Materials developed as a part of the *Skills development for research data* project:
https://www.ub.uio.no/english/about/projects/rdm-skills/

Finding and reusing research data

Agata Bochynska, PhD

Open Research and Digital Scholarship Center, University of Oslo Library

Ivana Malovic, PhD

Library of Medicine and Science

18. November 2022

Source: RDMKit, Elixir Europe https://rdmkit.elixir-europe.org/

Source: RDMKit, Elixir Europe https://rdmkit.elixir-europe.org/

Data discovery is finding and accessing data collected for a different purpose or by a different researcher or institution.

In the process of data discovery and reuse you are working with **secondary data**, as opposed to **primary data** that you would collect yourself.

Secondary Data Analysis: A Method of which the Time Has Come

Melissa P. Johnston
School of Library and Information Studies, University of Alabama, Tuscaloosa, AL, USA

pdf

Published
2017-05-28

Abstract

Technological advances have led to vast amounts of data that has been collected, compiled, and archived, and that is now easily accessible for research. As a result, utilizing existing data for research is becoming more prevalent, and therefore secondary data analysis. While secondary analysis is flexible and can be utilized in several ways, it is also an empirical exercise and a systematic method with procedural and evaluative steps, just as in collecting and evaluating primary data. This paper asserts that secondary data analysis is a viable method to utilize in the process of inquiry when a systematic procedure is followed and presents an illustrative research application utilizing secondary data analysis in library and information science research.

How to Cite
JOHNSTON, Melissa P.. Secondary Data Analysis: A Method of which the Time Has Come. **Qualitative and Quantitative Methods in Libraries**, [S.l.], v. 3, n. 3, p. 619-626, may 2017. ISSN 2241-1925. Available at: <http://www.qqml-journal.net/index.php/qqml/article/view/169>. Date accessed: 24 nov. 2021.

Citation Formats
ABNT
APA
BibTeX
CBE

Open science means transparency and **knowledge-sharing** in research processes to make knowledge **accessible** across academic groups, sectors and national boundaries. The concept of open science encompasses the entire research process […].

- The Research Council of Norway. Policy for open science 2020

"**Open Science** is becoming the modus operandi for carrying out research and innovation by **sharing knowledge, data and tools** as early as possible, in open collaboration with all relevant knowledge actors and society."

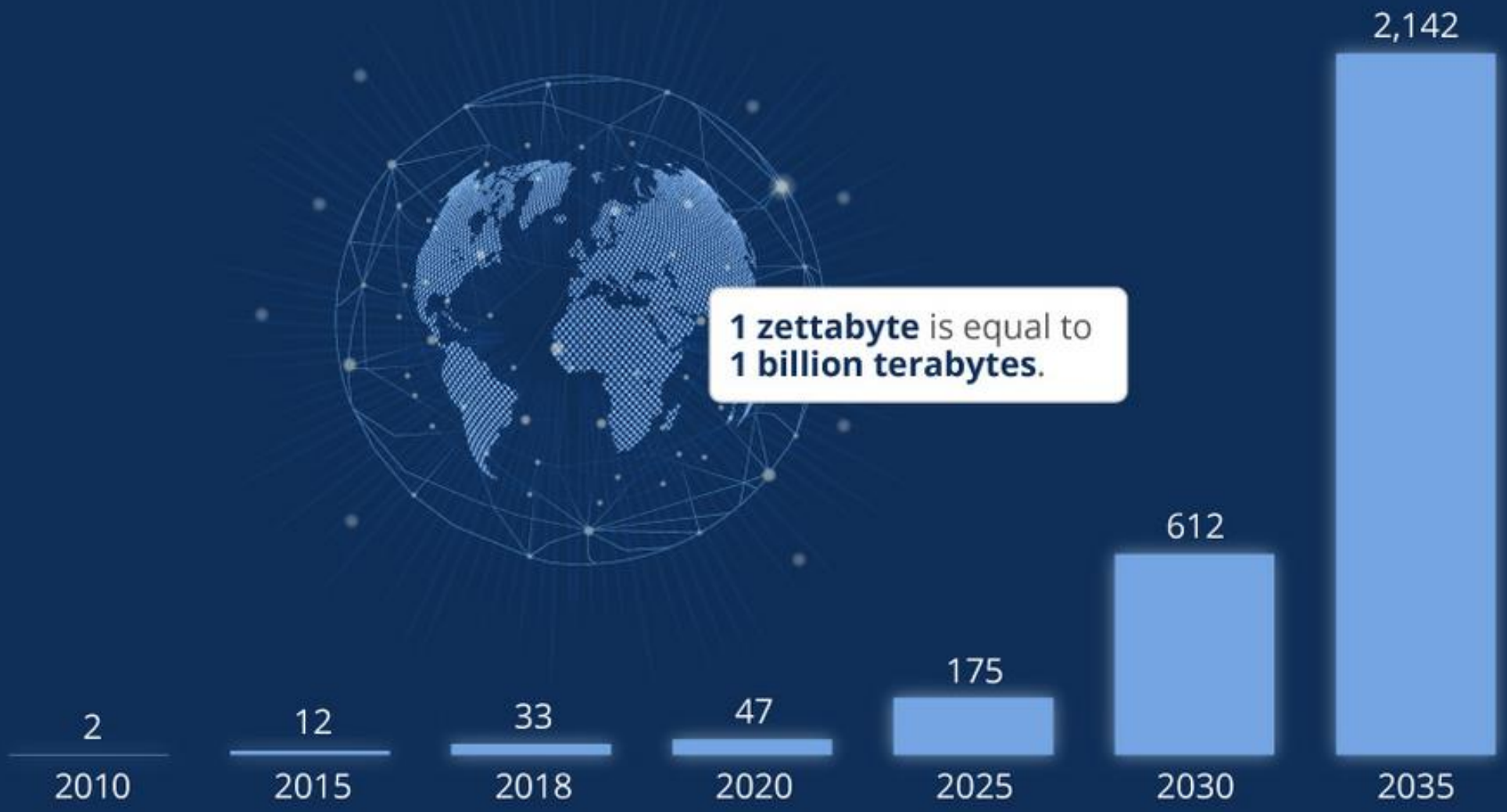
ec_rtd_horizon-europe-strategic-plan-2021-24.pdf (europa.eu)

More data sharing – more data to discover!

Global Data Creation is About to Explode

Actual and forecast amount of data created worldwide 2010-2035 (in zettabytes)

1 zettabyte is equal to
1 billion terabytes.

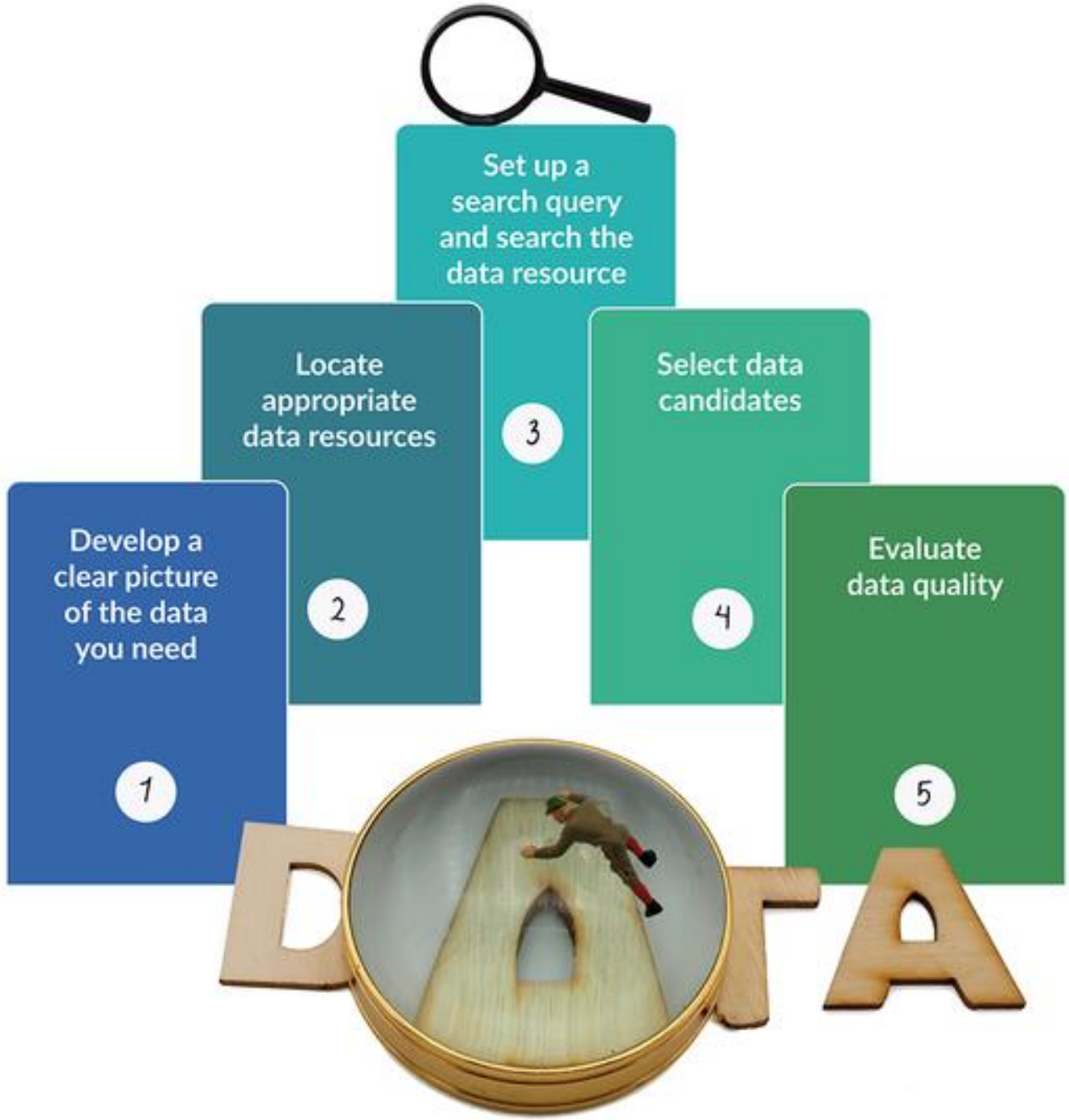


Source: Statista Digital Economy Compass 2019

Increasing need for data reuse

- High **costs** of primary data collection
- **Redundancy** or similarity in different sets of primary data
- High demands for **storage space** by increasing amount of data
- Promoting transparency, reproducibility and replication in research

