UiO : University of Oslo Library

Data Discovery

Agata Bochynska and Elin Stangeland
Open Research group, University of Oslo Library
Ivana Malovic
Library of Medicine and Science

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

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

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

Data discovery is finding, accessing, and reusing 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.

GO TO: menti.com

TYPE IN: 3733 1292

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)

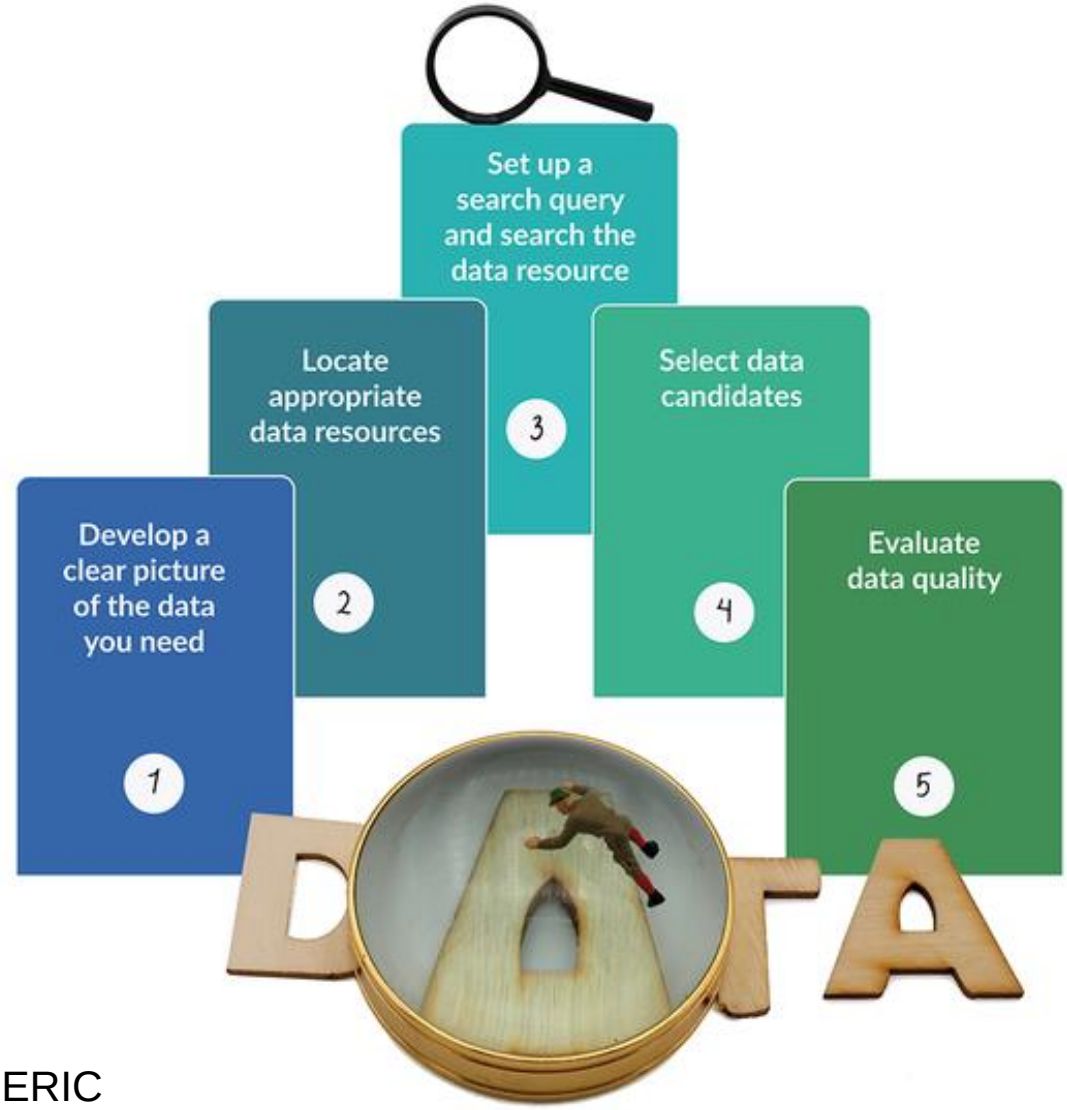
1 zettabyte is equal to
1 billion terabytes.

