



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

Data Discovery and Reuse

Agata Bochynska, PhD

Open Research group, University of Oslo Library

Ivana Malovic, PhD

Library of Medicine and Science

27. October 2021





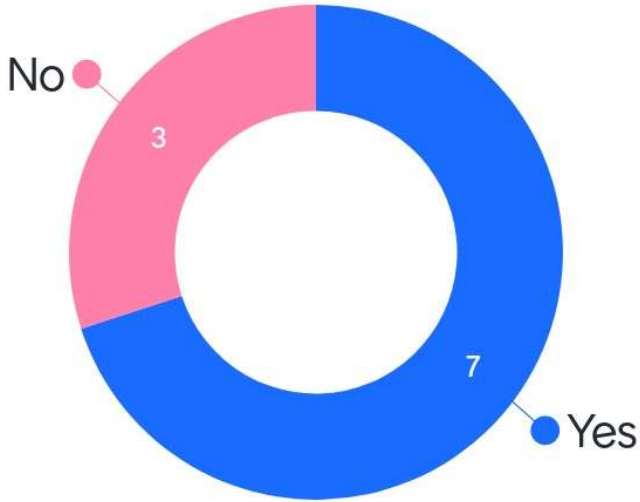
Data discovery is finding and accessing data collected for a different purpose or by a different researcher or institution.

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

GO TO: menti.com

TYPE IN: 1416 1185

Have you ever done a secondary data search and/or analysis?



Why do you think data reuse is important?

validation

saves resources, time and money

benchmarking

Save from reinventing the wheel

More data to your sample

Save a lot of time

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

The Research Council Policy for Open Science

In effect from 2020



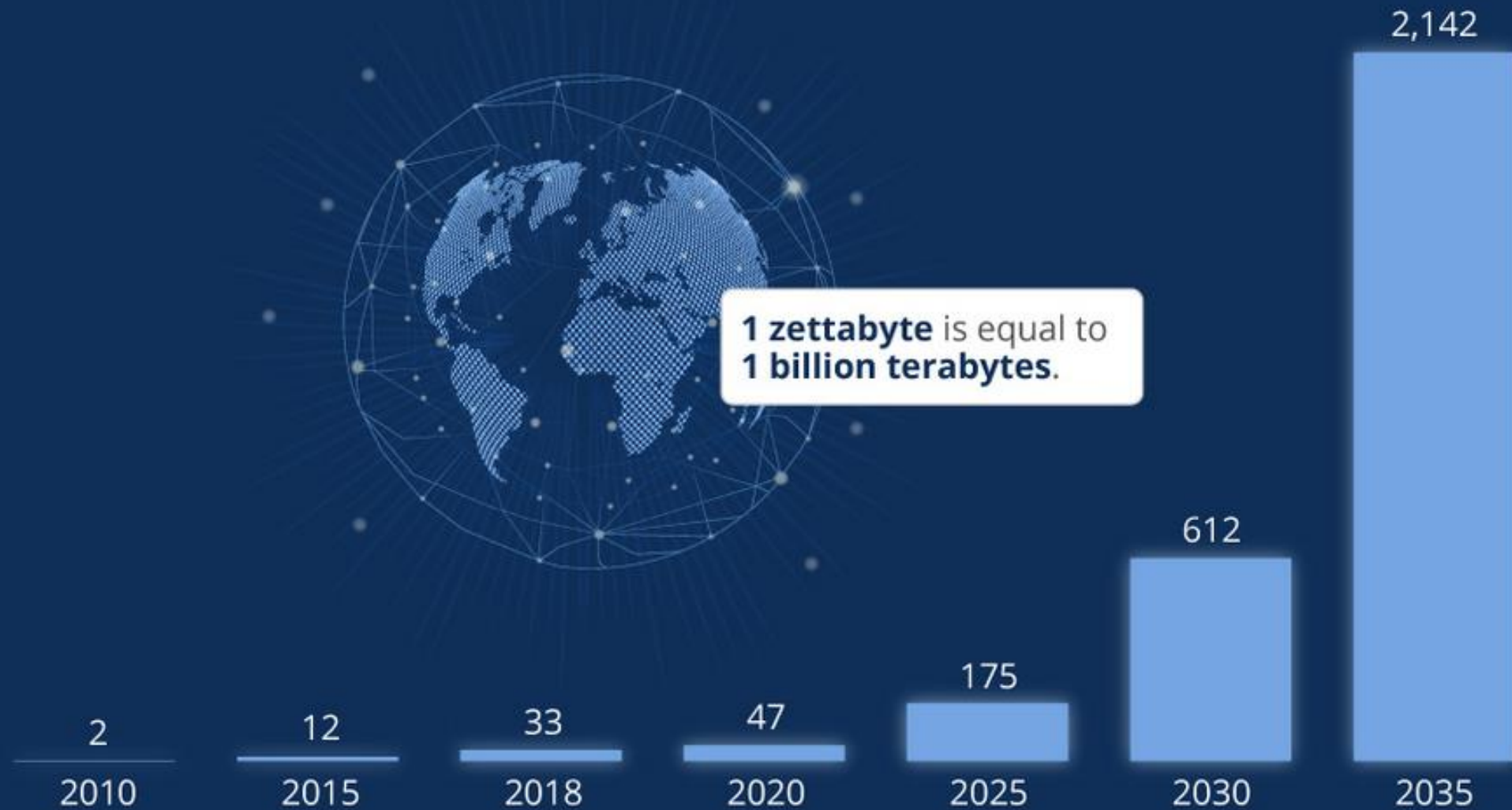


“**Open Science** is becoming the modus operandi for carrying out research and innovation by **sharing knowledge, data and tools** as early as possible, in open collaboration with all relevant knowledge actors and society.”

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)



