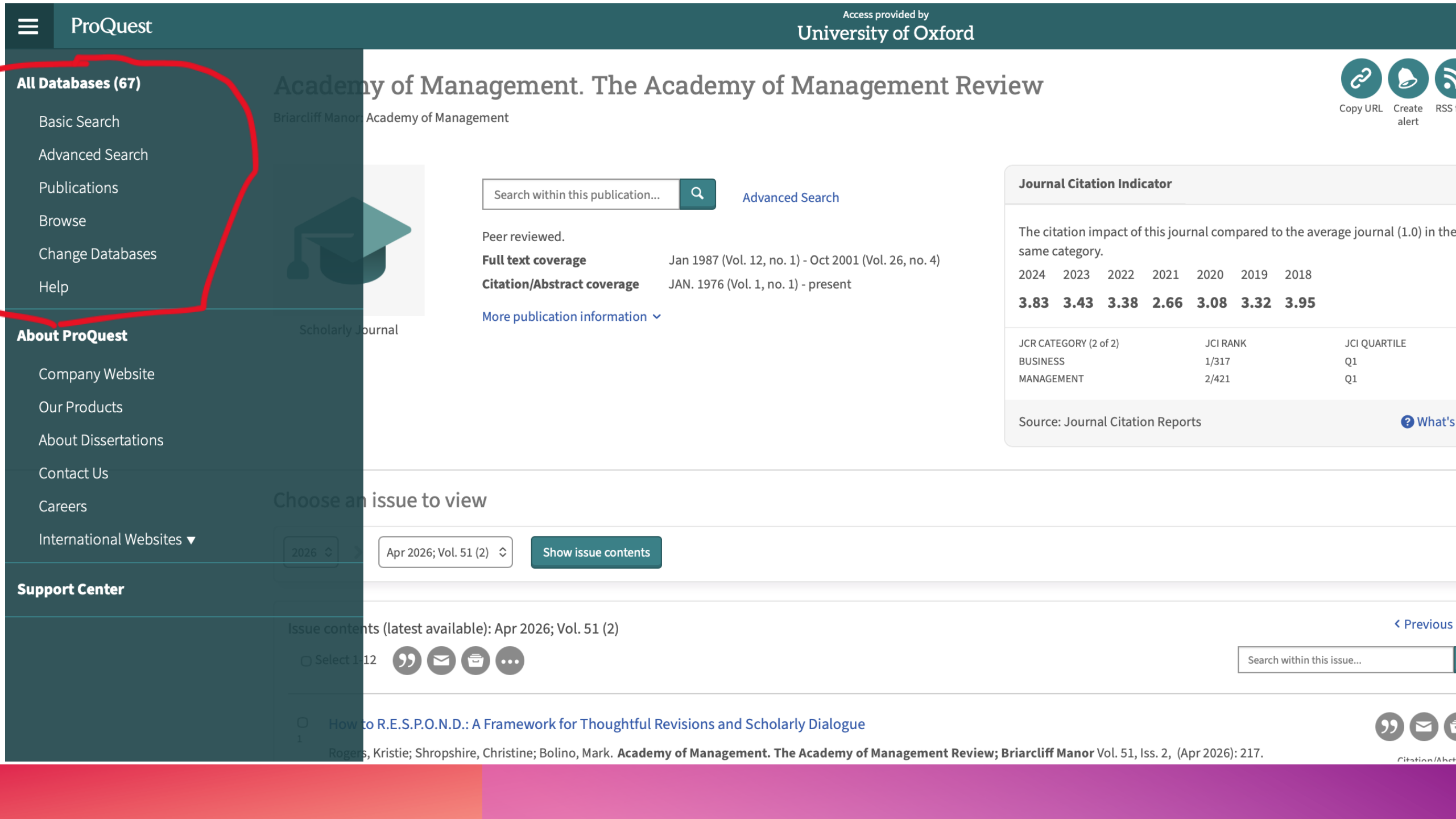
Select 1-12

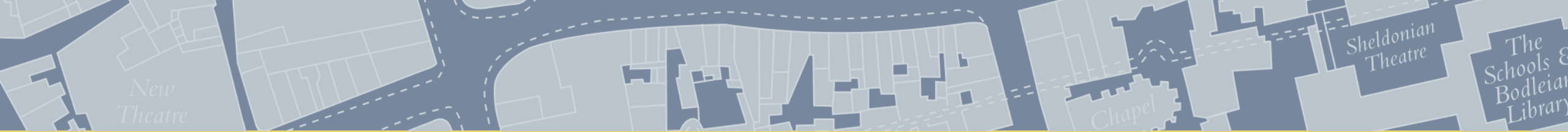
[< Previous](#)

How to R.E.S.P.O.N.D.: A Framework for Thoughtful Revisions and Scholarly Dialogue

Rogers, Kristie; Shropshire, Christine; Bolino, Mark. **Academy of Management. The Academy of Management Review; Briarcliff Manor** Vol. 51, Iss. 2, (Apr 2026): 217.

Citation/Abt





Sign in to get complete results and to request items

 Sign in

 DISMISS

## Announcements

- ✔ An interim solution providing access to electronic Legal Deposit content collected before October 2023 is available on Bodleian Libraries reading room PCs via the eLD SOLO Browser. Electronic Legal Deposit content published after that date is not currently available; see our LibGuide for alternative sources. Other British Library services remain unavailable, including the UK Web Archive. If you are unable to read an electronic Legal Deposit item due to a disability, please contact the Disability Librarian.
- ✔ The accessibility of SOLO has been improved to make it easier to navigate when zoomed in or using a screen reader. Further accessibility improvements pending. See our accessibility statement for further details.

## Getting Help

If you have any questions or need help making the most of the extensive electronic resources available to you then please:

- > use our Live Chat service
- > contact us via [reader.services@bodleian.ox.ac.uk](mailto:reader.services@bodleian.ox.ac.uk)

## Useful links

- > [Bodleian Libraries Home Page](#)
- > [Subject & Research Guides](#)
- > [Find A Library](#)
- > [Journal Search | BrowZine](#)
- > [Databases A-Z](#)
- > [Special Collections Catalogues](#)
- > [Digital Bodleian](#)
- > [Exam Papers](#)
- > [ORA \(Oxford University Research Archive\)](#)
- > [ORLO \(Oxford Reading Lists Online\)](#)



## Databases A-Z

Find the best library databases for your research.

