Open Reproducible Research

Concepts, challenges, and solutions

Markus Konkol, Research Software Engineer @MarkusKonkol

52north

Learning Goals

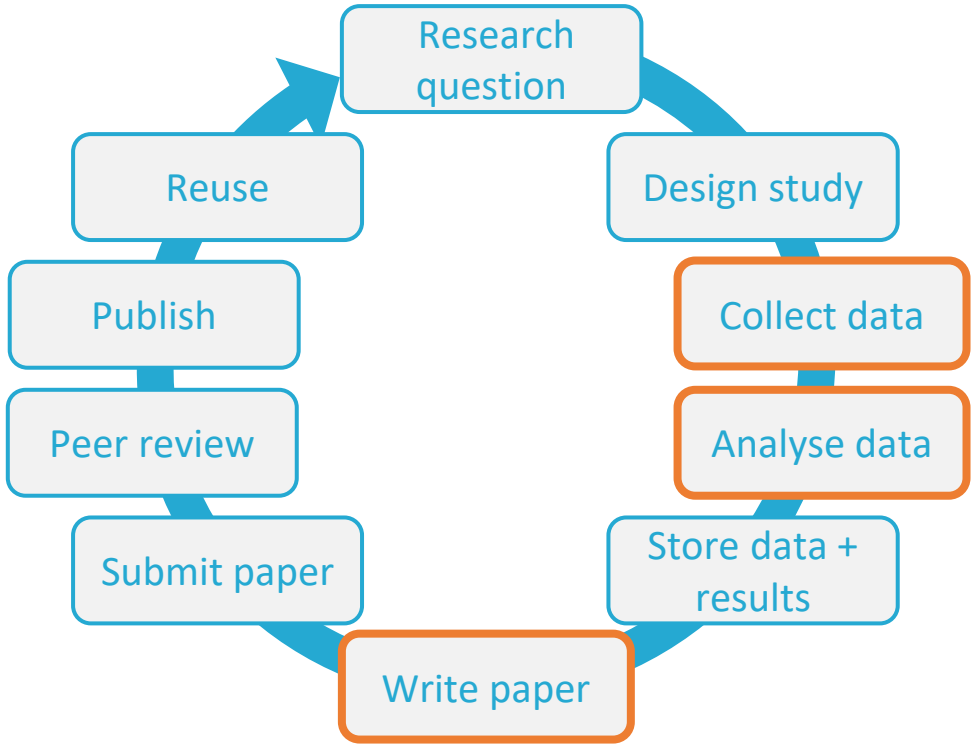
Upon completion of this lecture, you will be able to

- Articulate what Open Reproducible Research is
- Understand which obstacles impede Open Reproducible Research
- Apply Open Reproducible Research principles to your own work
- Choose appropriate tools to publish Open Reproducible Research

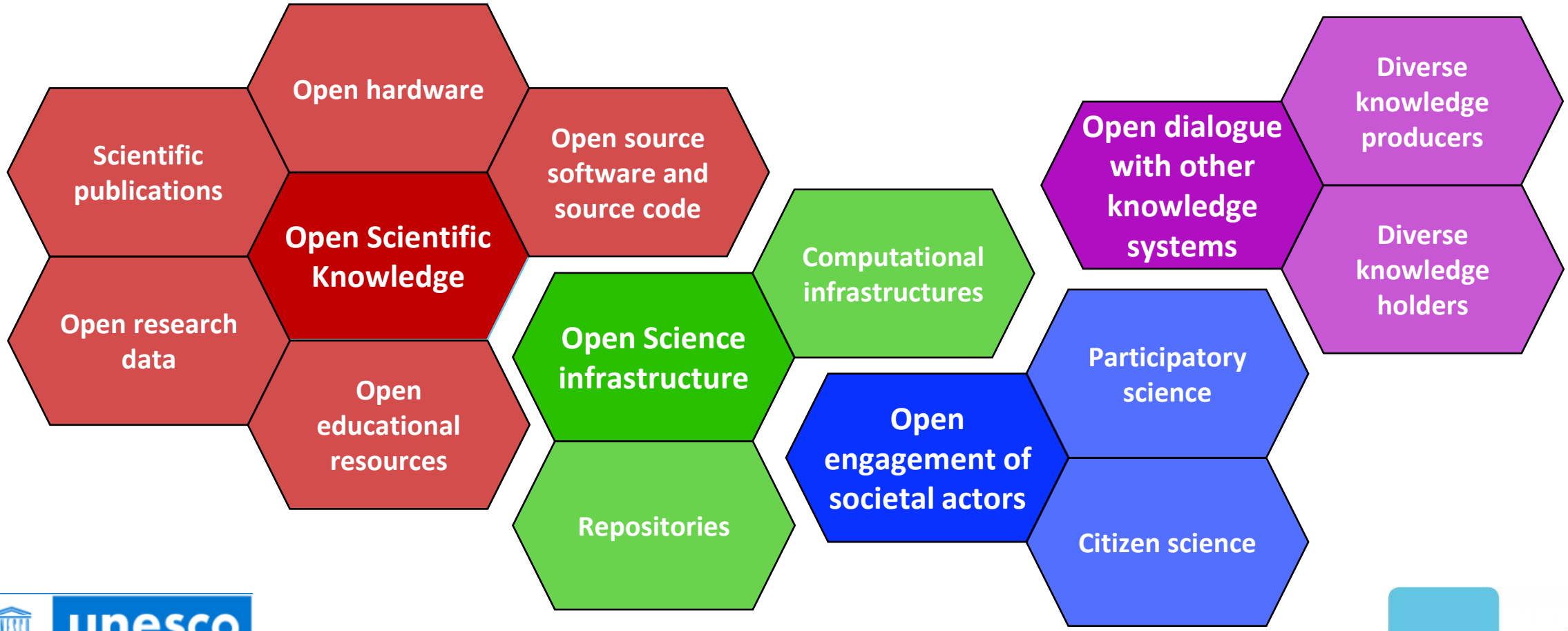
Agenda

1. Introduction to Open Reproducible Research
2. The reproducibility crisis
3. Obstacles impeding reproducibility
4. Five recommendations for ORR
5. Principles and best practices
6. Opportunities coming with ORR

Introduction to Open Reproducible Research



Introduction to Open Reproducible Research



UNESCO Recommendation on Open Science

Introduction to Open Reproducible Research

"[..] an article about a computational result is advertising, not scholarship. The actual scholarship is the [...] complete set of instructions which generated the figures.."

(Claerbout's claim)

Further reading: Buckheit and Donoho (2010)

"The problem is that most modern science is so complicated, and most journal articles so brief, it's impossible for the article to include details of many important methods and decisions made by the researcher as he analyzed his data on his computer."

Further reading: Marwick (2015)

"From time to time over the past few years, I've politely refused requests to referee an article on the grounds that it lacks enough information for me to check the work."

Further reading: Stark (2018)

52north

Introduction to Open Reproducible Research

Reproducible Research refers to achieving the **same results** (e.g., tables, figures, numbers) as reported in the paper by using the **same source code and data**.

In **Open Reproducible Research**, these materials are **publicly accessible**.

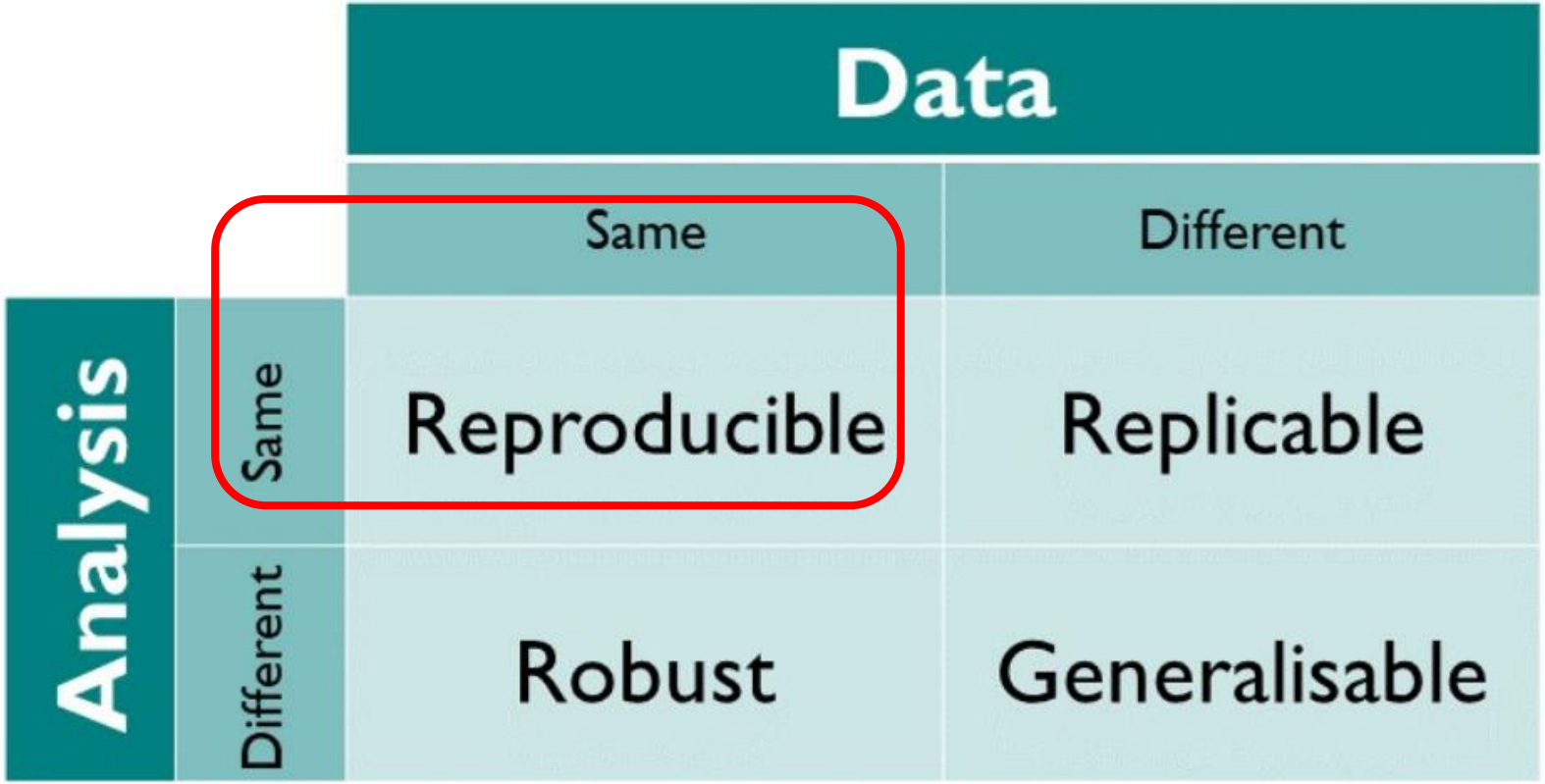
Replicable Research refers to coming to **similar conclusions** based on the **same analysis**, but **newly collected data**.

Reproducibility & Replicability are both essential for **transparent**, **verifiable**, and **reusable** scientific work.

Further reading: Goodman et al. (2016)

52north

Introduction to Open Reproducible Research



Further reading: The Turing Way

52north

Introduction to Open Reproducible Research

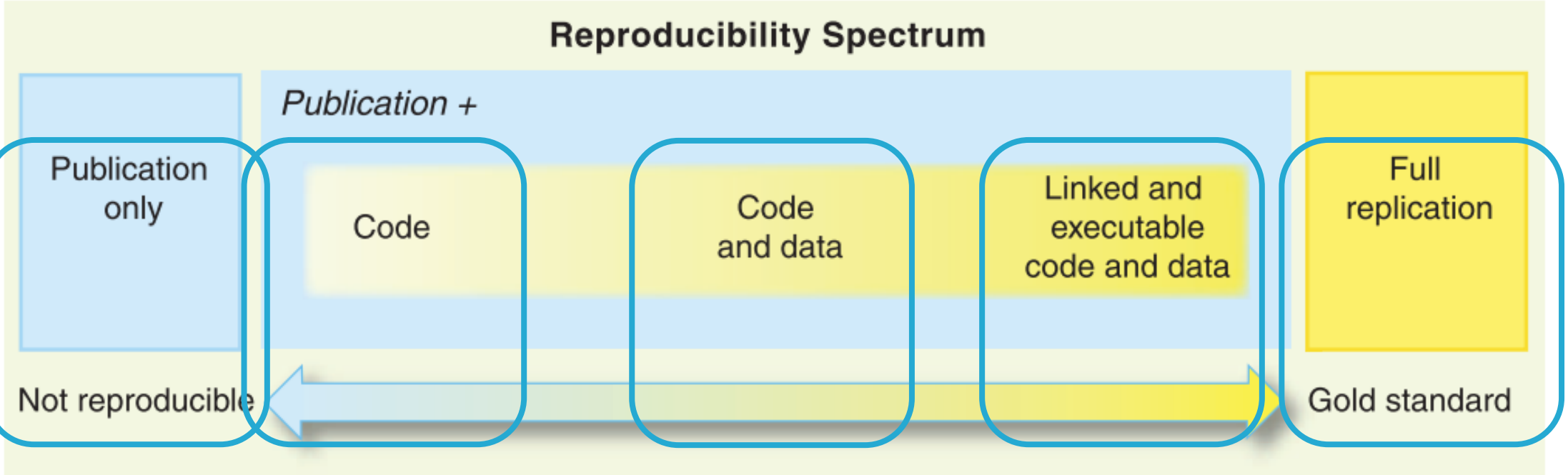


Fig. 1. The spectrum of reproducibility.