@StatistaCharts

Source: Statista Digital Economy Compass 2019

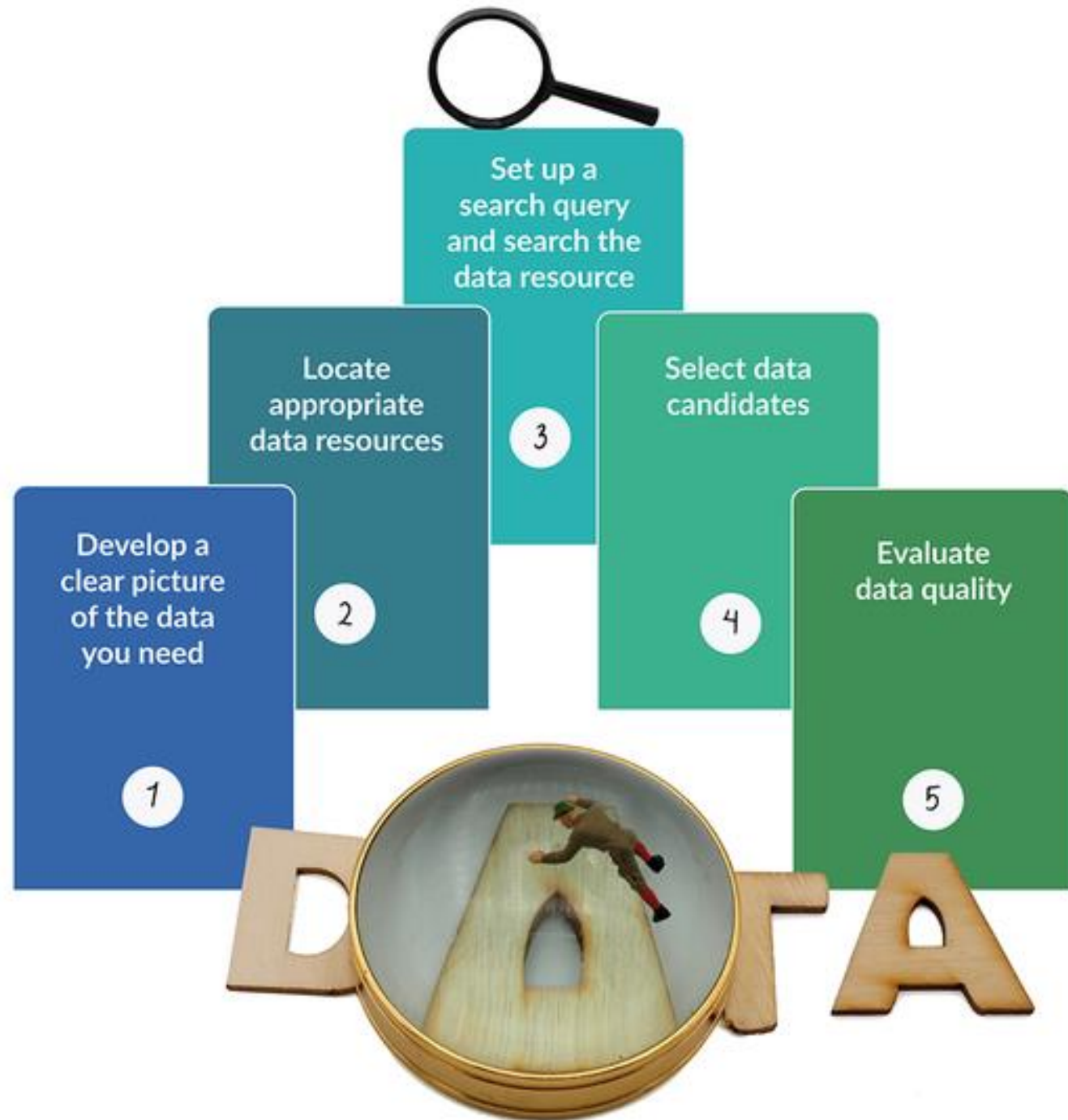
statista

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

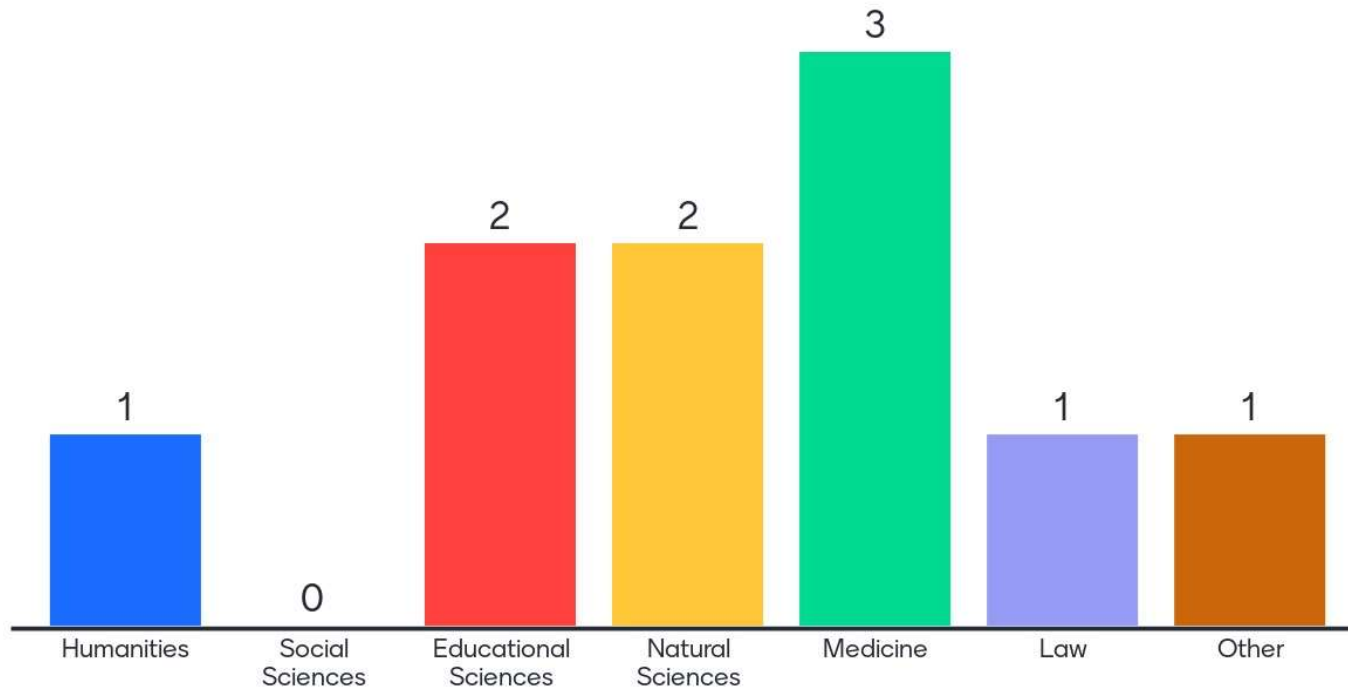


Develop a clear picture of
the data you need

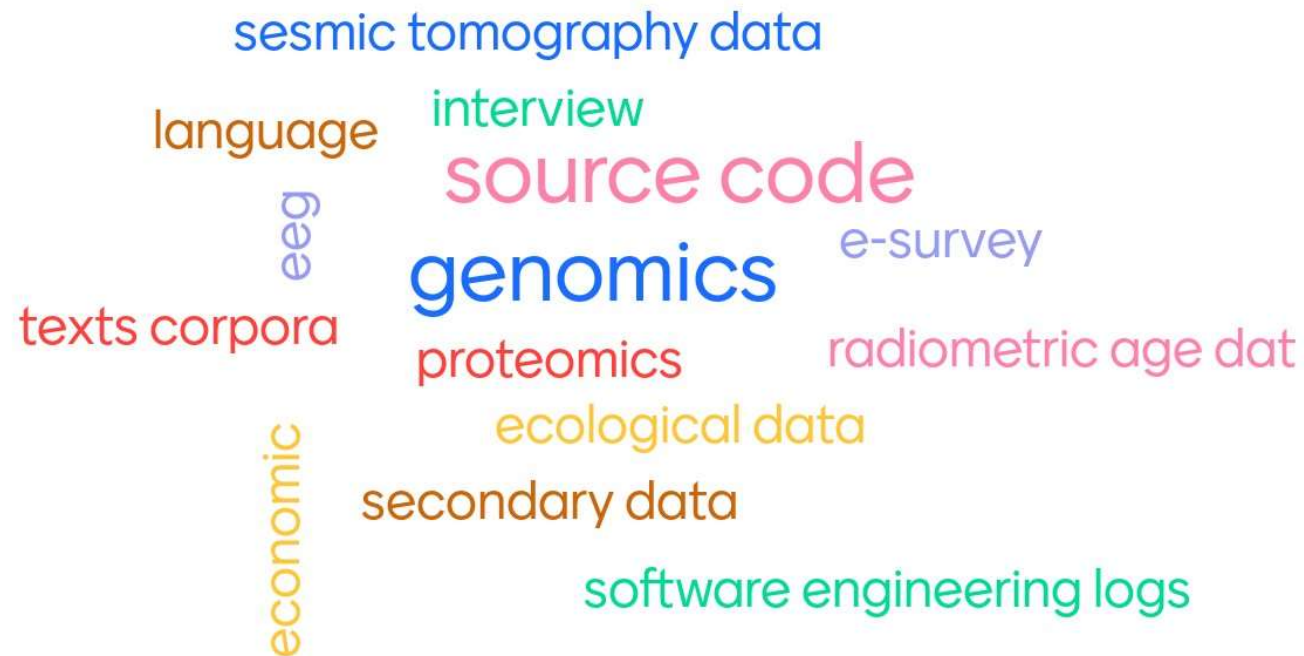
GO TO: menti.com