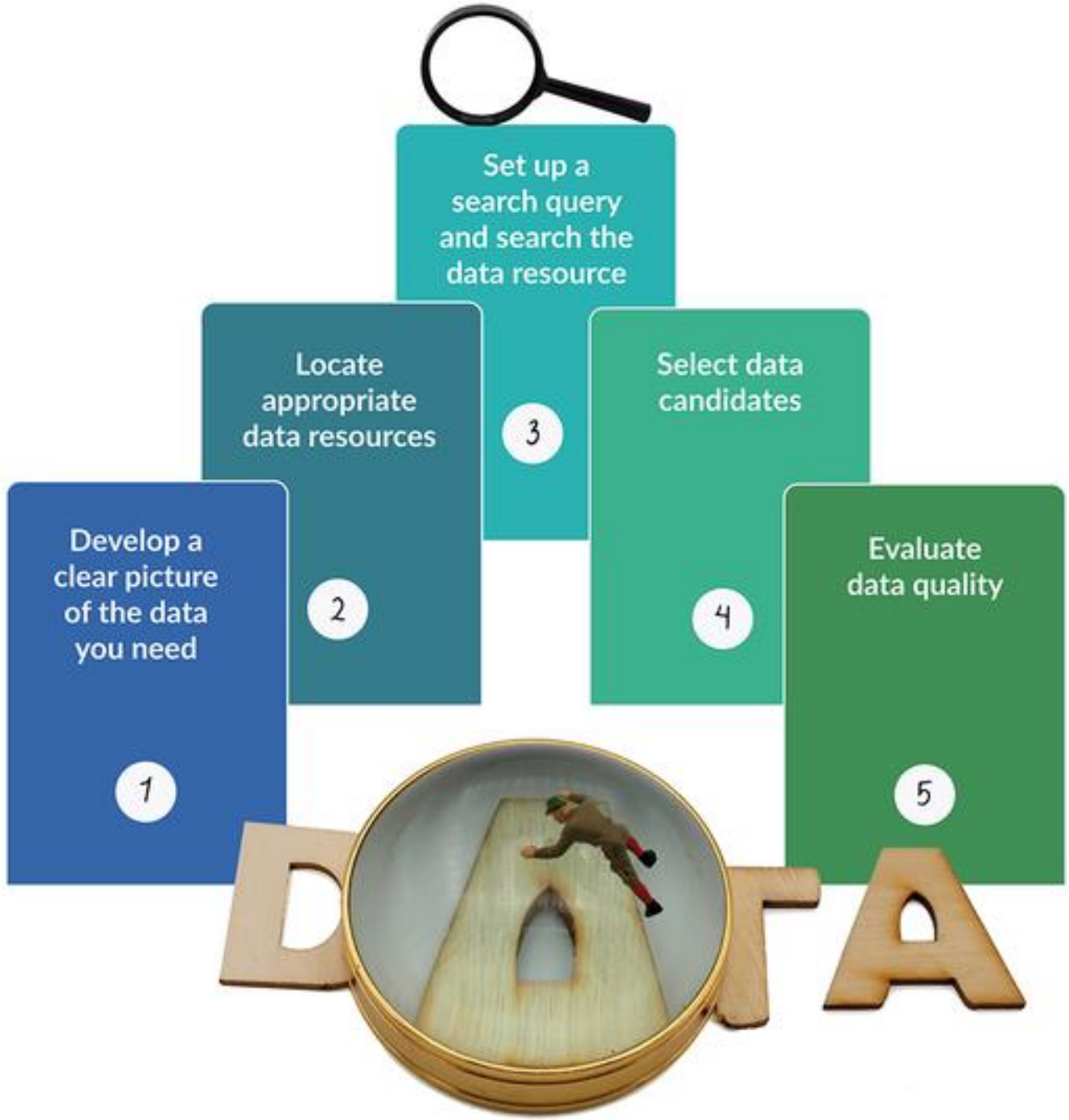
Data discovery: how-to


Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

Develop a clear picture of the data you need

Deciding on what kind of data you need

- What is the **theme/domain** you study?
- What is your **research question**?
- What are the **constructs/concepts** and how you will operationalize them?
- What is your **theory**?
- What **study** will you perform?
- What specific characteristics should the data have?


Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

Locate appropriate data resources

Where do I look for the data?

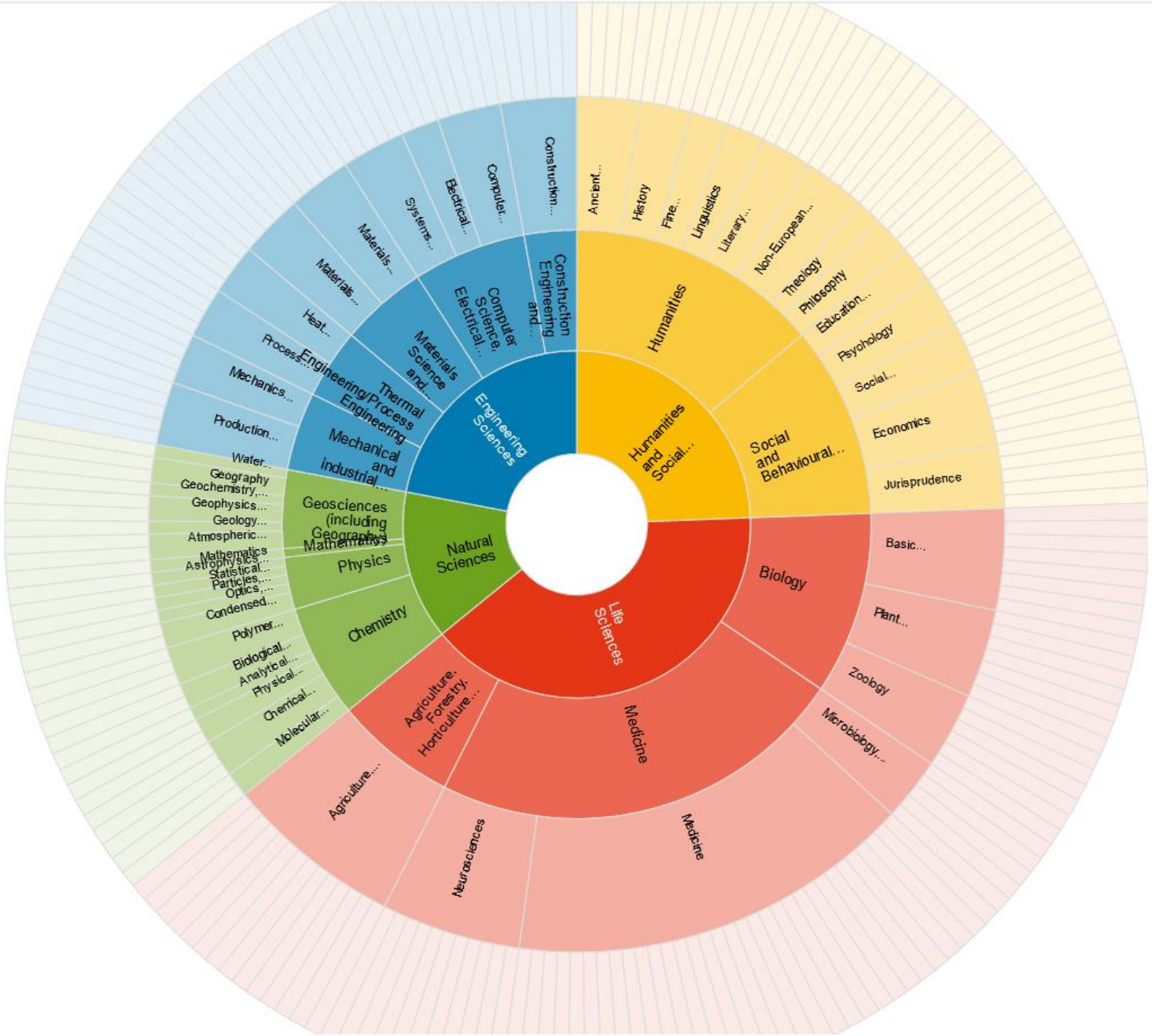
- Discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