Further reading: Peng (2011)

52north

Introduction to Open Reproducible Research

Lessons learned:

1. Source code to run the analysis and research data are as important as the scientific article.
2. Open Reproducible Research: Same research data, same source code, same results.
3. Reproducibility is not necessarily a binary concept but rather a spectrum.
4. Reproducibility + replicability + robustness + generalisability describe different concepts.

Why is unreproducible research a problem?

- The analysis is not fully transparent and easily understandable.
  - Difficult/impossible to describe analysis in pure text.
  - Access to source code can be a shortcut.
- The analysis is not verifiable.
  - Reviewers need to trust the results.
  - Investigating the analysis in any case a complex task.
- The analysis is not reusable.
  - Waste of time and money (duplication of efforts).
  - Waste of opportunities for collaborations and credit.

Further reading: Gil et al. (2016)

52north

Five ‘selfish’ reasons to do reproducible research

Reason number 1: reproducibility helps to avoid disaster

Reason number 2: reproducibility makes it easier to write papers

Reason number 3: reproducibility helps reviewers see it your way

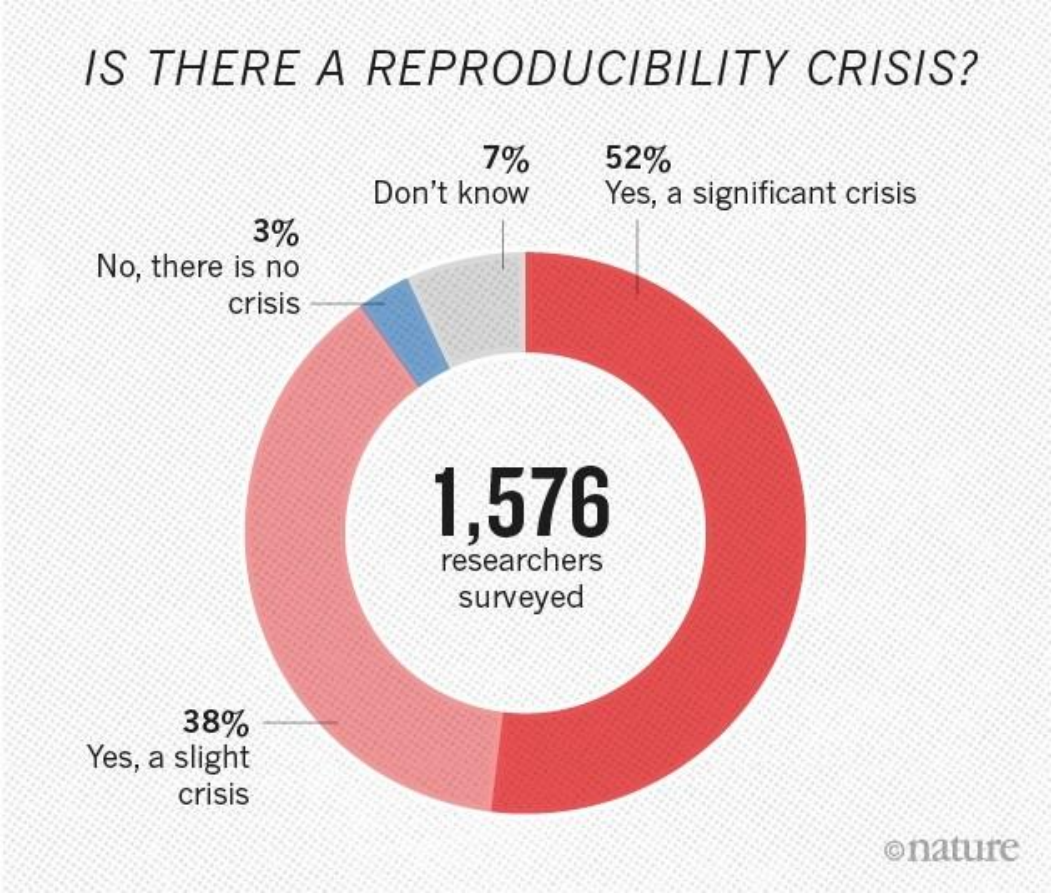
Reason number 4: reproducibility enables continuity of your work

Reason number 5: reproducibility helps to build your reputation

Further reading: Markowetz (2015)

52north

The Reproducibility Crisis



Reproducible Research refers to achieving the **same results** (e.g., tables, figures, numbers) as reported in the paper by using the **same source code and data**.
In **Open Reproducible Research**, these materials are **publicly accessible**.

Further reading: Baker (2016)

52north

The Reproducibility Crisis

COMPUTER SCIENCE

Artificial intelligence faces reproducibility crisis

Unpublished code and sensitivity to training conditions make many claims hard to verify

By **Matthew Hutson**

Last year, computer scientists at the University of Montreal (U of M) in Canada were eager to show off a new speech recognition algorithm, and they wanted to compare it to a benchmark, an algorithm from a well-known scientist. The only problem: The benchmark's source code wasn't published. The researchers had to recreate it from the (AAAI) in New Orleans, Louisiana, reproducibility was on the agenda, with some teams diagnosing the problem—and one laying out tools to mitigate it.

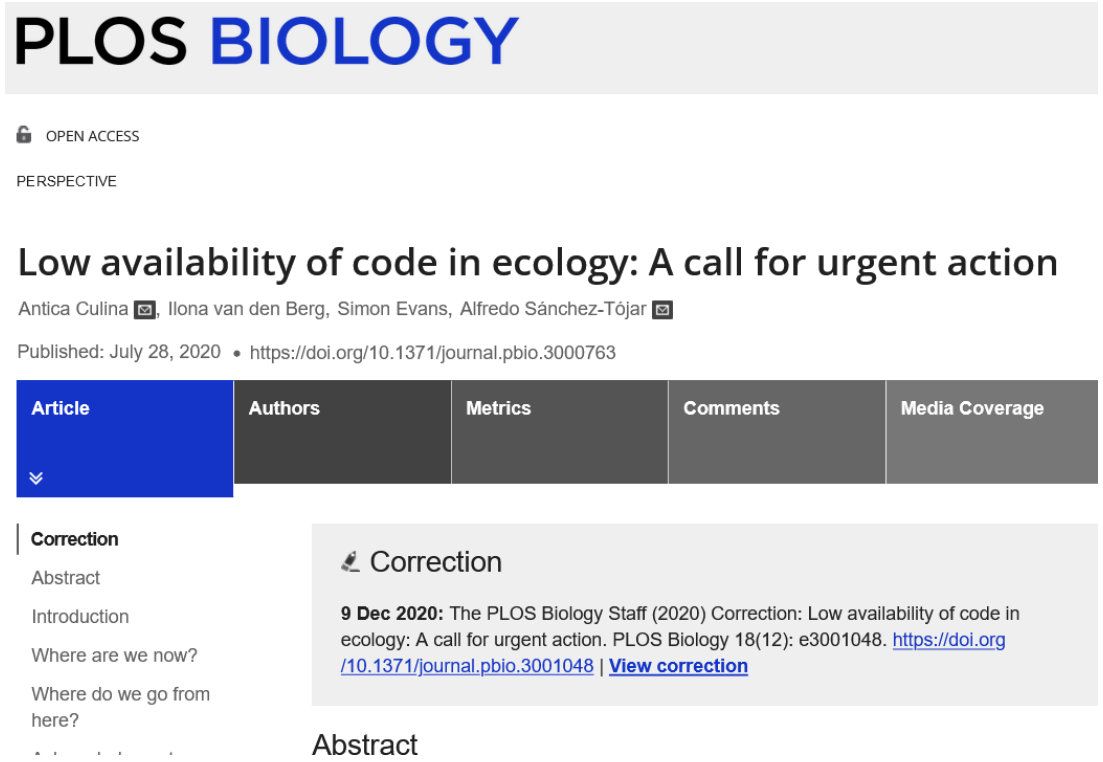
The most basic problem is that researchers often don't share their source code. At the AAAI meeting, Odd Erik Gundersen, a computer scientist at the Norwegian University of Science and Technology in Trondheim, reported the results of a survey of 400 algorithms presented in papers at two

- Checked 400 papers for reproducibility.
- 6% included source code of the algorithms.
- 30% included test data.
- 54% included a limited summary of the source code (a.k.a. pseudocode).
- Reasons for unavailable source code:
  - Source code is work in progress.
  - A company owned the source code.
  - Hidden to keep competitive advantage.
  - Stolen or lost.

Further reading: Hutson (2018)

52north

The Reproducibility Crisis



PLOS BIOLOGY

OPEN ACCESS

PERSPECTIVE

Low availability of code in ecology: A call for urgent action

Antica Culina, Ilona van den Berg, Simon Evans, Alfredo Sánchez-Tójar

Published: July 28, 2020 • https://doi.org/10.1371/journal.pbio.3000763

Article | Authors | Metrics | Comments | Media Coverage

Correction

Abstract

Introduction

Where are we now?

Where do we go from here?

Correction

9 Dec 2020: The PLOS Biology Staff (2020) Correction: Low availability of code in ecology: A call for urgent action. PLOS Biology 18(12): e3001048. https://doi.org/10.1371/journal.pbio.3001048 | **View correction**

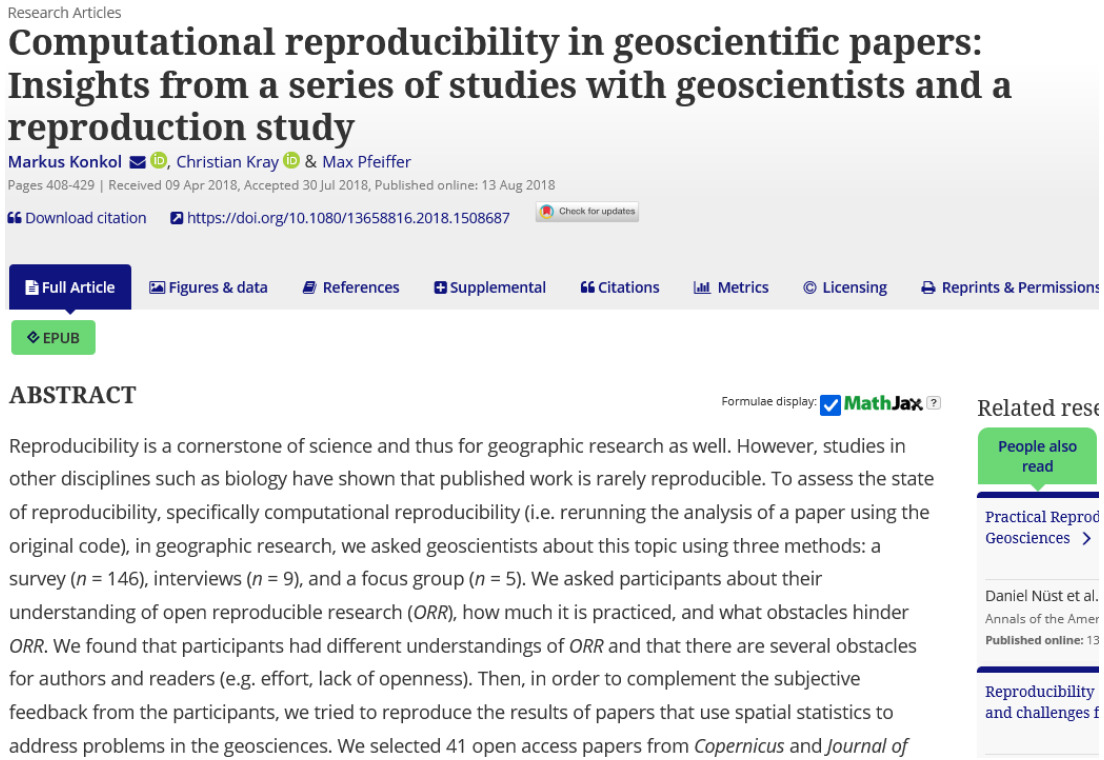
Abstract

- Checked 346 papers for reproducibility.
- 27% had source code attached.
- 79% were accompanied by data.
- Is data more important than code?
  - NO! The analysis is the context of the data.
  - **Paper, data, and analysis belong together.**
  - Needed to achieve transparency, verifiability, and reusability

Further reading: Culina et al. (2020)

52north

The Reproducibility Crisis



Research Articles

Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study

Markus Konkol, Christian Kray & Max Pfeiffer

Pages 408-429 | Received 09 Apr 2018, Accepted 30 Jul 2018, Published online: 13 Aug 2018

Download citation https://doi.org/10.1080/13658816.2018.1508687 Check for updates

Full Article | Figures & data | References | Supplemental | Citations | Metrics | Licensing | Reprints & Permissions

