Data Management Plan

➢ *a formal document that outlines how data will be handled during and after a research project.*
(Harvard Biomedical Data Management)

Why plan?

• *Because you (and others) might want to know*

  where you're going

• *It is a requirement (funders or journals)*

• *Because it saves money in the long run*

• *Because it leads to better quality research*

  and enables high quality curation and reuse
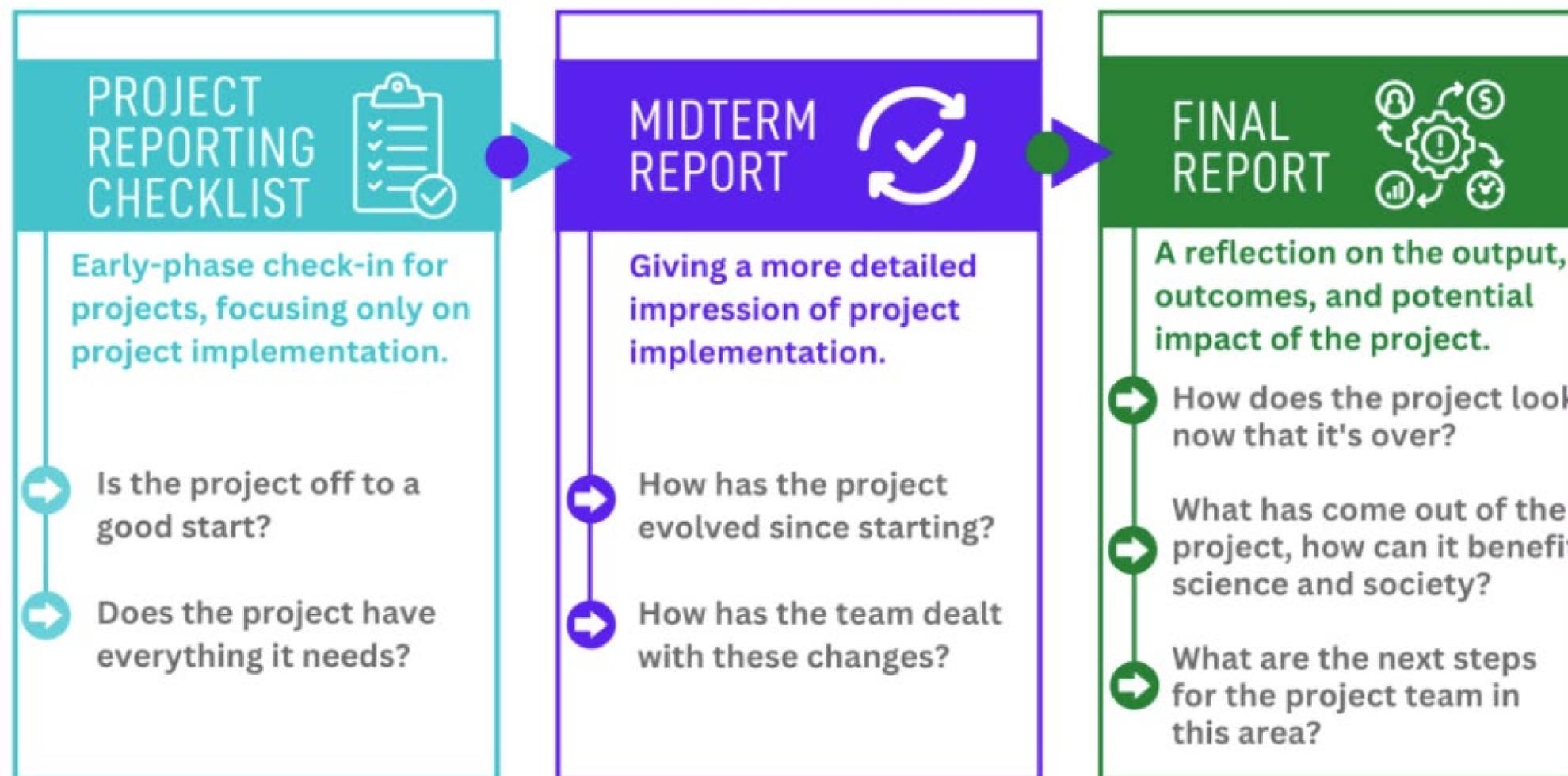
Data Management Plans today

- Who reads them?

- Where are they stored?

- What is the added value to the PI/institution/to the community?