TYPE IN: 1416 1185

What research area would you be looking for your data in?

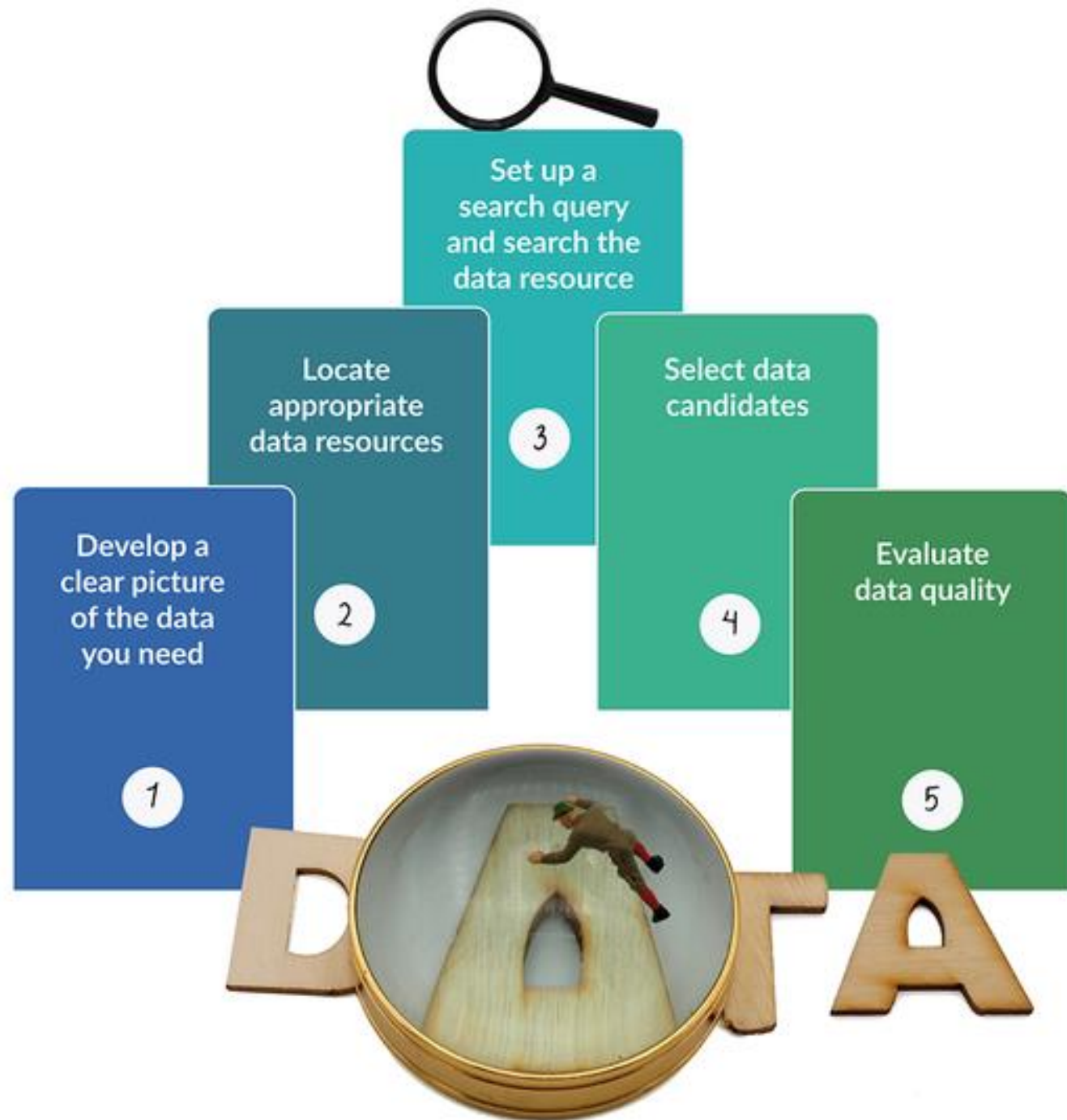


What types of data you think you would look for in your project?



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?



Locate appropriate data resources

Where do I look for the data?