EPUB

ABSTRACT

Formulae display: MathJax

Reproducibility is a cornerstone of science and thus for geographic research as well. However, studies in other disciplines such as biology have shown that published work is rarely reproducible. To assess the state of reproducibility, specifically computational reproducibility (i.e. rerunning the analysis of a paper using the original code), in geographic research, we asked geoscientists about this topic using three methods: a survey (n = 146), interviews (n = 9), and a focus group (n = 5). We asked participants about their understanding of open reproducible research (*ORR*), how much it is practiced, and what obstacles hinder *ORR*. We found that participants had different understandings of *ORR* and that there are several obstacles for authors and readers (e.g. effort, lack of openness). Then, in order to complement the subjective feedback from the participants, we tried to reproduce the results of papers that use spatial statistics to address problems in the geosciences. We selected 41 open access papers from *Copernicus* and *Journal of*

Related rese

People also read

Practical Reprod Geosciences

Daniel Nüst et al. Annals of the Amer Published online: 13

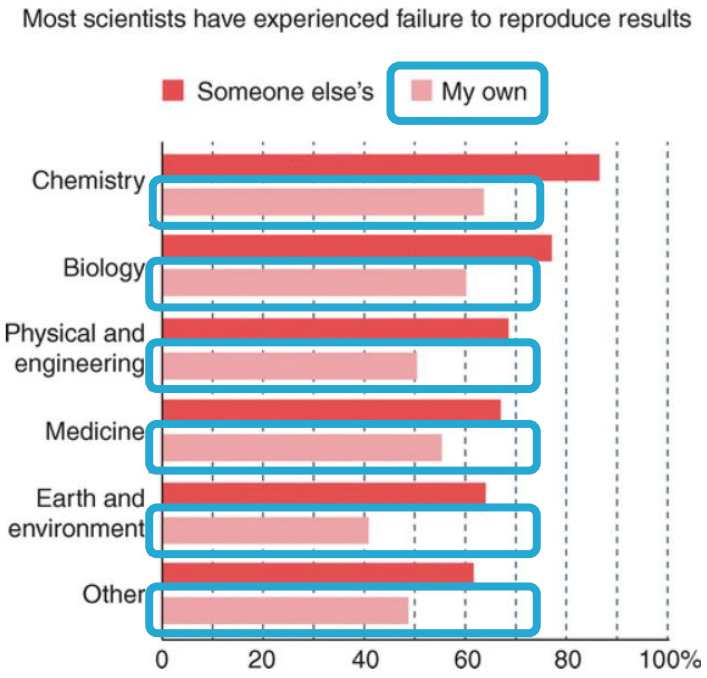
Reproducibility and challenges f

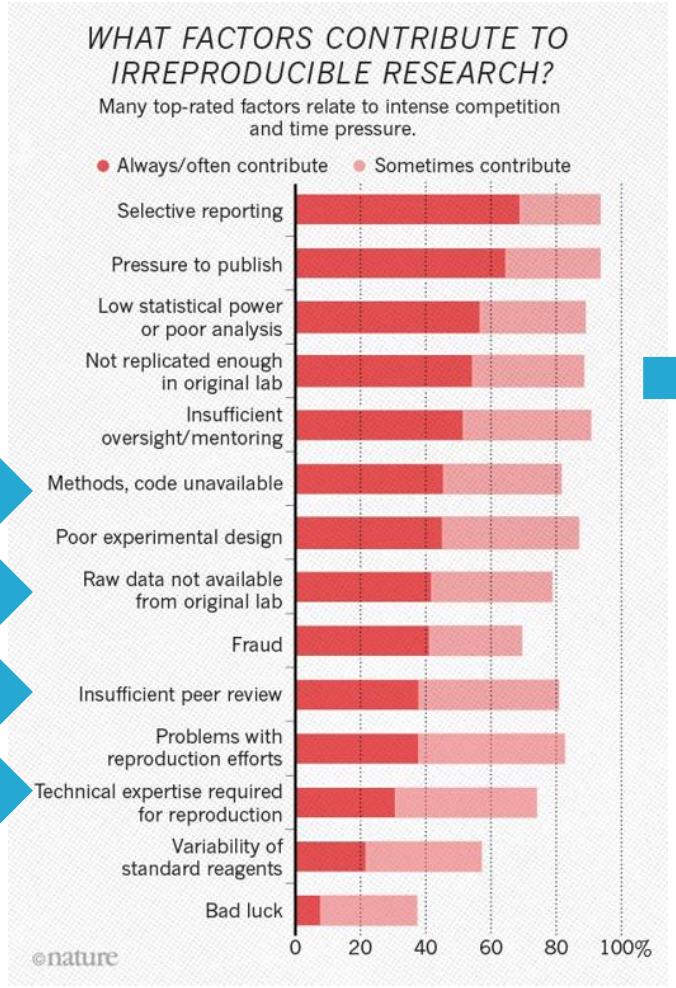
- Checked 41 papers that had code and data attached for executability and reproducibility.
- 2 out of 41 papers were executable + reproducible.
  - Several technical issues.
  - Several content-related differences.
  - More on that later…

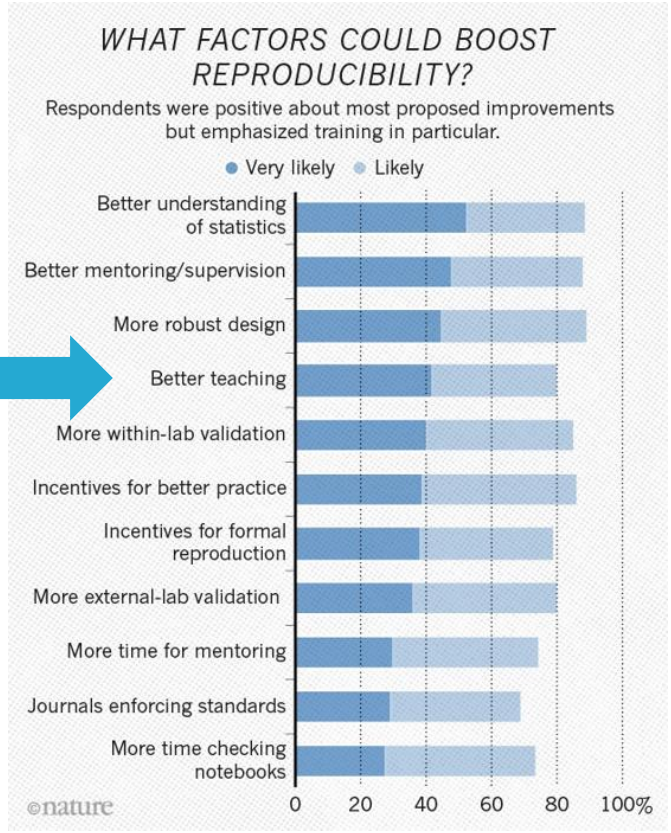
Further reading: Konkol et al. (2018)

52north

The Reproducibility Crisis







Further reading: Baker (2016)

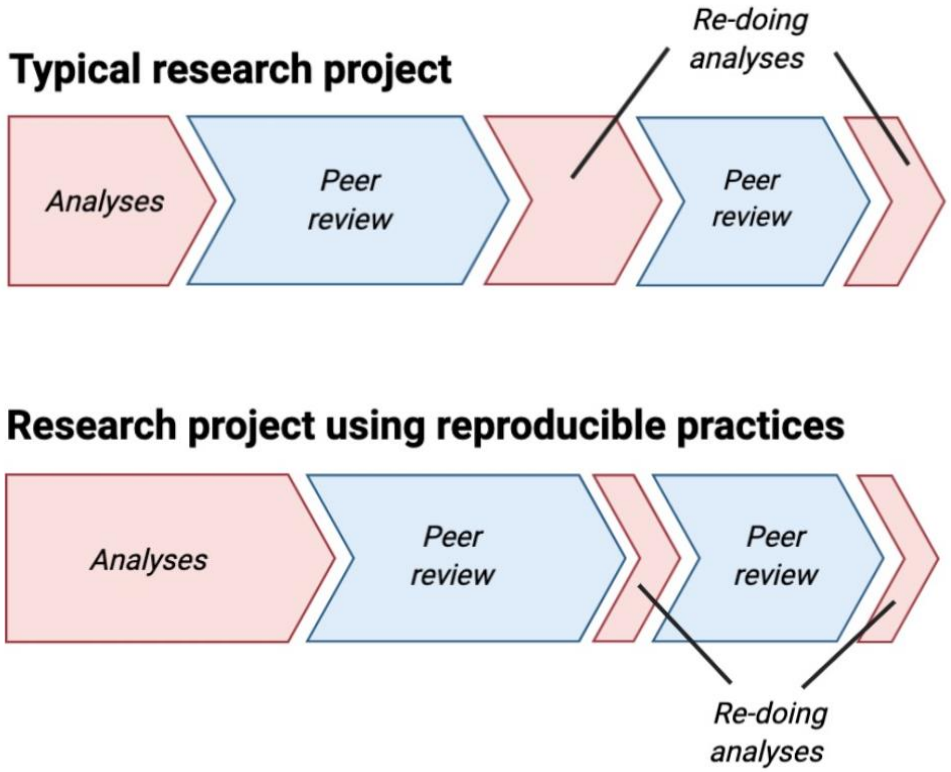
The Reproducibility Crisis

Lessons learned:

1. Reproducibility is needed to get transparent, verifiable, and reusable research results.
2. Reproducibility is beneficial for all + yourself (“Five ‘selfish’ reasons…”).
3. Many research articles do not have the source code and data attached.
4. Access to source code and data does not necessarily mean that the results are reproducible.
5. Are unreproducible research results necessarily wrong? Unclear…

Obstacles impeding ORR

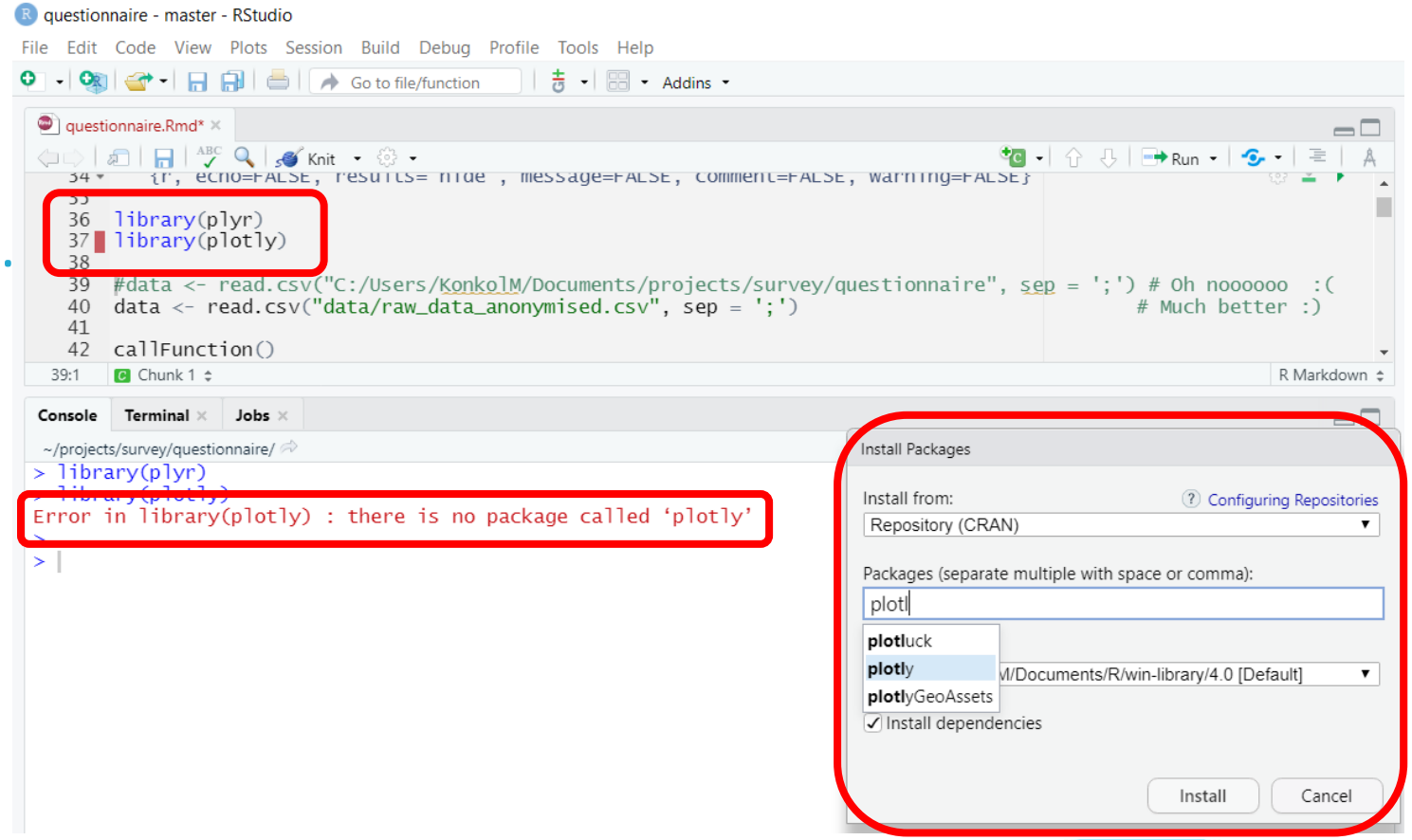
"Working reproducibly costs too much time, I need to publish papers!"