- **Discipline-specific repositories**
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Search for discipline-specific repositories

https://www.re3data.org/


Engineering Sciences
Humanities and Social...
Natural Sciences
Life Sciences
Construction Engineering and...
Computer Science, Electrical...
Materials Science and...
Thermal Engineering/Process
Mechanical and Industrial...
Humanities
Social and Behavioural...
Geosciences (including Geography)
Mathematics
Physics
Chemistry
Biology
Agriculture, Forestry, Horticulture...
Medicine
Construction...
Compute...
Electrical...
Systems...
Materials...
Materials...
Heat...
Process...
Mechanics...
Production...
Water...
Ancient...
History
Fine...
Linguistics
Literary...
Non-European...
Theology
Philosophy
Education...
Psychology
Social...
Economics
Jurisprudence
Geography
Geochemistry,...
Geophysics...
Geology...
Atmospheric...
Mathematics
Astrophysics...
Statistical...
Particles,...
Optics,...
Condensed...
Polymer...
Biological...
Analytical...
Physical...
Chemical...
Molecular...
Basic...
Plant...
Zoology
Microbiology...
Agriculture...
Neurosciences
Medicine

Filter

Reset all

Subjects

- Humanities and Social Sciences (4)
  - Humanities (4)
    - Linguistics (4)
      - General and Applied Linguistics (4)
      - Individual Linguistics (1)
  - Social and Behavioural Sciences (1)
    - Education Sciences (1)
- Engineering Sciences (1)
  - Computer Science, Electrical and System Engineering (1)
    - Computer Science (1)
      - Artificial Intelligence, Image and Language Processing (1)

Content Types

Countries

API

Data access

Data access restrictions

Database access

Database licenses

Data licenses

Data upload

Data upload restrictions

Enhanced publication

Institution responsibility type

Institution type

Keywords

Metadata standards

PID systems

Provider types

Quality management

Repository languages

Software

Syndications

Repository types

Versioning

Search...

Search

Toogle short help

← Previous 1 Next →

Sort by

Found 4 result(s)

Eurac Research CLARIN Centre

ERCC

Subject(s)	Linguistics, Humanities and Social Sciences, Humanities, General and Applied Linguistics, Artificial Intelligence, Image and Language Processing, Computer Science, Computer Science, Electrical and System Engineering, Engineering Sciences
Content type(s)	Databases, Scientific and statistical data formats, Structured text, Software applications
Country	Italy, European Union

The Eurac Research CLARIN Centre (ERCC) is a dedicated repository for language data. It is hosted by the Institute for Applied Linguistics (IAL) at Eurac Research, a private research centre based in Bolzano, South Tyrol. The Centre is part of the Europe-wide CLARIN infrastructure, which means that it follows well-defined international standards for (meta)data and procedures and is well-embedded in the wider European Linguistics infrastructure. The repository hosts data collected at the IAL, but is also open for data deposits from external collaborators.

Michigan Corpus of Academic Spoken English

MICASE

Subject(s)	Humanities and Social Sciences, Humanities, Linguistics, General and Applied Linguistics, Individual Linguistics
Content type(s)	Structured text, Audiovisual data
Country	United States

MICASE provides a collection of transcripts of academic speech events recorded at the University of Michigan. The original DAT audiotapes are held in the English Language Institute and may be consulted by bona fide researchers under special arrangements. Additional access: https://lsa.umich.edu/eli/language-resources/micase-micusp.html

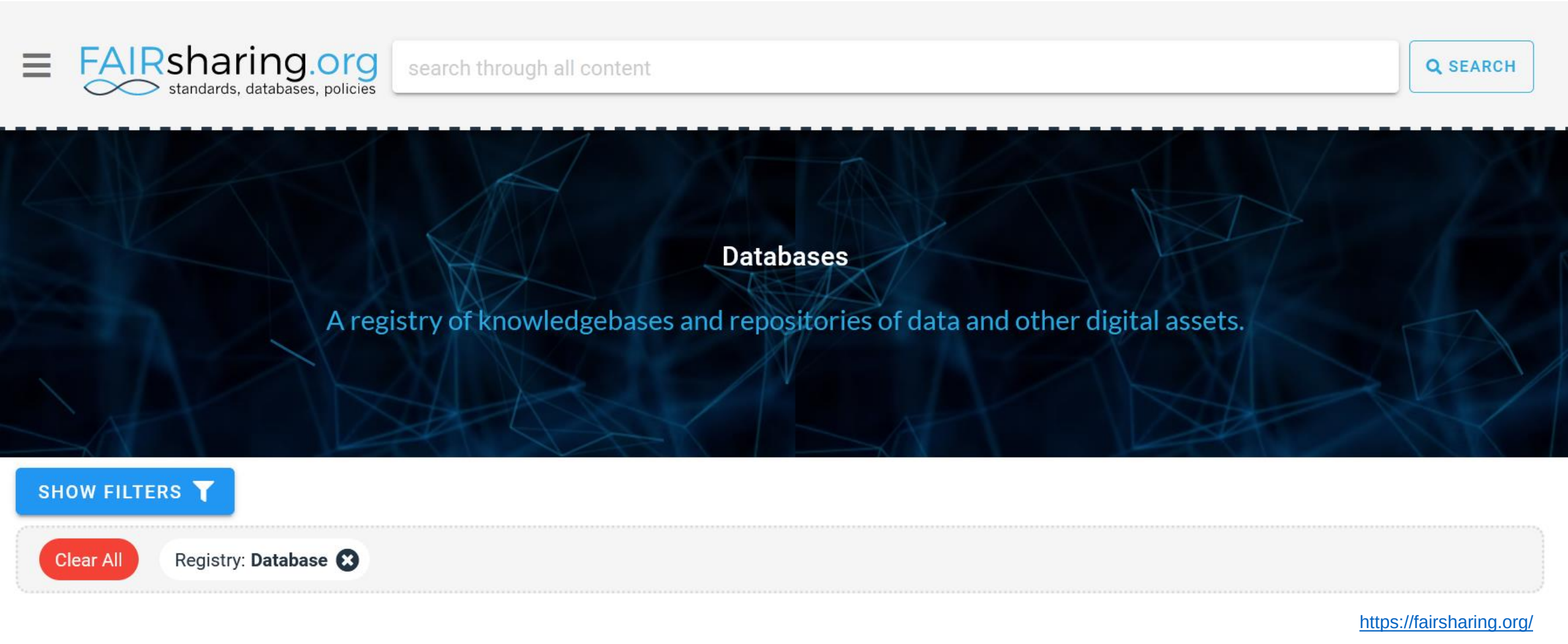
The University of Pittsburgh English Language Institute Corpus

PELIC

Subject(s)	Linguistics, General and Applied Linguistics, Humanities, Humanities and Social Sciences
Content type(s)	Standard office documents, Archived data, Source code, Images
Country	United States

The University of Pittsburgh English Language Institute Corpus (PELIC) is a 4.2-million-word learner corpus of written texts. These texts were collected in an English for Academic Purposes (EAP) context over seven years in the University of Pittsburgh's Intensive English Program, and were produced by over 1100 students with a wide range of linguistic backgrounds and proficiency levels. PELIC is longitudinal, offering greater opportunities for tracking development in a natural classroom setting.

Search for discipline-specific repositories



https://fairsharing.org/

Example: NIPH (FHI)

Content A-Z Contact us Norsk nettsted

NIPH
Norwegian Institute of Public Health

Search in NIPH Menu

Frontpage > Research & Access to data > Access to data

Access to data

Researchers can apply for access to data from health registries and health studies, as well as biological material from the biobanks. Here you will find guidelines and electronic application forms.

Les på norsk Get the latest news

HOW DO I APPLY FOR ACCESS?

ARTICLE

How to apply for access to data

The Norwegian Institute of Public Health (NIPH) can provide access to data from health registries and population-based health surveys once an application for data is approved.

