

# Data Stewardship

## Why & How?

21.03.19



Yasemin Turkyilmaz-van der Velden  
Data Steward TU Delft  
[y.turkyilmaz-vandervelden@tudelft.nl](mailto:y.turkyilmaz-vandervelden@tudelft.nl)

# Outline

- Who am I?
- Data stewardship-Why?
  - Motivations for open science
  - Reproducibility crisis in science
  - FAIR data principles
- Data stewardship-How?

# Who am I?

## Master of Science

Molecular Mechanisms of Disease

Radboud University Nijmegen

**Virology & Molecular immunology**



**Stem cell biology & Transcription**

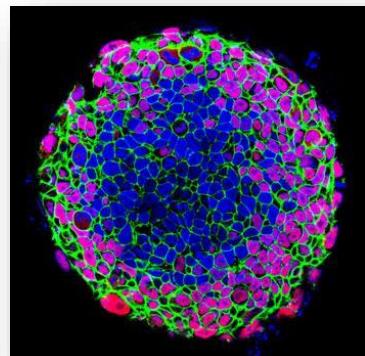


Image by Zhen Ma, UC Berkeley

## PhD candidate

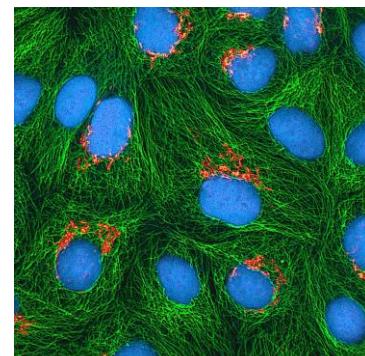
Department of Molecular Genetics

Erasmus MC Rotterdam

**UV-induced DNA damage & repair**



**Proteomics & Microscopy**



# Open Science-Why?



# Academic publishing



# Academic publishing



# Academic publishing



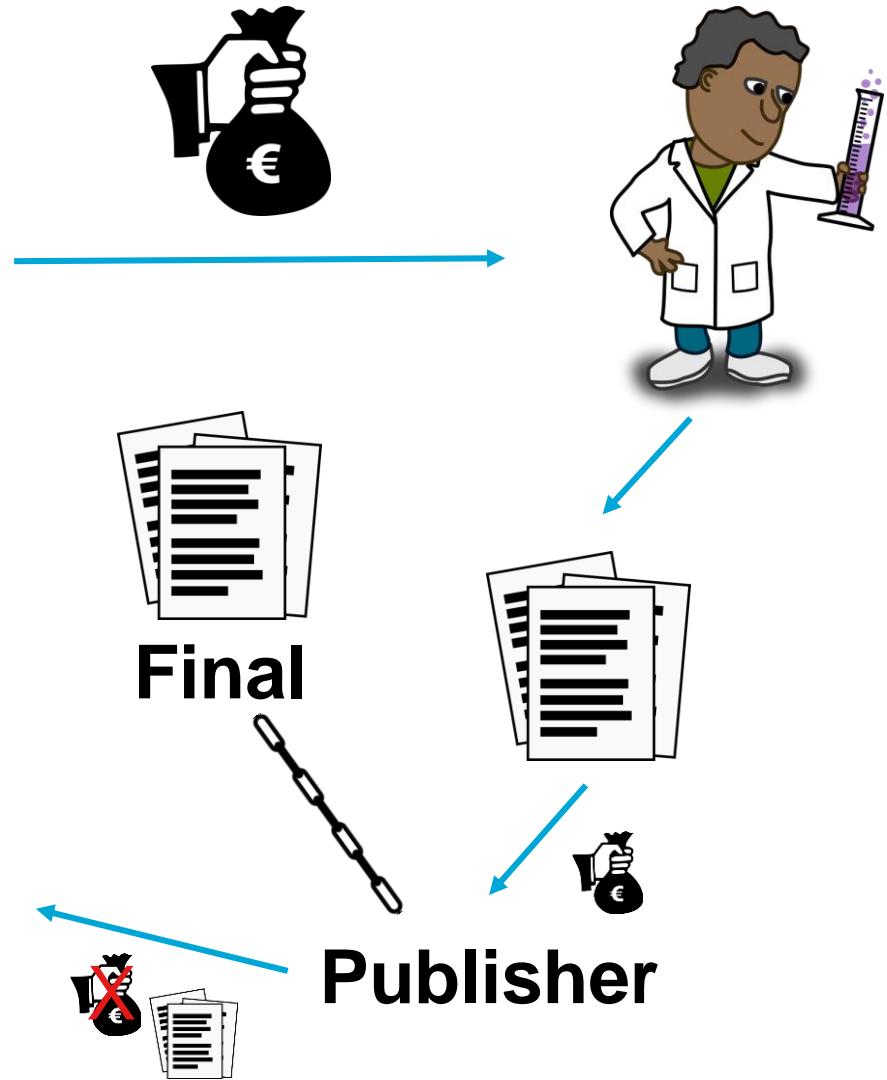
**Publisher**

# Academic publishing

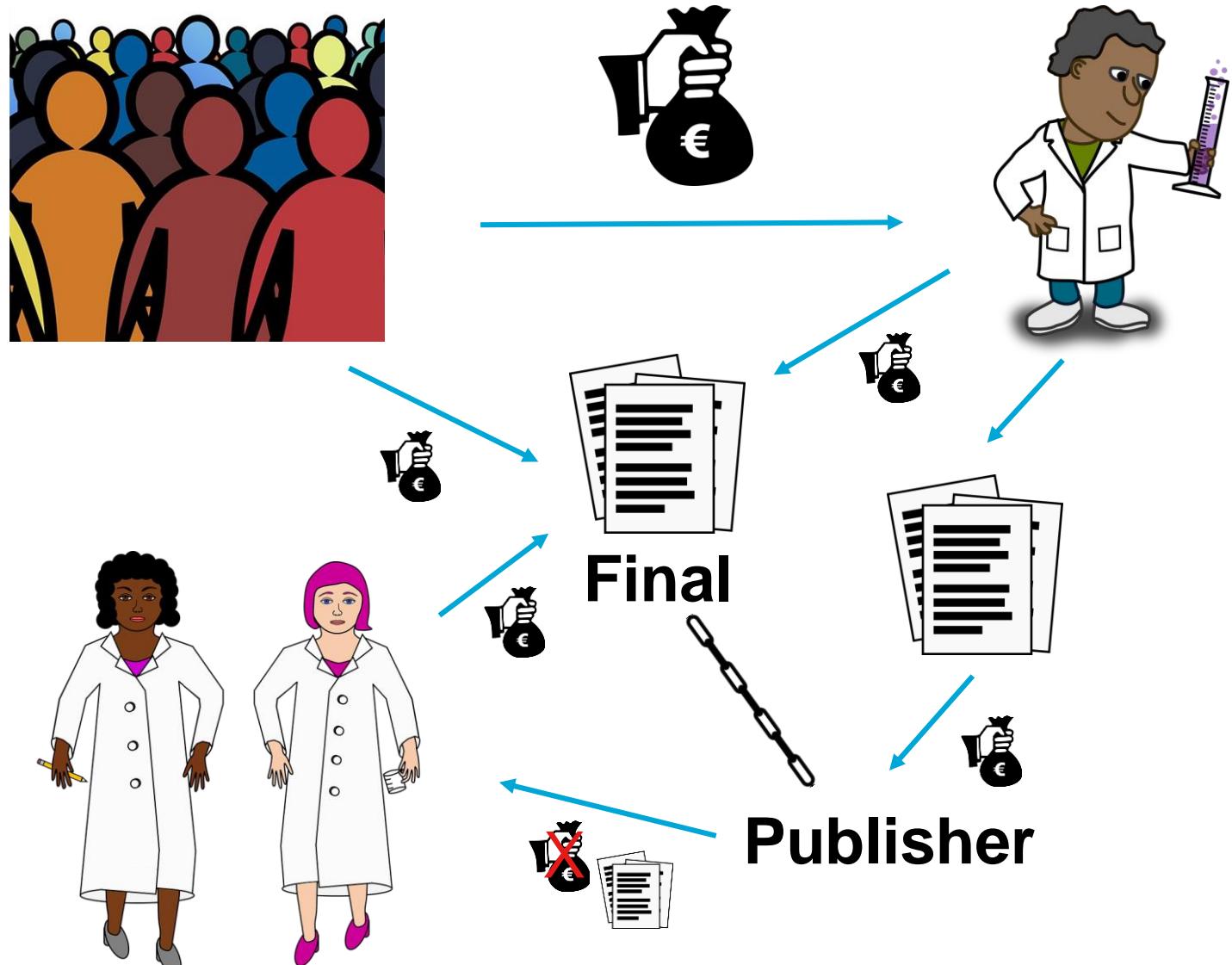


Publisher

# Academic publishing



# Academic publishing



# Profit margin of publishers



<https://www.macrotrends.net/stocks/charts/AAPL/apple/profit-margins>

<https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science>