[All](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) <#>

### 61 Databases found for Economics

[Clear Filters/Browse All Databases](#)

#### A

[AREAER Online](#) 

[more...](#)

[ASSIA Applied Social Sciences Index and Abstracts](#)

[more...](#)

#### B

[BACI](#) 

[more...](#)

[Bodleian Data Service](#) 

[more...](#)

[British Mercantile Trade Statistics 1662-1809 \(trial until 4 March 2026\)](#)

[more...](#)

[LIVE CHAT \(click to open\)](#)

#### Use of Oxford e-resources

Oxford subscribes to a wide range of e-resources, e-journals and e-books to support your research and study. Use of these is governed strictly by licence agreements.

Remember:

- E-resources may be used for study and research but NOT for commercial purposes
- You may NOT share your password or pass content from e-resources to anyone else
- Systematic downloading or viewing of excessive amounts of data is strictly forbidden

Failure to comply may result in:

- Referral to the University Proctors
- Publishers withdrawing content from the whole university

## Economics: Journals & Working papers

Subjects: [Development Studies](#), [Economics](#)[Home](#)[Books](#)[Journals & Working papers](#)[Data & Statistics](#)[Online news sources and archives](#)[Theses and Dissertations](#)

### Introduction

[Introduction](#)[Definitions](#)

A journal is a regular publication (monthly, yearly, etc.) in which you will find academic and research articles. The articles present current research and are critiqued by experts before publication, so you can be confident of their quality. The majority of journals held by the Bodleian Libraries in both physical and electronic formats can be searched via [SOLO](#), as can the individual articles within ejournals.

Working Papers are pre-publication versions of academic articles, book chapters, technical papers, or reviews. Authors may release working papers to share ideas or to receive feedback on their work. This is done before the author submits the final version of the paper to a peer reviewed journal or conference for publication. Working paper can be cited by other peer-reviewed work.

On this page you will find guidance on how to search for and access print and ejournals.

### Physical journals

[Introduction](#)[Principal Collections](#)

The tabs at the top of this section list Oxford University libraries with print journals of interest to those studying Economics. A lot of journals are available online but some are still in print, especially those published before the introduction of computers and online journals, and they have not all been digitised.

#### Help with print journals

For those wishing to learn more about searching for journals in Oxford, we recommend the following guide:

- [SOLO: Search Oxford Libraries Online guide](#)

A guide for students and researchers at the University of Oxford, or those visiting, who seek support in using the Bodleian Libraries resource discovery tool, SOLO.

### Recommend a journal

If the Bodleian Libraries don't have the journal or article you are looking for, you can make a recommendation by completing the form below ([Oxford Single-Sign On](#) required).

- [Recommend a purchase](#)

### Working papers

Working (Research / Discussion) Papers, pre-publication versions of technical papers, academic articles, book chapters, or reviews, are a great way

# FURTHER OPTIONS FOR SEARCHING AND USE

- More research guides
  - Journal search - [https://solo.bodleian.ox.ac.uk/discovery/jsearch?vid=44OXF\\_INST:SOLO](https://solo.bodleian.ox.ac.uk/discovery/jsearch?vid=44OXF_INST:SOLO)
  - Newspapers Libguide - <https://libguides.bodleian.ox.ac.uk/newspapers>
  - (General) Guidance for Researchers - <https://libguides.bodleian.ox.ac.uk/research-guides>
  - Bodleian Data Service - <https://www.bodleian.ox.ac.uk/collections-and-resources/data>

# OUTSIDE OF BODLEIAN COLLECTIONS

- Free to use text and data archives (intended for non-commercial academic work)
  - UK Data Archive/Service (ESRC) – <https://www.data-archive.ac.uk>
  - Eurostat - <https://ec.europa.eu/eurostat>
  - ICPSR - <https://libguides.bodleian.ox.ac.uk/research-guides>
  - More examples via Bodleian Data Service - <https://www.bodleian.ox.ac.uk/collections-and-resources/data>
- **Open Data** portals such as World Bank, OECD etc also listed on libguides
- **Restricted Data** is another option
- Discussed in library iSkills training on data sources and use

- Find Data
- Search & Compare Variables
- Data-related Publications
- Data Collections & Projects

[Home](#) / Find Data at

## Find Data

Find studies, variable

## ICPSR

and more



## We make finding data easy

ICPSR has a LOT of data. To find what you're looking for, use tools like the Bibliography and variable searches, or dive into a thematic collection to get started. Use the search bar to get results at the study level (data), variable level (question or field in the data), and publications related to each data set.

People search for data to meet a variety of needs — a given population, geographic region, time period, or topic. The rich information accompanying ICPSR data (metadata) makes these types of searches possible. Happy exploring!

## [Social, Behavioral, & Economic COVID Coordinating Center \(SBE CCC\)](#)

A center for communication and collaboration on COVID-19 related research, SBE CCC streamlines information-sharing between our NIH-funded consortium members, the behavioral and social science community, and the public.

---



## [Social Media Archive at ICPSR](#)

The Social Media Archive (SOMAR) is a revolutionary initiative and data resource that makes social media data accessible and useful to researchers like never before. Our archive contains a wide range of data collected from large-scale social media platforms such as Twitter, Facebook, Instagram, and Reddit, as well as smaller, more specialized data sets focused on specific research topics. These data sets have been collected and curated by researchers from around the world, and they cover various topics such as political communication, online behavior, and social networks.

---



## [U.S. Census Data](#)

Preserves and disseminates survey instruments, specifications, data dictionaries, codebooks, and other materials provided by the U.S. Census Bureau.

---



## [Western Economic Association International](#)

The ICPSR Western Economics Association International (WEAI) Data Hub is a comprehensive resource offering access to datasets, educational materials, and support for economic research.

---

# Learning Hub

## Enhance your data skills and teaching

### [New to using data](#)

Best practice and training for researchers new to accessing and using data in our collection. Includes advice and tools to correctly cite data; student-specific information on our Dissertation Award for undergraduates; and more.