- A registry of discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

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

Discipline-specific repositories

Search

Browse ▾

Suggest

Resources ▾

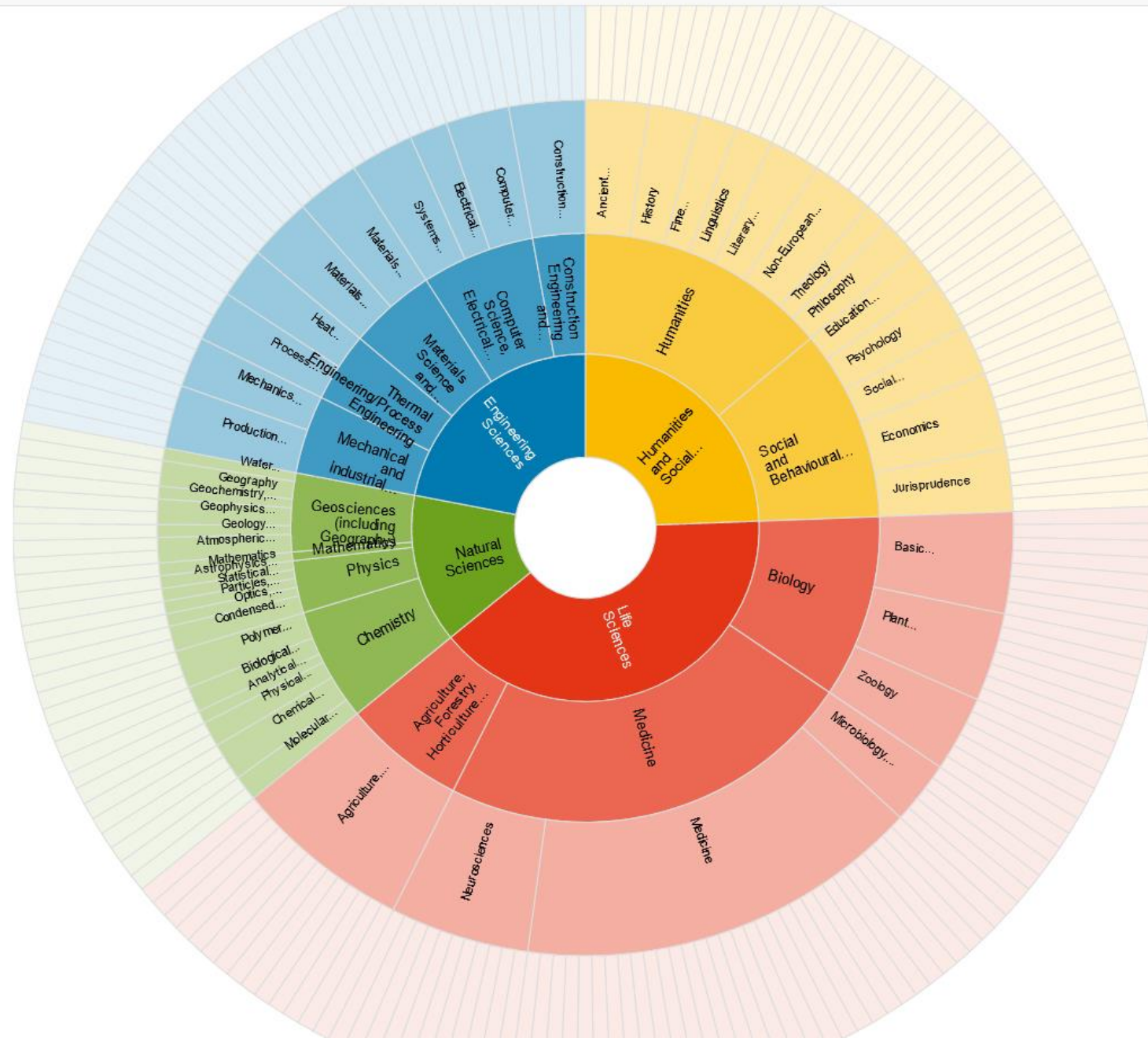
Contact



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

🔍 Search



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

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

Toggle 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.

Filter

[Reset all](#)**Subjects** **Humanities and Social Sciences (3)****Social and Behavioural Sciences (3)****Psychology (3)**

General, Biological and Mathematical Psychology (1)








Developmental and Educational Psychology (3)

Social Psychology, Industrial and Organisational Psychology (2)

Differential Psychology, Clinical Psychology, Medical Psychology, Methodology (2)

Social Sciences (2)

Empirical Social Research (1)

Content Types **Countries** **AID systems** **API** **Certificates** **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** 

Filter

[Reset all](#)

Subjects ▾

Humanities and Social Sciences (3)

Social and Behavioural Sciences (3)

Psychology (3)

General, Biological and Mathematical Psychology (1)

Developmental and Educational Psychology (3)

Social Psychology, Industrial and Organisational Psychology (2)

Differential Psychology, Clinical Psychology, Medical Psychology, Methodology (2)

Social Sciences (2)

Empirical Social Research (1)

Content Types ▾

Countries ▾

AID systems ▾

API ▾

Certificates ▾

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

[Toggle short help](#)

← Previous 1 Next →

Sort by ▾

Found 3 result(s)

Databrary



Subject(s) Developmental and Educational Psychology Psychology Social and Behavioural Sciences Humanities and Social Sciences

Content type(s) Audiovisual data Images Software applications

Country United States

Databrary is a data library for researchers to share research data and analytical tools with other investigators. It is a web-based repository for open sharing and preservation of video data and associated metadata in the area of developmental sciences. The project aims to increase the openness in scientific research and dedicated to transforming the culture of developmental science through building a community of researchers empowering them with an unprecedented set of tools for discovery. Databry is complemented by Datavyu (an open source video-coding software) as well as Labnanny (a data management system to enable data-sharing). So any contributor can share raw digital video files, other data streams.

PsychData



Research Data for Psychology

Subject(s) Psychology Social Sciences Social and Behavioural Sciences Humanities and Social Sciences Developmental and Educational Psychology

Social Psychology, Industrial and Organisational Psychology Differential Psychology, Clinical Psychology, Medical Psychology, Methodology

Content type(s) Scientific and statistical data formats Plain text Standard office documents

Country Germany

Goal of the psychology data archive PsychData is the documentation and long-term archiving of research data from all areas of psychology and the social sciences, using specially created metadata and to provide use of the data for scientific purposes such as secondary analysis and reanalysis. Psychdata contains all areas of psychology, in particular data sets from clinical, developmental, educational, gero-, and work and organizational psychology stemming from longitudinal studies, major surveys, and test development.

Repository of Psychological Instruments in Serbian



REPOPSI

Subject(s) Humanities and Social Sciences Social and Behavioural Sciences Psychology General, Biological and Mathematical Psychology Developmental and Educational Psychology

Social Psychology, Industrial and Organisational Psychology Differential Psychology, Clinical Psychology, Medical Psychology, Methodology Social Sciences Empirical Social Research

Content type(s) Standard office documents Databases Scientific and statistical data formats Plain text

Country United States Serbia

The Repository of Psychological Instruments in Serbian (REPOPSI; <https://osf.io/5zb8p/>), run by the Laboratory for Research of Individual Differences at the University of Belgrade and hosted on the Open Science Framework, is an open-access repository of psychological instruments. REPOPSI is a collection of psychological measures, scales, tests, and other research instruments commonly used in social and behavioral science research. Documented are

Bjerknes Centre

for Climate Research

BCDC Bjerknes Climate Data Centre



ABOUT

DATA SEARCH

PROJECTS

SUBMIT

PARTNERS

NEWS

CONTACT

REMOTE SENSING

PALEO

MODEL OUTPUT

OCEANOGRAPHY

ATMOSPHERE

DATA PRODUCTS

NEWS

Currently there are no events.

<https://www.bcdc.no/>

>> 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*.



OPEN POLAR

The Global Open Access Portal for
Research Data and Publications on
the Arctic and Antarctic

Search this website...



HOME

ABOUT

SUGGEST & CONTRIBUTE

RELATED PROJECTS

OPEN POLAR TEAM

FEEDBACK & CONTACT

NEWS



Check out and Search
the Open Polar



Suggest and contribute to
the Open Polar



Team

Tweets by @Open_Polar

Open Polar Retweeted



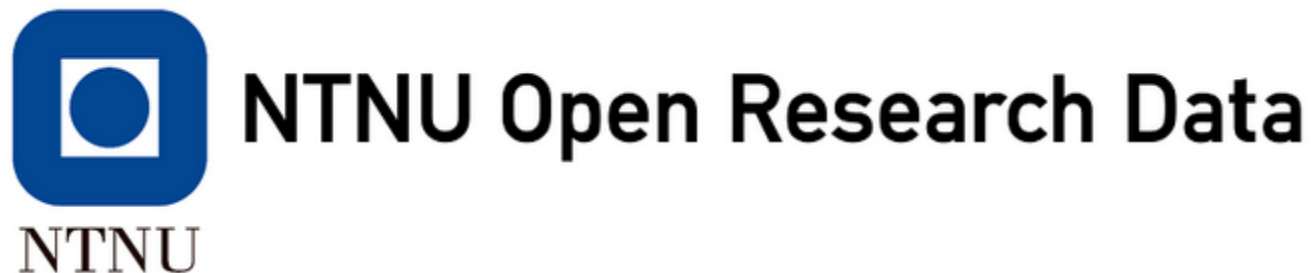
Where do I look for the data?