nationaal  
platform  
**open**  
science

**Home**

**Open Science**

**National Platform Open Science**

**Themes** ▾

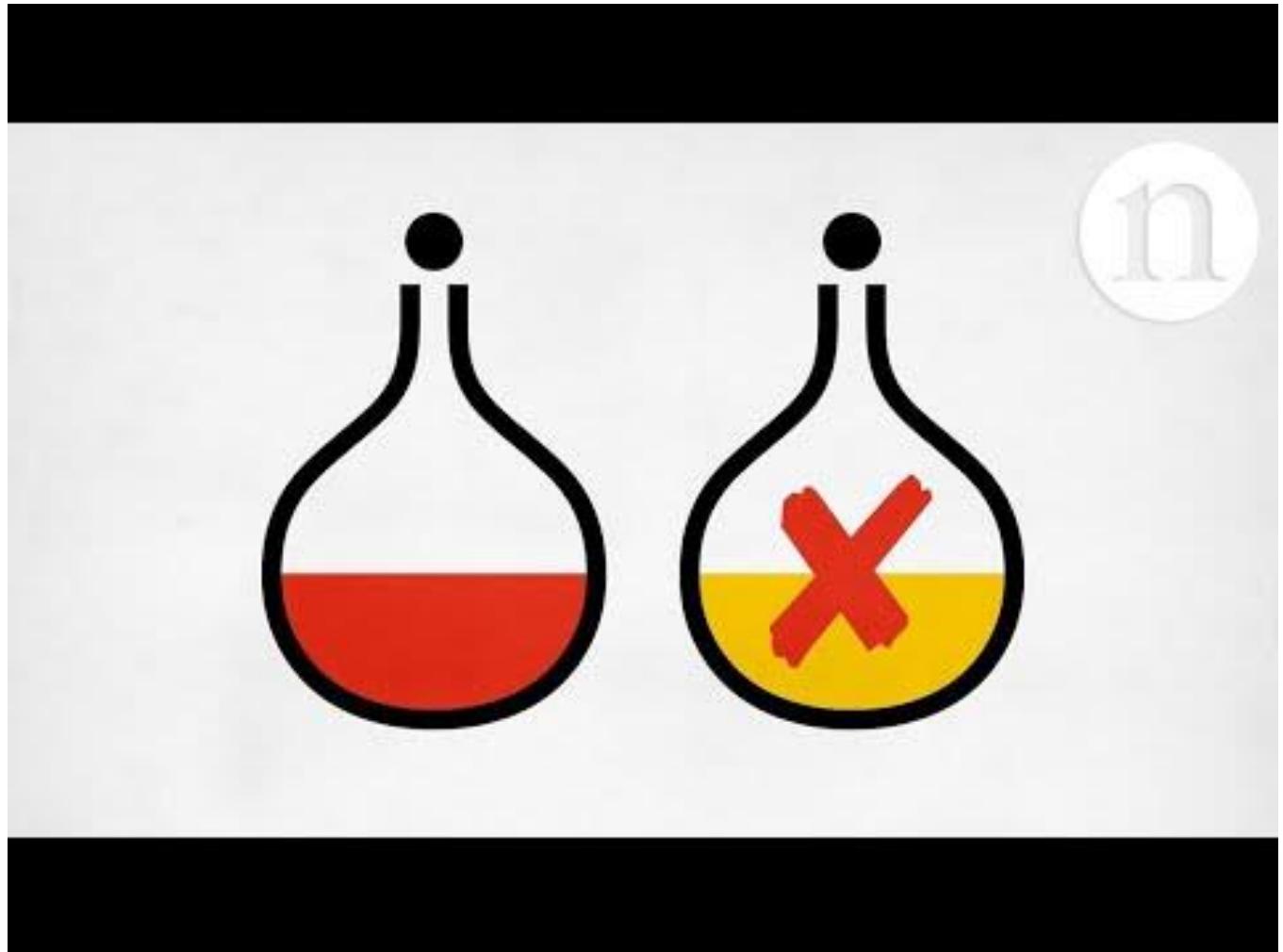
# 100% open access publishing

100% open access in 2020 requires universities to fill in how they will make open access publications part of the evaluation system in the intervening years.

## Open access in 2020

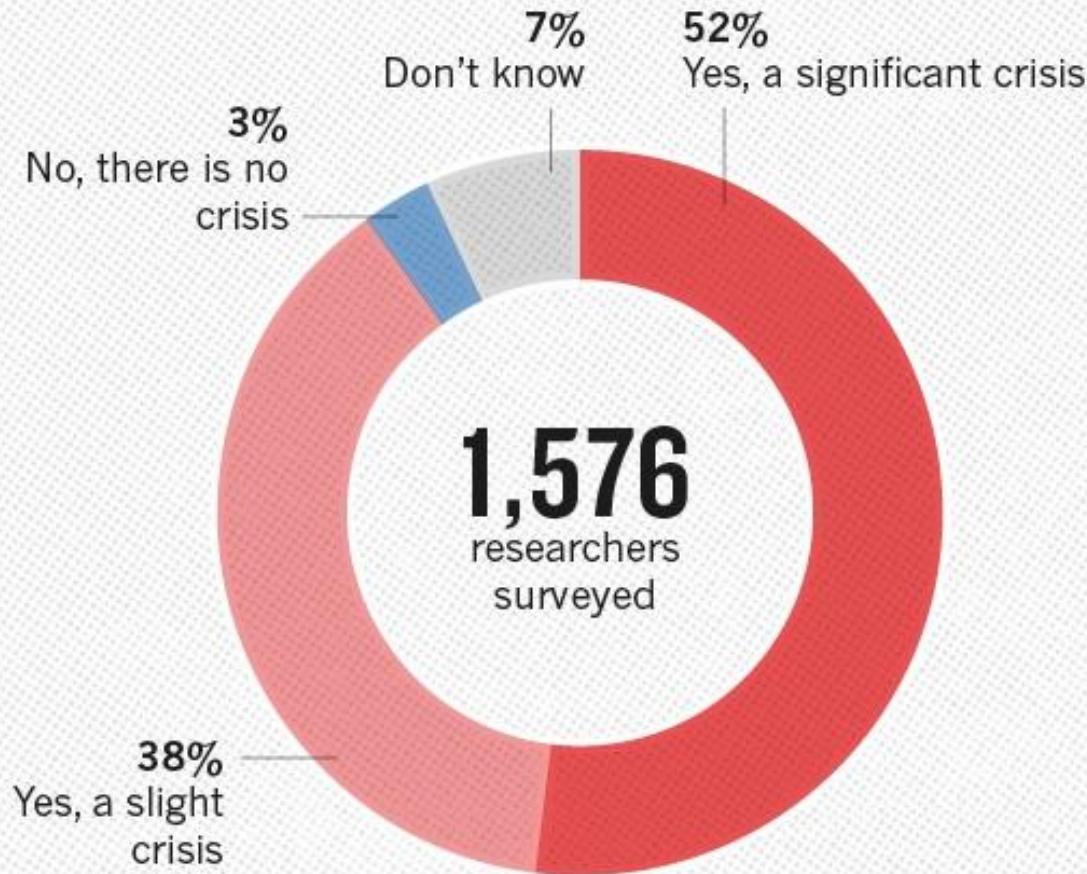
The Netherlands' ambition is to achieve 100% open access in 2020. The leading principle in this regard is that publicly-funded research results should also be accessible to the public at no extra cost.