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

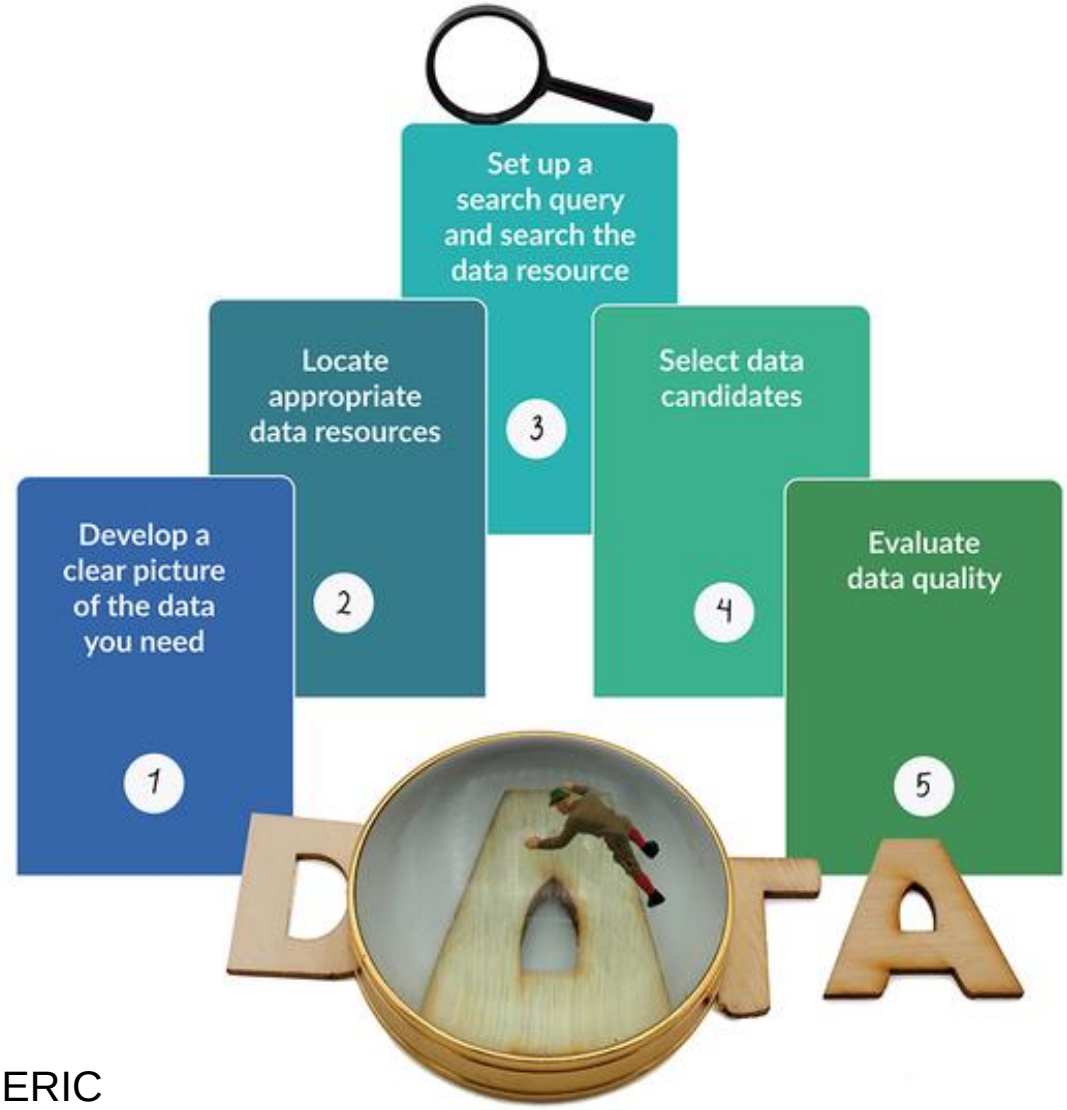
Develop a clear picture of the data you need

GO TO: menti.com

TYPE IN: 3733 1292

Deciding on what kind of data you need

- What is the theme/domain you study?
- What is your research question?
- What are the constructs you want to work with?
- How will you operationalize the constructs?
- What is your theory?
- What study will you perform?
- What specific characteristics should the data have?
- Do you have other preconditions?