Further reading: Quintana (2020)

Technical obstacles impeding ORR

- Three categories of technical issues.
- *Minor issues* rather easy to solve.
- Example error: Library not found but available in repository.

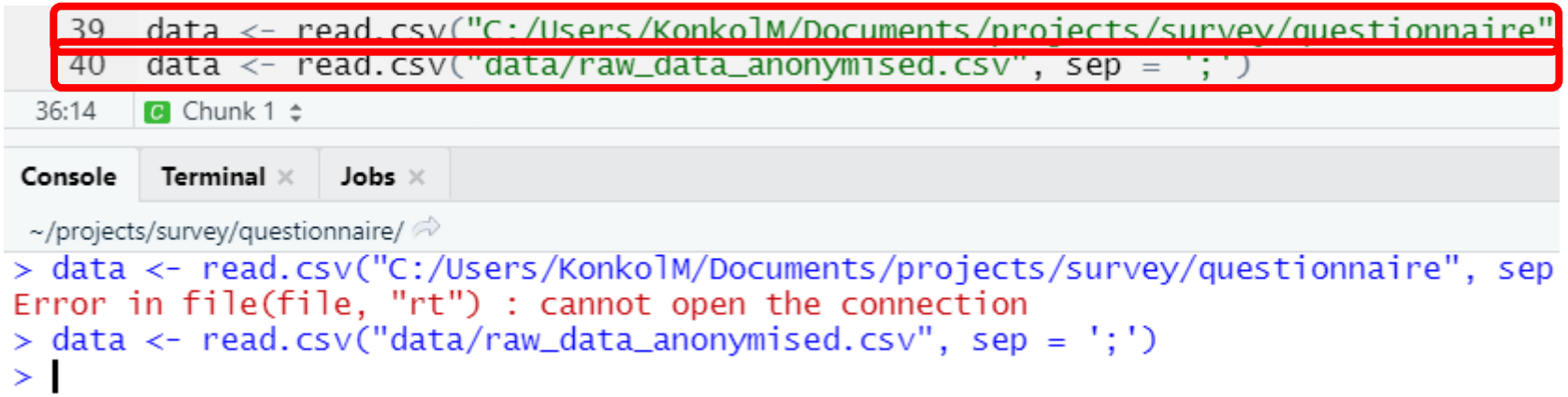


Further reading: Konkol et al. (2018)

52north

Technical obstacles impeding ORR

- *Substantial* issues require more effort.
- Example error: Wrong file directory.
- Solution: Use relative instead of absolute file paths.



Further reading: Konkol et al. (2018)

52north

Technical obstacles impeding ORR

- *Severe* issues require time, knowledge about the programming language, and understanding of the source code.
- Example error: *cannot open file dataABC.csv. No such file or directory.*
  - Was the file available in the folder? No ☹
  - Was the file created by the source code? No ☹
  - Contact author → get **missing** source code snippet that produced dataABC.csv.
- Solution: Ask a colleague to run your analysis.

19:00–20:00 Dinner

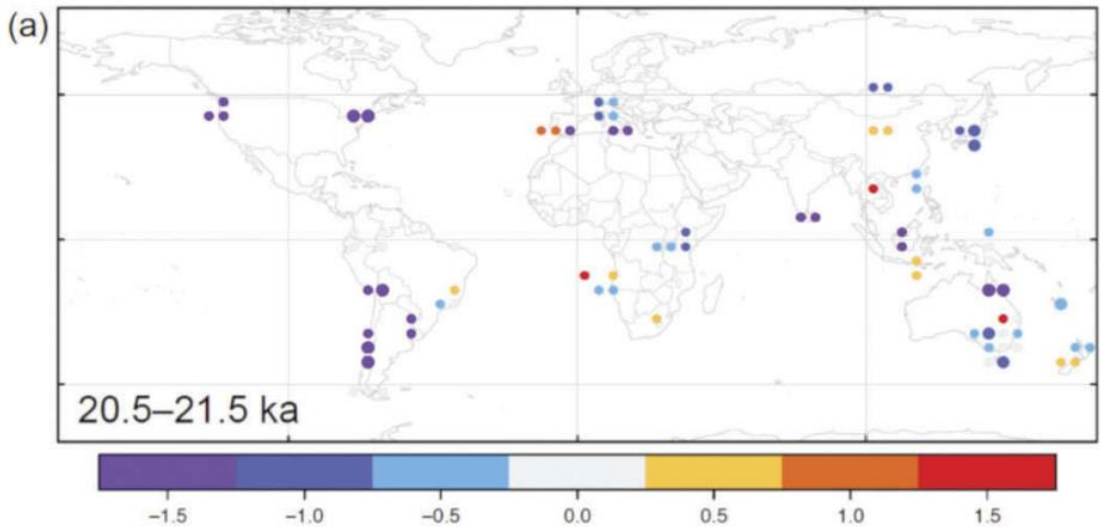
20:00–21:30 **Reproducibility Hackathon**

Further reading: Konkol et al. (2018)

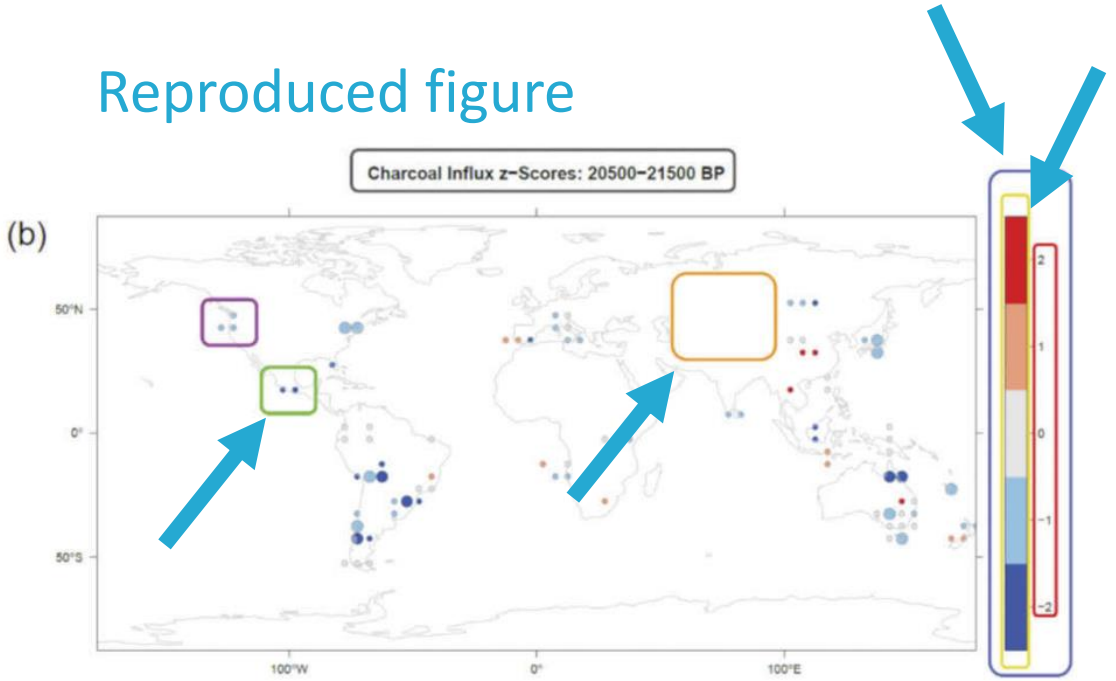
52north

Technical obstacles impeding ORR

Original figure



Reproduced figure



Further reading: Konkol et al. (2018)

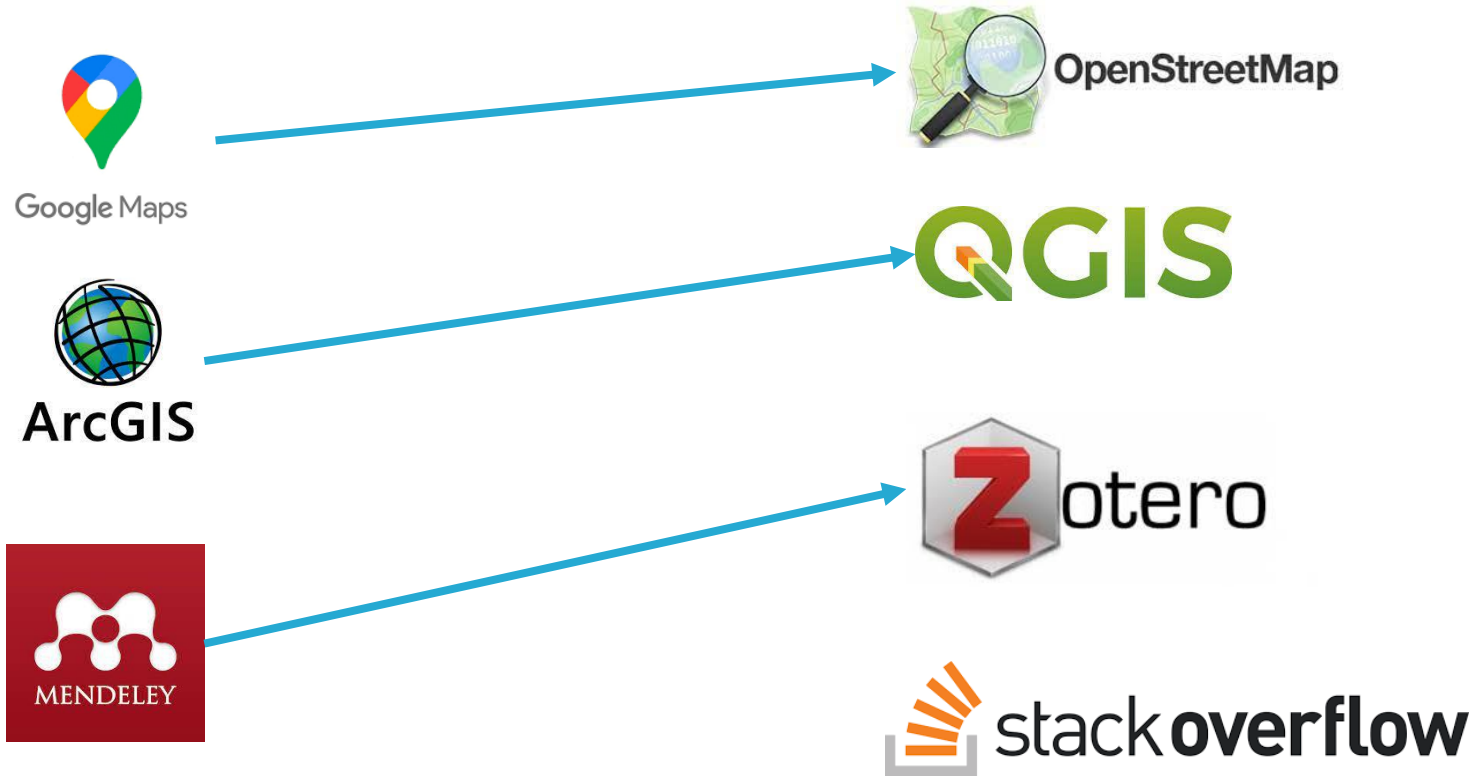
Technical obstacles impeding ORR

Lessons learned:

- Technical issues are of different complexity.
- *Minor* issues are easy to solve and require less time.
- *Substantial* issues require knowledge of the programming language.
- *Severe* issues require knowledge of the programming language and understanding of the source code.
- Reproduced figures can have differences related to the design + the numbers.

Five recommendations for ORR

Recommendation 1: Use open source software instead of commercial software.



Further reading: Hasselbring et al. (2020)

52north

Five recommendations for ORR

Recommendation 2: Learn a scripting language.

- Scripts describe every step of an analysis
- Human-readable description of what the code does
- Others can understand
  - What has been done
  - How it has been done

- Not reproducible
- No step-by-step description
- No control over the algorithms

Further reading: Alston and Rick (2020), Lasser (2020)

52north

Five recommendations for ORR

Recommendation 3: Learn a computational notebook format.

Further reading: Data science in Python & Jupyter, R Markdown Introduction

52north

Five recommendations for ORR

Recommendation 3: Learn a computational notebook format.

```
# Results

This section summarises the results of the survey.
Each subsection reports on the numbers, free text answers, and briefly discusses the results.
An overall discussion is provided in the next chapter.
A link to the dataset and the R Markdown file underlying the results is available in the supplements.