# Reproducibility crisis in science

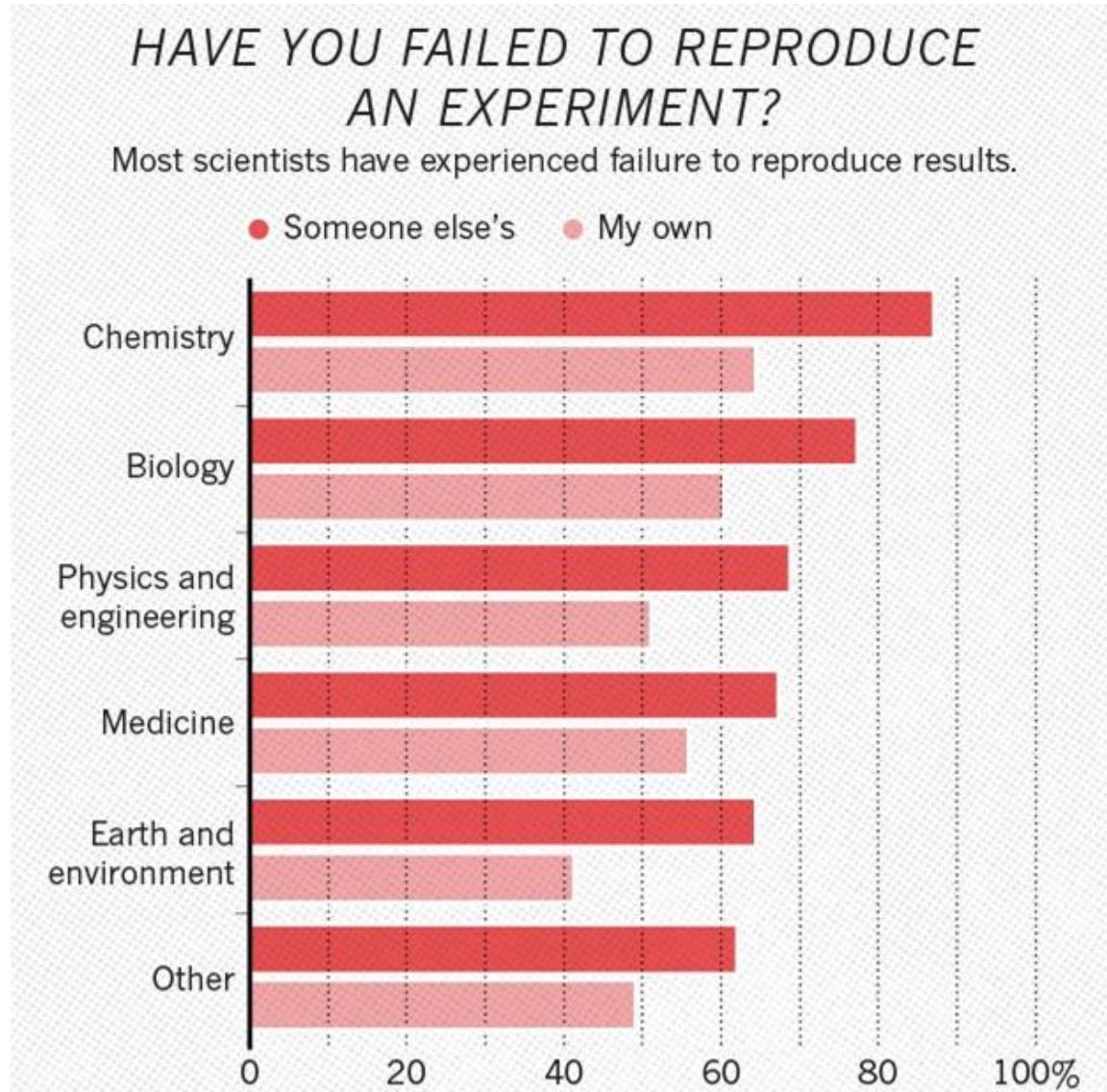


[https://www.youtube.com/watch?v=j7K3s\\_vi\\_1Y](https://www.youtube.com/watch?v=j7K3s_vi_1Y)