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

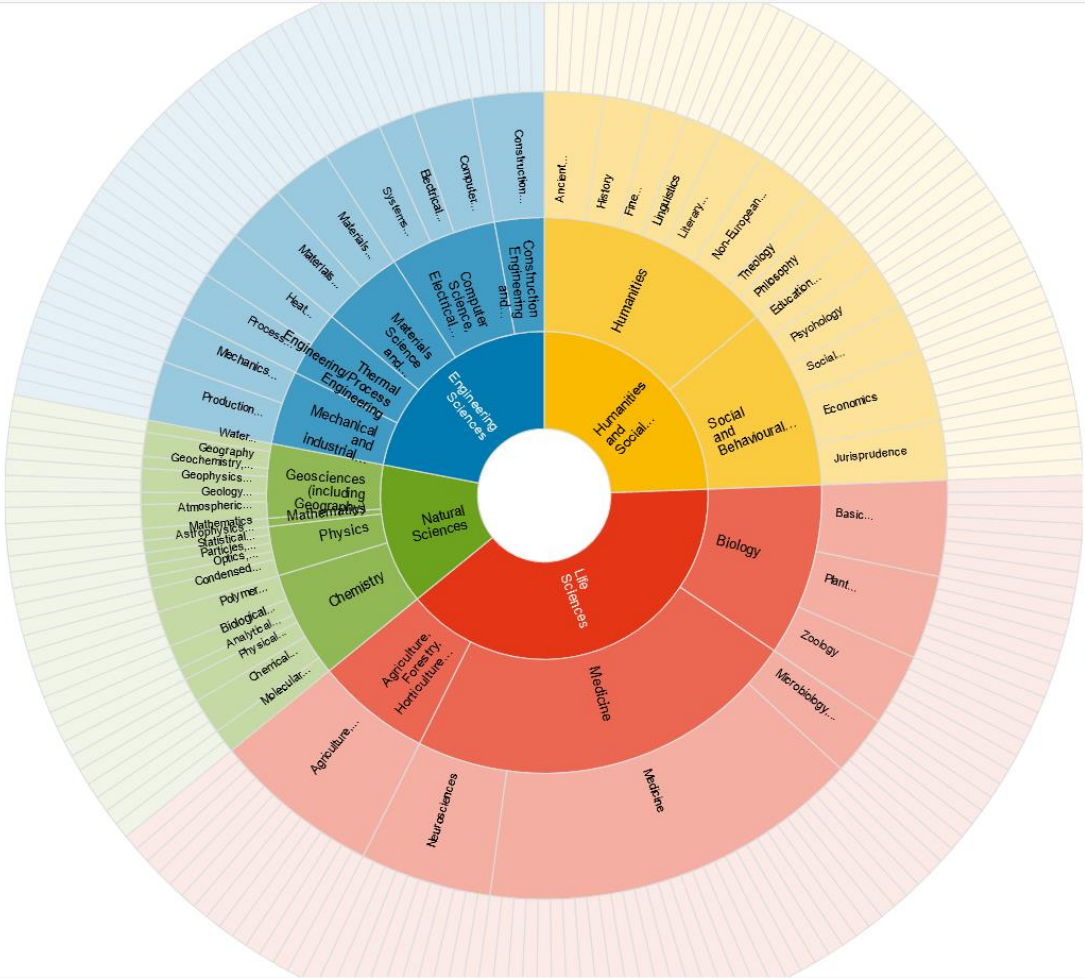
Locate appropriate data resources

Where do I look for the data?

- A registry of data repositories
- A search engine or (meta)data aggregator
- A data catalogue
- A data journal

Discipline-specific repositories

https://www.re3data.org/



Browse by subject

Filter

Reset all

Subjects

- Humanities and Social Sciences (3)
  - Humanities (1)
    - Ancient Cultures (1)
      - Classical Archaeology (1)
      - Egyptology and Ancient Near Eastern Studies (1)
    - History (1)
    - Fine Arts, Music, Theatre and Media Studies (1)
  - Social and Behavioural Sciences (1)
- Life Sciences (85)
  - Biology (85)
    - Basic Biological and Medical Research (85)
      - Biochemistry (7)
      - Biophysics (2)
      - **Cell Biology (85)**
      - Structural Biology (13)
      - General Genetics (33)
      - Bioinformatics and Theoretical Biology (26)
      - Anatomy (1)
    - Plant Sciences (10)
      - Plant Ecology and Ecosystem Analysis (1)
      - Plant Biochemistry and Biophysics (3)
      - Plant Cell and Developmental Biology (1)
      - Plant Genetics (7)
    - Zoology (13)
      - Animal Ecology, Biodiversity and Ecosystem Research (1)
      - Animal Genetics, Cell and Developmental Biology (12)
  - Medicine (45)
    - Microbiology, Virology and Immunology (21)
      - Metabolism, Biochemistry and Genetics of Microorganisms (5)

Search... Search

Toogle short help

← Previous 1 2 3 4 Next →

Sort by

Found 85 result(s)

STRING

Known and Predicted Protein-Protein Interactions

Subject(s)	Life Sciences · Biology · Cell Biology · General Genetics · Bioinformatics and Theoretical Biology · Basic Biological and Medical Research
Content type(s)	Raw data · Structured graphics · Plain text · Archived data · Networkbased data · other
Country	Germany · European Union · Switzerland · Denmark

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources: - Genomic Context - High-throughput Experiments - (Conserved) Coexpression - Previous Knowledge STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable.

IMEx

The International Molecular Exchange Consortium

Subject(s)	Cell Biology · Microbiology, Virology and Immunology · Basic Biological and Medical Research · Biology · Life Sciences · Medicine
Content type(s)	Scientific and statistical data formats · Networkbased data · Structured text · Structured graphics · other
Country	United Kingdom · European Union · United States · Italy · India · Canada · France · Singapore · International

REMOTE SENSING
PALEO
MODEL OUTPUT
OCEANOGRAPHY
ATMOSPHERE
DATA PRODUCTS

NEWS

Currently there are no events.