- A registry of discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

- A registry of 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?

- A registry of discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

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

Search engines

Google



Dataset Search

Search for Datasets



Try [coronavirus covid-19](#) or [education outcomes site:data.gov](#).

[Learn more](#) about Dataset Search.

Search engines

The screenshot displays the BASE search engine interface. At the top left is the BASE logo. On the top right, there are buttons for 'Login' and 'English' with a dropdown arrow. Below the logo, there are navigation tabs: 'Basic search', 'Advanced search' (which is underlined), 'Browsing', and 'Search history'. The main content area is divided into two columns. The left column is titled 'Advanced Search' and contains several search criteria, each with a dropdown menu and a text input field: 'Entire Document', 'Title', 'Author', 'ORCID iD', 'Subject Headings', 'DOI', and '(Part of) URL'. At the bottom of this column, there are two more options: '10 Hits per page' and 'Boost open access documents'. The right column is titled 'Document Type' and lists various document types, each with a checkbox: 'All', 'Text' (with sub-items: Book, Book part, Journal/Newspaper, Article contribution, Other non-article, Conference object, Report, Review, Course material, Lecture, Manuscript, Patent, Thesis, Bachelor thesis, Master thesis, Doctoral and postdoctoral thesis), 'Musical notation', 'Image/Video' (with sub-items: Still image, Moving image/Video), 'Software', 'Map', 'Dataset', and 'Audio'.



Where do I look for the data?


- A registry of discipline-specific repositories
- General-purpose repositories
- A search engine or (meta)data aggregator
- A data journal

Where do I look for the data?

- A registry of 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](#) ▶

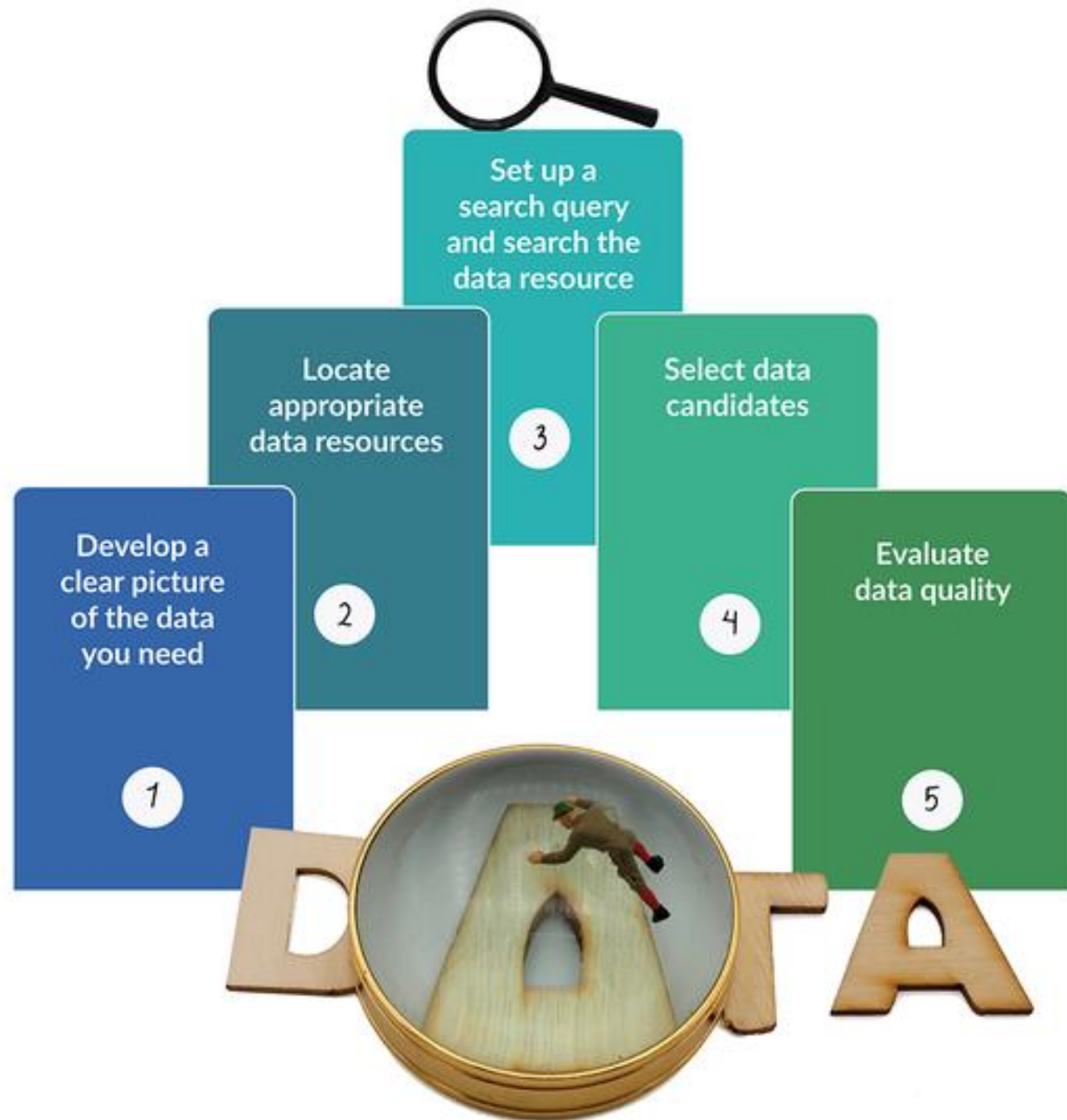


Co-Editors-in-Chief: Katherine Royse & Jian Peng
Impact factor: 2.714
2019 Journal Citation Reports (Clarivate Analytics): 74/200 (Geosciences, Multidisciplinary) 42/93 (Meteorology & Atmospheric Sciences)
Online ISSN: 2049-6060



[LATEST ISSUE >](#)

Volume 7, Issue 2
November 2020



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: 1416 1185

What would be an example search query for the data you need?

monolingual AND (bilingual OR multilingual)

"automated program repair" AND security

Liver endothelial cells OR LEC OR LECs OR LSEC
OR LSECs OR liver sinusoidal endothelial cells OR
liver scavenger endothelial cells

humanrights AND business

Genetic variation AND caryophyllaceae

Louisville hotspot AND track AND bathymetry
AND Cretaceous AND seamount

Metagenomics AND ectomychorrhiza

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?)



searchRxiv

Sharing search strategies for better evidence synthesis

Latest search strings

Covid-19 vaccination response in people taking immunosuppressants.

Covid-19 vaccination response in people taking immunosuppressants, especially Azathioprine.

Systematic reviews in high impact general internal medicine journals.

The search uses the 5 highest impact General and Internal Medicine journals that publish systematic reviews from the 2011 Journal Citation Reports in ...

Metal mining impacts on social ecological systems in boreal and Arctic regions.