## IS THERE A REPRODUCIBILITY CRISIS?



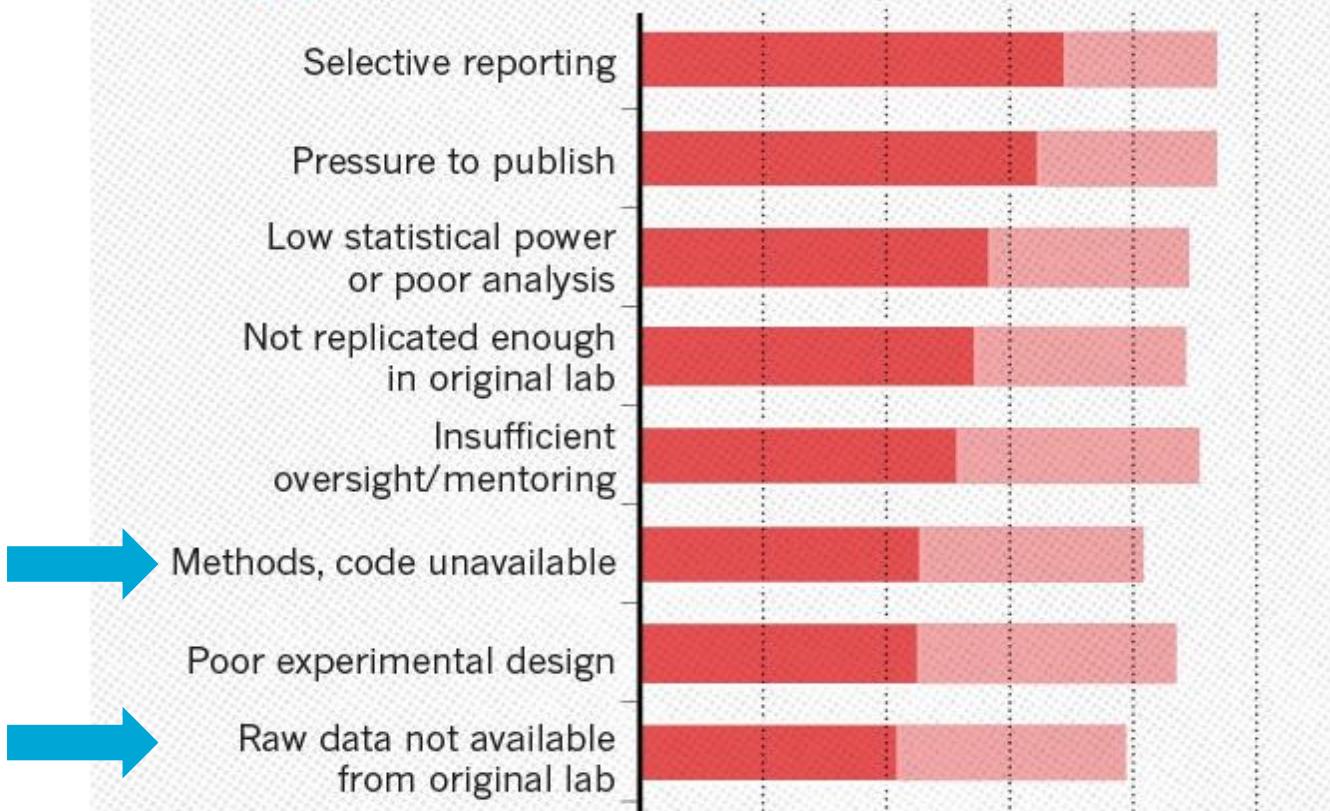
©nature



## WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

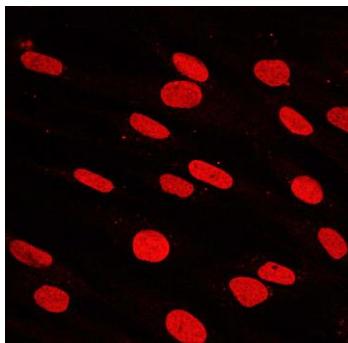
Many top-rated factors relate to intense competition and time pressure.

- Always/often contribute
- Sometimes contribute



# A close look to the research data life cycle

Raw data

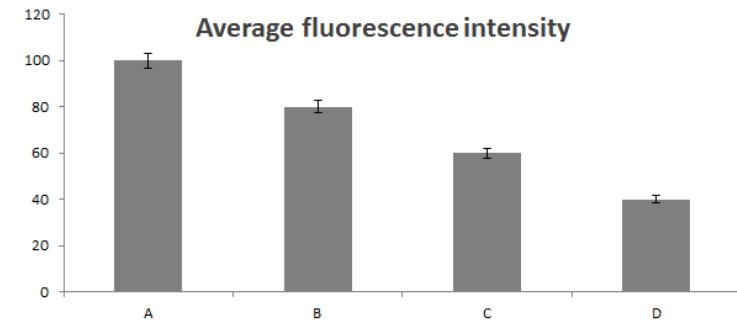


Intermediate data

Fluorescence intensity				
A	B	C	D	
98	82	58	39	
102	80	59	36	
100	75	61	37	
97	85	58	41	
96	81	60	35	
101	81	62	37	
101	77	56	43	
101	85	56	37	
98	85	57	39	
95	75	61	43	

Final data

Average fluorescence intensity



- Are the published final data available for validation, reproduction or reuse?