>> BCDC Home

DATA PUBLICATION HIGHLIGHTS

12 November 2018

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

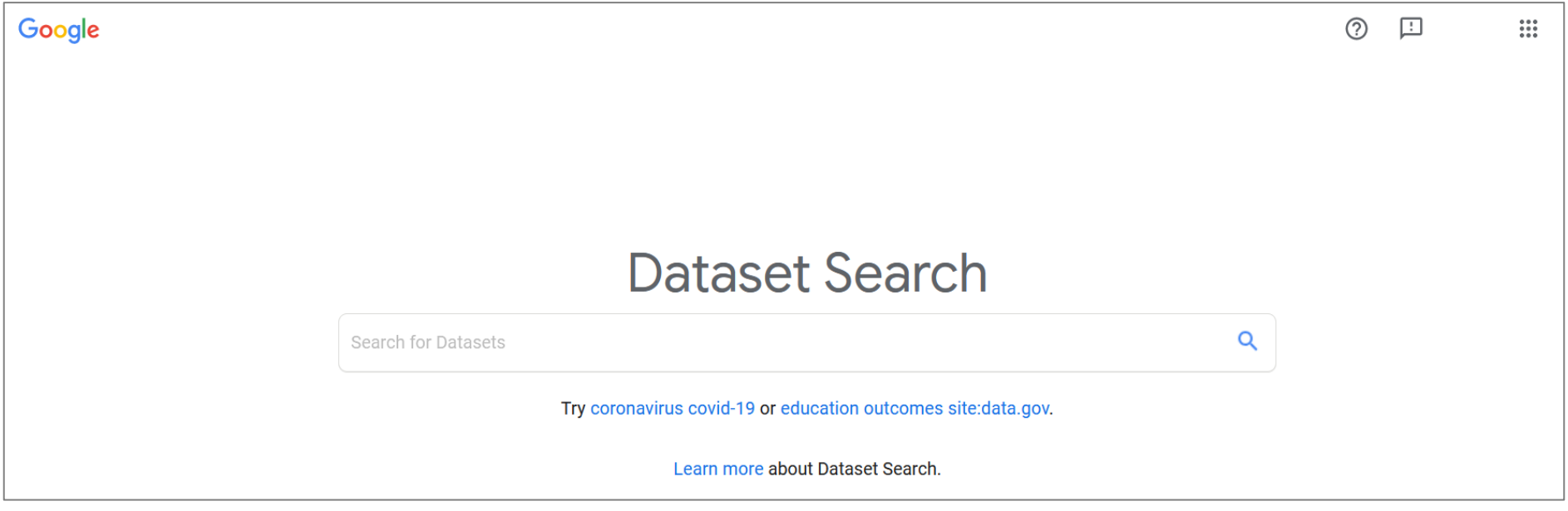
Sessford et al. have performed a high-resolution, multi-proxy analysis of the sediment core GS15-198-36CC and recently published their results.

The core was retrieved from the northern bank of the Greenland-Iceland Ridge, and exhibits a 30-year temporal resolution during the Dansgaard-Oeschger events 8-5 (40-30 ka). Multiple proxy records were measured for this time period: magnetic susceptibility, oxygen and carbon isotopes for *N. pachyderma* and *C. neoteretis*, absolute abundance of benthic species *C. neoteretis* and *E. excavatum*, and trace element ratios of benthic species *C. neoteretis*.

