LOVE DATA WEEK

Welcome to this webinar!

It’s part of the Nordic love data week

See full schedule on:

https://www.nordiclovedataweek.org/agenda/

slides will be available on the website above after the presentation

Where will my data end up?

Rosa Lönneborg, Research data coordinator & data repository manager KTH

So you have received a grant –congrats!

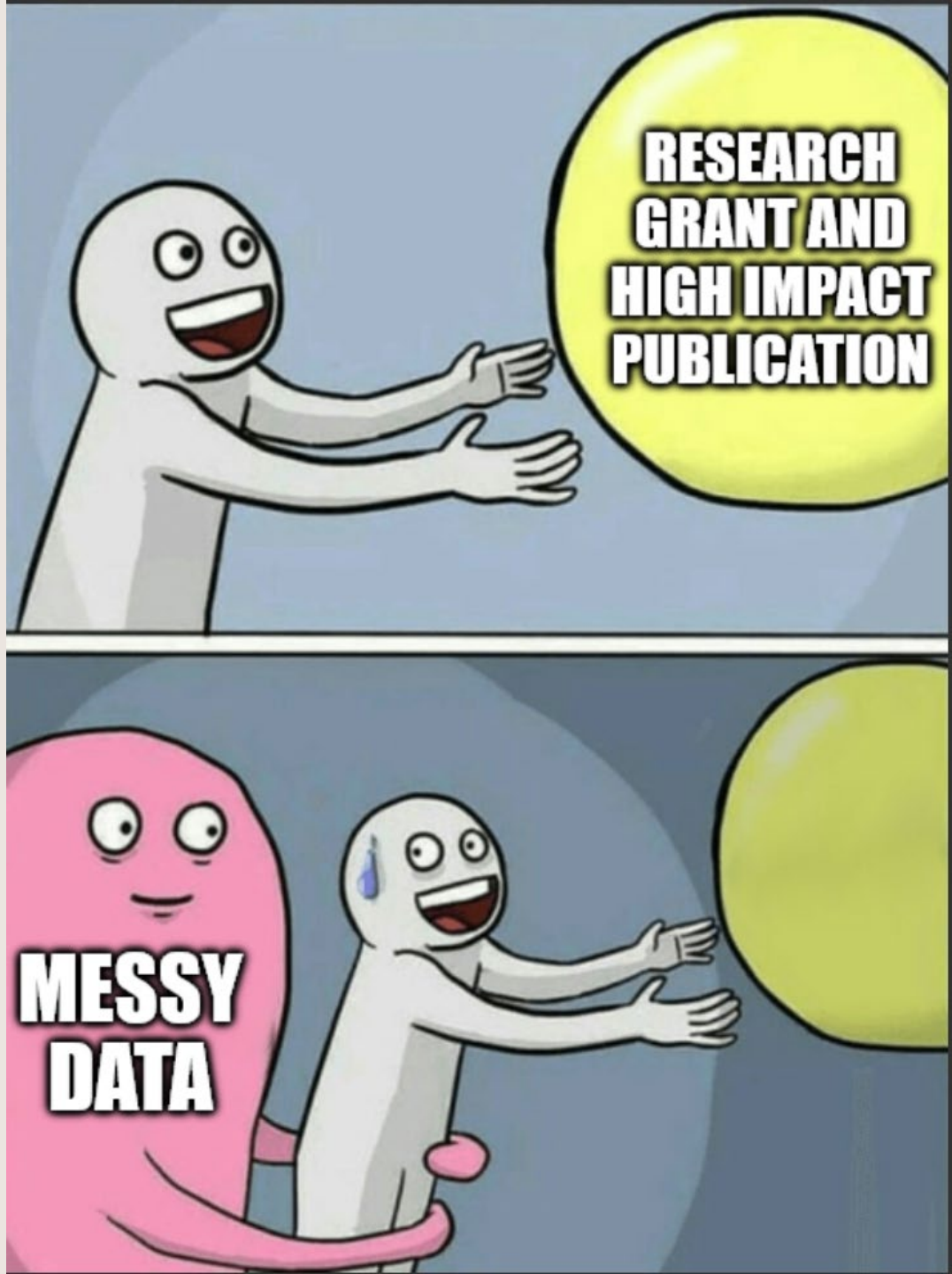
But…the funder requires you to deposit data to a ”trusted repository”

And you have submitted a manuscript –congrats!