# Datasets available ‘on request’ are not available

Current Biology 24, 94–97, January 6, 2014 ©2014 Elsevier Ltd All rights reserved <http://dx.doi.org/10.1016/j.cub.2013.11.014>

Report

## The Availability of Research Data Declines Rapidly with Article Age

- Data availability decreases by **17% per year**
- Chance of email address working decreases by **7% per year**

<http://dx.doi.org/10.1016/j.cub.2013.11.014>

# Datasets available ‘on request’ are not available

Current Biology 24, 94–97, January 6, 2014 ©2014 Elsevier Ltd All rights reserved <http://dx.doi.org/10.1016/j.cub.2013.11.014>

Report

## The Availability of Research Data Declines Rapidly with Article Age

- Data availability decreases by **17% per year**
- Chance of email address working decreases by **7% per year**

What’s the alternative to sharing ‘on request’?

<http://dx.doi.org/10.1016/j.cub.2013.11.014>

# What is a repository?

A place where things can be stored and shared



# Repositories for datasets

**re3data.org**  
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://www.re3data.org/>

General purpose

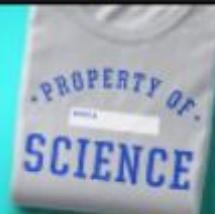


Discipline-specific

European Genome-phenome Archive



Read *Science*.  
Wear science.



Join now for a **free tee**  
plus 50 issues of *Science*.

SHARE

REPORT



0



D. G. E. Gomes<sup>1,2</sup>, R. A. Page<sup>1</sup>, I. Geipel<sup>1</sup>, R. C. Taylor<sup>1,3</sup>, M. J. Ryan<sup>1,4</sup>, W. Halfwerk<sup>1,5,\*</sup>

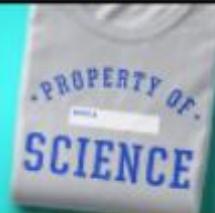
[+ See all authors and affiliations](#)



0

Science 16 Sep 2016:  
Vol. 353, Issue 6305, pp. 1277-1280  
DOI: 10.1126/science.aaf7934

Read *Science*.  
Wear science.



Join now for a **free tee**  
plus 50 issues of *Science*.

SHARE

REPORT



0



D. G. E. Gomes<sup>1,2</sup>, R. A. Page<sup>1</sup>, I. Geipel<sup>1</sup>, R. C. Taylor<sup>1,3</sup>, M. J. Ryan<sup>1,4</sup>, W. Halfwerk<sup>1,5,\*</sup>

[+ See all authors and affiliations](#)

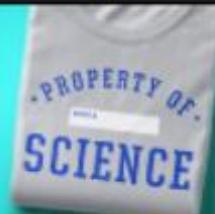


0

Science 16 Sep 2016:  
Vol. 353, Issue 6305, pp. 1277-1280  
DOI: 10.1126/science.aaf7934

404 NOT FOUND

Read *Science*.  
Wear science.



Join now for a **free tee**  
plus 50 issues of *Science*.

SHARE

REPORT



0



D. G. E. Gomes<sup>1,2</sup>, R. A. Page<sup>1</sup>, I. Geipel<sup>1</sup>, R. C. Taylor<sup>1,3</sup>, M. J. Ryan<sup>1,4</sup>, W. Halfwerk<sup>1,5,\*</sup>

[+ See all authors and affiliations](#)



0

Science 16 Sep 2016:  
Vol. 353, Issue 6305, pp. 1277-1280  
DOI: 10.1126/science.aaf7934

Raw data are available at the Dryad Data Repository ([dx.doi:10.5061/dryad.5gk8j](https://doi.org/10.5061/dryad.5gk8j)).

**Data from: Bats perceptually weight prey cues across sensory systems when hunting in noise**

Gomes DGE, Page RA, Geipel I, Taylor RC, Ryan MJ, [Halfwerk W](#)

Date Published: September 21, 2016

DOI: <https://doi.org/10.5061/dryad.5gk8j>

When using this data, please cite the original publication:

Gomes DGE, Page RA, Geipel I, Taylor RC, Ryan MJ, Halfwerk W (2016) Bats perceptually weight prey cues across sensory systems when hunting in noise. *Science* 353(6305): 1277-1280.  
<https://doi.org/10.1126/science.aaf7934>

Additionally, please cite the Dryad data package:

Gomes DGE, Page RA, Geipel I, Taylor RC, Ryan MJ, [Halfwerk W](#) (2016) Data from: Bats perceptually weight prey cues across sensory systems when hunting in noise. Dryad Digital Repository. <https://doi.org/10.5061/dryad.5gk8j>

[Cite](#) | [Share](#)

# Data Documentation

## Human readable

- Readme files with info about:
  - Methods used for data collection and analysis
  - Data-specific information (parameters, variables, column headings, symbols used, etc.)