### [Data skills modules](#)

There is a wealth of data available for reuse in research and reports. These free, interactive tutorials are designed for anyone who wants to start using secondary data. They show you how to get started with finding good quality data, understanding it and starting your analyses.

### [Students](#)

Students can access most of the UK Data Service's collection of social, economic and population data. Find resources to help you find and use our data during your studies including the UK Data Service dissertation resources.

### [Survey data](#)

Survey data, including data from long-running surveys, series and longitudinal studies, are a major part of social science

### [International data](#)

Our international macrodata contain socio-economic time series data aggregated to a country or regional level for a range of

### [Qualitative data](#)

Qualitative research gives a voice to the lived experience, offering researchers a deeper insight into a topic or individuals'

# Search our data catalogue

## Economics

<https://datacatalogue.ukdataservice.ac.uk/themes/economics>

Economics is a social science concerned with the factors that determine the production, distribution, and consumption of goods and services.

As a theme, economics covers various topics including economic activity, public spending, fiscal policy, taxation, trade and productivity. The UK Data Service holds a variety of data collections related to economics, from key government and longitudinal surveys to international indicators. Explore our library of [impact case studies](#) demonstrating how our data have been used. Find out [how to search our catalogue](#).

### Some key datasets

**IMF World Economic Outlook, 1980-2026**



**OECD Economic Outlook: Statistics and Projections**



**OECD Main Economic Indicators**



**English Longitudinal Study of Ageing**



**Labour Force Survey**



**1970 British Cohort Study**



**Annual Population Survey**



**IMF International Finance Statistics, 1920-2021**



# RESTRICTED DATA

- Provided by archives and statistical agencies
- Restricted due to issues of confidentiality or content sensitivity
  - Economic Microdata e.g. EU Labour Force Survey
  - Proprietary e.g. Gallup World Poll Data
- Involves an application process
  - Bureaucratic
  - Time delays
  - More overt security around data handling and management

# PAUSE AND RECAP

- This accounts for searching, finding and accessing strategies for Bodleian content
- All content discussed so far supplied and licensed as ‘human readable’ resources i.e. ...
  - Content collected as part of search, select and download strategies to read and analyse
  - Involves identification of relatively small number of key resources to build on or replicate
  - Deals with ‘reasonable’ amounts of content
  - Essentially manual in nature
- Different challenges involved with newer machine readable (aka TDM) techniques

# COMPUTATIONAL METHODS

- Text and Data Mining or computational methods involve
  - Large scale downloading of content
  - Systematic downloading of content
  - Creation of a very large corpus
  - Using indirect, automated machine analysis - beyond the abilities of an individual or group
  - Revealing patterns, clusters, themes
  - Identifying new relationships in the content
- Some Bodleian Collections are suitable - **but most are not!**
  - Copyright, Licensing and Terms and Conditions may rule it out

## Databases A-Z

Find the best library databases for your research.

Economics (61) ▾ All Database Types ▾ All Vendors / Providers ▾ Search for Databases Go

[All](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) <#>

### 61 Databases found for Economics

Clear Filters/Browse All Databases

#### A

[AREAER Online](#)  [more...](#)

[ASSIA Applied Social Sciences Index and Abstracts](#) [more...](#)

#### B

[BACI](#)  [more...](#)

[Bodleian Data Service](#)  [more...](#)

[British Mercantile Trade Statistics 1662-1809 \(trial until 4 March 2026\)](#) [more...](#)

#### LIVE CHAT (click to open)

#### Use of Oxford e-resources

Oxford subscribes to a wide range of e-resources, e-journals and e-books to support your research and study. Use of these is governed strictly by licence agreements.

#### Remember:

- E-resources may be used for study and research but NOT for commercial purposes
- You may NOT share your password or pass content from e-resources to anyone else
- Systematic downloading or viewing of excessive amounts of data is strictly forbidden

#### Failure to comply may result in:

- Referral to the University Proctors
- Publishers withdrawing content from the whole university

# ACQUIRING TDM CONTENT

- Even so building up large volumes of text or data is possible
- Our key concern is its done
  - Effectively – making best use of Bodleian collections or other data sources available to you
  - Legally – in terms of copyright and licensing
  - Orderly – without disruption of supply frameworks

# EFFECTIVELY

- Know what's available using;
  - Libguides / SOLO/ Bodleian Data Service etc.
  - Use TDM libguide - <https://libguides.bodleian.ox.ac.uk/TDM>
  - Subject Consultant network
- Store data securely using appropriate tools
  - OneDrive and other options
  - See Research Data Oxford - <https://www.ox.ac.uk/research/support/data-computing/research-data/tools-services-and-training-for-rdm>

# LEGALLY

- Be aware of licencing conditions of content being accessed
  - Do they have an open use creative commons licence or similar
  - Are there restrictions around academic / non-commercial use
  - Is computational analysis permitted
  - See 'Finding Open Access Media' - <https://libguides.bodleian.ox.ac.uk/OAandCC>
- Conditions usually mentioned on SOLO or libguides
- Seek advice from me or others if unsure!

# ORDERLY

- Instances growing of subscription data downloaded for TDM leading to disruption of access
- Cases where downloads limits are imposed
- Access may be cut completely e.g.
  - Bloomberg – Business data
  - Factiva - News
- Why is this happening?
  - Licence conditions being broken
  - Unusual access activity (i.e. large downloads) interpreted as security risk by vendor

# ARRANGING AN API

- Use of an API (Application Programming Interface) is an acceptable way to compile a large dataset
- At its most **general** see an API as a gateway, contract or arrangement
  - Usually organized by publishers on an individual *in advance*
  - Gives access to a dataset outside usual licensing arrangement (i.e. another form of restricted data)
  - Requires consideration of data security
  - Often needs input from an academic library
- Most commonly used for **text mining** of journal or newspaper content
  - Anyone here today used one yet?

# SUPPORTING API USE

- An API is popular with researchers planning to use Python or other coding techniques
- Popular code libraries then support the analysis
  - NLTK
  - SpaCy (Python)
  - Tidytext (R)
- What is library role?
  - Making you aware of API availability from T&F, Sage etc.
  - Providing library approval or technical assistance
  - See TDM libguide



## Text and Data Mining: APIs

Social Sciences and Humanities

[Home](#)[TDM-ready collections](#)[APIs](#)[Data scraping vs TDM](#)[TDM software](#) [About](#)

<https://libguides.bodleian.ox.ac.uk/c.php?g=720843&p=5364962>

On this page you can find an API definition, a brief description of how APIs are used and how they can be accessed by the Bodleian card holders.