- How can another researcher use them?
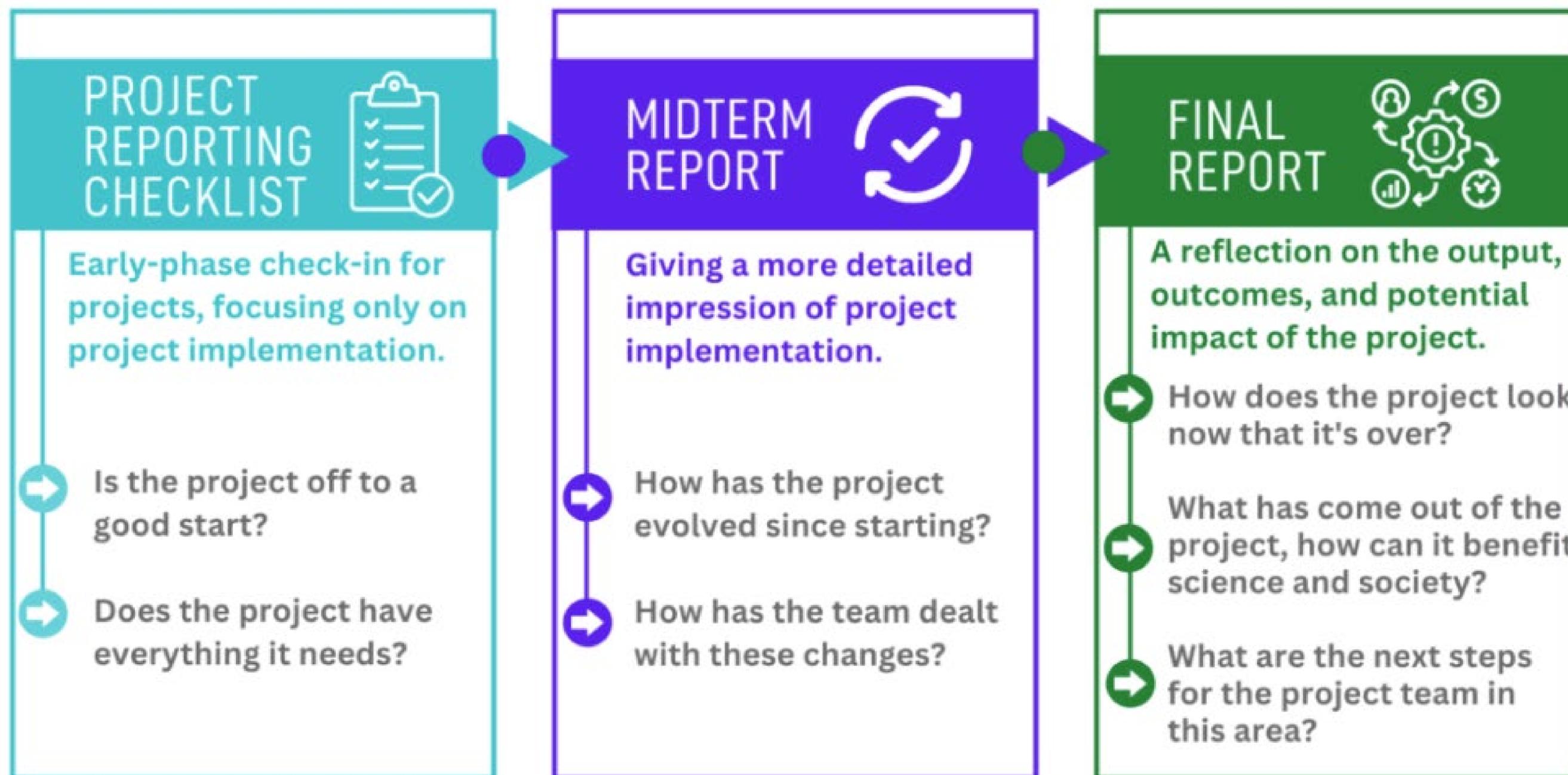
Data Management Plan

- Typically, a requirement from funders (EU, some national agencies); institutional commitment varies

- « Heavy » templates:
    - Around ~ 10 pages (after filling the questions out)
    - Linear structure, the researchers need to go through questions -> Tend to ask the same question several times from different perspectives (not dataset-specific)
    - Difficult for researchers to think of all their datasets in relation to the questions
    - Simply an administrative task?

Example

- Fonds National de la Recherche (FNR) report
- Scientific report: light and to the point
- DMP template : ~10 pages (to be submitted when doing the reporting to the FNR)

# Fonds National de la Recherche (FNR) - Luxembourg National Research Fund

PROJECT REPORTING CHECKLIST

Early-phase check-in for projects, focusing only on project implementation.

➡ Is the project off to a good start?

➡ Does the project have everything it needs?

MIDTERM REPORT

Giving a more detailed impression of project implementation.

➡ How has the project evolved since starting?

➡ How has the team dealt with these changes?

FINAL REPORT

A reflection on the output, outcomes, and potential impact of the project.

➡ How does the project look now that it's over?

➡ What has come out of the project, how can it benefit science and society?

➡ What are the next steps for the project team in this area?