# Data Documentation

## Human readable

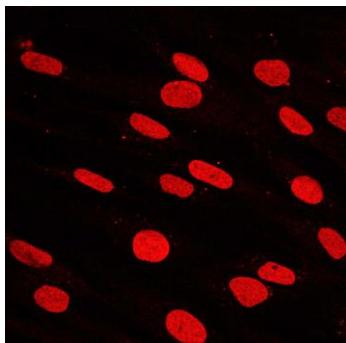
- Readme files with info about:
  - Methods used for data collection and analysis
  - Data-specific information (parameters, variables, column headings, symbols used, etc.)

## Machine readable

- Metadata with defined fields:
  - Title, date, creator(s), keywords..
  - Disciplinary standards if possible

# A close look to the research data life cycle

Raw data

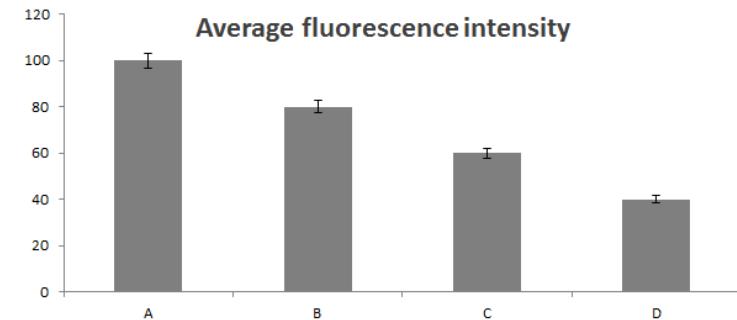


Intermediate data

Fluorescence intensity				
A	B	C	D	
98	82	58	39	
102	80	59	36	
100	75	61	37	
97	85	58	41	
96	81	60	35	
101	81	62	37	
101	77	56	43	
101	85	56	37	
98	85	57	39	
95	75	61	43	