Updated 09.12.2020

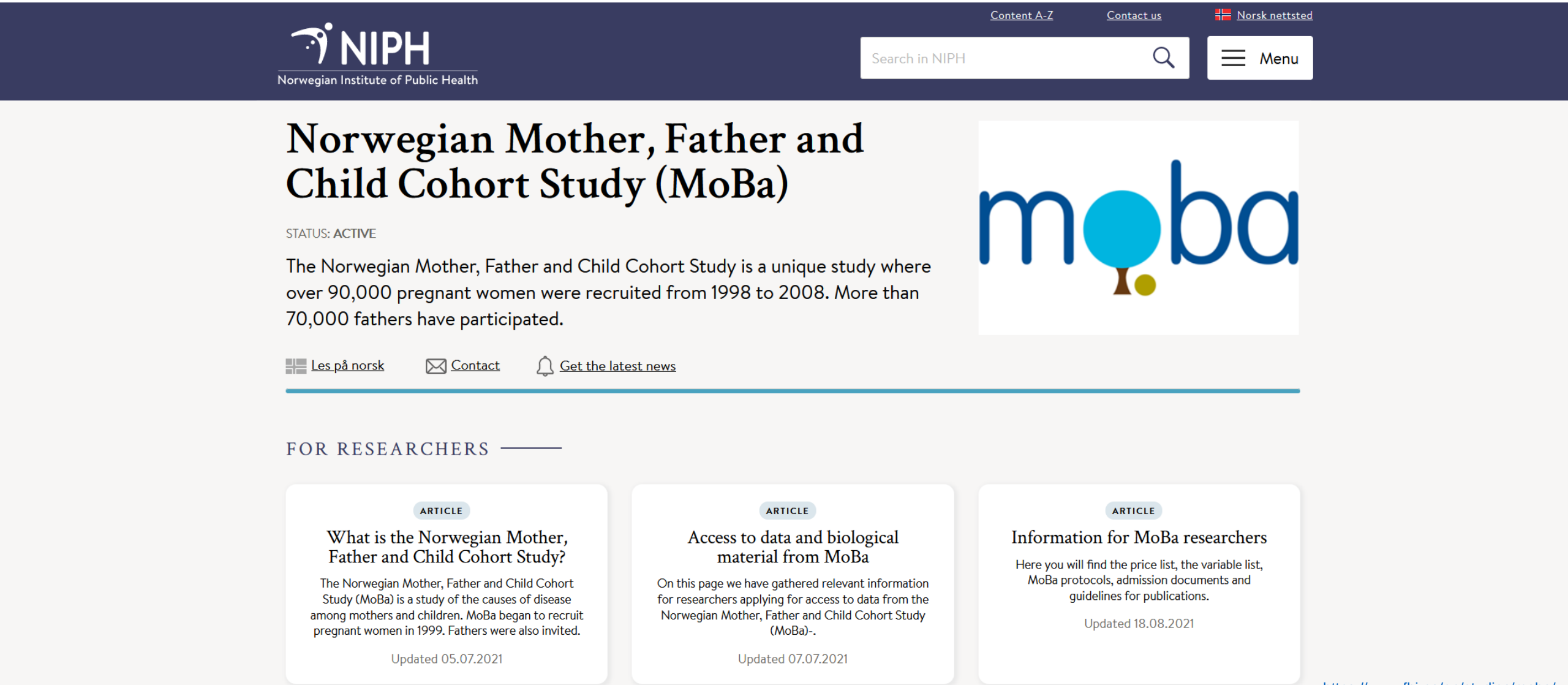
ARTICLE

Application form for access to data

For applications for access to data or biological samples from studies or mandatory national health registries at the NIPH.

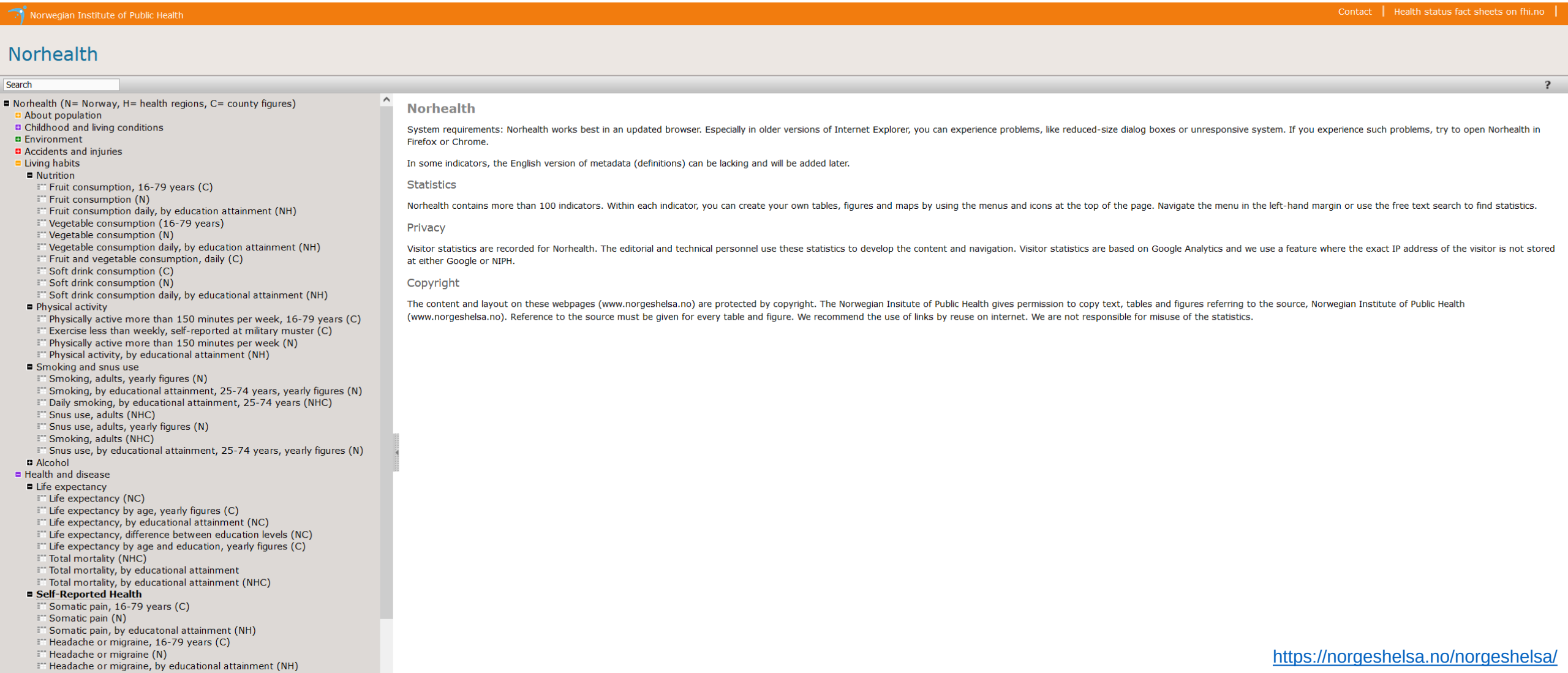
Updated 09.12.2020

Example: NIPH (FHI) – Big Data



https://www.fhi.no/en/studies/moba/

Example: NIPH (FHI) - Statistics



https://norgeshelsa.no/norgeshelsa/

https://www.who.int/data/collections

Example: WHO

World Health Organization

Health Topics | Countries | Newsroom | Emergencies | Data | About WHO

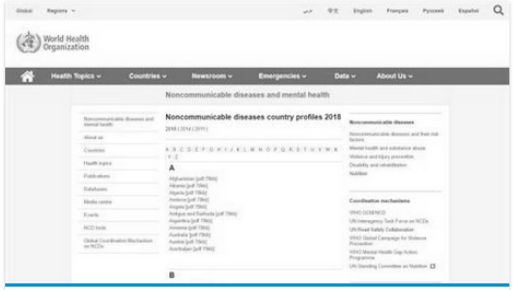
Data / WHO data collections

Data collections

The World Health Organization manages and maintains a wide range of data collections related to global health and well-being as mandated by our Member States.

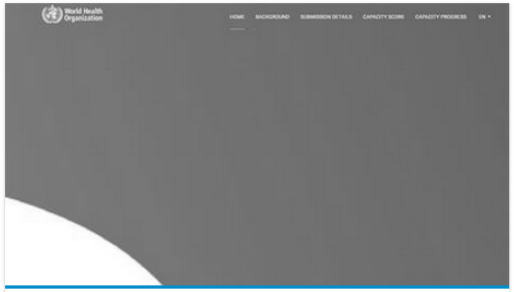
Explore our key health data products and resources from across the organization.

Search



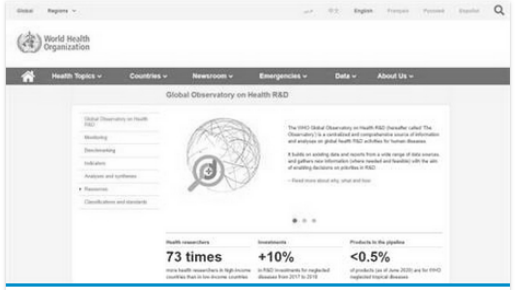
Noncommunicable diseases profiles

Number of lives that can be saved by implementing WHO 'best buys', Risk of premature death, National targets, Risk factors, National Systems Response



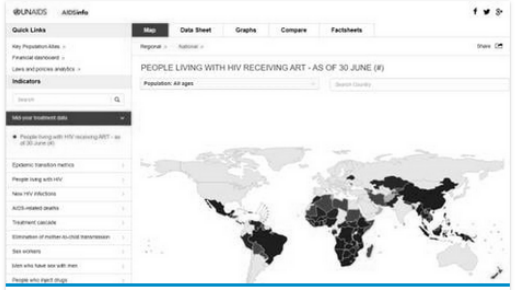
e-SPAR

Electronic State Parties Self-Assessment Annual Reporting Tool (e-SPAR) is a web-based platform proposed to support State Parties of the International Health Regulations (IHR) to fulfil their obligation to