TIME IS RUNNING OUT

What we provide

- The basics on RDM
- Annotated DMP templates, with examples!
  - They can use it through dmponline platform or
  - By adapting the template to their case
  - Includes standard texts for e.g. policy, links to further info, and guidance

- Trainings
  - One on one
  - Through the doctoral schools at Uni.lu (with ECTS!!)
  - Through ULearn, the training platform offered to all Uni.lu staff
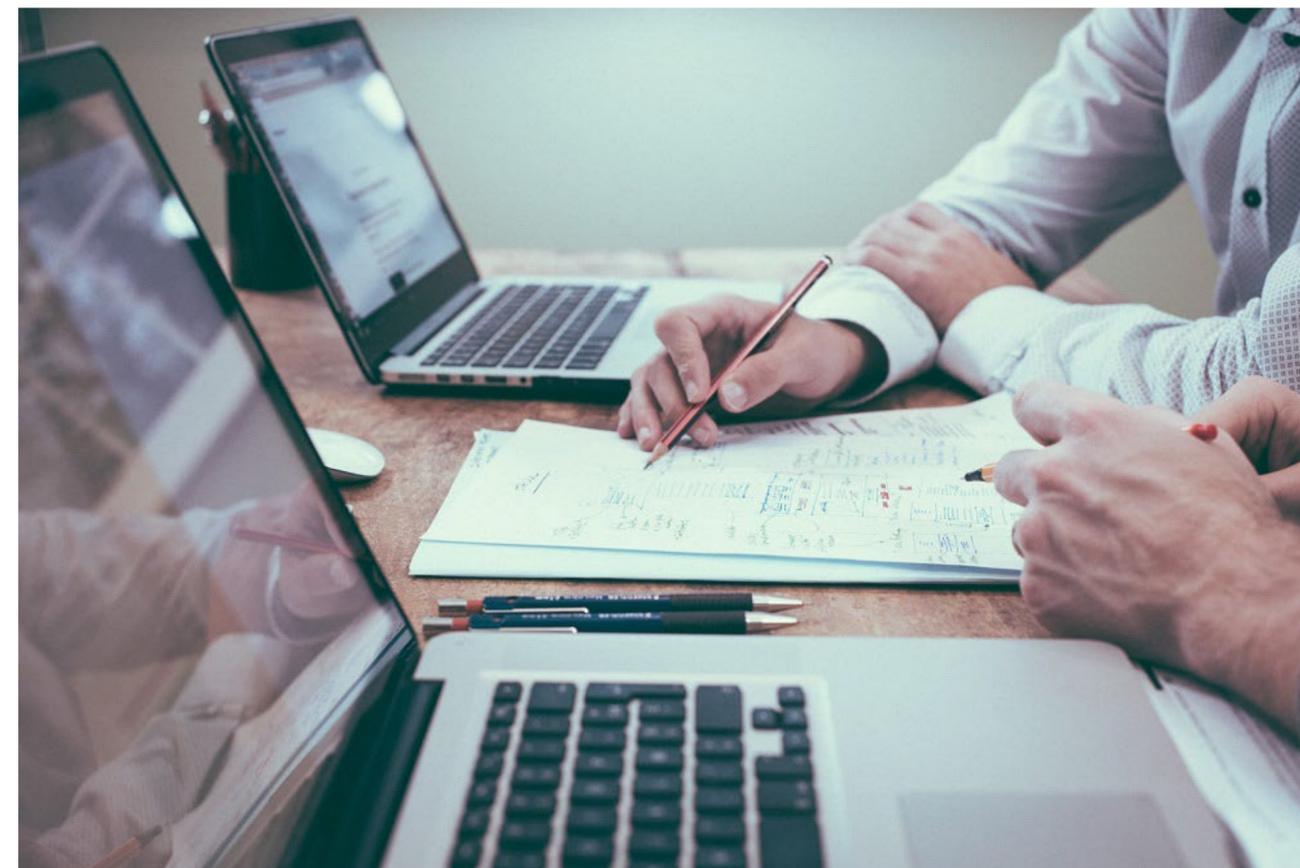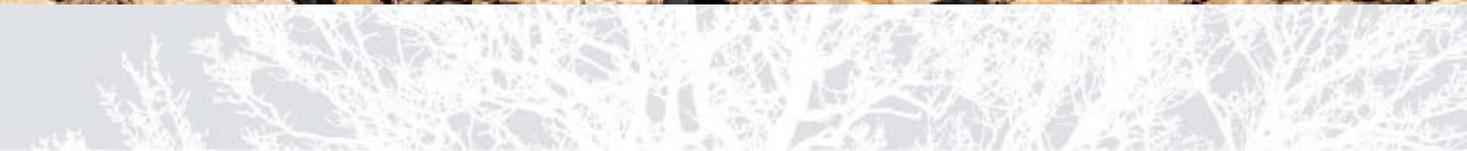
What we provide

Love My Data 11 - 15 March 2024

- bring awareness of research data management (RDM), create a momentum for caring for data and raise the data community spirit

  - Design a Data Management Sticker Competition!
  - RDM Crossword fun
  - Spring clean your data day
  - Data repair cafe
  - LEGO® Metadata for Reproducibility Game

We need to bring the researcher perspective on data and data management: best practices

- **Data illiteracy**!
  - What is research data? Most do not know (it's not part of their studies) or even understand when they begin their research or work in disciplines where research data is unconventional

- **Research management illiteracy** [Bachelors and masters are not research-oriented]
  - There is no formal training
  - Often no logbooks exist
  - Everything depends on the supervisor/field/group culture
  - Lack of/conflicting institutional/national policy → hence no consequence