But…the journal requires you to deposit data to a ”high quality repository”

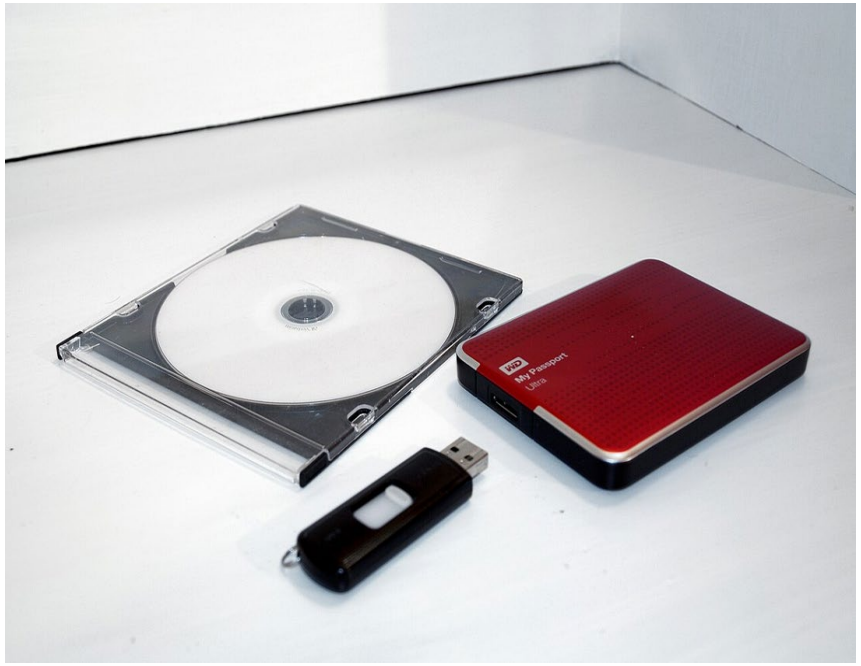
Why? I'm already stressed out

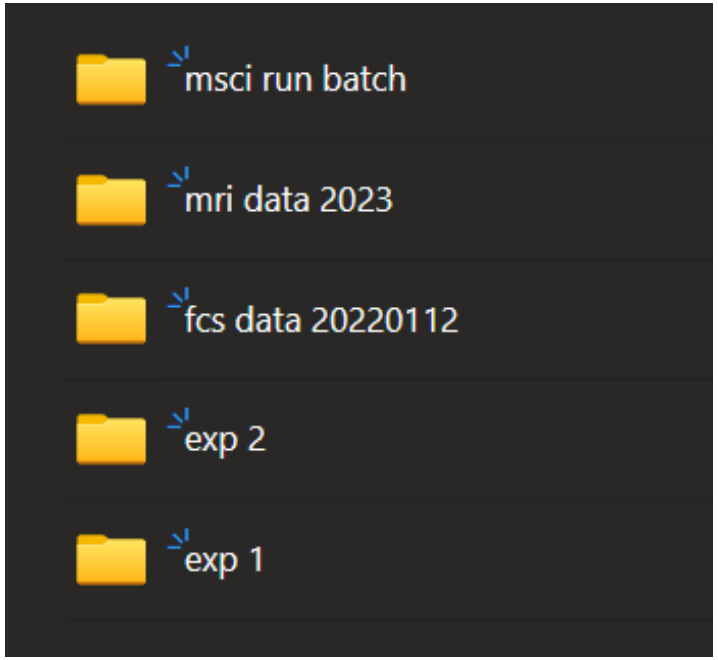
I don't have time to fix the data now and what is this #!? about a repository anyways?



Collecting and analysing data, publishing results… but where does the data end up?