The global observatory for health research and development

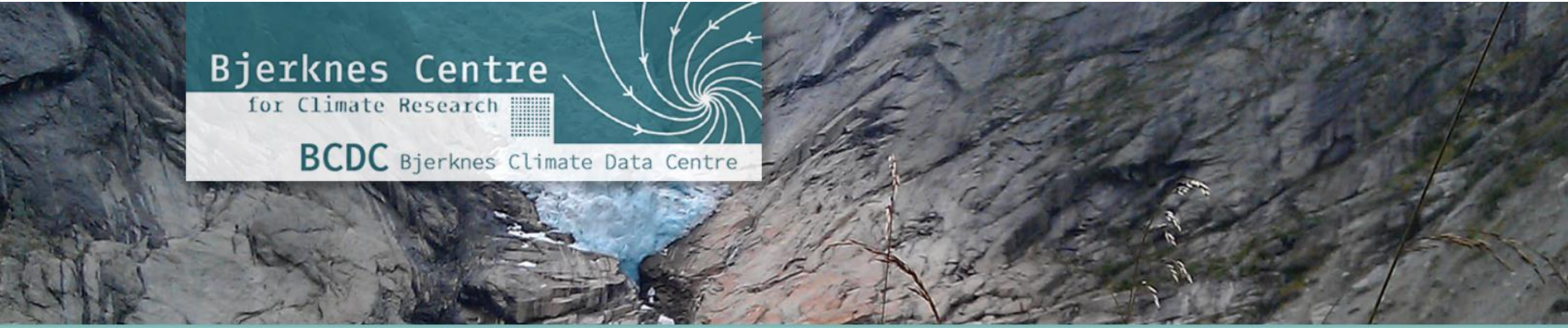
The WHO Global Observatory on Health R&D is a centralized and comprehensive source of information and analyses on global health R&D



HIV laws and policies database

Laws and Policies Analytics is a platform to view data on HIV-related laws and policies in countries compiled from official sources and reported by both national authorities and civil society to UNAIDS and the World

Example: BCDC



ABOUT DATA SEARCH PROJECTS SUBMIT PARTNERS NEWS CONTACT

REMOTE SENSING
PALEO
MODEL OUTPUT
OCEANOGRAPHY
ATMOSPHERE
DATA PRODUCTS

>> BCDC Home

DATA PUBLICATION HIGHLIGHTS

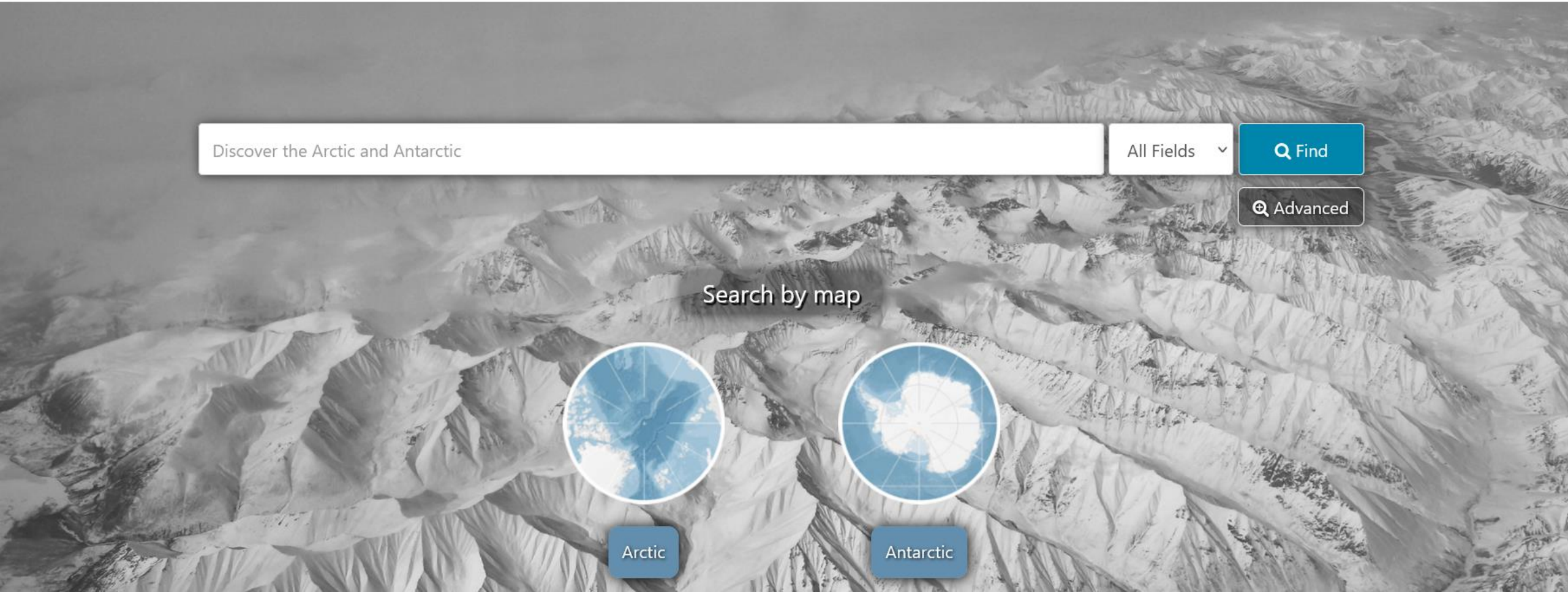
12 November 2018

High-Resolution Benthic Mg/Ca Temperature Record of the Intermediate Water in the Denmark Strait Across Dansgaad-Oeschger Stadial-Interstadial Cycles

https://www.bcdc.no/

https://openpolar.no/

Example: OPEN POLAR



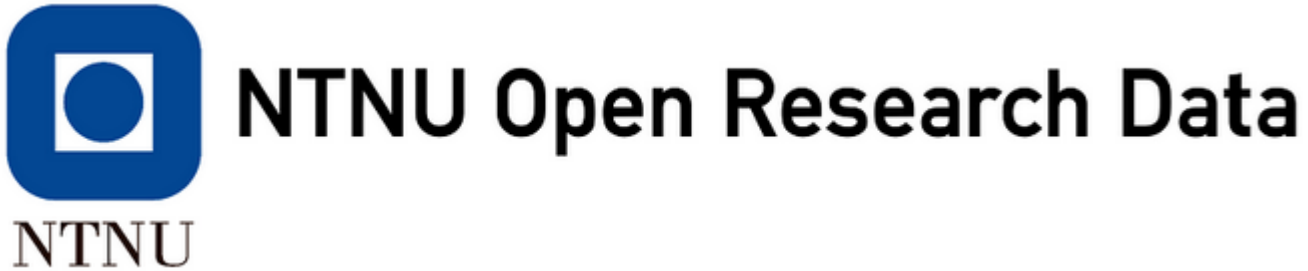
Where do I look for the data?

- Discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

- Discipline-specific repositories
- **General-purpose repositories**
- A search engine or (meta)data aggregator
- A data journal

General purpose repositories



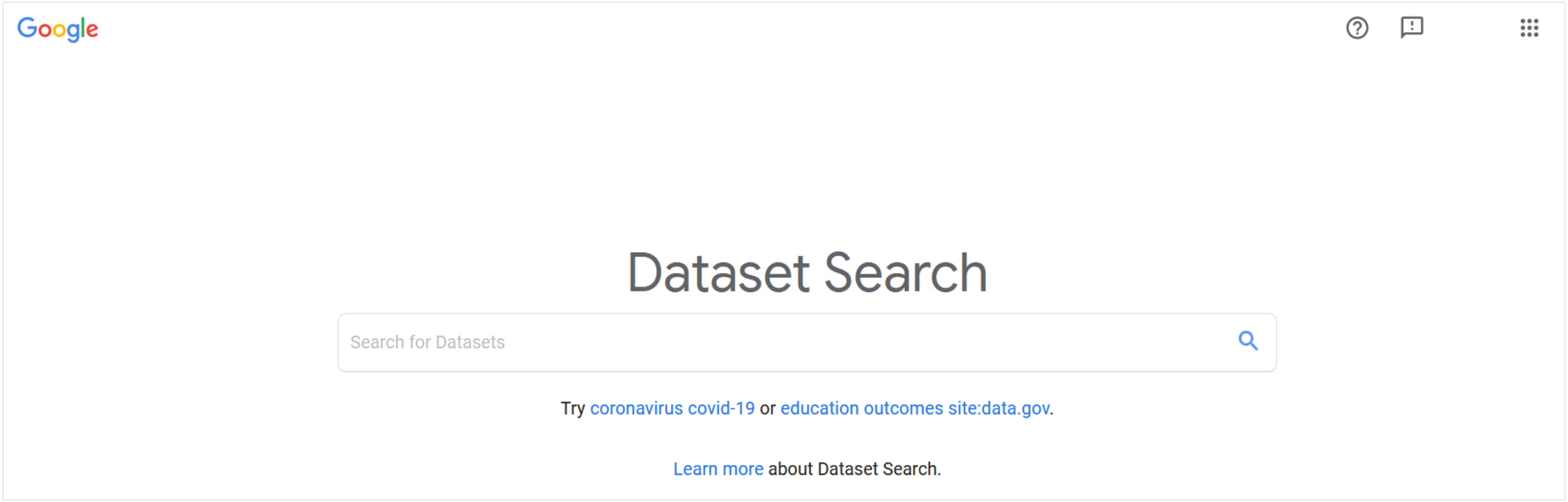
Where do I look for the data?

- Discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

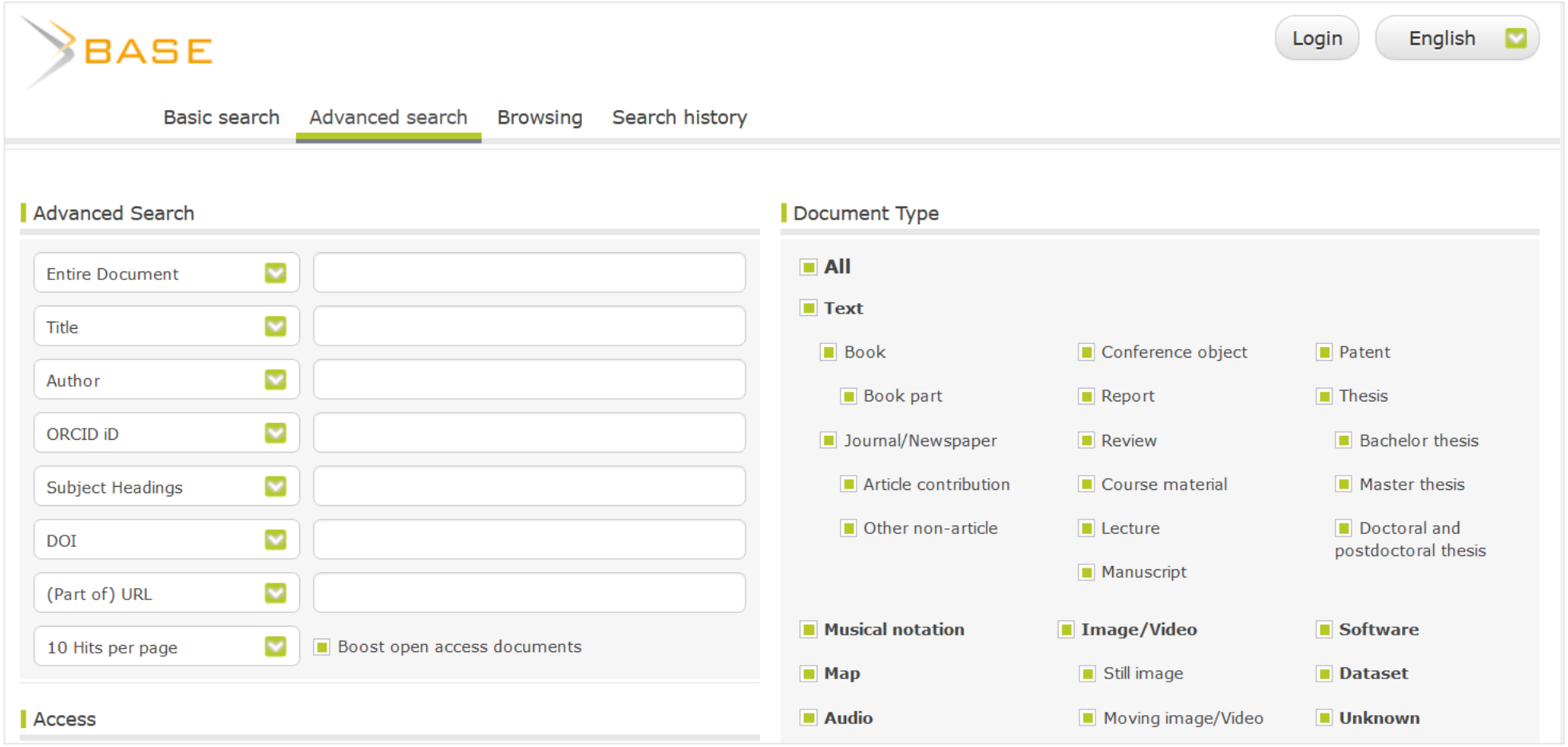
- Discipline-specific repositories
- General-purpose repositories
- **A search engine or (meta)data aggregator**
- A data journal

Search engines



https://datasetsearch.research.google.com/

Search engines



UNIVERSITY
OF OSLO

https://www.base-search.net/Search/Advanced

Where do I look for the data?

- Discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

- Discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- **A data journal**

Data journals

scientific **data**

View all journals | Search | Login

Explore content | **Journal information** | **Publish with us** | **Sign up for alerts** | **RSS feed**

nature › scientific data › about

About

- Principles
- Open Access
- FAQ

About

Scientific Data is a peer-reviewed, open-access journal for descriptions of scientifically valuable datasets, and research that advances the sharing and reuse of scientific data.

Read our key principles ▸

Geoscience Data Journal Open Access — RMetS

Co-Editors-in-Chief: Katherine Royse & Jian Peng

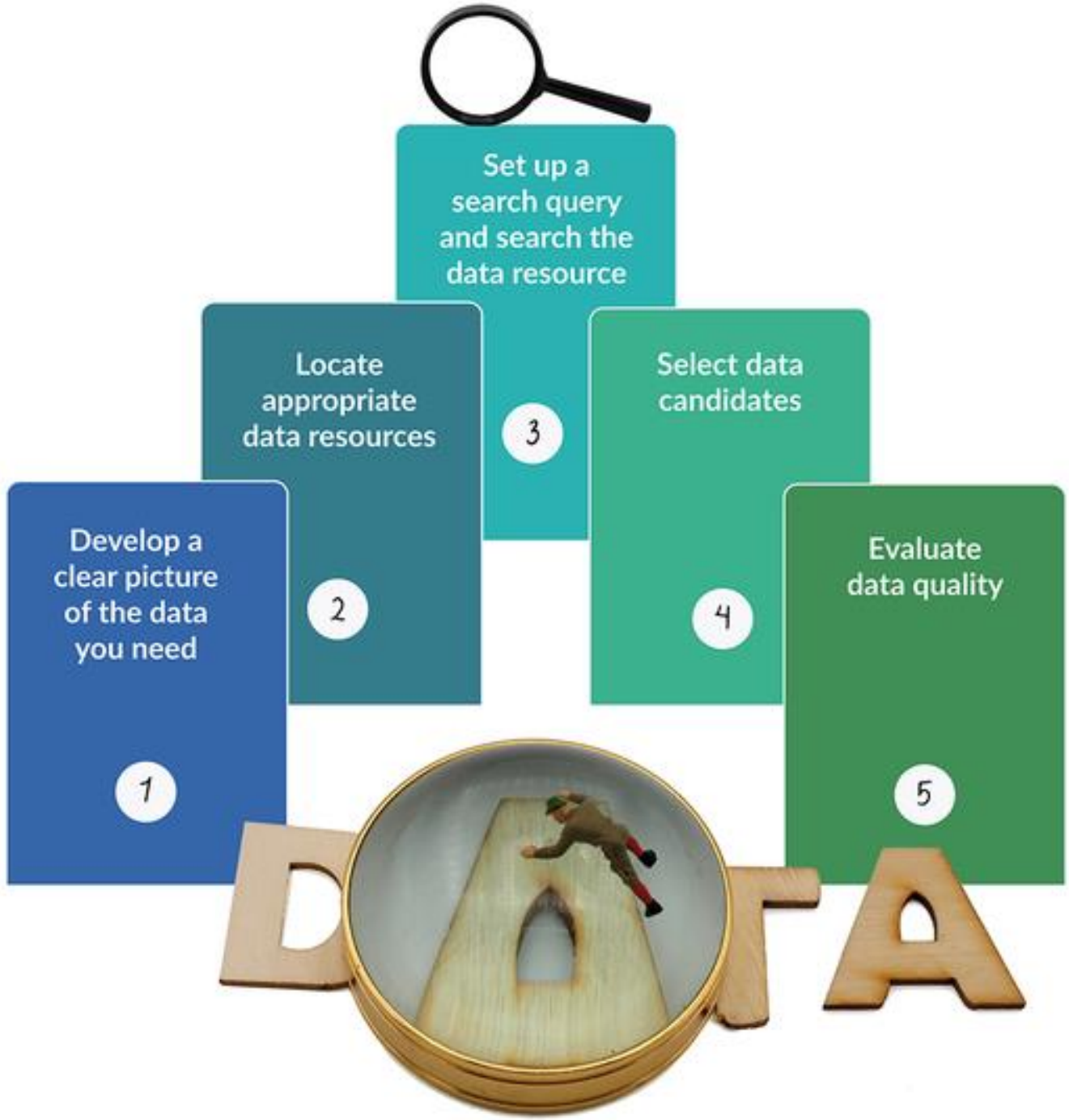
Impact factor: 2.714

2019 Journal Citation Reports (Clarivate Analytics): 74/200 (Geosciences, Multidisciplinary) 42/93 (Meteorology & Atmospheric Sciences)

Online ISSN: 2049-6060

LATEST ISSUE >

Volume 7, Issue 2
November 2020


Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

Set up a search query and search the data resource

How to search the data resource?

- Familiarize yourself with the structure of the data resource
- Register yourself as a user
- Learn how the data repository advanced search functions work
- Ask for help!
  - Ask your subject librarian: https://www.ub.uio.no/english/using/guidance/index.html
  - Consult information pages: https://sokogskriv.no/en/searching/

How to set up search queries?

Choose **keywords**

- Use the terms from your discipline
- Focus on main concepts
- Think of possible synonyms

Use **boolean operators** (if allowed)

- Terms such as AND, OR

In general search engines (e.g. Web of Science) add «data» or «dataset» to the search query or choose the type of document in the filters.