Research management is not standard Project management

- In research - protocols are to be developed

- Research plan is fluid; things change and move, any plan is not closed, research plans are not reliable. Very often there are:
  - No standards
  - No meta data standards
  - …

What we propose, a hands-on workshop where they map their data

- Start the discussion by identifying the data → what data are they using?
- And then, map the data, map how FAIR the data is, **each dataset**
    - **F**indable
    - **A**ccessible
    - **I**nteroperable
    - **R**eusable
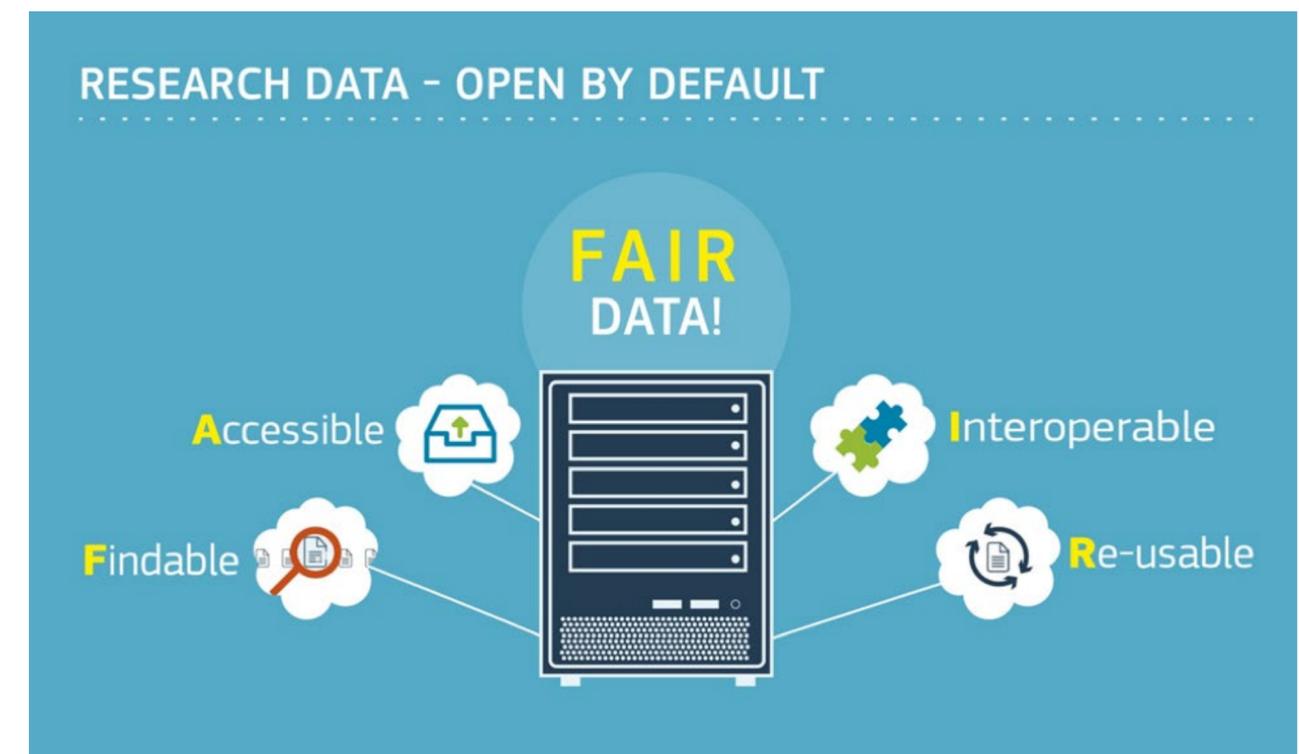
openaire.eu

PROJECT LIFECYCLE - FAIR Data Mapping TEMPLATE (Name/Dept:)

Project phase & Title of dataset	Description (words)	Findable	Interoperable	Accessible	Reusable
		Dataset: Type & format	MetaData (1. administrative metadata, 2. citation metadata 3. structural metadata)	Describe how usage with other data/specific applications/machines.	Access/storage/security/costs. How access is guaranteed.	Describe how data can be reused and replicated
Exploratory phase						
Dataset 1 Name						
Dataset 2 Name						
Dataset 3 Name						
Dataset 4 Name						
Dataset 5 Name						
Advanced phase	In this phase, you are 1+ years into the project and you will describe your existing or new datasets, taking inspriation of what you have described previously, and what has changed.					