[APIs](#)

### What is an API?

In technical terms an 'application processing interface' (API) is any mechanism set up to enable one digital system to request, retrieve, or manipulate data or from another system without direct human intervention. It's a digital intermediary, which allows a software tool such as a data analysis script or visualisation platform to:

- Request specific data (e.g., texts, metadata, images),
- Submit data for processing (e.g., linguistic analysis, geocoding),
- Interact with databases or archives programmatically,
- Automate repetitive research tasks.

In terms of research practice an API may be seen as a special data access agreement between researcher and data supplier. One that provides far larger volumes of data that are usually intended to be handled by a browser-based human-readable interface. For example instead of manually copying information from websites, use of an API provides direct structured access (often in formats like JSON or XML) to the underlying machine-

[Access to TDM APIs](#)

Process for arranging an individual TDM API varies depending on the publisher/vendor but there are some common questions and issues.

1. Vendor may ask whether "institutional arrangements for a TDM API are already in place" before arranging individual access. Currently the only arrangements that provide institution wide API access relate to ProQuest content (via TDM Studio), Gale Premium Collection (via Gale DSL), and Hathi Trust (via HTRC). All other vendors require an individual application.
2. Vendor may request a letter of support from the academic library funding the resource subscription before agreeing to arrange an API.
3. A free at point of use TDM API may be offered and allowed as part of the existing *institutional licence* but with certain restrictions (i.e. amount and frequency of requests, what content is covered, intended use) Less restricted options may be offered at additional cost. This cannot be funded by the library.

#### For example:

Basic TDM API from Springer is being set up free of charge as it is covered by Jisc license with the following stipulations:

- Rate Limits: Supports 100 hits per minute;
- One key per account;
- 8 query parameters and constraints available (e.g., DOI, subject, publication year)

# TYPICAL TDM TECHNIQUES

- Cluster Analysis
- Anomaly Detection
- Visualizations
  - Topic Modelling
  - Sentiment Modelling
- Any others we need to know about??
  - Keep up informed via consultations

# SUPPORTING TDM TECHNIQUES

- Library role is NOT to teach these, but..
  - Point to resources that inform them
  - Consider how library collections are being used
- We can promote code or non-code TDM approaches
  - Where coding proficient researcher want to find sources
  - Where non-coding researchers want TDM platforms (i.e. include data, API and codeless tools)
- The Bodleian provides two TDM friendly subscriptions from;
  - Gale
  - Proquest
- Friendly in that all content is licenced for TDM and includes inbuilt API

# GALE DIGITAL SCHOLAR LAB

## What Is the Gale Digital Scholar Lab?



The Lab is a single research platform where you can apply natural language processing tools to raw text data (OCR) from your institution's Gale Primary Sources holdings, or from uploaded OCR. Gale Digital Scholar Lab is organized in three broad steps: **Build, Clean, and Analyze**. These steps support newcomers and experienced users alike as they interpret both Gale Primary Sources and their own documents. An integrated Learning Center provides instructional tutorial videos and explanations throughout. The six built-in analysis tools are: Ngrams, Sentiment Analysis, Topic Modeling, Named Entity Recognition, Document Clustering, Parts of Speech.

# CONTENT

## Gale Primary Sources content

Provides access to millions of pages of content spanning many centuries and geographic regions, including:

- 17th and 18th Century Burney Collection
- 19th Century UK Periodicals
- Any Archives Unbound collections held by Oxford
- British Library Newspapers
- The Economist Historical Archive, 1843-2011
- Eighteenth Century Collections Online (ECCO)
- The Financial Times Historical Archive, 1888-2010
- The Illustrated London News Historical Archive, 1842-2003
- The Making of Modern Law: Legal Treatises, 1800-1926
- The Making of the Modern World 1450-1850
- Nineteenth Century U.S. Newspapers
- The Times Digital Archive
- Times Literary Supplement Historical Archive
- U.S. Declassified Documents Online.

## GALE PRIMARY SOURCES

- Cross-search all available Gale digital archives from a single interface
- Explore history through digitized, text-searchable primary sources
- Discover new connections using Topic Finder and Term Frequency



All



Newspapers



Legal Studies

Search across all selected Gale Primary Sources databases (33 of 33)

Start your research

Publication date(s)

All Dates  Before  On  After  Between

Include documents with no known publication date.

[View all limiters in Advanced Search >>](#)

Browse the Gale Primary Sources databases provided by your institution

Select All Deselect All



ARCHIVES OF  
SEXUALITY AND GENDER

[Archives of Sexuality and Gender](#)

Research a robust and significant collection of primary sources for the historical study of sex, sexuality, and gender.



BRITISH LITERARY  
MANUSCRIPTS ONLINE

[British Literary Manuscripts Online](#)

Explore a digitized collection of manuscripts of British authors that includes poems, plays, novels, diaries, and more.



DAILY MAIL HISTORICAL  
ARCHIVE, 1896-2004

[Daily Mail Historical Archive](#)

Search 100+ years of this major British newspaper, including access to the Atlantic Editions.

# THE LAB

Library Menu: Bodleian Libraries of the University of Oxford

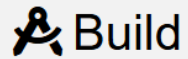
English

GALE DIGITAL SCHOLAR LAB

Build Clean Analyze My Research

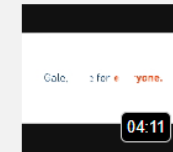


Add Note Search History



## Build

Build Content Sets by finding documents in your institution's Gale Primary Sources holdings or by uploading your own documents.