Discipline-general repositories

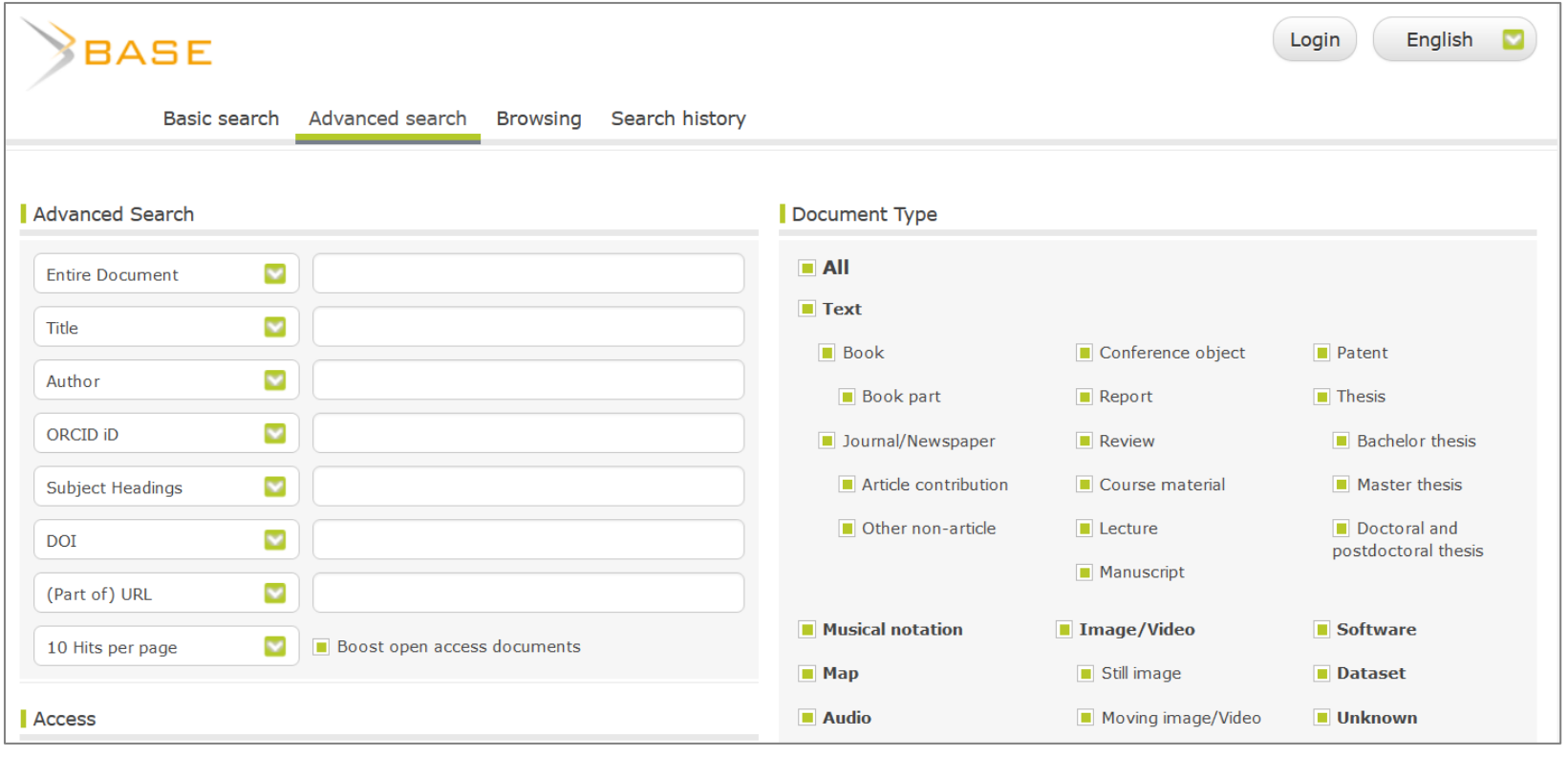
UiT Open Research Data

Search engines



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

Search engines



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

Data journals

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

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

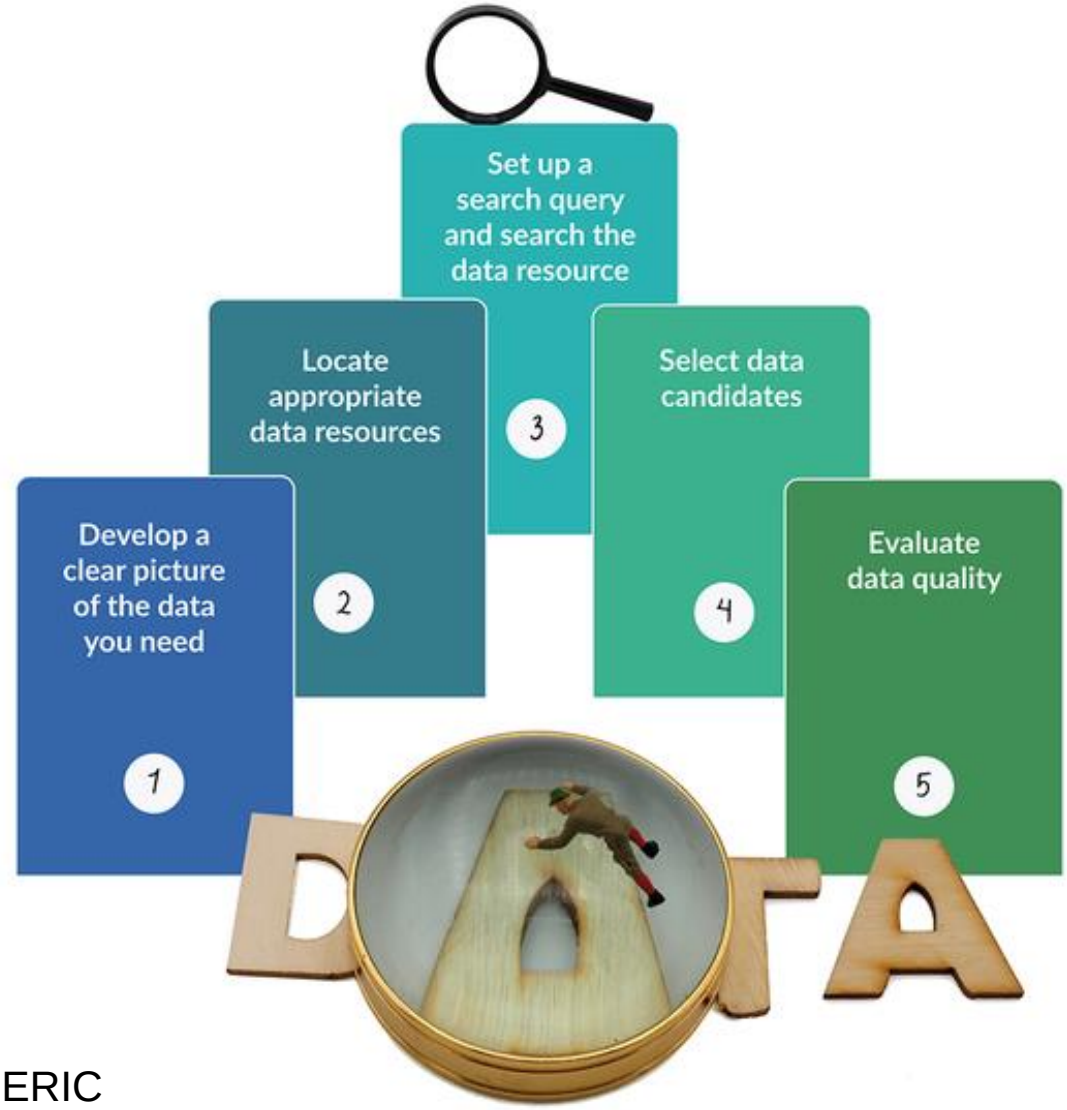
Open Access

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 ▸


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.