PROJECT LIFECYCLE - FAIR Data Mapping: Project by XXXX (Department: CDKJDK)

	Findable			Interoperable	Accessible	Reusable
Project phase & Title of dataset	**Description (words)**	**Dataset: Type & format**	**MetaData (1. administrative metadata, 2. citation metadata 3. structural metadata)**	**Describe usage with other data/specific applications/machines.**	**Access/storage/security/costs. How access is guaranteed.**	**Describe how data can be reused and replicated**
Exploratory phase						
Dataset 1	Project A1.1. Qualitative interviews of farmers undertaken in 2020 by xxx. These are transcribed word documents and there is a need to explore if recordings exist of interviews. I will source these datasets from xxx, a public archive.	1. Word documents with transcriptions of interviews. Digital documents shared on USB stick (due to size of documents). 2. Explore if voice recordings exist in MP3 format and if these can be shared due to GDPR and how (TBC). Source: xxx Archives Luxembourg.	Descriptive metadata: The dataset and file will be named A1.1 Transcriptions farmer interviews 2015 SOS FAIM and stored during the research period in a research folder accessible by Research Team only on Univesity server Atlas. The file will be password protected. A README txt file will describe this dataset, who has access, where it was sourced and give a summary of how it was used in the study. In our Research Team we use EndNote to catalogue each data source for our project, which acts as our database of record for all datasets. This also enables us to later reference the datasets in our publications and research output. Admin metadata: This will be included once analysis has been completed to describe languge(s) of interviews, media types created of the data for analysis and license needed to obtain the data from xxx.	Reused data from previous research, transcribed interviews in word documents (digital format). Unknown if we can source original voice recordings. Need to explore GDPR and informed consent procedures. Transcribed files contain same formatted structured and the same interview questions. Ideally the interviews can be coded for analysis, should GDPR allow it. As we use Endnote to catalogue all datasets, our research output will include a bibliography of all used qualified datasets catalogued with Endnote metadata, e.g. Title, Author, Year published (or produced), Type of document (dataset, publication, book etc.), Source, keyword fields etc.	The Research Team will have access to this research dataset via a secure department drive on Atlas University server. The file is password protected. Access is guaranteed to the Research Team whilst stored on Atlas including regular automatic backups. To open the dataset the Research Team use a password protection system known by the Research Team. No costs are associated with the storage of these data.	The metadata gives all neccesary information on the sourcing of these data and the storage of these data, we will also describe in a report the methodology of the project. This makes the project replicable.

What we propose…

- Communicate the whole message through best practices in research, bring researchers in!
  - Asking questions such as:
    - How many of you have lost their data? (Sharing practical experiences)
    - Have you ever tried to understand the notes (can be study notes) of a colleague?
    - Show them data on the reproducibility crisis!
    - Explain what licenses are in practice, it's a complicated subject that requires attention.
    - We show them the rules for their articles, what they can re-use as authors and how, based on the licenses the journal provides

PLoS One. 2020; 15(4): e0230416.

Published online 2020 Apr 22. doi: 10.1371/journal.pone.0230416

PMCID: PMC7176083

PMID: 32320428

The citation advantage of linking publications to research data

12.2% of BMC publications provided DAS containing a link to data in a repository. We also find an association between articles that include statements that link to data in a repository and up to 25.36% (± 1.07%) higher citation impact on average, using a citation prediction model. We discuss

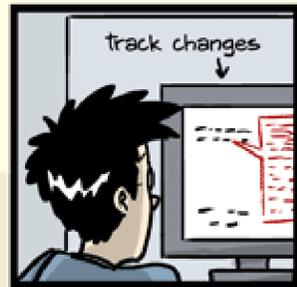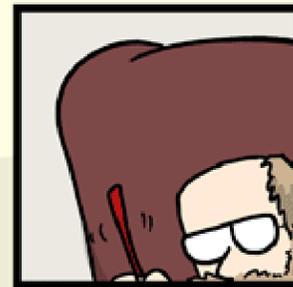**Decline can mean 80% of data are unavailable after 20 years.**

"In their parents' attic, in boxes in the garage, or stored on now-defunct floppy disks — these are just some of the inaccessible places in which scientists have admitted to keeping their old research data. Such practices mean that data are being lost to science at a rapid rate, a study has now found."