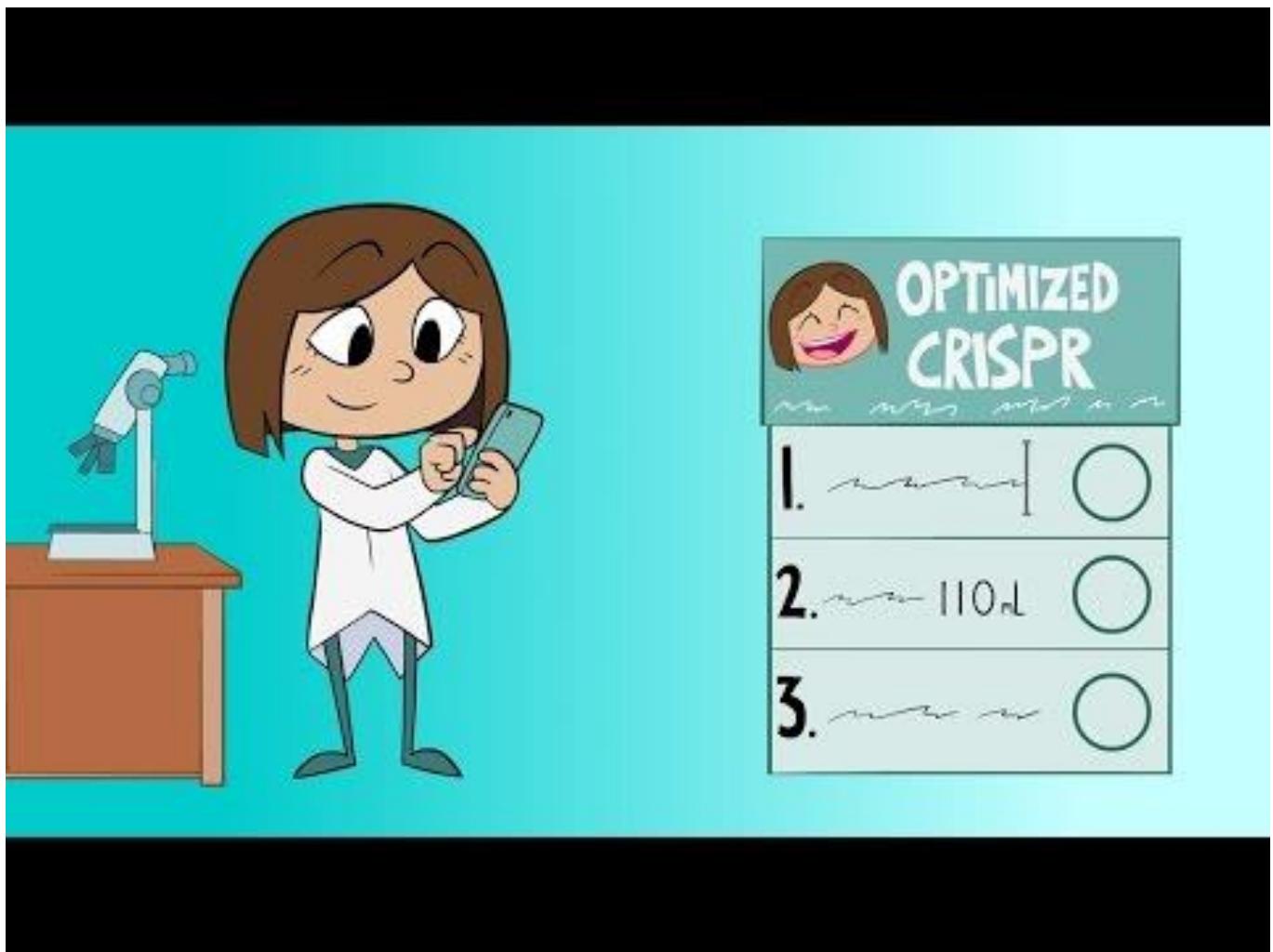
Final data

Average fluorescence intensity

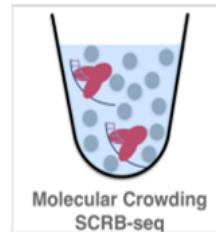


- Are the published final data available for validation, reproduction or reuse?
- What about experimental methods and measurement parameters?

# What is protocols.io?



<https://www.youtube.com/watch?v=84B8P6BAOgM>

[Steps](#)[Abstract](#)[Guidelines](#)[Materials](#)[Forks](#)[More](#)[RUN](#)[COPY / FORK](#) [EXPORT](#) 

Version 2

May 22, 2018

Working

## mcSCRB-seq protocol

Nature Communications

Johannes Bagnoli<sup>1</sup>, Christoph Ziegenhain<sup>1</sup>, Aleksandar Janjic<sup>1</sup>, Lucas Esteban Wange<sup>1</sup>, Beate Vieth<sup>1</sup>, Swati Parekh<sup>1</sup>, Johanna Geuder<sup>1</sup>, Ines Hellmann<sup>1</sup>, Wolfgang Enard<sup>1</sup>

<sup>1</sup>Ludwig-Maximilians-Universität München

[dx.doi.org/10.17504/protocols.io.p9kdr4w](https://doi.org/10.17504/protocols.io.p9kdr4w)

Human Cell Atlas Method Development Community



Aleksandar Janjic

Ludwig-Maximilians-Universität München



### BEFORE STARTING

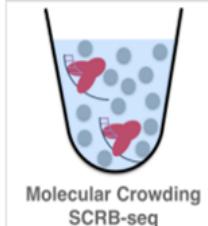
Wipe bench surfaces with RNase Away and keep working environment clean.

#### Preparation of lysis plates

- 1 Prepare **Lysis Buffer** according to the number of plates to be filled.

	A	B	C	
--	---	---	---	--

[dx.doi.org/10.17504/protocols.io.p9kdr4w](https://doi.org/10.17504/protocols.io.p9kdr4w)

[Steps](#)[Abstract](#)[Guidelines](#)[Materials](#)[Forks](#)[More](#)[RUN](#)[COPY / FORK](#) [EXPORT](#) 

## mcSCRB-seq protocol

Nature Communications

Johannes W. Bagnoli, et al (2018) Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nature Communications* 9:2937. doi: 10.1038/s41467-018-05347-6

Johannes Bagnoli<sup>1</sup>, Christoph Ziegennain<sup>1</sup>, Aleksandar Janjic<sup>1</sup>, Lucas Esteban Wangen<sup>1</sup>, Beate Vieth<sup>1</sup>, Swati Parekh<sup>1</sup>, Johanna Geuder<sup>1</sup>, Ines Hellmann<sup>1</sup>, Wolfgang Enard<sup>1</sup>

Version 2

<sup>1</sup>Ludwig-Maximilians-Universität München

Version 1 - Mar 19, 2018

7504/protocols.io.p9kdr4w

Version 2 - May 22, 2018 •

Atlas Method Development Community



Aleksandar Janjic

Ludwig-Maximilians-Universität München



### BEFORE STARTING

Wipe bench surfaces with RNase Away and keep working environment clean.

#### Preparation of lysis plates

- 1 Prepare **Lysis Buffer** according to the number of plates to be filled.



[dx.doi.org/10.17504/protocols.io.p9kdr4w](https://dx.doi.org/10.17504/protocols.io.p9kdr4w)

# FAIR Data Principles



# FAIR Data Principles



[http://ec.europa.eu/research/press/2016/pdf/opendata-infographic\\_072016.pdf#view=fit&pageMode=none](http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf#view=fit&pageMode=none)

- Requirement of NWO, ZonMw, Horizon 2020



# Data Stewardship-How?



# What does all of this mean for you?

- Awareness about importance of research data management
- Training about:
  - How to securely save and backup data?
  - How to work with confidential (commercial or personal sensitive) data?
    - Access control, encryption, anonymization...
  - File naming and version control
  - Data documentation
  - Data analysis skills (learning Python, R, ...)
  - ...

# Coming back to the bigger picture...



On road to reproducible research!

# Thank you Questions?