```{r, echo=FALSE, results="hide", message=FALSE, comment=FALSE, warning=FALSE}
likertPlot = function(dataSet, questions, x, y, scale, ordered){
  plotlevels <- scale
  subset = dataSet[,x:y]

  sapply(subset, class)
  sapply(subset, function(x) { length(levels(x)) } )

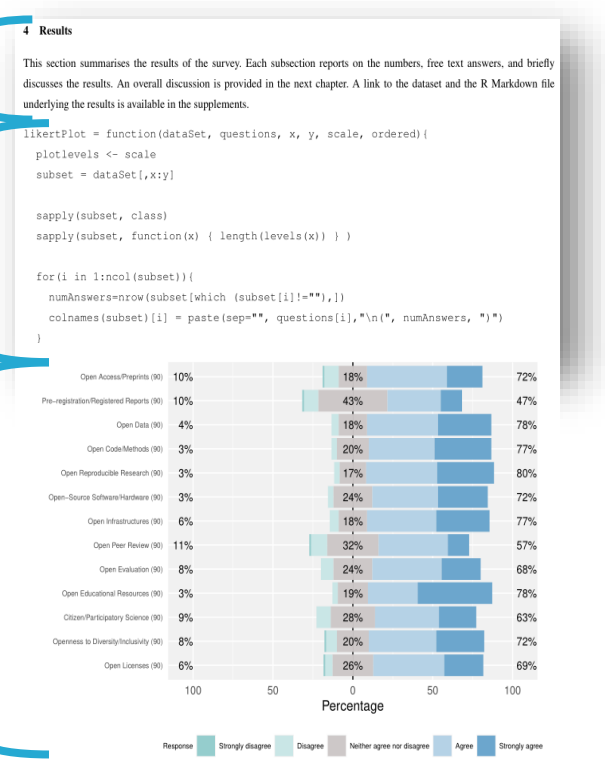
  for(i in 1:ncol(subset)){
    numAnswers=nrow(subset[which (subset[i]!=""),])
    colnames(subset)[i] = paste(sep="", questions[i],"\n(", numAnswers, ")")
  }

  for(i in seq_along(subset)) {
    subset[,i] <- factor(subset[,i], levels=plotlevels)
  }
```

Text

Code

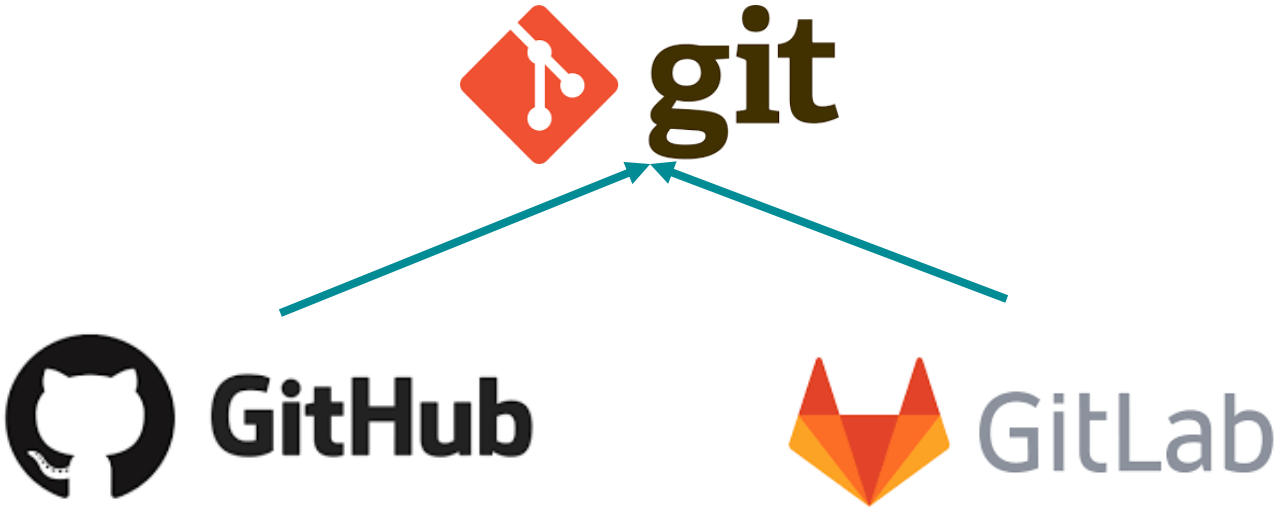
Output



Further reading: Data science in Python & Jupyter, R Markdown Introduction
```

Five recommendations for ORR

Recommendation 4: Learn a collaborative software development tool.



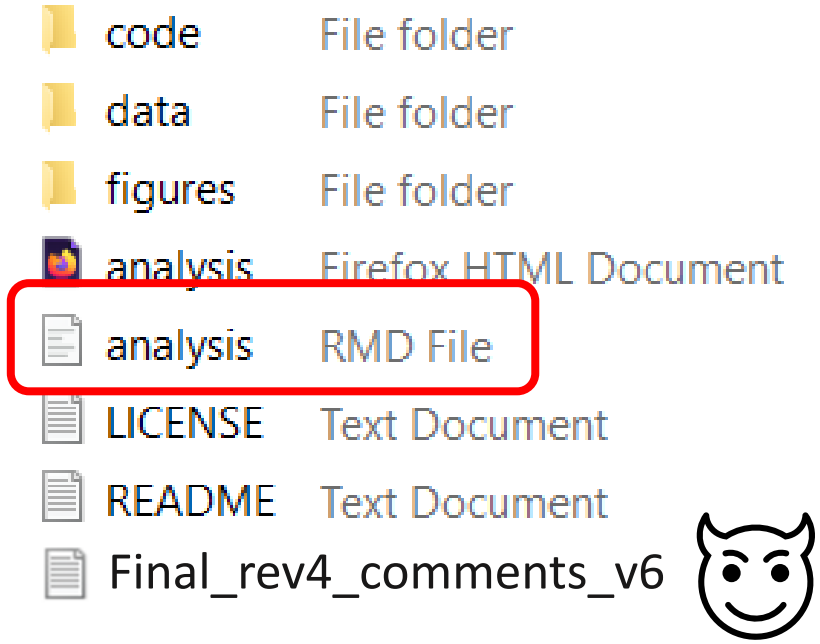
Further reading: Version Control with Git (software carpentry)

52north

Five recommendations for ORR

Recommendation 5: Document your source code.

- Create a clean workspace with a hierarchical folder structure and name files properly.
- Include a README text file to explain the code.
  - What does the software?
  - How can I install it?
  - Are there any computational requirements (e.g., operating system)?
  - How can I use it?
  - How long does the analysis take?
- Add a LICENSE, e.g., MIT License, APACHE License, or GNU.

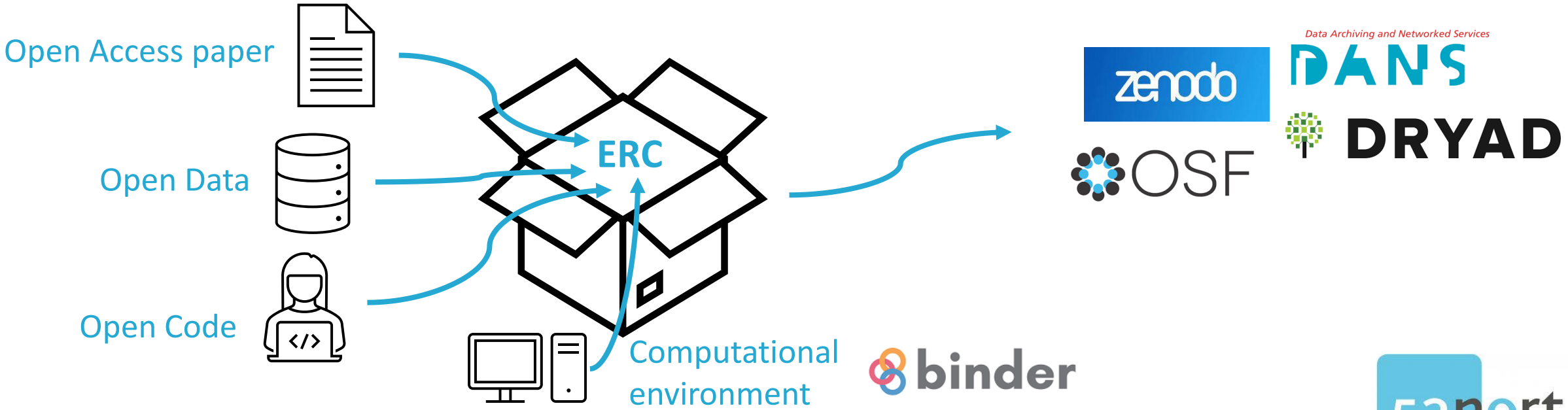


Further reading: Hettne et al. (2020)

Principles and best practices

Share the **scientific paper**, **research data**, **source code**, and details of the **computational environment** that generate published findings in **open trusted repositories**.

- Such a package is also known as *Executable Research Compendium* (ERC).



Further reading: Nüst et al. (2017), Stodden et al. (2016), Sandve et al. (2013)

52north

Principles and best practices

Insert a persistent identifier (e.g., DOI) in the published article that links to the data and source code underlying the results

- Example: *“Research data and source code supporting this publication is available on [name of the repository] and accessible via the following DOI: [doi to repository]”*

If legitimate reasons to restrict access to the materials apply to your work, mention it.

- Example: *“Research data and source code supporting this publication is not available due to [indicate reasons, e.g., licenses, data on human subjects, private or sensitive data etc.]*

Further reading: Nüst et al. (2021), Stodden et al. (2016), Sandve et al. (2013)

52north

Principles and best practices

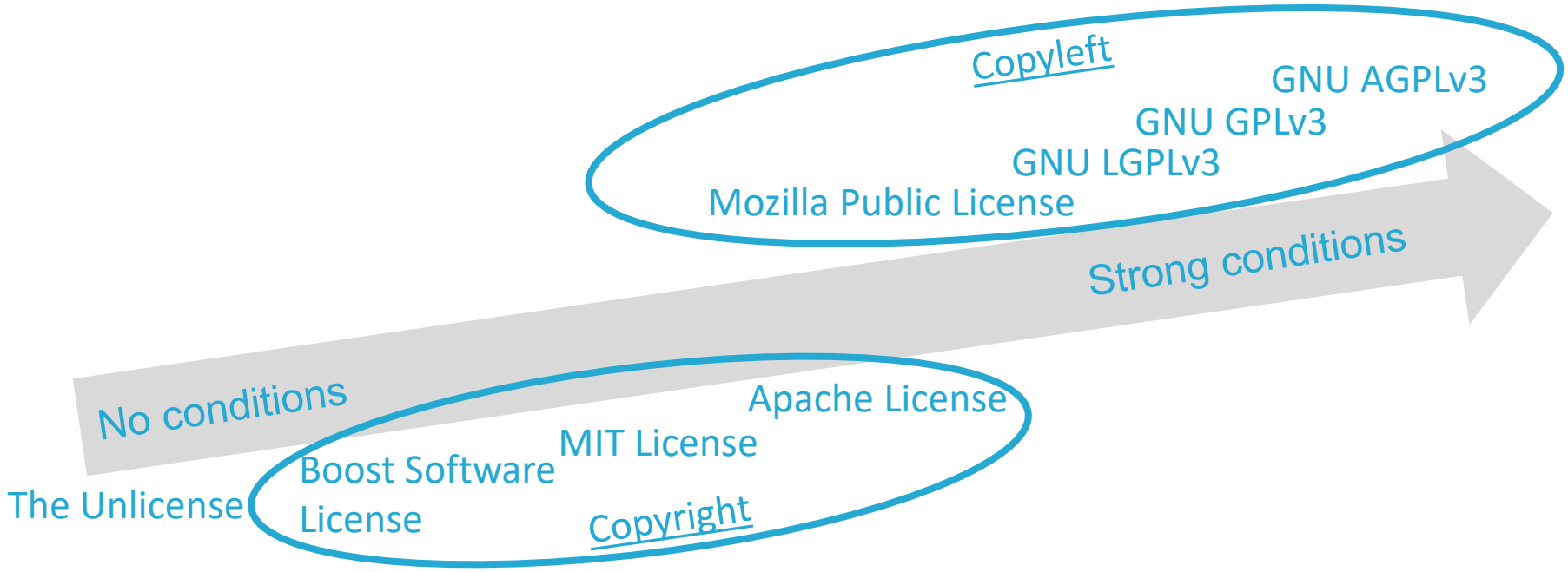
To enable credit for shared materials, citation should be standard practice.

- Example: *Statistics were done using R 3.5.0 (R Core Team, 2018), the rstanarm (v2.13.1; Gabry & Goodrich, 2016) and the psycho (v0.3.4; Makowski, 2018) packages. The full reproducible code is available in Supplementary Materials.*

Further reading: Makowski (2018)

Principles and best practices

Use open licensing when publishing source code.



Further reading: Stodden et al. (2016), Choose an open source license, Open Source Initiative

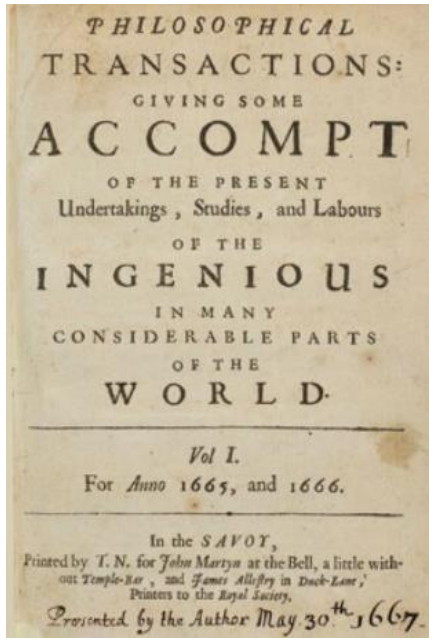
Practical

Principles and best practices

Lessons learned:

- The Executable Research Compendium is a package including the paper, source code, data, and the computational environment.
- The scientific article should contain a persistent identifier that links to these materials.
- It is important to cite reused software, for example, packages.
- Source code should be released under an open license.