### Video: Building a Content Set

Additional topics:

- [Search Strategies](#)
- [Understanding Search Results](#)
- [View All Build Help »](#)

### Search

- peter kemp
- peter kenyon
- peter kelly
- peter kennedy
- peter ker
- peter kerr
- peter kendall
- peter kellner
- peter keenan
- peter kern
- peter kell

y, OCR confidence, and many more

### Upload

Drag and Drop .txt or .csv Files here or [Browse](#)

Get Upload Template

Create a Text Document

Manage All Uploads

# BUILDING A CONTENT SET

Library Menu: Bodleian Libraries of the University of Oxford English

**GALE DIGITAL SCHOLAR LAB** Build Clean Analyze My Research

Search: peter kemp Add Note Add To Content Set Remove From Content Set Search History

7,054 Results **Sort by:** Relevance

Search Terms: Basic Search: peter kemp [Revise Search](#)

Select All

<input type="checkbox"/> <b>Peter Kemp reflects</b> OCR Confidence: Not Captured ?Peter Kemp reflects on Edmund White's short stories, page 11 ?Peter Kemp reflects on Edmund White's short stories, page 11...	<table border="1"> <tbody> <tr> <td><b>Publication</b></td> <td>The Sunday Times.</td> </tr> <tr> <td><b>Publication Date</b></td> <td>March 12, 1995</td> </tr> <tr> <td><b>Pages</b></td> <td>1</td> </tr> <tr> <td><b>Archive</b></td> <td>The Sunday Times Historical Archive</td> </tr> <tr> <td><b>Source library</b></td> <td>Times Newspapers Limited</td> </tr> <tr> <td><b>Content Type</b></td> <td>Newspaper</td> </tr> <tr> <td><b>Document Type</b></td> <td>Front matter</td> </tr> </tbody> </table>	<b>Publication</b>	The Sunday Times.	<b>Publication Date</b>	March 12, 1995	<b>Pages</b>	1	<b>Archive</b>	The Sunday Times Historical Archive	<b>Source library</b>	Times Newspapers Limited	<b>Content Type</b>	Newspaper	<b>Document Type</b>	Front matter
<b>Publication</b>	The Sunday Times.														
<b>Publication Date</b>	March 12, 1995														
<b>Pages</b>	1														
<b>Archive</b>	The Sunday Times Historical Archive														
<b>Source library</b>	Times Newspapers Limited														
<b>Content Type</b>	Newspaper														
<b>Document Type</b>	Front matter														
<input type="checkbox"/> <b>Peter Kemp</b> OCR Confidence: Not Captured PETER KEMP PETER KEMP Peter Kemp, DSO, MC, author, soldier and war reporter, died on October 30 aged 55. He was born in Bombay on August 19, 1915. PETER KEMP was in many ways a throwback to an earlier era. He...	<table border="1"> <tbody> <tr> <td><b>Publication</b></td> <td>The Times.</td> </tr> <tr> <td><b>Publication Date</b></td> <td>November 6, 1993</td> </tr> <tr> <td><b>Pages</b></td> <td>1</td> </tr> <tr> <td><b>Archive</b></td> <td>The Times Digital Archive</td> </tr> <tr> <td><b>Source library</b></td> <td>Times Newspapers Limited</td> </tr> <tr> <td><b>Content Type</b></td> <td>Newspaper</td> </tr> <tr> <td><b>Document Type</b></td> <td>Obituary</td> </tr> </tbody> </table>	<b>Publication</b>	The Times.	<b>Publication Date</b>	November 6, 1993	<b>Pages</b>	1	<b>Archive</b>	The Times Digital Archive	<b>Source library</b>	Times Newspapers Limited	<b>Content Type</b>	Newspaper	<b>Document Type</b>	Obituary
<b>Publication</b>	The Times.														
<b>Publication Date</b>	November 6, 1993														
<b>Pages</b>	1														
<b>Archive</b>	The Times Digital Archive														
<b>Source library</b>	Times Newspapers Limited														
<b>Content Type</b>	Newspaper														
<b>Document Type</b>	Obituary														
<input type="checkbox"/> <b>M'Lean Beats Peter Kemp</b> Author: [By Cable to the Herald] OCR Confidence: 51% M LEAN BEATS PETER KEMP[BY CABLE TO THE HERALD.] M LEAN BEATS PETER KEMP[BY CABLE TO THE HERALD.] [] Sydney, Dec. 15.—A sculling match between Peter Kemp and McLean over the Parramatta course, for £300 a side and...	<table border="1"> <tbody> <tr> <td><b>Publication</b></td> <td>The New York Herald (European Edition).</td> </tr> <tr> <td><b>Publication Date</b></td> <td>December 16, 1890</td> </tr> <tr> <td><b>Pages</b></td> <td>1</td> </tr> <tr> <td><b>Archive</b></td> <td>International Herald Tribune Historical Archive, 1887-2013</td> </tr> <tr> <td><b>Source library</b></td> <td>The New York Times Company</td> </tr> <tr> <td><b>Content Type</b></td> <td>Newspaper</td> </tr> <tr> <td><b>Document Type</b></td> <td>Article</td> </tr> </tbody> </table>	<b>Publication</b>	The New York Herald (European Edition).	<b>Publication Date</b>	December 16, 1890	<b>Pages</b>	1	<b>Archive</b>	International Herald Tribune Historical Archive, 1887-2013	<b>Source library</b>	The New York Times Company	<b>Content Type</b>	Newspaper	<b>Document Type</b>	Article
<b>Publication</b>	The New York Herald (European Edition).														
<b>Publication Date</b>	December 16, 1890														
<b>Pages</b>	1														
<b>Archive</b>	International Herald Tribune Historical Archive, 1887-2013														
<b>Source library</b>	The New York Times Company														
<b>Content Type</b>	Newspaper														
<b>Document Type</b>	Article														

**Filter Your Results**

Archives	Content Type	Document type
Publication Date	Publication title	Publication Sections
Publication country or territory	Publication state/province	Publication city
Languages	Subjects	Author - items by
Source library	Illustrated works	OCR confidence range
<a href="#">Search Within</a>		

**Show Documents Added to**

Select Content Set

[Manage](#) [Analyze](#)

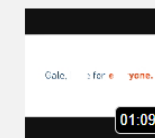
# CLEANING



## Clean

Select Cleaning Configuration to Edit

Default Cleaning Configuration



Video: Cleaning a Content Set

Additional topics:

- [Creating a Clean Configuration](#)
- [Applying During Analysis](#)
- [View All Clean Help »](#)

### Stop words

Set the words you want the Analysis Tools to ignore. [Choose a Starter List](#)

a  
about  
above  
across  
after  
afterwards  
again  
against  
all  
almost  
alone  
along  
already  
also  
although  
always  
am  
among

### Text Correction

Options for automatic text correction that will be applied before each Analysis

- Turn on all options
- Text Modification**
  - All lower case
- Characters**
  - Remove all extended ASCII characters
  - Remove all number characters
- Special Characters**
  - Remove all special characters
  - [Set specific characters >](#)
- Punctuation**
  - Remove all punctuation
  - [Set specific punctuation >](#)

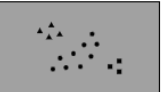
### Configuration Notes

Space to make notes or describe the purpose of this configuration.