GO TO: menti.com

TYPE IN: 3733 1292

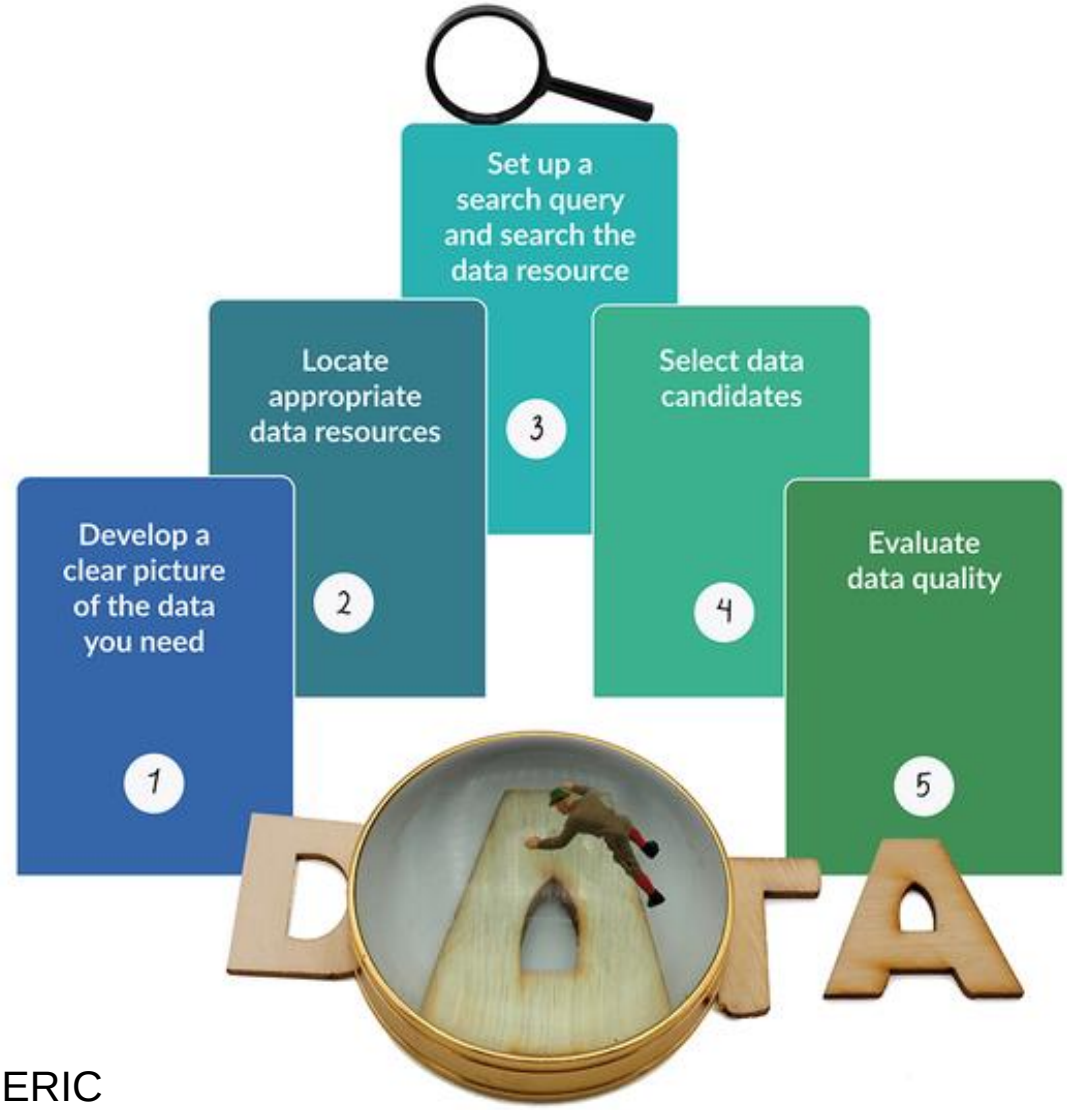
Adjusting your search: you might have to broaden or narrow down your scope

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?)