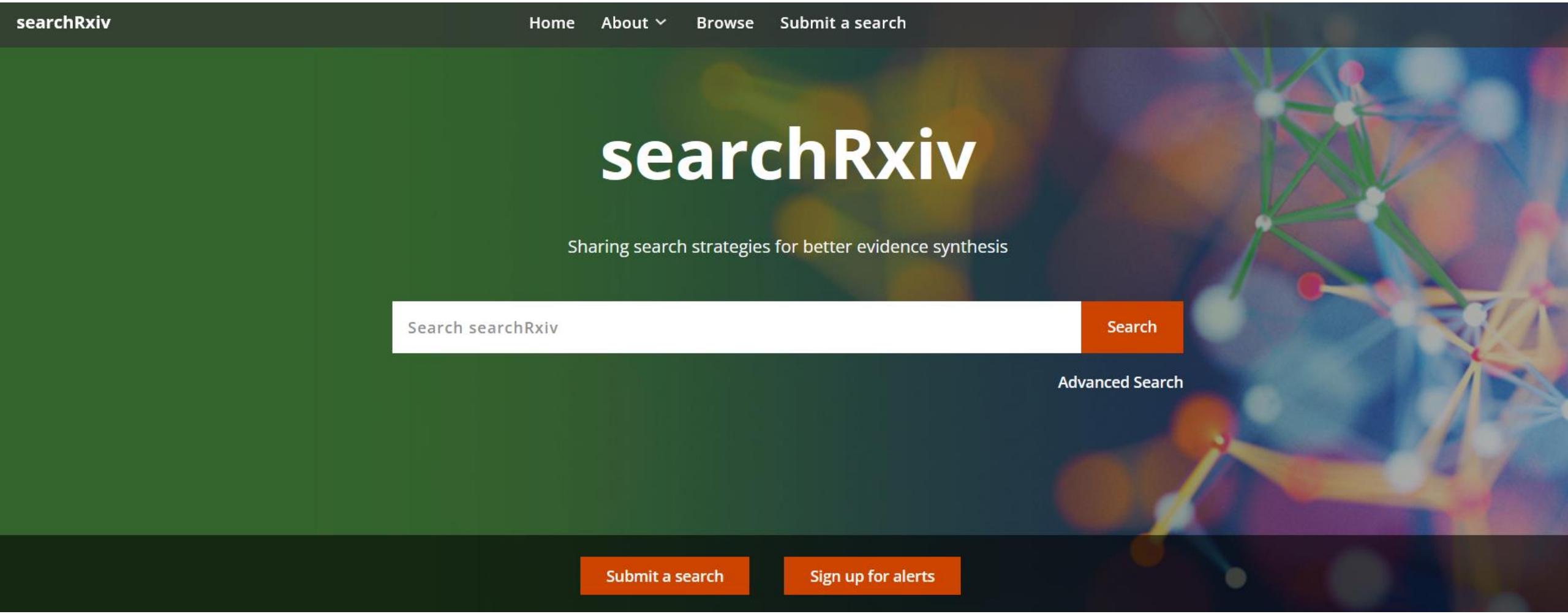
Adjusting your search: you might have to broaden or narrow down your scope

If your search is too narrow:

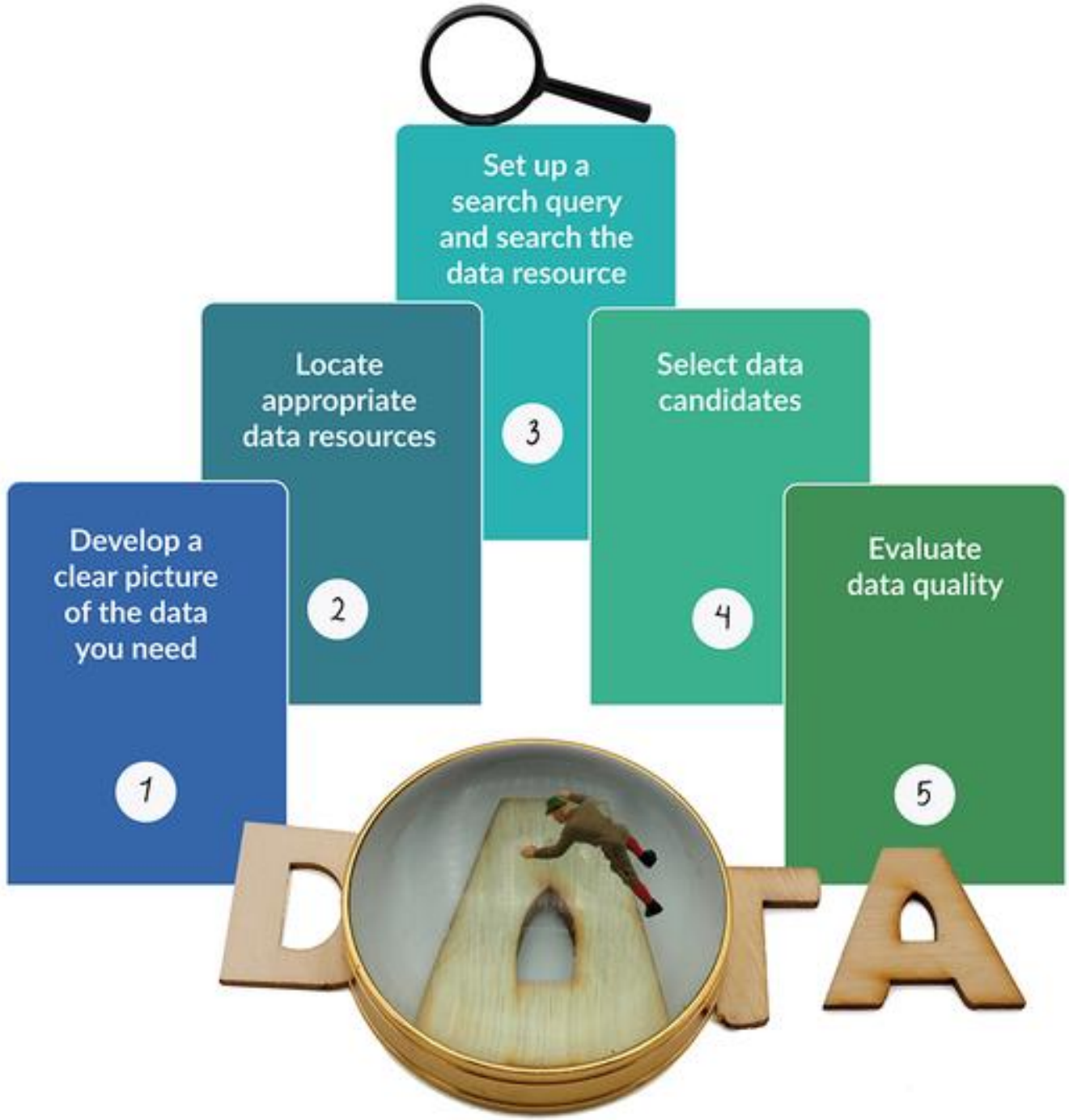
- Check your spelling
- Use more general search terms
- Turn off some of the filters you applied
- Use more synonyms

If your search is too broad:

- Use more specific search terms
- Use more search terms
- Use more filters
- Check the use of boolean logic (is it applied correctly?)