Background: Mining activities, including prospecting, exploration, construction, operation,

Twitter feeds



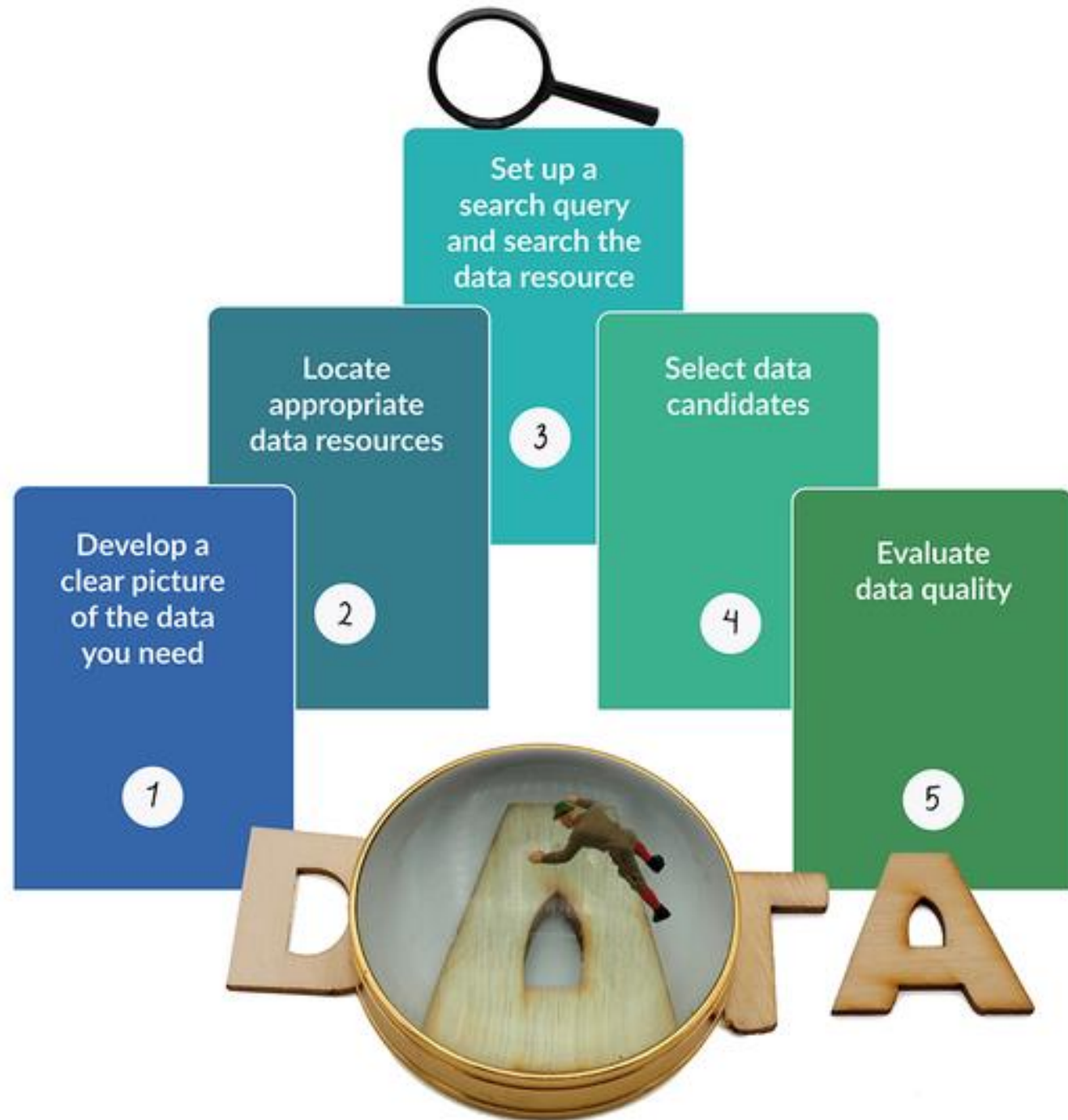
CABI Research Insights

[Follow](#)

CABI's official account for the research community. Follow us for the latest news, insights and discussions to support your research journey.

CABI Research Insights @CABI_Research · 21 Oct

What happens after you submit your grant proposal? 😊

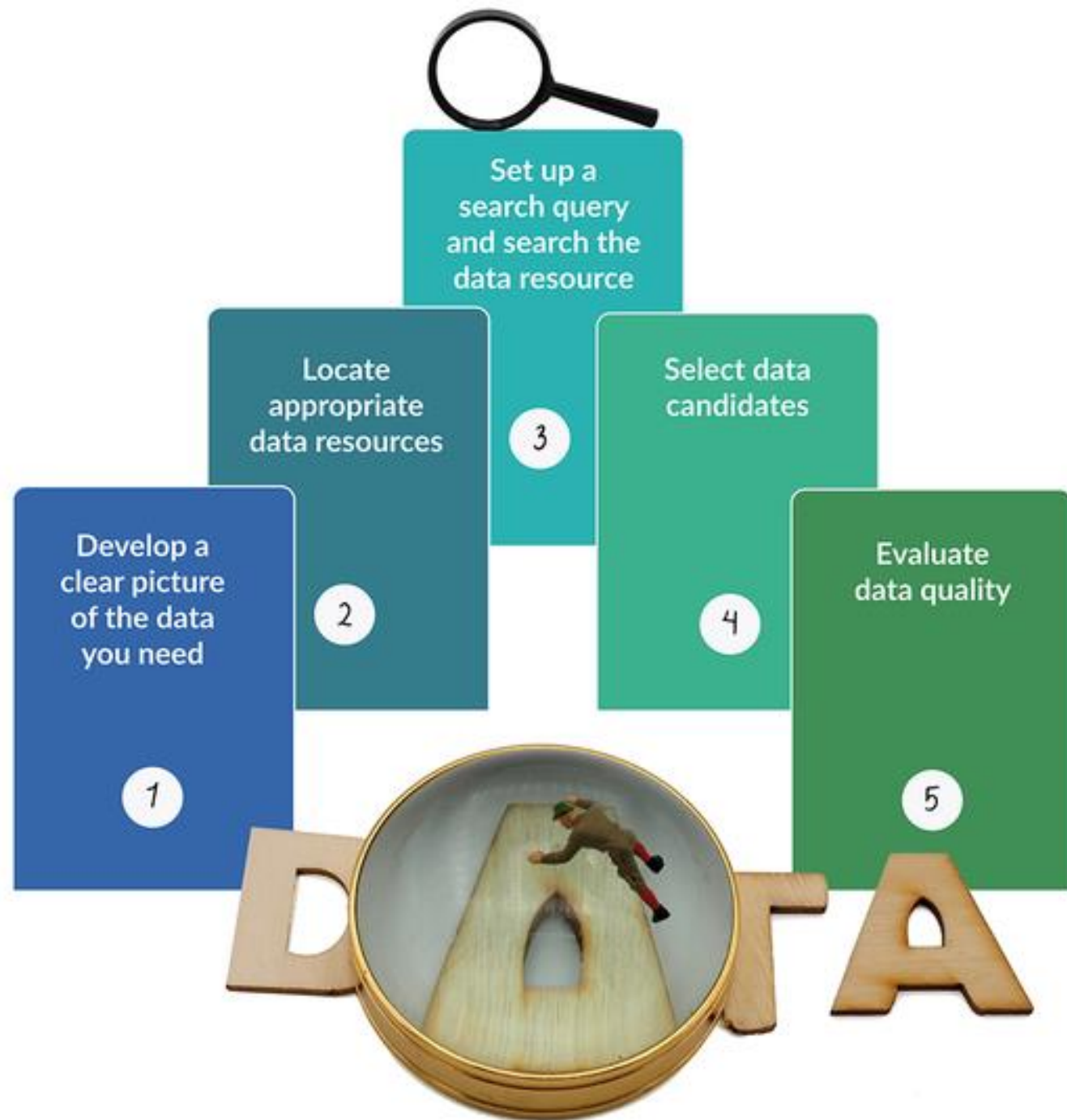


Select data candidates

Can I use these data?