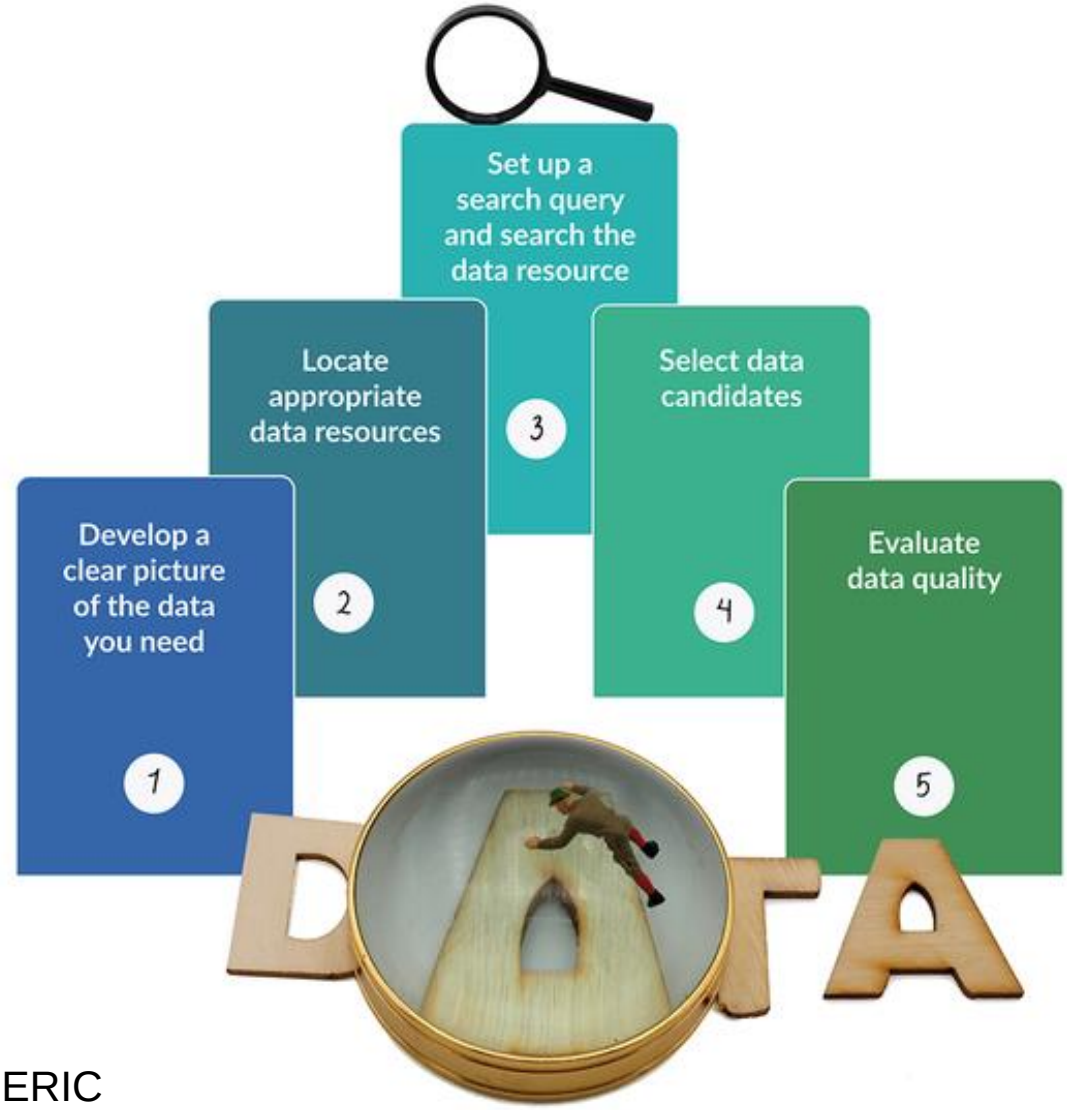

Develop a clear picture of the data you need
1
Locate appropriate data resources
2
Set up a search query and search the data resource
3
Select data candidates
4
Evaluate data quality
5
DATA

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?

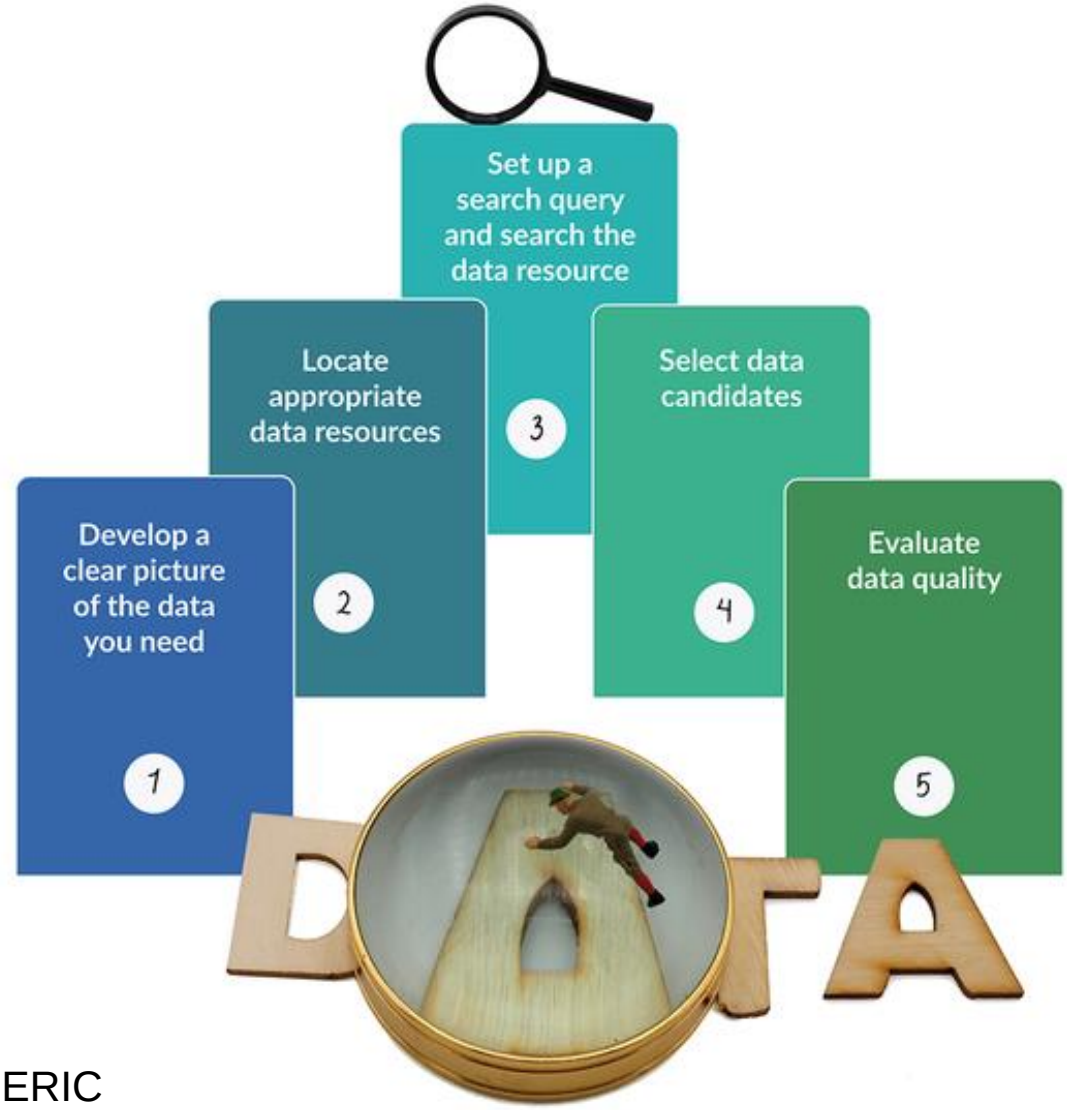
*Check dataset documentation very carefully!