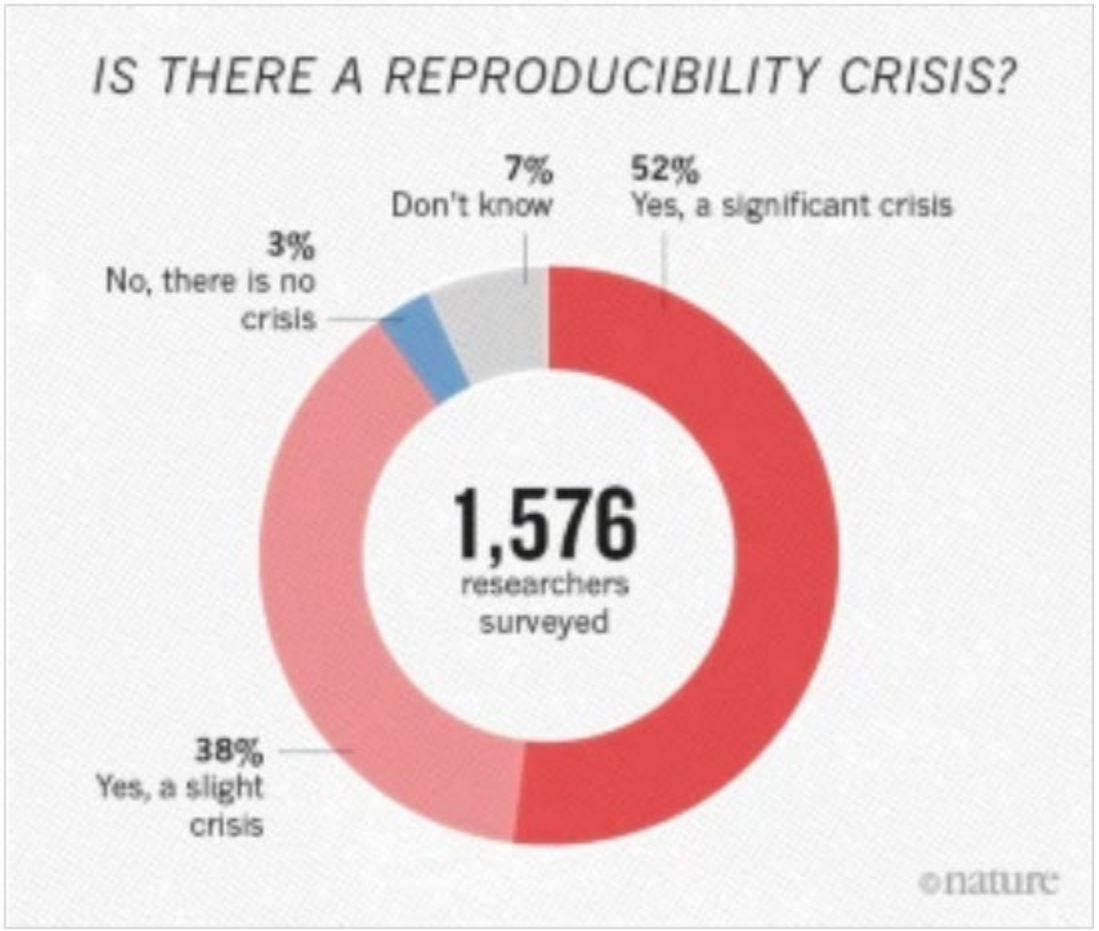
Fairly common end-stations for data collected or generated in research are places like this:

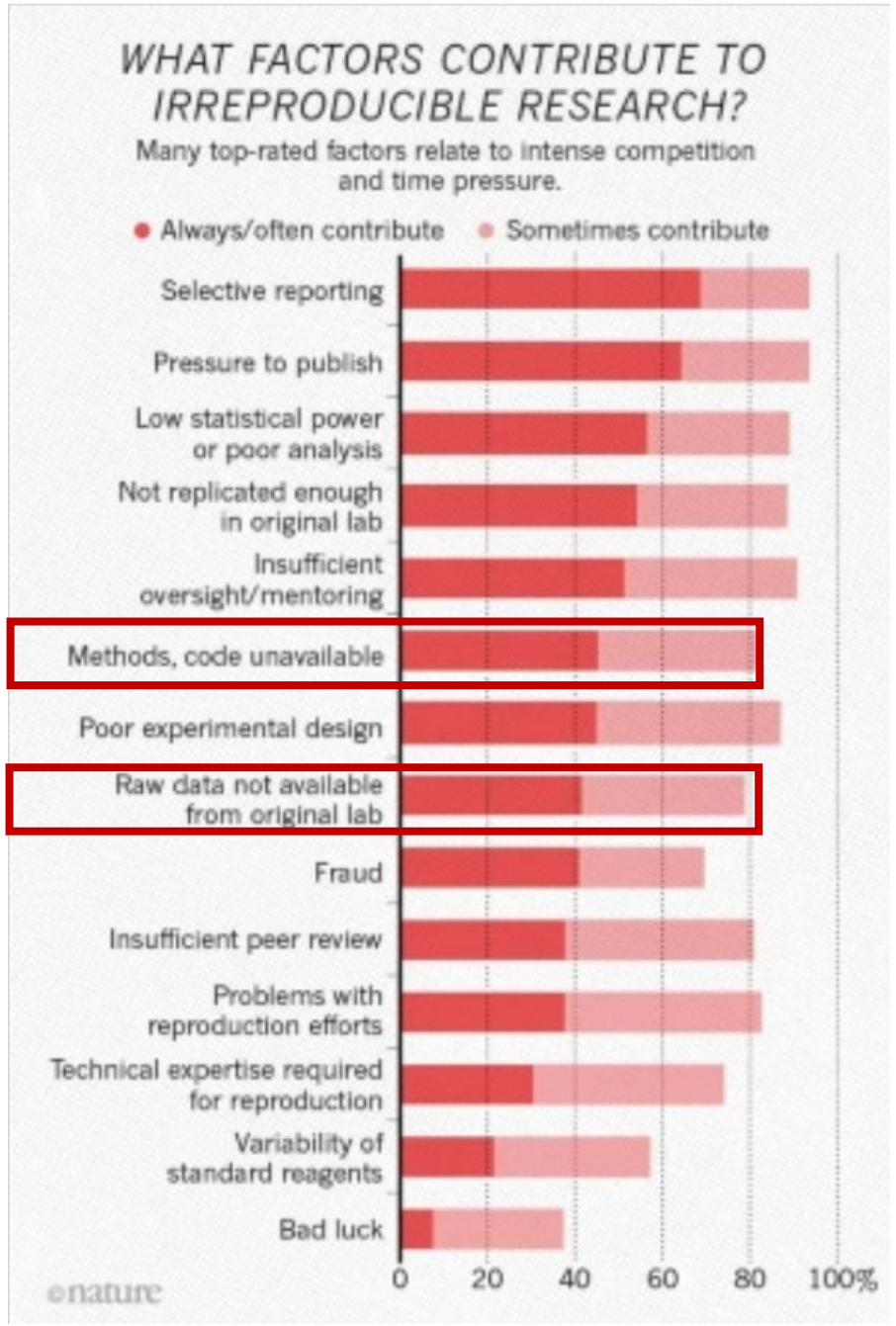




But what happens if someone wants to access the data five years later?

Reproducibility in science



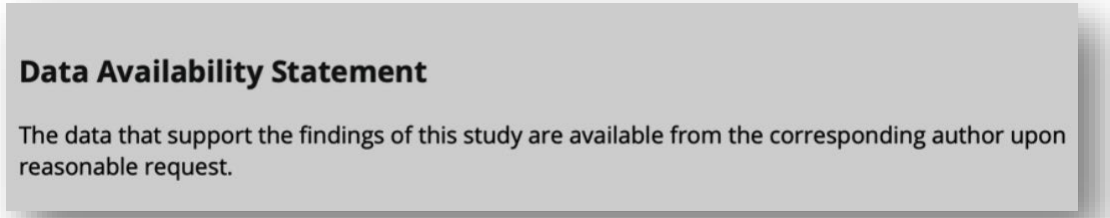


Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016) https://doi.org/10.1038/533452a

What is a data repository?

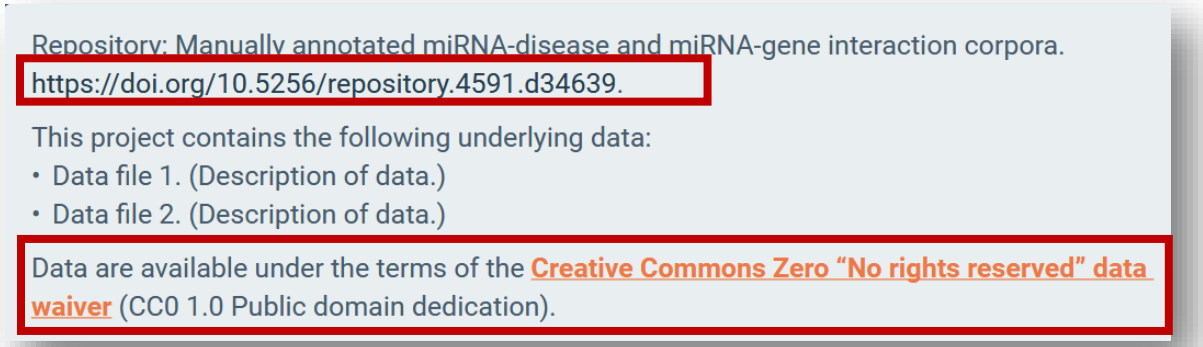
- A data repository is *a place to hold data*, share data, and organize data in a logical manner over time.
- It usually also means that a searchable index of the data is accessible to the public via an open standard protocol
- The main purpose of a research data repository:
  - Data underlying scientific results can be found later so that scientific findings can be reviewed, validated and potentially reproduced

What is a ”high quality” / ”trusted” repository?



Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.



Repository: Manually annotated miRNA-disease and miRNA-gene interaction corpora. https://doi.org/10.5256/repository.4591.d34639.

This project contains the following underlying data:

- Data file 1. (Description of data.)
- Data file 2. (Description of data.)

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

- EU Horizon Europe quality criteria for trusted repositories in short*.
- The repository should provide :
  - Unique and persistent IDs for datasets, persons and organisations.
  - License information
  - Information about grants/funded projects
  - Open and machine-readable metadata
- Note that this is quality criteria for the **repository**, not for the **data**

*Full report:

https://doi.org/10.5281/zenodo.7728016

How do I do it?

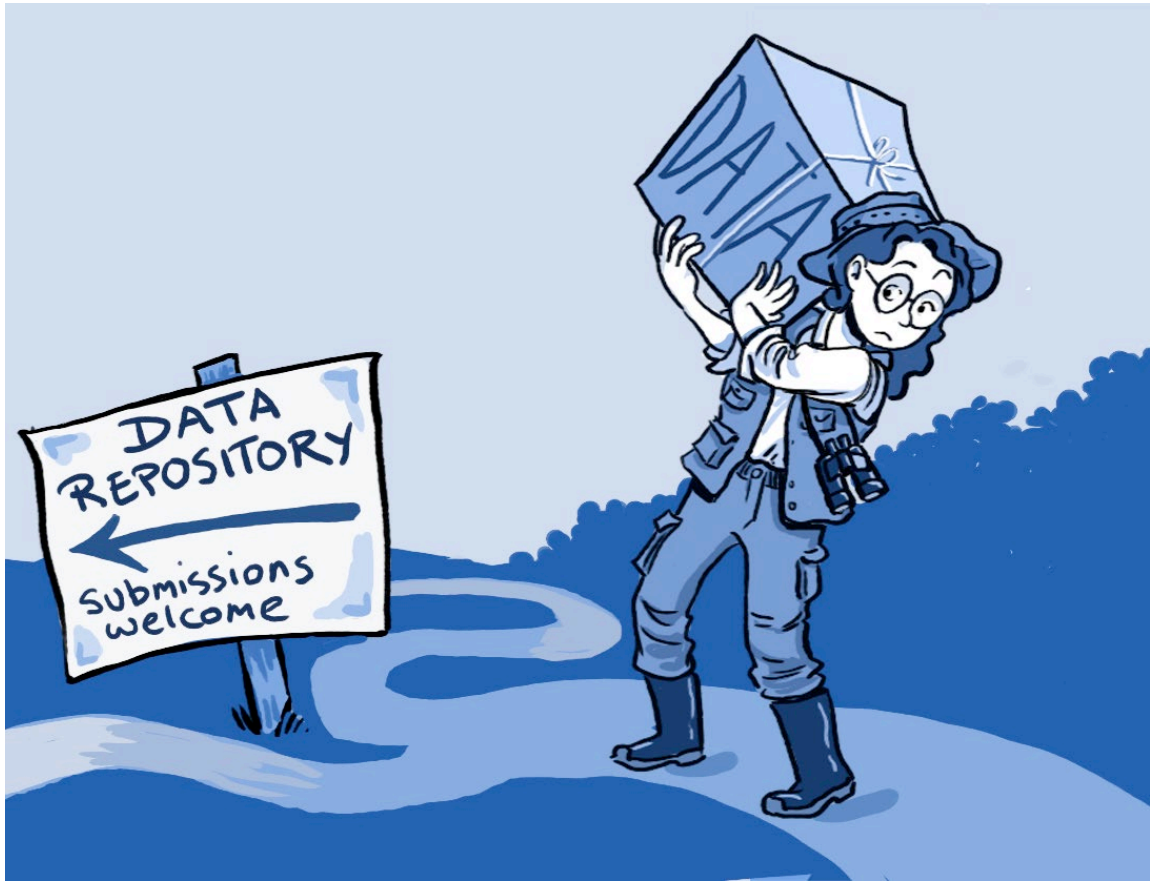
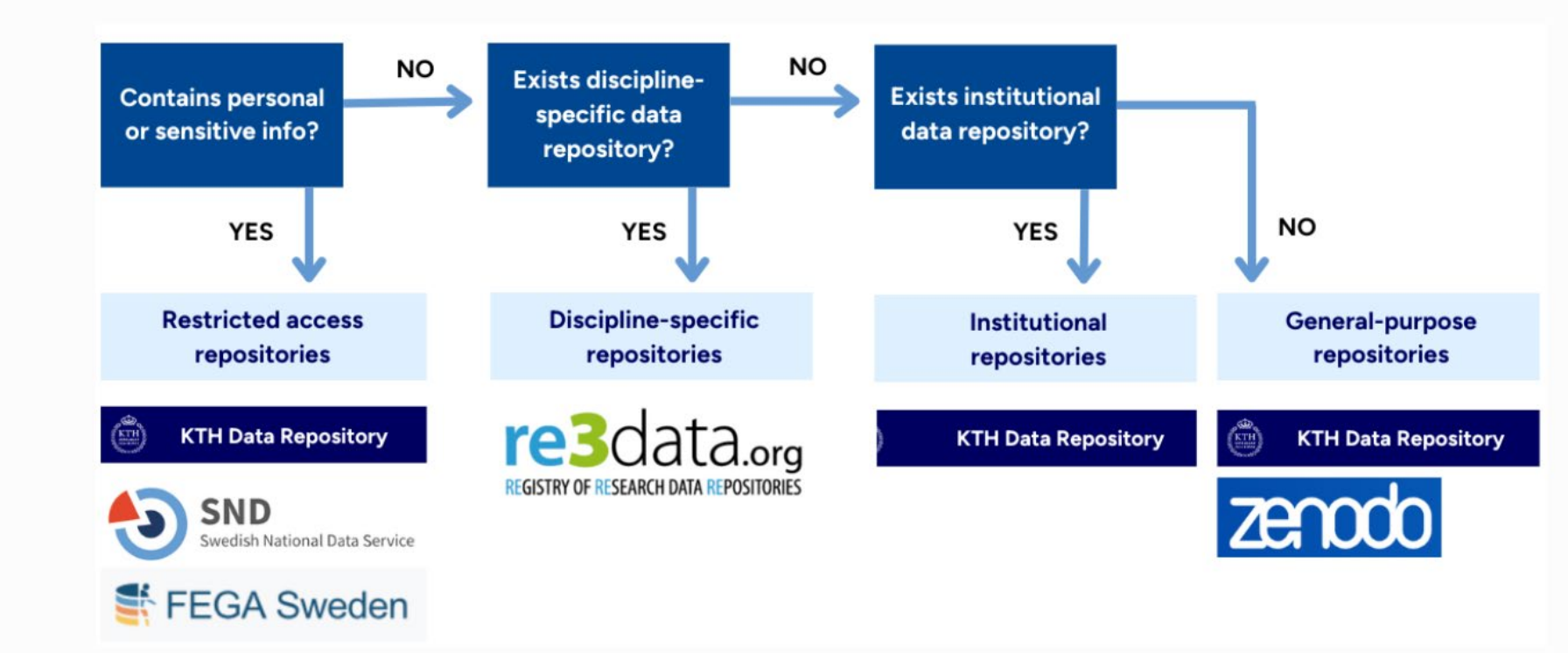


Image: Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, et al. (2014), CC BY 4.0 via Wikimedia Commons

- There are **a lot of** different data repositories –each have their own specific guidelines
- Submitting data usually requires some preparation by you. Examples:
  - Clear description of variables
  - Organized folder structure + readme-file
  - Standard nomenclature /terms to describe context and content
  - De-identification of personal information
- Ask for help in good time! Research data support at university or someone experienced in your research community

How do I select a good repository? Short guide



- Different types comes with different pro's and con's
- Check if funder or journal have specific guidelines on what type of repository to use

Some nice examples of where data can end up– let's have a look and discuss them!

- Domain specific – restricted access (social science)

https://researchdata.se/en/catalogue/collection/national-som

- Domain specific – open access (ocean data)
  https://doi.org/10.1594/PANGAEA.849829
- General purpose repository – general purpose data (both academic and industrial use)

https://doi.org/10.5281/zenodo.3384388

- A data portal – visualising and putting data in context - Human protein atlas

https://www.proteinatlas.org/

- Institutional repository – mixed access

https://doi.org/10.71775/kth.h0774-f3p44