https://www.nature.com/articles/nature.2013.14416

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.
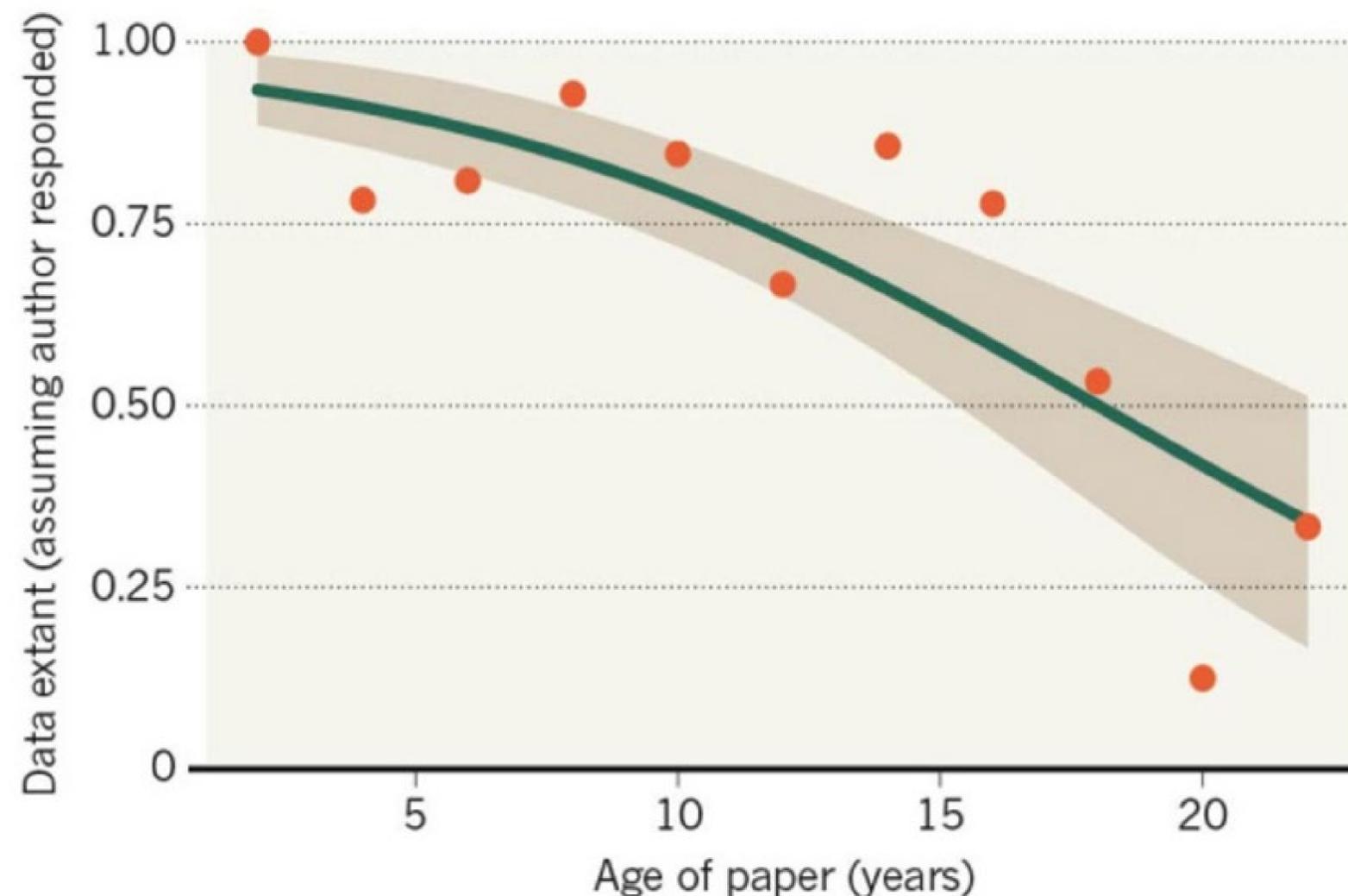
Horror Story #1: I was working on Chapter 2 of my thesis and had made a lot of great progress. I'd written around 20 pages and was feeling excited about how much work I had completed. Then—out of nowhere—I was visited by everyone's worst nightmare: the blue screen of death. My hard drive was done. While we were able to recover many files, Chapter 2 was not one of them and I didn't have a back-up copy.
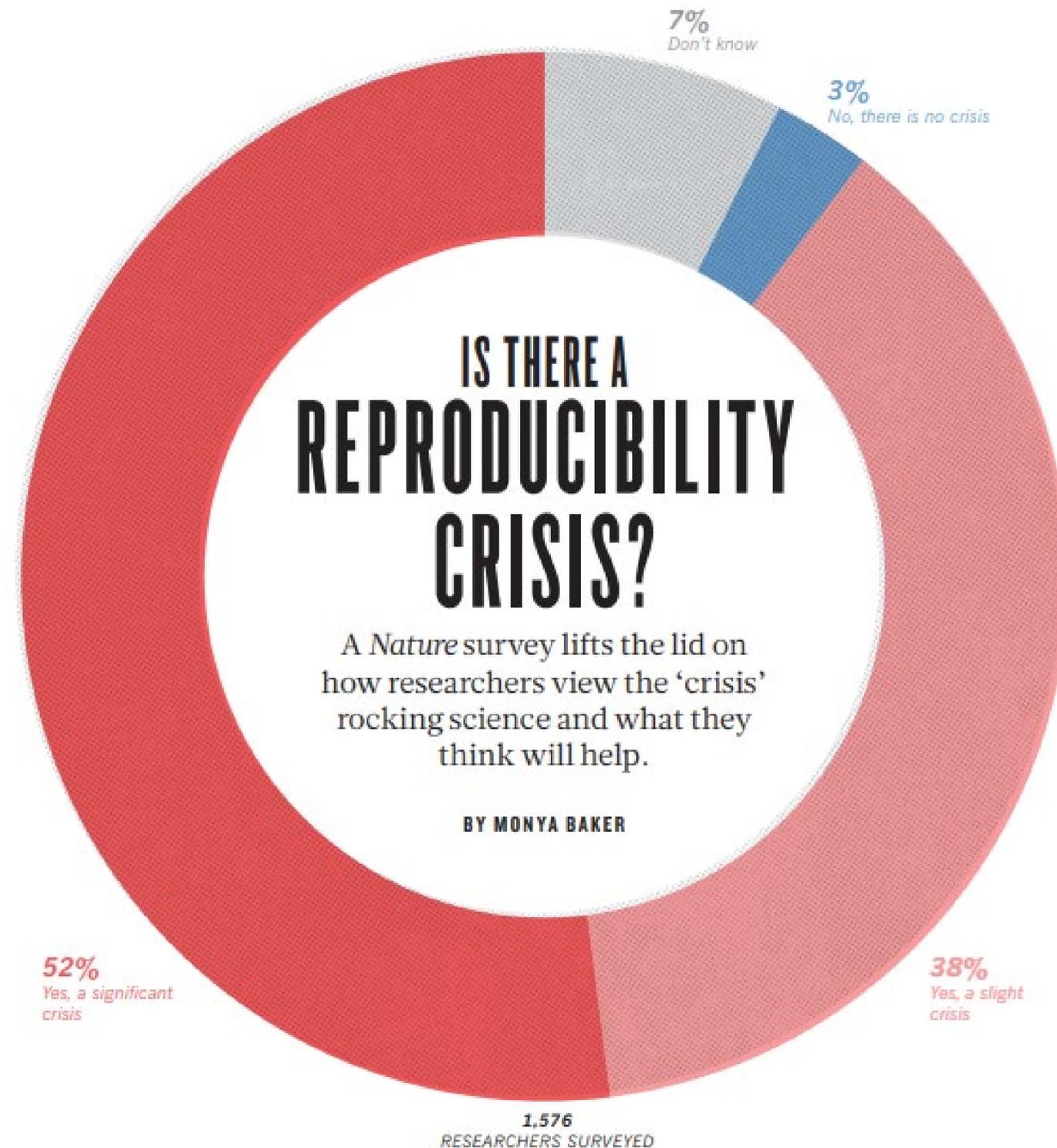
Solution: Back up your files. It's really that simple. I always chose not to do it and my excuse was "It takes too long." It really doesn't, though. Think about it this way: it takes around two minutes to back up your files, but it could take days to rewrite a paper that has been lost.