Opportunities



PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.
Presented by the Author May 30th 1667

Article

AGN as potential factories for eccentric black hole mergers

https://doi.org/10.1038/s41586-021-04333-1
Received: 8 October 2020
Accepted: 10 December 2021
Published online: 9 March 2022
Check for updates

J. Samsing[1✉], I. Bartos[2], D. J. D'Orazio[3], Z. Haiman[4], B. Kocsis[4,5], N. W. C. Leigh[6,7], B. Liu[1], M. E. Pessah[1] & H. Tagawa[8]

There is some weak evidence that the black hole merger named GW190521 had a non-zero eccentricity[1,2]. In addition, the masses of the component black holes exceeded the limit predicted by stellar evolution[3]. The large masses can be explained by successive mergers[4,5], which may be efficient in gas disks surrounding active galactic nuclei, but it is difficult to maintain an eccentric orbit all the way to the merger, as basic physics would argue for circularization[6]. Here we show that active galactic nuclei disk environments can lead to an excess of eccentric mergers, if the interactions between single and binary black holes are frequent[5] and occur with mutual inclinations of less than a few degrees. We further illustrate that this eccentric population has a different distribution of the inclination between the spin vectors of the black holes and their orbital angular momentum at merger[7], referred to as the spin–orbit tilt, compared with the remaining circular mergers.

black holes that eventually merge in active galactic nuclei (AGN) disks can be brought into the disk through gas capture from the surrounding nuclear star cluster[8] or can be produced through in situ star formation[9–11]. Once a black hole is in the disk, it will undergo radial migration[9] and can, as a result, pair up with another black hole to form a binary[12,13]. Recent studies show that interactions between such migrating binary black holes and other single black holes in the AGN disk, referred to as binary–single interactions, likely provide the main pathway for bringing binaries to merger[5,14–17] (Fig. 1). Despite progress on characterizing such interactions[18], the inclusion of gravitational-wave emission during the interactions, which has been shown to be essential in resolving mergers with non-zero eccentricity that form in stellar clusters[19,20], remains unexplored. Observationally, GW190521 is among the first gravitational-wave sources with indications of an AGN disk origin[5,20]. It is sensible to inquire whether its apparent non-zero eccentricity[1], as well as its observed approximately 90° spin–orbit tilt[7], could arise naturally as a distinct signature, characteristic of dynamically induced AGN disk mergers.

With this motivation, we explore how binary black holes merge through binary–single interactions in AGN disk environments when gravitational-wave emission is included in the dynamics by means of the 2.5-post-Newtonian (2.5-PN) term[21] (see Methods). To approach this complex astrophysical situation in a systematic way, we focus here on quantifying the unique signatures that might be associated with the roughly 2D disk-like environment of the AGN disk compared with the usual 3D interactions found in stellar clusters[18]. For this, we perform controlled experiments of initially circular black hole binaries interacting with singles incoming on an orbital plane that is inclined relative to the binary orbital plane by an angle ψ, for which $\psi = 0$ corresponds to a coplanar interaction (Fig. 4). For a given scattering, we study the interacting black holes merge while they are all bound and interacting[20] (Fig. 1); 2-body merger: the binary black hole survives the three-body interaction, but merges before undergoing its next interaction[22].

Figure 2 shows the probability for merger as a function of black hole binary semimajor axis, a. As shown in this figure, restricting the interactions to be coplanar leads to a notable enhancement of mergers; for example, for a binary with a semimajor axis $a = 1$ AU, the fraction of 3-body mergers is about 100 times larger in the 2D disk case compared with the 3D cluster case. As outlined in the Methods, this enhancement is due to the difference in eccentricity distributions of the dynamically assembled binaries, $P(e)$, that follows $\propto e/\sqrt{1-e^2}$ in the 2D case, compared with $\sim 2e$ in the 3D case[22–24]. Our analytic approximations (see Methods) for both the 2-body and the 3-body merger probabilities, p_2 and p_3, respectively, are also included in Fig. 2. Assuming the equal-mass limit, the corresponding ratio of probabilities between the 2D and the 3D cases is

$$\frac{p_2^{(2D)}}{p_2^{(3D)}} \approx 10^3 \times \left[\frac{m}{20\,M_\odot}\right]^{-6/14} \left[\frac{a}{1\,\mathrm{AU}}\right]^{21/47} \left[\frac{t_{\mathrm{BH}}}{10^3\,\mathrm{years}}\right]^{1/7}, \quad (1)$$

$$\frac{p_3^{(2D)}}{p_3^{(3D)}} \approx 10^2 \times \left[\frac{m}{20\,M_\odot}\right]^{-5/14} \left[\frac{a}{1\,\mathrm{AU}}\right]^{-3/14}. \quad (2)$$

in which m is the black hole mass and t_{BH} is the time between interactions scaled to a value characteristic for AGN disk models[5]. These results show that the effects from breaking the scattering isotropy become increasingly important at larger a and smaller m. Note further the weak dependence on t_{BH}.

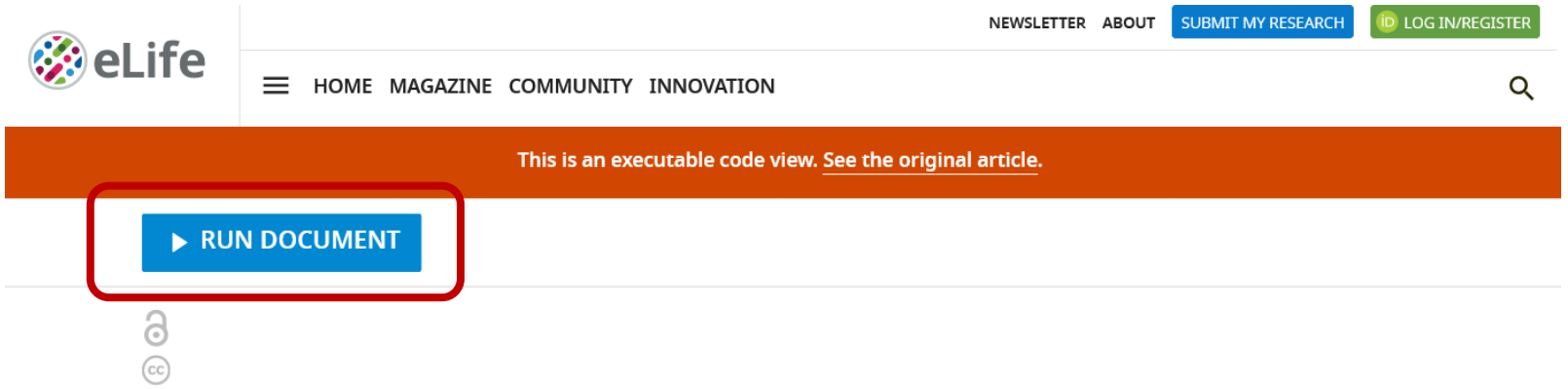
2022

1665

Further reading: Philosophical Transactions

52north

Opportunities



Replication Study: Transcriptional amplificatic cells with elevated c-Myc

L Michelle Lewis, Meredith C Edwards, Zachary R Meyers, C Conover Talbot, Haiping Hao, David Blum, Reproducibility I
Elizabeth Iorns, Rachel Tsui, Alexandria Denis, Nicole Perfito, Timothy M Errington
University of Georgia, Bioexpression and Fermentation Facility, Georgia, United States; Johns Hopkins University, Deep Sec

Further reading: Lewis et al. (2018)

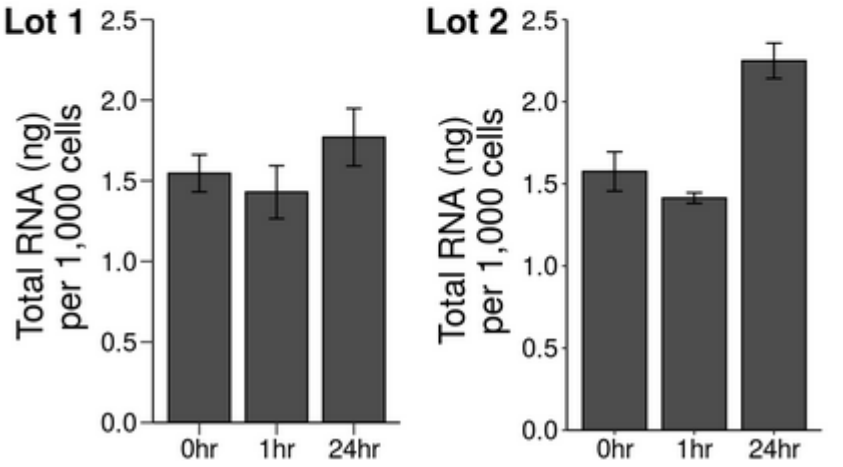
Opportunities

eLife

Induction of c-Myc in P493-6 cells and impact on total RNA levels.

P493-6 cells were grown in the presence of tetracycline (Tet) for 72 hr and switched into Tet-free growth medium to induce c-Myc expression. Cells were cultured in two separate lots of serum. **(A)** Representative Western blot using an anti-c-Myc antibody (top panels) or an anti-ß-Actin antibody (bottom panel). Two exposures of the anti-c-Myc antibody are presented to facilitate detection of c-Myc.

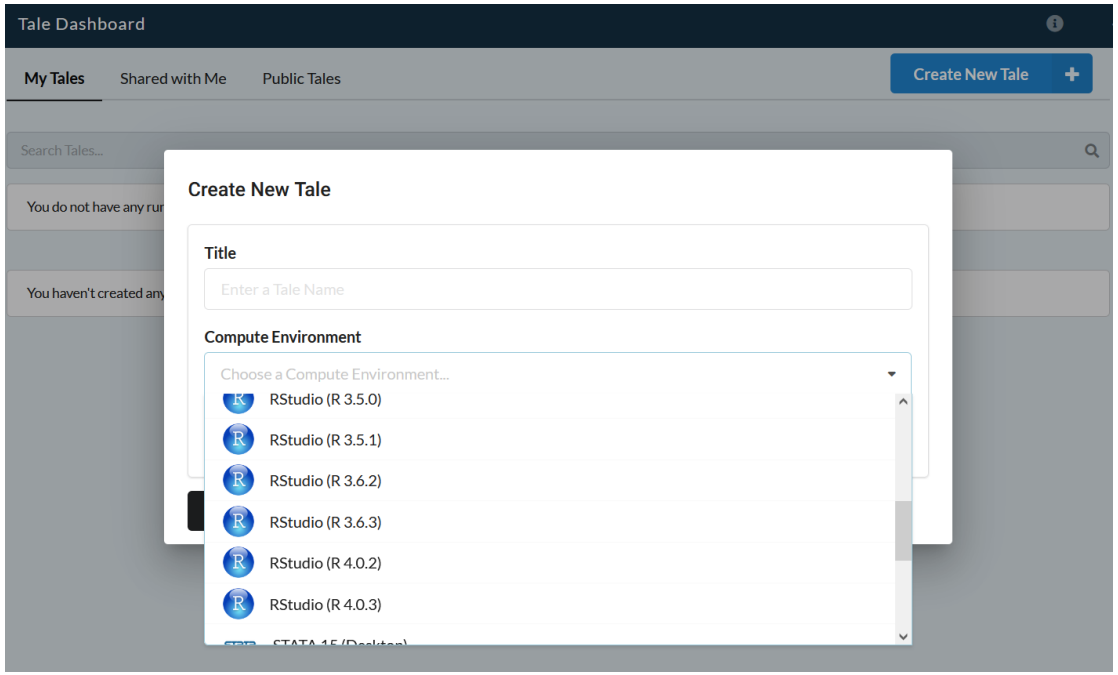
```
68   scale_x_discrete(labels = c("0hr", "1hr", "24hr")) +
69   theme(plot.margin = unit(c(1,1,1,2), "lines"),
70         axis.text.x = element_text(size=15, color = "black"),
71         axis.text.y = element_text(size = 15, color = "black"),
72         axis.title.y = element_text(size = 20),
73         axis.title.x = element_blank(),
74         panel.background = element_blank(),
75         axis.line.y = element_line(),
76         legend.position = "none",
77         axis.line.x = element_line())
78
79 Figure_1B <- plot_grid(plot.lot1, plot.lot2, labels = c("Lot 1", "Lot 2"), label_size = 20, hjust
80 Figure_1B
```