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

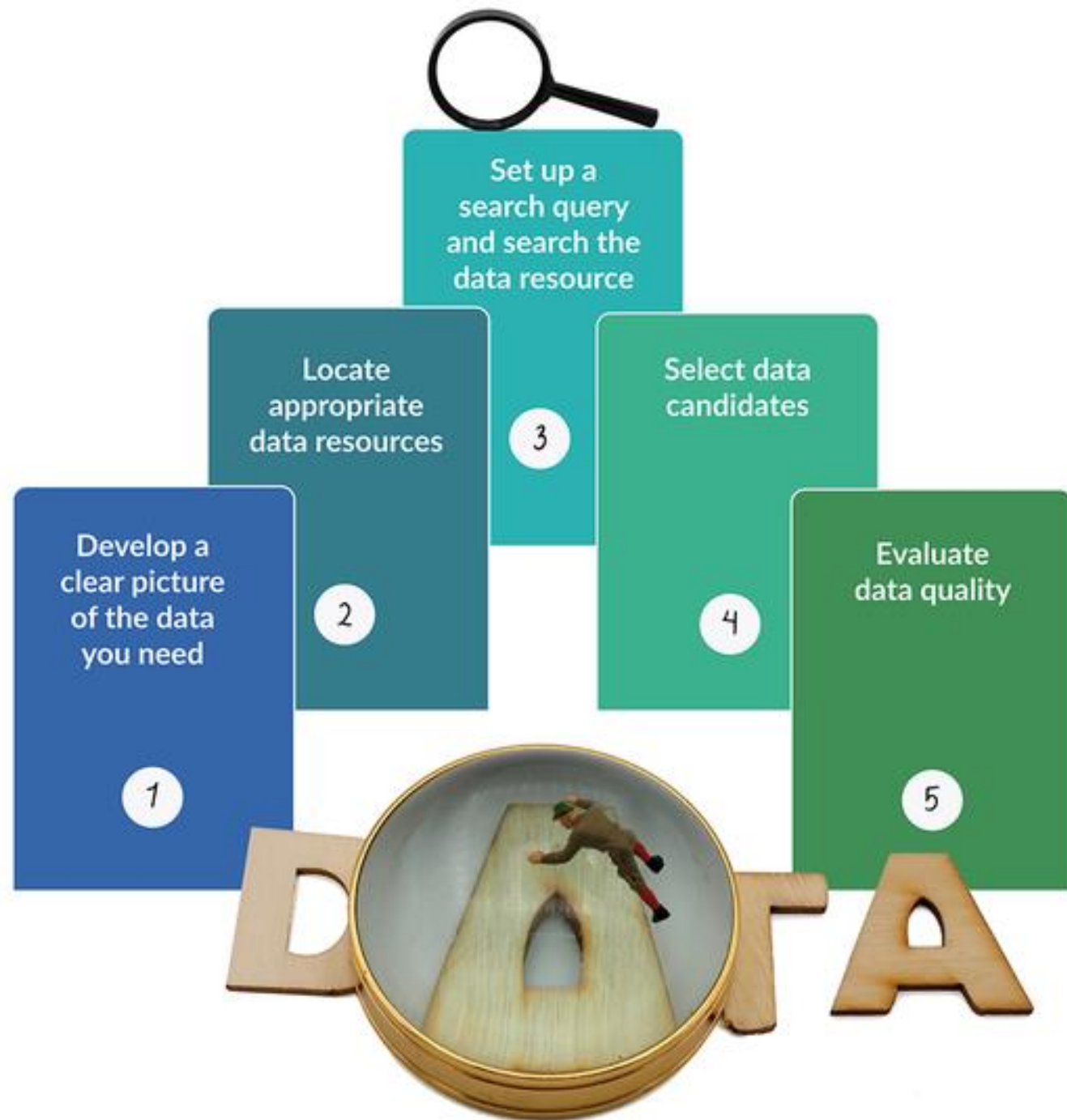
*Check dataset documentation very carefully!



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?



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>

Document what you find and
what you do!

Examples

Case 1

Dataset: Human Bodily Micromotion in Music Perception and Interaction



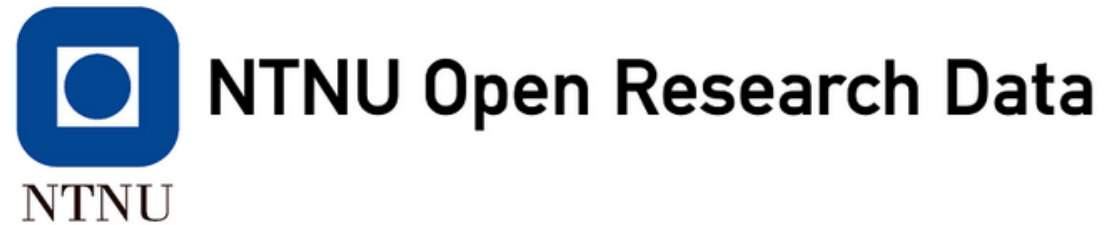
Case 2

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



Case 3

Discover qualitative data: NorFisk Dataset




Case 4

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>

Thank you!

Questions?

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

 OPEN ACCESS WEEK 2021

***It Matters How We Open Knowledge:
Building Structural Equity***

OCTOBER 25-31

Open Access Week 2021

Join us for a series of national and local events during the Open Access Week and learn about open research, open publishing, and data sharing!

Time: Oct. 25, 2021 9:00 AM–Oct. 29, 2021 4:00 PM

<https://www.uio.no/english/courses-events/events/all-libraries/2021/open-access-week-2021.html>

National Open Access Week events

Monday 25th - Friday 29th October

As part of this year's Open Access Week, several Norwegian academic libraries, together with [RDA \(Research Data Alliance\) Norway](#) and [ELIXIR Norway](#), present a series of events with topics related to open publishing, open science and data sharing. All events are digital and are open to anyone interested.

[Here you can find the full program for the national Open Access Week events in Norway and register for each webinar.](#)

Local events at the University of Oslo

Monday 25th October

11:00-12:30: [Shut Up and Write: Documentation \(on-site\)](#)

Is it difficult to set aside time to write proper documentation for your data collection? We offer a structured information and writing session for you!

13:00-14:30: [Shut Up and Write: Data Management Plan \(on-site\)](#)

Is it difficult to set aside time to write proper Data Management Plan for your research project? We offer a structured information and writing session for you!

Tuesday 26th October

13:00-14:30: [Sharing and archiving research data](#)

Are you keen on sharing your research data? Or does your publisher or funder require that your dataset is available? In this course we will introduce you to data sharing and how to do it yourself. The course is a part of a [series of research data management courses](#).

Wednesday 27th October

13:00-14:30: [Data discovery and reuse](#)

There are lots of data out there to discover and reuse. In this course we will introduce you to the concept of data discovery and show you the basics on how to work with secondary data. The course is a part of a [series of research data management courses](#).