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?

Document what you find and what you do!

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

https://search.nsd.no/

Case 1

Dataset: Human Bodily Micromotion in Music Perception and Interaction

NORSK SENTER FOR FORSKNINGSDATA

Case 2

Replication data for: What makes a word easy to acquire?

https://doi.org/10.1098/rstb.2015.0219

Case 3

Discover and reuse the data from literature!

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

BIOLOGICAL SCIENCES

Open Access

Check for updates

View PDF

Tools Share

Cite this article

Research article

How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model

Jostein Starrfelt and Lee Hsiang Liow

Published: 05 April 2016 https://doi.org/10.1098/rstb.2015.0219

Case 4

Discover qualitative data: NorFisk Dataset

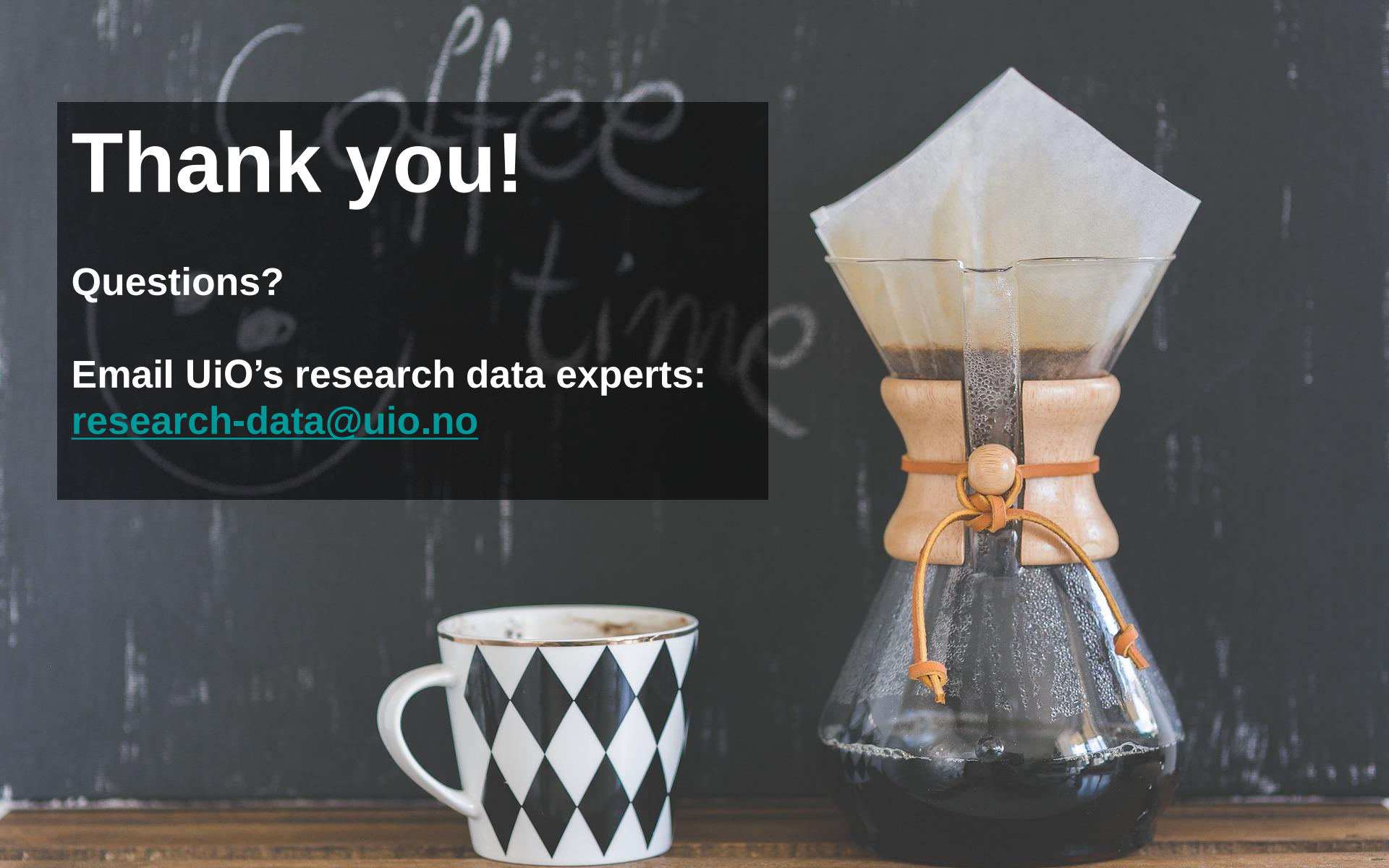
DataverseNO
Dataverse Network Norway

NTNU Open Research Data
NTNU

Thank you!

Questions?

Email UiO's research data experts:
research-data@uio.no



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