https://www.cabidigitallibrary.org/journal/searchrxiv

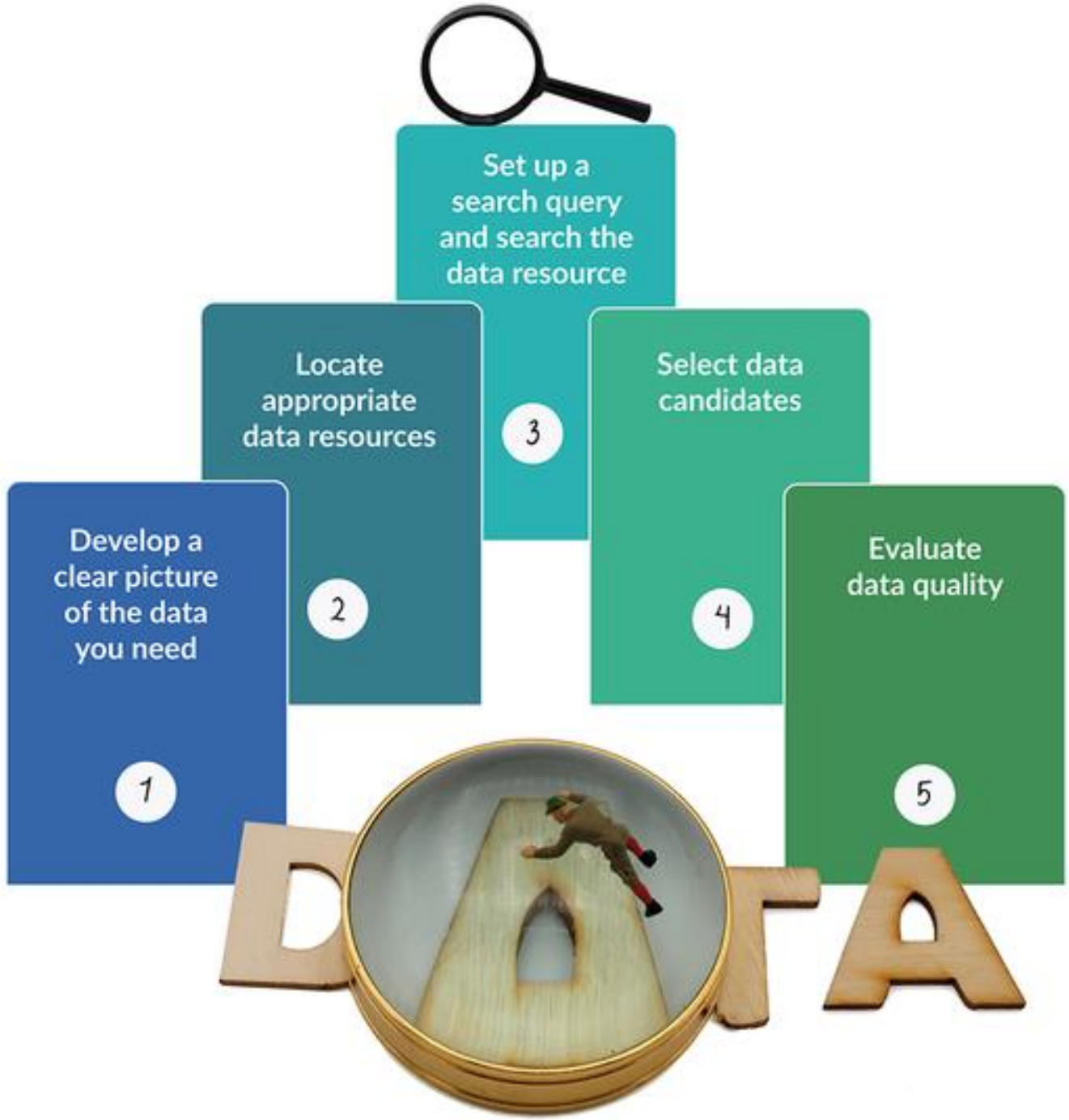

Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

Select data candidates

Can I use these data?

- Are the data relevant to your research questions?
- Are the concepts appropriate?
- Are the variables and the indicators appropriate?

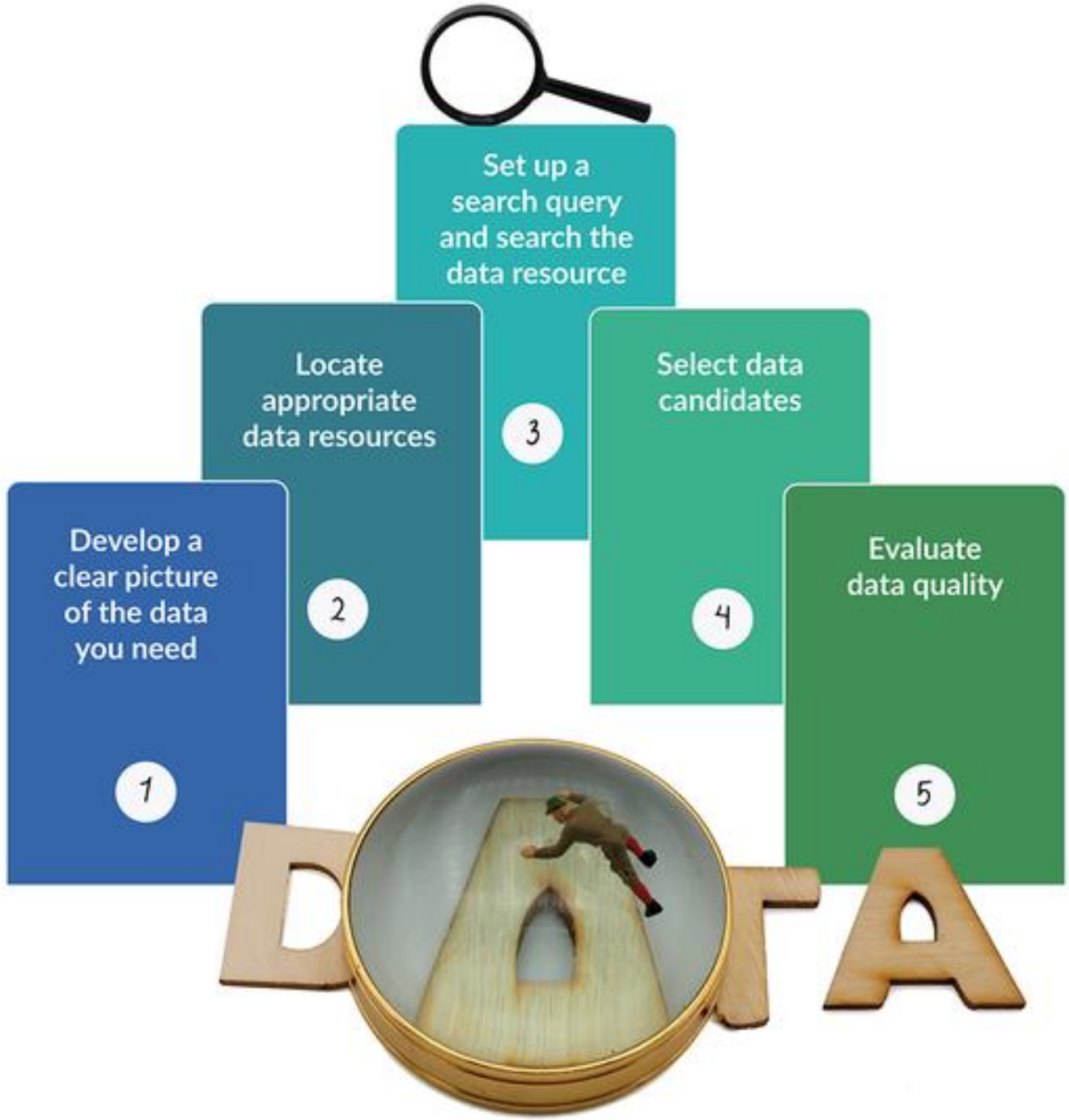
*Check dataset **documentation** (e.g. README files, data dictionaries or codebooks) very carefully!


Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

Evaluate data quality

What is the quality of the data?

- What information was collected, from whom, when and where?
- Who collected the data and when?
- Why was the data created? (research, social policy, marketing?)
- How was the data collected? (methodology)
- How was the data processed? Were there any changes in data?
- Is the data “clean” (were nonlogical and erroneous values deleted?)
- What quality assurance procedures were used? Did researchers use verified measurement tools?


Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

Other considerations

Access the data: is it free? Do I need to register? Is the access restricted? Do I need to apply to get access?

Data format: is the format of the files correct for your analyses? Do you need to transform the files or the dataset?

Missing data: are there any missing data in the dataset? How are you going to handle missing data?

Cite the data

Harvard citation style:
Author names. Year. Title of resource. [medium type]. Host institution name, Physical location. Date of access. Identifier

Vancouver citation style
Author names. Title of resource [medium type]. Host institution name: Physical location; Year of publication. [Date accessed]. Available from: Identifier

Cite the data

Harvard citation style example:
Scarrow, S., Webb, P., Poguntke, T., 2017, Political Party Database, 2011-2014, [data collection], UK Data Service, Accessed 17 October 2018. SN: 8265, http://doi.org/10.5255/UKDA-SN-8265-1

Solvang, Øystein; Stein, Jonas; Brattland, Camilla, 2020, "Covid-19 Municipal Level (Norway) Social Science Dataset", https://doi.org/10.18710/NMKI2B, DataverseNO, V2

Cite Dataset

EndNote XML

RIS

BibTeX

Learn about Data Citation Standards.

The dataset is a cross-sectional dataset covering social and public health data pertaining
Norwegian municipalities. The dataset was compiled from public register data and media
related fatalities is current as of ultimo July 2020. Data on other variables is from 2018, 2

Document what you find and what you do!

Examples

https://dataverse.no/

Case 1

Dataset: Covid-19 Municipal Level (Norway) Social Science Dataset

https://search.nsd.no/

Case 2

Dataset: Human Bodily Micromotion in Music Perception and Interaction

NORSK SENTER FOR FORSKNINGSDATA

UNIVERSITY OF OSLO

Materials developed as a part of the *Skills development for research data* project:
https://www.ub.uio.no/english/about/projects/rdm-skills/

Thank you!

Contact us at:

research-data@uio.no

18. November 2022

Open Science Lunch

Open Science Lunch

Each last Thursday of the month at 12:00 we invite you to join us for a lunch seminar to hear about how to make your research more open. We will discuss research transparency and visibility, open publishing, data sharing, and more!

Upcoming

Time and place: Nov. 24, 2022 12:00 PM–1:00 PM, Hybrid: Georg Sverdrups hus and Zoom

Open Science Lunch: CRediT your co-authors

Learn about CRediT - a new international standard for transparent assignment of individual research contributions.

https://www.ub.uio.no/english/courses-events/events/open-science-lunch/

Sources

CESSDA. The process of data discovery. https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/7.-Discover

UCL Data discovery & re-use: https://www.ucl.ac.uk/library/research-support/research-data-management/best-practices/how-guides/data-discovery-re-use

Gould Library: Data, Datasets and Statistical Resources https://gouldguides.carleton.edu/c.php?g=146834&p=964067

MacInnes, J. (2020). Secondary Analysis of Quantitative Data. In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), *SAGE Research Methods Foundations.* https://www.doi.org/10.4135/9781526421036870195

Learn how to use the boolean operators in search queries: https://www.youtube.com/watch?v=lEo96kOKGmA