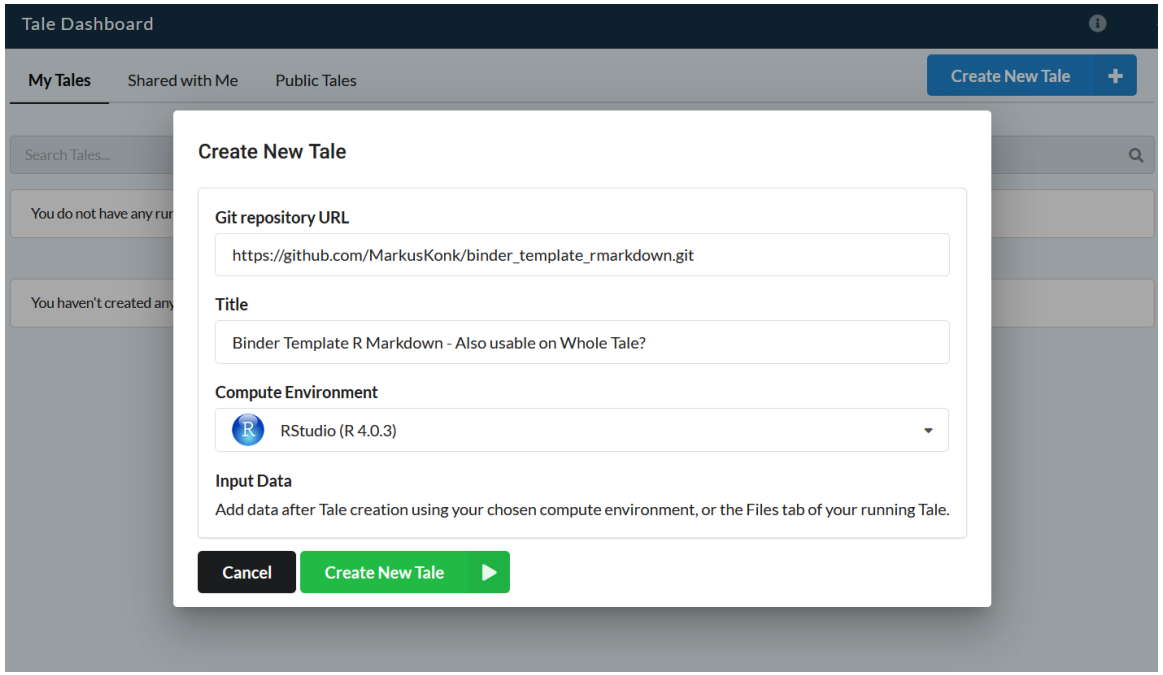


Further reading: Lewis et al. (2018)

52north

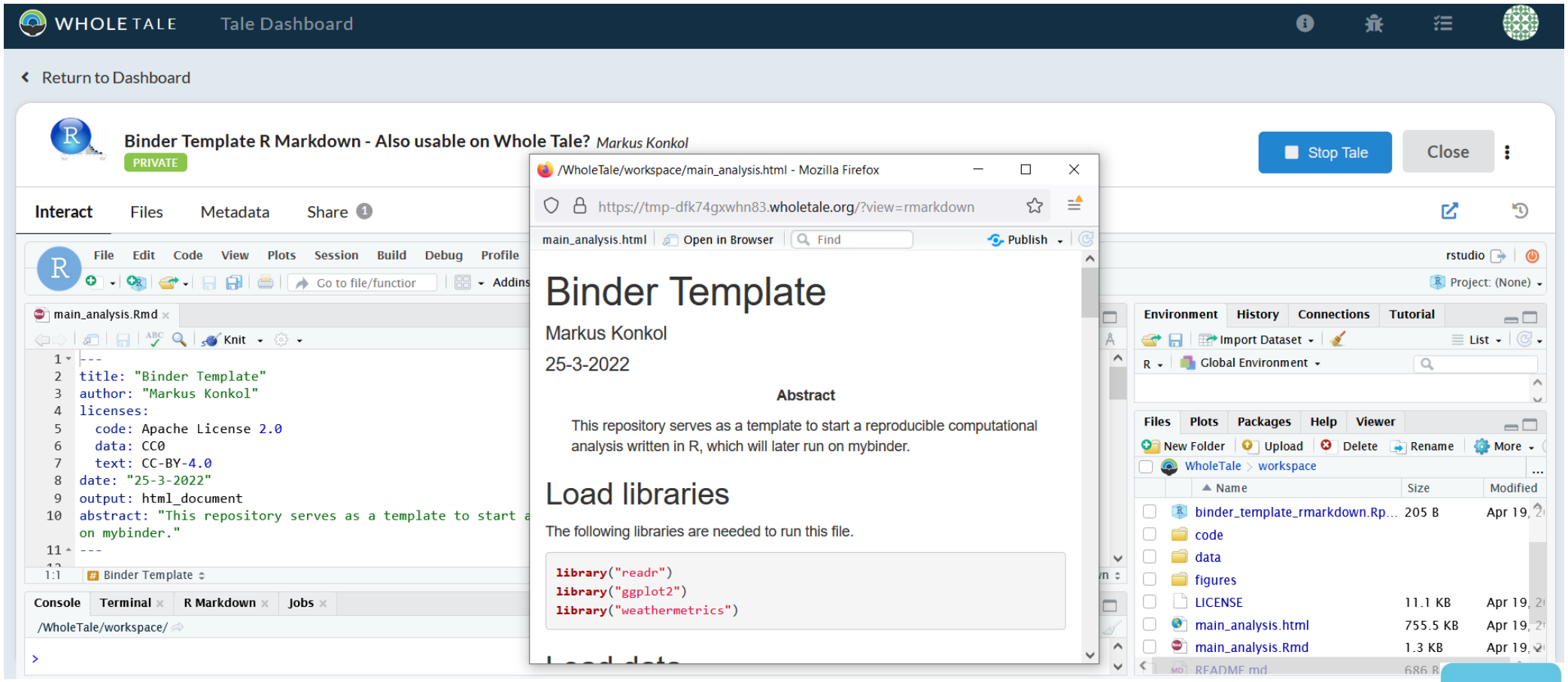
Opportunities





Further reading: Brinckman et al. (2019)

Opportunities



Further reading: Brinckman et al. (2019)

Opportunities

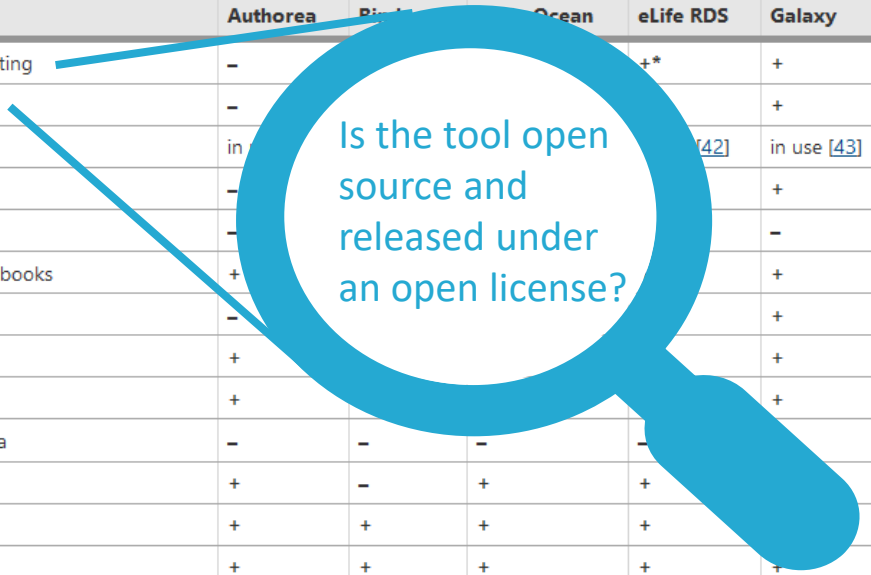
Further reading: Konkol et al. (2020)

Opportunities

Table 2 Overview of which application supports the corresponding criteria. (N/D = no data)

From: Publishing computational research - a review of infrastructures for reproducible and transparent scholarly communication

	Authorea		Ocean	eLife RDS	Galaxy	Gigantum	Manuscripts	o2r	REANA	Repro Zip	Whole Tale
Free self-hosting	–			+*	+	–	+	+	+*	+	+
Open license	–				+	+/–	+	+	+	+	+
In use	in			[42]	in use [43]	–	–	–	in use [44]	in use [31]	–
Grant-based	–				+	–	N/D	+	+	+	+
R Markdown	–				–	+	–	+	–	–	+
Jupyter Notebooks	+				+	+	–	–	+	+	+
Extensible	–				+	–	–	–	+	+	+
Upload	+				+	–	+	+	–	–	+
Copyright	+				+	+	N/D	+	N/D	N/D	+
Sensitive data	–	–	–	–		–	–	–	–	–	–
Discovery	+	–	+	+		–	–	+	–	–	+
Inspection	+	+	+	+		+	+	+	–	–	+
Execution	+	+	+	+	+	+	+	+	+	+	+
Manipulation	+	+	+	+	+	+	+	+	+	+	+
Substitution	–	–	–	–	–	–	–	+	–	+	–
Download	+	+	+	+	+	+	+	+	–	+	+
Modify/Delete after publishing	–	+	–	–	+	+	+	–	+	+	–
Shared via DOI	+	–	+	+	–	–	–	–	–	–	+
Shared via URL	+	+	+	+	+	+	+	+	–	+	–



Further reading: Konkol et al. (2020)

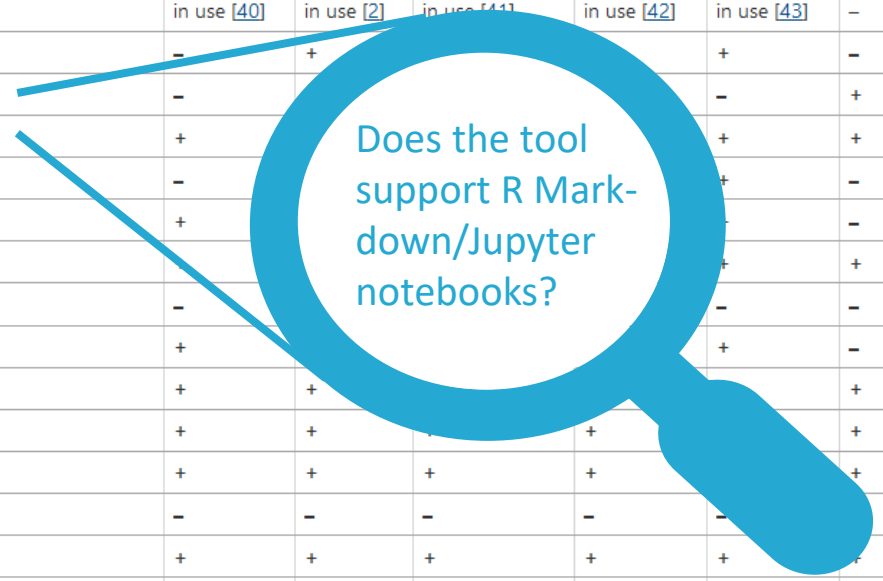
52north

Opportunities

Table 2 Overview of which application supports the corresponding criteria. (N/D = no data)

From: Publishing computational research - a review of infrastructures for reproducible and transparent scholarly communication

	Authorea	Binder	Code Ocean	eLife RDS	Galaxy	Gigantum	Manuscripts	o2r	REANA	Repro Zip	Whole Tale
Free self-hosting	–	+	–	+*	+	–	+	+	+*	+	+
Open license	–	+	–	+	+	+/–	+	+	+	+	+
In use	in use [40]	in use [2]	in use [41]	in use [42]	in use [43]	–	–	–	in use [44]	in use [31]	–
Grant-based	–	+			+	–	N/D	+	+	+	+
R Markdown	–				–	+	–	+	–	–	+
Jupyter Notebooks	+				+	+	–	–	+	+	+
Extensible	–				+	–	–	–	+	+	+
Upload	+					–	+	+	–	–	+
Copyright					+	+	N/D	+	N/D	N/D	+
Sensitive data	–				–	–	–	–	–	–	–
Discovery	+				+	–	–	+	–	–	+
Inspection	+	+				+	+	+	–	–	+
Execution	+	+		+		+	+	+	+	+	+
Manipulation	+	+	+	+		+	+	+	+	+	+
Substitution	–	–	–	–	–		–	+	–	+	–
Download	+	+	+	+	+		+	+	–	+	+
Modify/Delete after publishing	–	+	–	–	+	+	+	–	+	+	–
Shared via DOI	+	–	+	+	–	–	–	–	–	–	+
Shared via URL	+	+	+	+	+	+	+	+	–	+	–



Further reading: Konkol et al. (2020)