Configuration description/notes

# ANALYSIS

Document Clustering 💡


  
Document Similarity

---

**Run Details**

Default Setup ▼ Edit Run

Named Entity Recognition 💡



  
Entities Found

---

**Run Details**

Default Setup ▼ Edit Run

Ngrams 💡



   
Word Cloud    Term Frequency

---

**Run Details**

(Mon Feb 24 10:12:50 UTC 2025) ▼ View New Setup Run Time: 1:06

Parts of Speech 💡




   
Parts Comparison    Pie Chart

---

**Run Details**

Default Setup ▼ Edit Run

Sentiment Analysis 💡



    
Sentiment Scores    Sentiment Over Time    Sentiment By Timeframe

---

**Run Details**

(Mon Feb 24 10:03:34 UTC 2025) ▼ View New Setup Run Time: 0:27

Topic Modeling 💡

   
Topic Proportion    Topics

---

**Run Details**

Default Setup ▼ Edit Run



# GETTING STARTED

- Follow the instructions in the [Description on SOLO](#) or Databases A-Z:
- Click on *Link to Database*
- Click on *Institution Credentials...Recommended*
- Proceed with SSO
- Click on *Log In / Create Account*
- Click on *Use University Credentials* (preferred – Google log in also available)
- Select *Personal Workspace*

# GALE DIGITAL SCHOLAR LAB - RESOURCES

- [Quick Start Guide](#) (video)
- [Learning Center](#) (documentation & videos – need to be logged into your workspace)
- [Support](#) (LibGuide, webinars & tutorials)



# TDM STUDIO

## Introduction

TDM Studio is a text and data mining tool created by ProQuest.

It allows programmatic analysis of published content from the millions of pages of news and scholarly publications provided through **current** university ProQuest subscriptions.



# NEWS CONTENT AND TDM

- Factiva and Nexis provide news content for 'traditional' research
- Factiva does not allow large scale downloading for TDM
- Nexis allows API access but at a substantial cost
- So we have added a new Proquest News subscription to support TDM
- May be used within TDM Studio
  - Either start at Solo or Database A-z entry for International Newsstream Collection
  - Or add it when within TDM Studio



DATABASE

## International Newsstream Collection

ProQuest (Firm)

[Online Access >](#)



[TOP](#)

[EXPORT OPTIONS](#)

[VIEW ONLINE](#)

[DETAILS](#)

[LINKS](#)

[TAGS](#)

### Export Options



QR



FAVOURITES/ORLO



EXCEL



CITATION



PRINT



PERMALINK



EMAIL



BIBTEX



ENDNOTE



MENDELEY



REFWORKS



RIS

### View Online

#### Full text availability

[Link to Database](#)



### Details

**Title** [International Newsstream](#) Collection

**Publisher Details** ProQuest (Firm)

**Description** [International Newsstream](#) Collection provides access to over 1,700 print, digital, and broadcast sources, combining five regional collections from around the world. With an extensive archive from the 1980s to current coverage, researchers gain insights into global events, regional developments, and [international](#) relations.

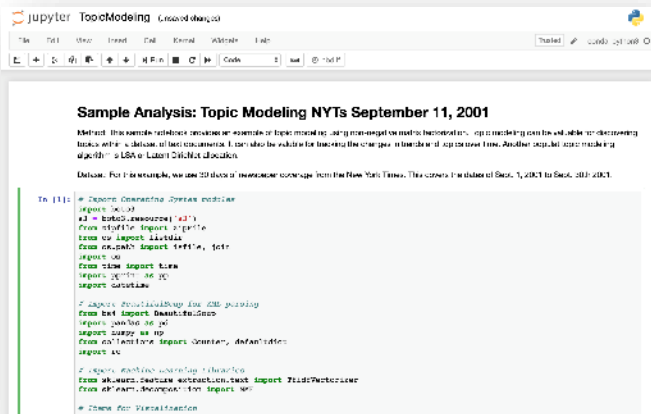
This subscription will initially run for a 1 year trial period (until 30 April 2027). Please send feedback to [John Southall](#).

**Subject** Full text  
[Business + Management Studies >](#)  
[Business + Management Studies - News >](#)  
[History >](#)  
[History - Modern, 1800- >](#)  
[Internet + Media Studies >](#)  
[Latin American Studies >](#)

# DATA ANALYSIS

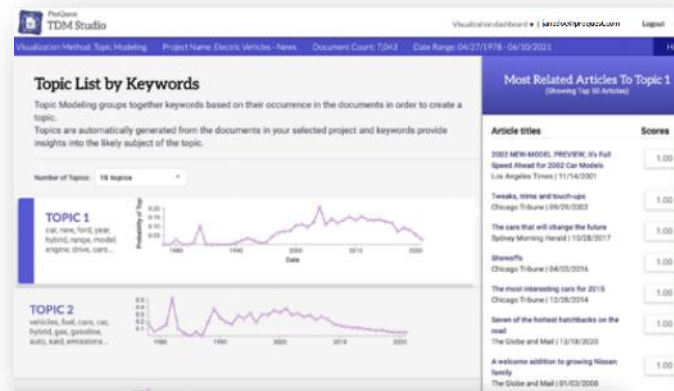
- TDM Studio provides three options for data analysis:

## Workbench



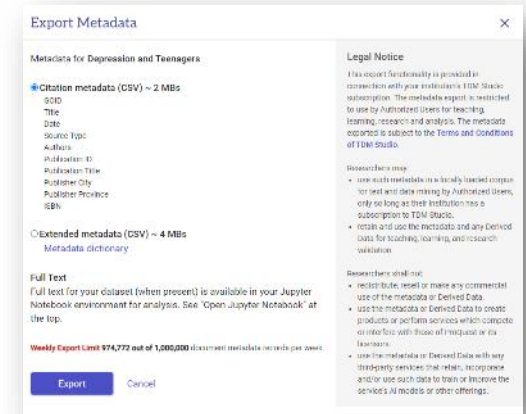
Cloud-based environment with Python and R libraries, preloaded scripts, and learning resources.