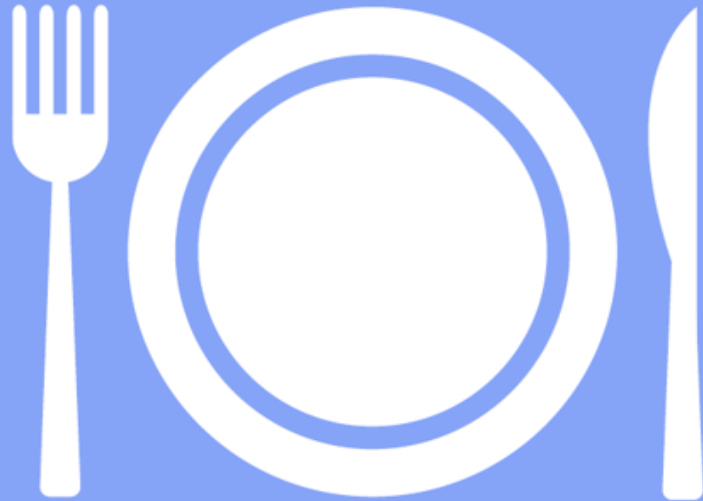
Thursday 28th October

12:00-13:00: [Open Science Lunch: Open your research with ORCID](#)

Join us for our monthly online Open Science Lunch and learn about what ORCID is and how it can help make your research more open. Read more about Open Science Lunch series and upcoming events [here](#).

14:00-15:00: [ReproducibilityTea Journal Club](#). Join us for our bi-weekly, online discussion of articles about open science practices, research reproducibility and meta-research.

You can [sign up for our mailing list](#) to receive regular updates or [get in touch directly](#) if you have any questions.



Open Science Lunch 2021

Organized by **Digital Scholarship Center** and **Open Research**
University of Oslo Library

Each **last Thursday of the month at 12.00** we invite you to join us virtually for an **online open lunch** to hear about **how to make your research more open**.

August 26th

Agata Bochynska

The benefits of preregistering your research

September 30th

Daniel Kalvø

A short guide to Plan S

October 28th

Ivana Malovic

Open your research with ORCID

November 25th

Elin Stangeland

Open and FAIR data



**UNIVERSITY
OF OSLO**

Workshops

- Shut Up and Write: Documentation (online)
 - Monday 01/11 11:00-12:30, Zoom

<https://www.ub.uio.no/english/courses-events/courses/other/research-data/time-and-place/suaw-doc-ws-uio-fall2021-2.html>

- Shut Up and Write: Data Management Plan (online)
 - Monday 01/11 13:00-14:30, Zoom

<https://www.ub.uio.no/english/courses-events/courses/other/research-data/time-and-place/suaw-dmp-ws-uio-fall2021-2.html>

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=IEo96kOKGmA>

SCIENCE

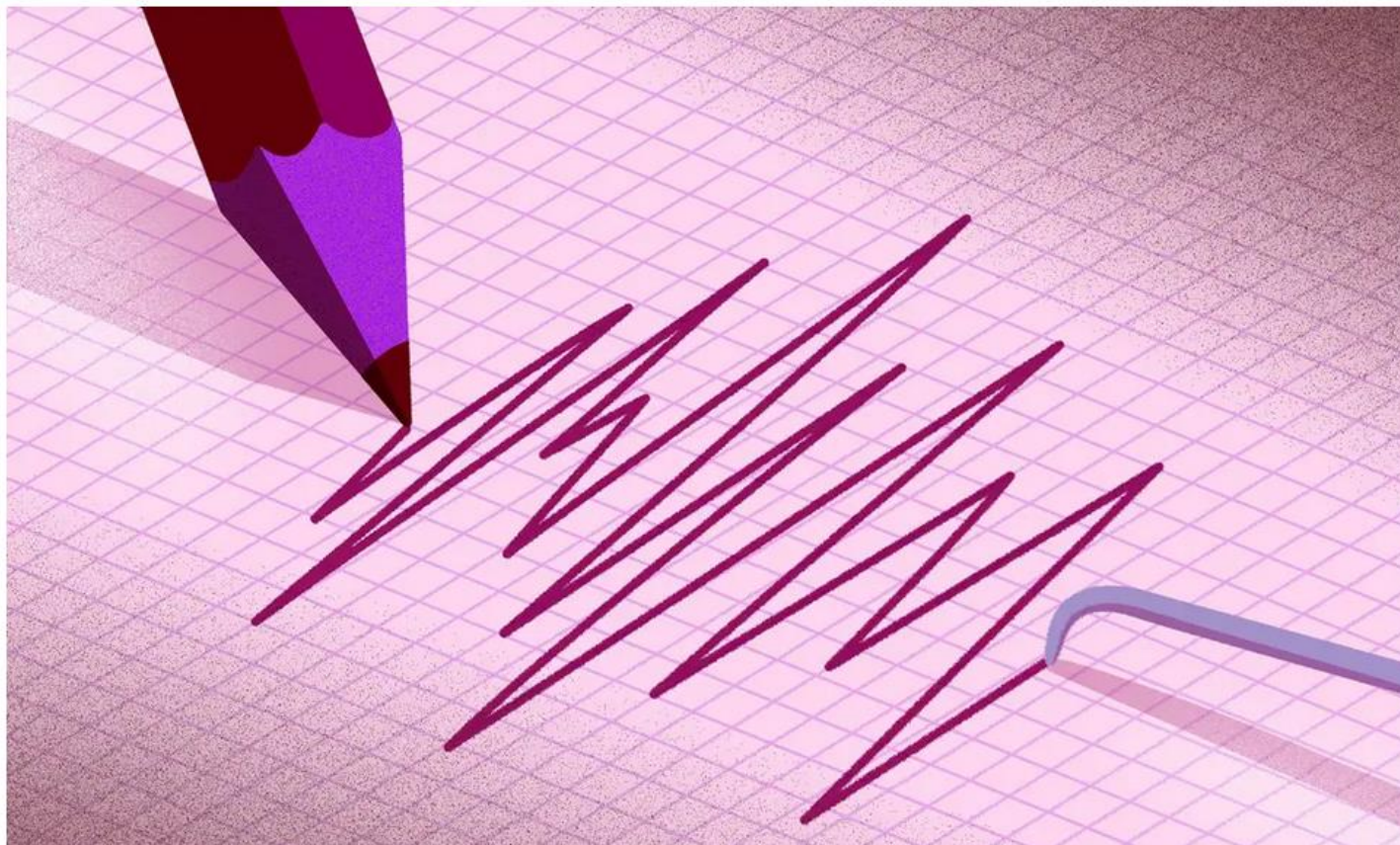
A Famous Honesty Researcher Is Retracting A Study Over Fake Data

Renowned psychologist Dan Ariely literally wrote the book on dishonesty. Now some are questioning whether the scientist himself is being dishonest.



Stephanie M. Lee
BuzzFeed News Reporter

Posted on August 20, 2021, at 2:40 p.m. ET



"When the researchers published their 2020 update, they posted the data from their 2012 paper for the first time. **Publicly sharing data was once a rarity** in science but is slowly becoming more commonplace amid calls for greater transparency."