Horror Story #2: $125 million NASA Mars probe, lost in space in 1999. JPL engineers did not take into consideration that the units had been converted, i.e., the acceleration readings measured in English units of **pound-seconds^2** for a metric measure of force called **newton-seconds^2**. In a sense, the spacecraft was lost in translation.

Solution: Communication! Research is a collaborative endeavor, and many stakeholders are involved: researchers, technicians and laboratory staff, support staff, administrative staff, data stewards…..

1,500 scientists lift the lid on reproducibility

https://www.nature.com/articles/533452a

7%
Don't know

3%
No, there is no crisis

IS THERE A
REPRODUCIBILITY
CRISIS?

A *Nature* survey lifts the lid on
how researchers view the 'crisis'
rocking science and what they
think will help.

BY MONYA BAKER

52%
Yes, a significant crisis

38%
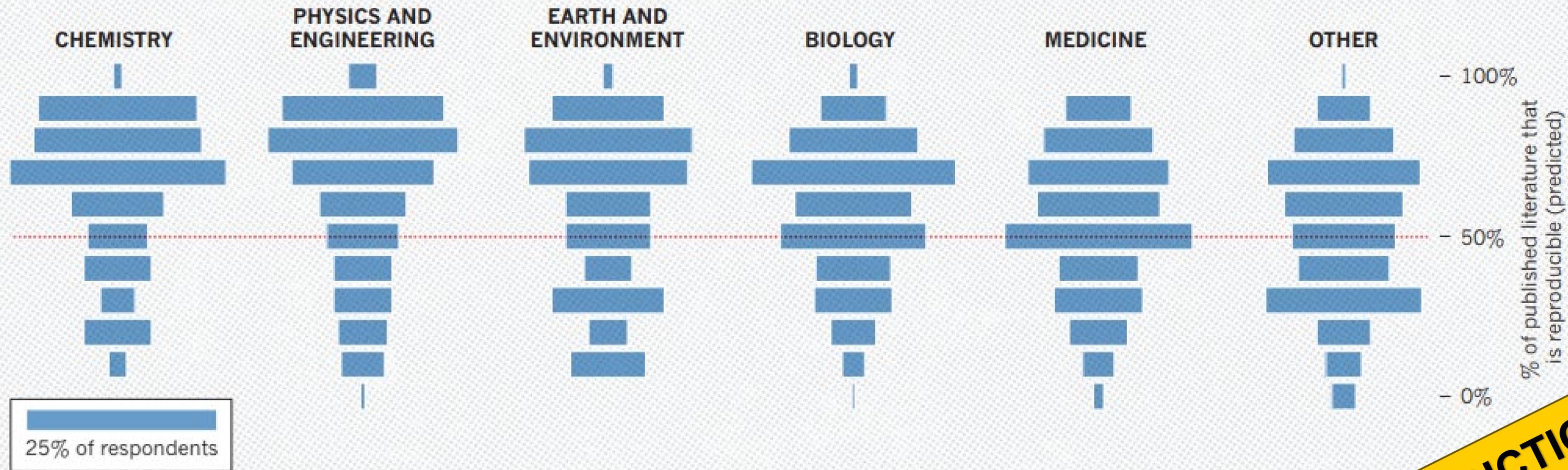*Yes, a slight crisis*

1,576
RESEARCHERS SURVEYED

A 'CRISIS' IN NUMBERS

Nature surveyed 1,576 scientists online to get their thoughts on reproducibility in their field and in science in general. See **go.nature.com/2vjr4y** for more charts and access to the full data.