## Visualization Dashboard



Analyze without coding using built-in geographic, topic, and sentiment analysis tools.

## Metadata Export



No coding needed—export up to 1 million metadata records per week to use in preferred external tool such as Excel.

# CONTENT

## Content types:

- Dissertations – 6 million records
- Current and historical newspapers
- Congressional hearings
- Scholarly journals
- Historic periodicals
- Primary sources
- History Vault

## Subjects:

- Art
- Business
- Health & Medicine
- History
- Literature
- Science & Technology
- Social Sciences

## 180+ Databases:

- U.S. Newsstream
- ProQuest Central
- ABI Inform Collection
- The Wall Street Journal
- Annual Register
- History Vault
- Nursing & Allied Health



# DATASETS

Search across TDM Studio content and via the dataset creation page.

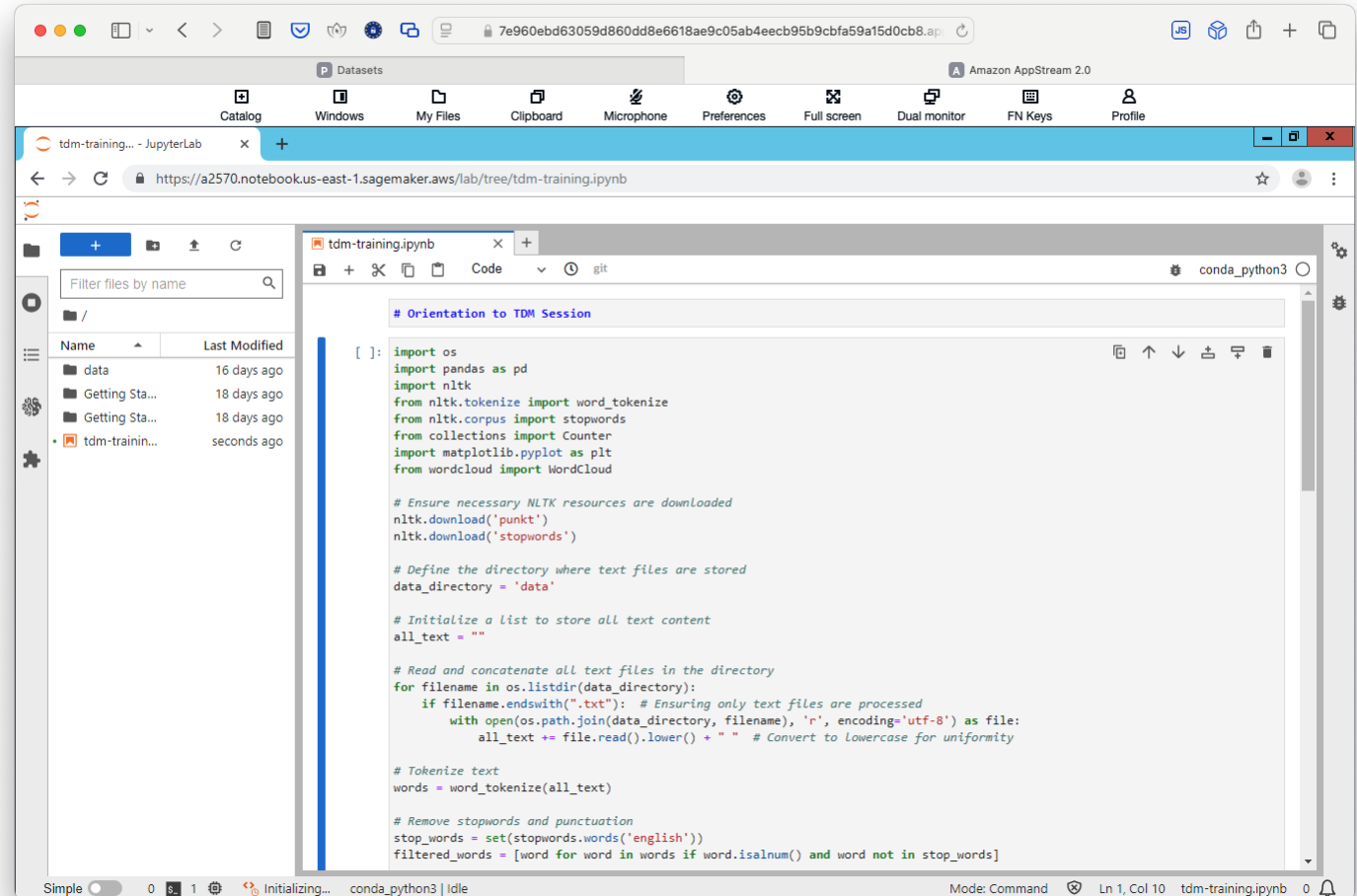
## Benefits:

- **Search All Content:** Search across all available TDM Studio content – each dataset can be up to 2 million documents.
- **Easy Filtering:** Apply filters like publication and date on one page without restarting.
- **Quick Access to Leading Publications:** Instantly find popular publications for faster results.

The screenshot displays the TDM Studio dataset creation interface. The search query is "AI AND GPU AND LLM". The page shows 854 documents with a "Save Dataset" button indicating a maximum of 2 million documents. The left sidebar contains filters for "Applied filters" (ABI/INFORM Global, ABI/INFORM Trade & Industry, 1 Year), "Publication date" (1 Year, 5 Years, 10 Years, 50 Years, with date range 2/21/2024 to 2/21/2025), "Publication title" (View All), "Databases" (View All), and "Source Type" (Wire Feeds: 528, Trade Journals: 112). The main content area lists several documents, including "Global Semiconductor 20 Feb 25\_EN", "Nebius Group N.V. announces fourth quarter and full-year 2024 financial results", "Japan Telecommunications 19 Feb 25\_EN", "Japan Electronic 19 Feb 25\_EN", "'Where are the aliens?': Elon Musk explains Grok 3 mission, says xAI wants to answer the biggest questions [International-News]", "ETtech Explainer: Meet AI 'Saba' trained on Arabic, Tamil, Malayalam [Startups]", "Valentine's Day: Wall Street & the AI Bull Market on the Rocks?", and "WEEKLY RECAP: DATACENTERDYNAMICS LTD. NEWS THIS PAST WEEK FEB 13, 2025".

# WORKBENCH

- Cloud-based environment
- Python and R libraries  
(including PyTorch and TensorFlow)
- Preloaded scripts
- Learning resources



The screenshot displays a JupyterLab environment within a browser window. The interface includes a top navigation bar with various tool icons, a file browser on the left, and a central code editor. The code editor shows a Python notebook cell with the following code:

```
# Orientation to TDM Session

[ ]: import os
import pandas as pd
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from collections import Counter
import matplotlib.pyplot as plt
from wordcloud import WordCloud

# Ensure necessary NLTK resources are downloaded
nltk.download('punkt')
nltk.download('stopwords')

# Define the directory where text files are stored
data_directory = 'data'

# Initialize a list to store all text content
all_text = ""

# Read and concatenate all text files in the directory
for filename in os.listdir(data_directory):
    if filename.endswith(".txt"): # Ensuring only text files are processed
        with open(os.path.join(data_directory, filename), 'r', encoding='utf-8') as file:
            all_text += file.read().lower() + " " # Convert to lowercase for uniformity

# Tokenize text
words = word_tokenize(all_text)

# Remove stopwords and punctuation
stop_words = set(stopwords.words('english'))
filtered_words = [word for word in words if word.isalnum() and word not in stop_words]
```



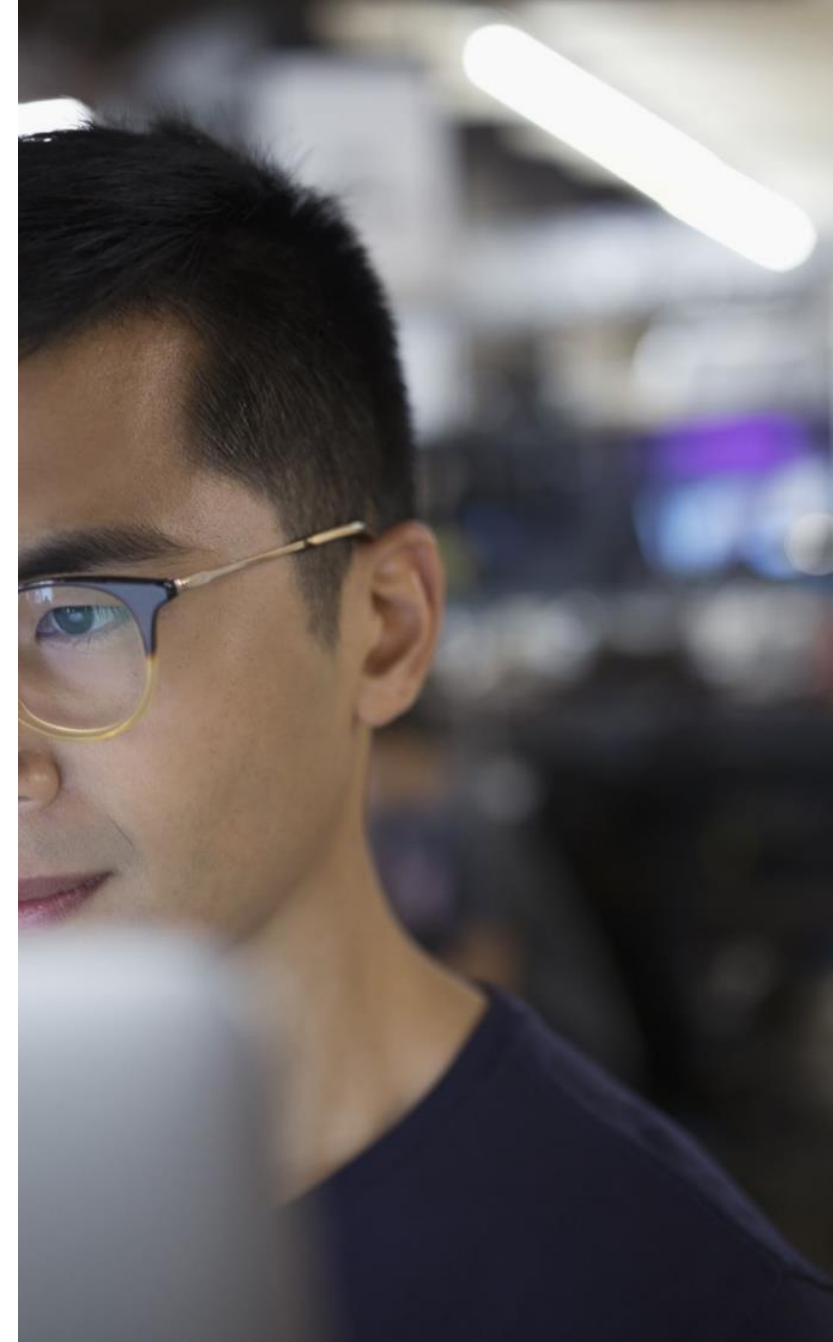
# GETTING STARTED

Follow the instructions in the [Description on SOLO](#):

- Click on *Link to Database*
- Click on *Create Account*
- Register using your *University of Oxford* email address
- Check your email to verify and then log in to TDM Studio
- Log in to TDM Studio with your new account details at: <https://tdmstudio.proquest.com/home>

# TDM STUDIO RESOURCES

- [Quick Start Guide](#)
- [TDM Studio Videos](#)
- [Documentation \(LibGuide\)](#)



# CLOSING

- Feel free to get in touch for support in the future
  - [john.southall@bodleian.ox.ac.uk](mailto:john.southall@bodleian.ox.ac.uk)
  - <https://www.bodleian.ox.ac.uk/collections-and-resources/data>