HOW MUCH PUBLISHED WORK IN YOUR FIELD IS REPRODUCIBLE?
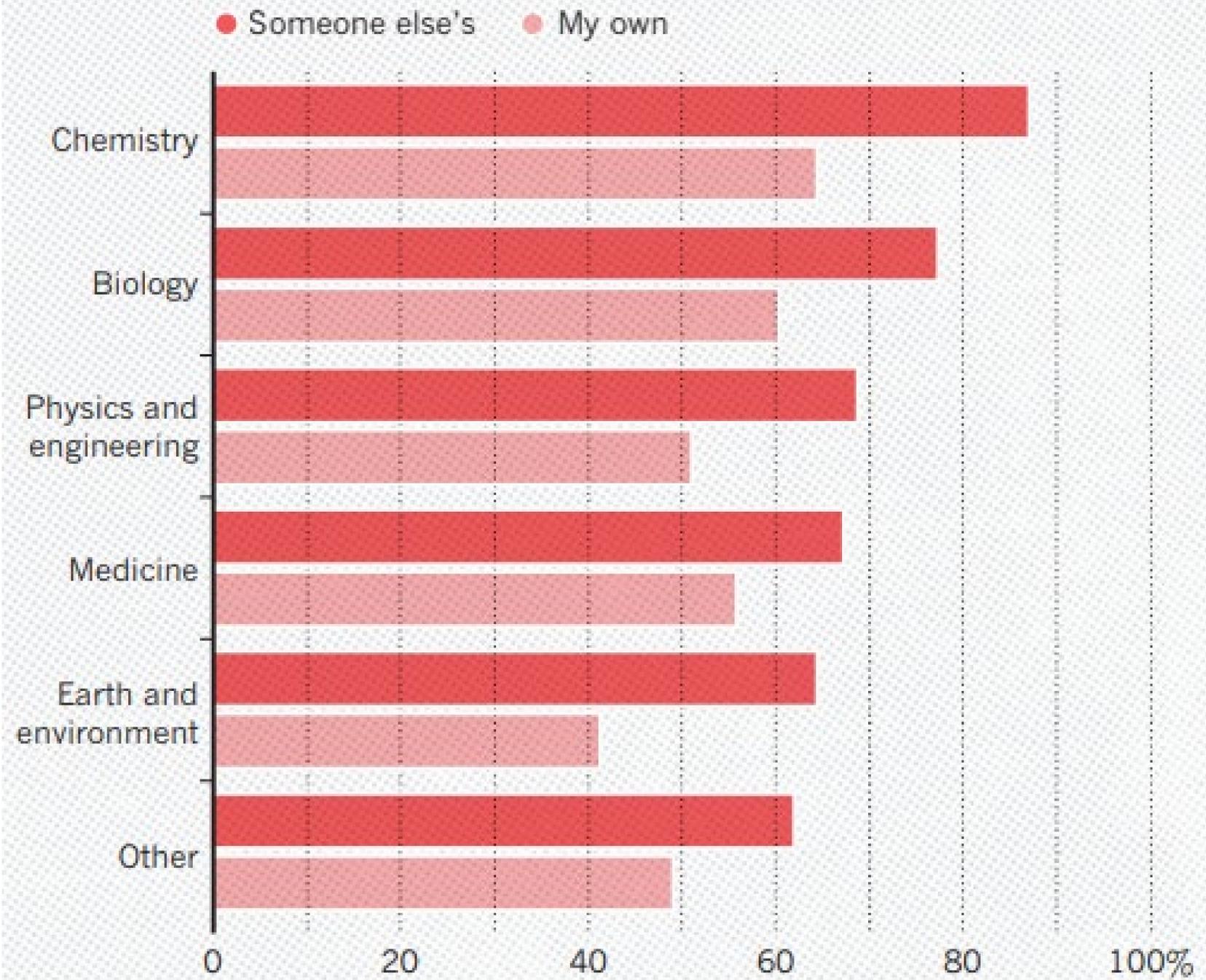Physicists and chemists were most confident in the literature.

CHEMISTRY · PHYSICS AND ENGINEERING · EARTH AND ENVIRONMENT · BIOLOGY · MEDICINE · OTHER

% of published literature that is reproducible (predicted)

100% · 50% · 0%

25% of respondents

https://www.nature.com/articles/533452a

"*I tell this story to illustrate the truth of the statement I heard long ago in the Army:*
plans are worthless, but planning is everything."

Dwight D. Eisenhower

THANK YOU!

Any questions?
lovemydata@uni.lu

EARMA

EUROPEAN ASSOCIATION OF
RESEARCH MANAGERS AND ADMINISTRATORS