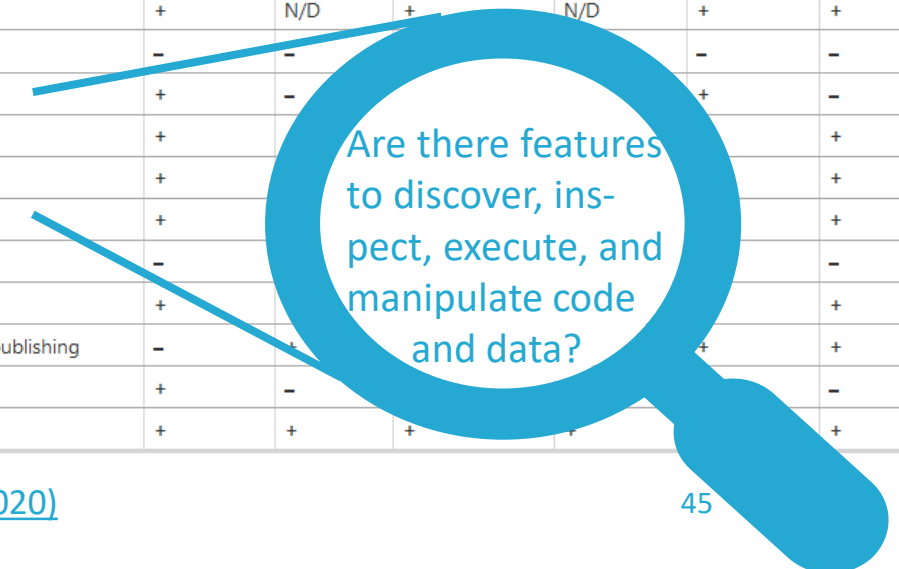
52north

Opportunities

Table 2 Overview of which application supports the corresponding criteria. (N/D = no data)

From: Publishing computational research - a review of infrastructures for reproducible and transparent scholarly communication

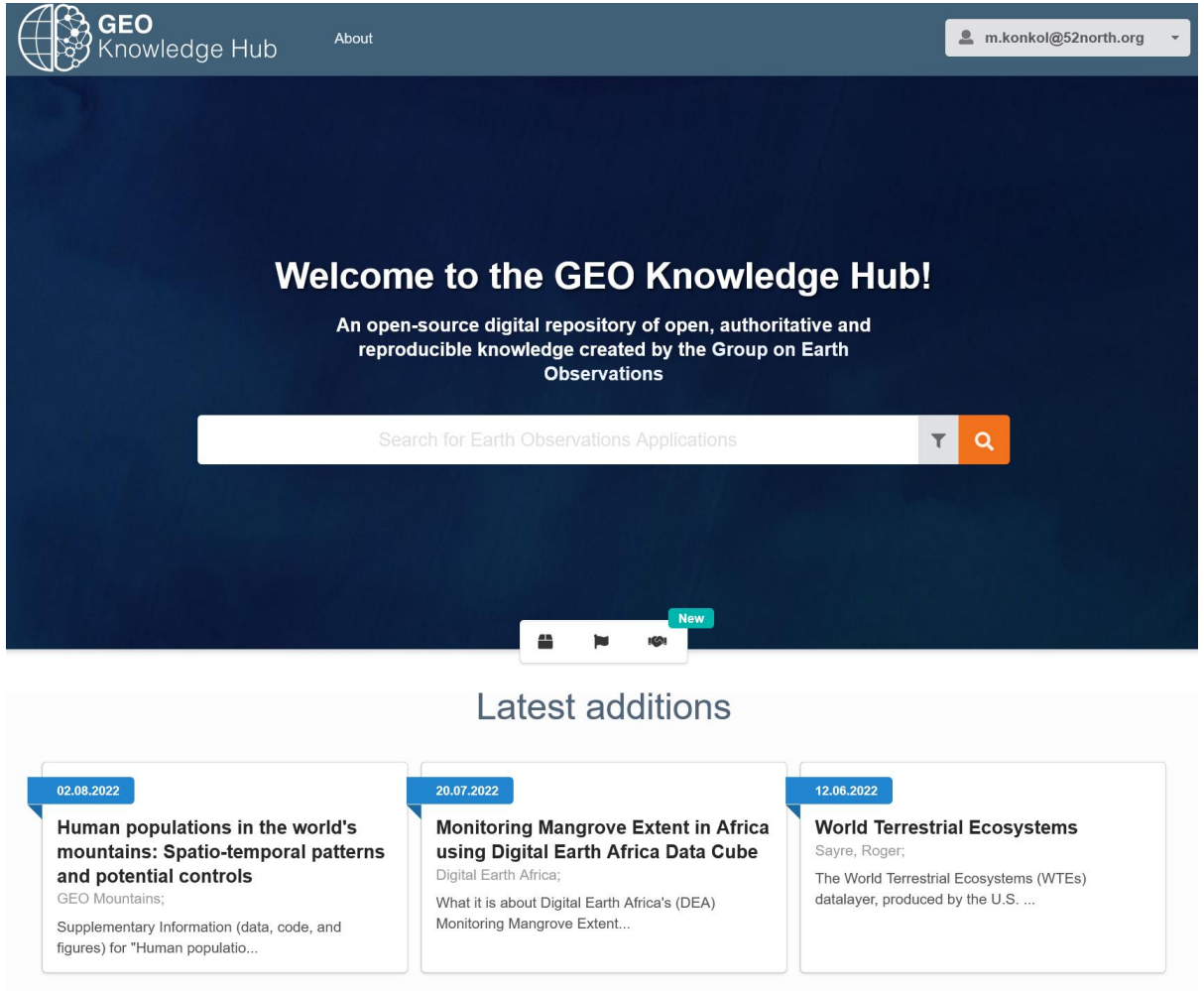
	Authorea	Binder	Code Ocean	eLife RDS	Galaxy	Gigantum	Manuscripts	o2r	REANA	Repro Zip	Whole Tale
Free self-hosting	–	+	–	+*	+	–	+	+	+*	+	+
Open license	–	+	–	+	+	+/–	+	+	+	+	+
In use	in use [40]	in use [2]	in use [41]	in use [42]	in use [43]	–	–	–	in use [44]	in use [31]	–
Grant-based	–	+	–	+	+	–	N/D	+	+	+	+
R Markdown	–	+	+	+	–	+	–	+	–	–	+
Jupyter Notebooks	+	+	+	+	+	+	–	–	+	+	+
Extensible	–	+	+	+	+	–	–	–	+	+	+
Upload	+	+	+	–	+	–	+	+	–	–	+
Copyright	+	N/D	+	N/D	+	+	N/D	+	N/D	N/D	+
Sensitive data	–	–			–	–	–	–	–	–	–
Discovery	+	–			+	–	–	+	–	–	+
Inspection	+					+	+	+	–	–	+
Execution	+					+	+	+	+	+	+
Manipulation	+					+	+	+	+	+	+
Substitution	–					–	–	+	–	+	–
Download	+					+	+	+	–	+	+
Modify/Delete after publishing	–				+	+	+	–	+	+	–
Shared via DOI	+	–				–	–	–	–	–	+
Shared via URL	+	+	+	+		+	+	+	–	+	–



Further reading: Konkol et al. (2020)

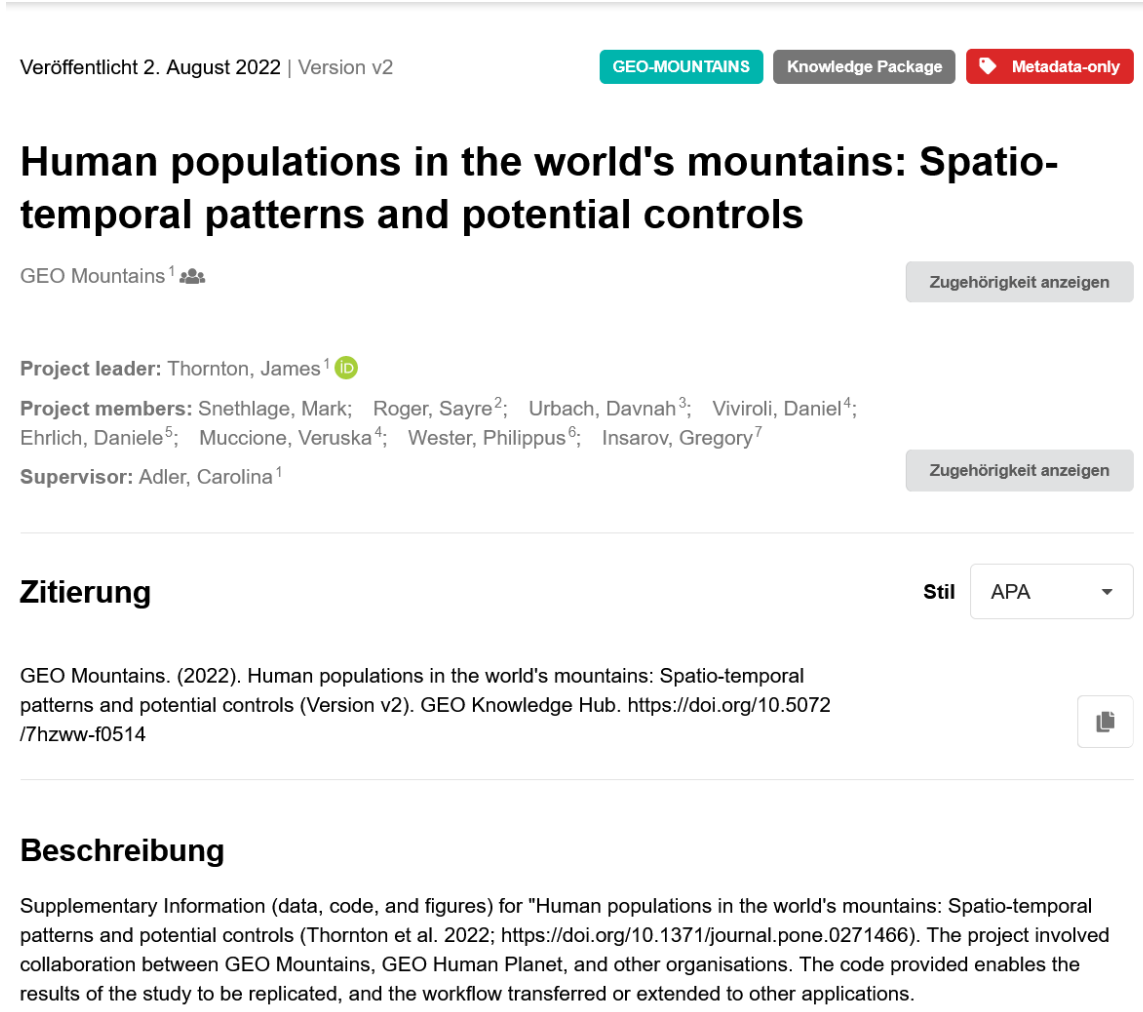
52north

GEO Knowledge Hub



Further reading: GEO Knowledge Hub,

GEO Knowledge Hub



Veröffentlicht 2. August 2022 | Version v2

GEO-MOUNTAINS Knowledge Package Metadata-only

Human populations in the world's mountains: Spatio-temporal patterns and potential controls

GEO Mountains[1]

Zugehörigkeit anzeigen

Project leader: Thornton, James[1]

Project members: Snethlage, Mark; Roger, Sayre[2]; Urbach, Davnah[3]; Viviroli, Daniel[4]; Ehrlich, Daniele[5]; Muccione, Veruska[4]; Wester, Philippus[6]; Insarov, Gregory[7]

Supervisor: Adler, Carolina[1]

Zugehörigkeit anzeigen

Zitierung

Stil APA

GEO Mountains. (2022). Human populations in the world's mountains: Spatio-temporal patterns and potential controls (Version v2). GEO Knowledge Hub. https://doi.org/10.5072/7hzww-f0514

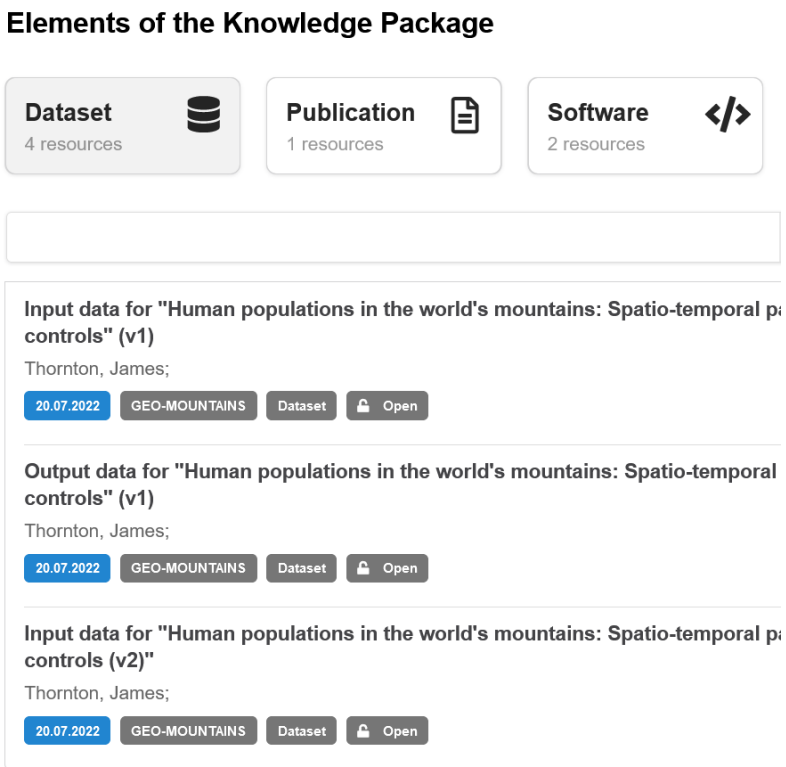
Beschreibung

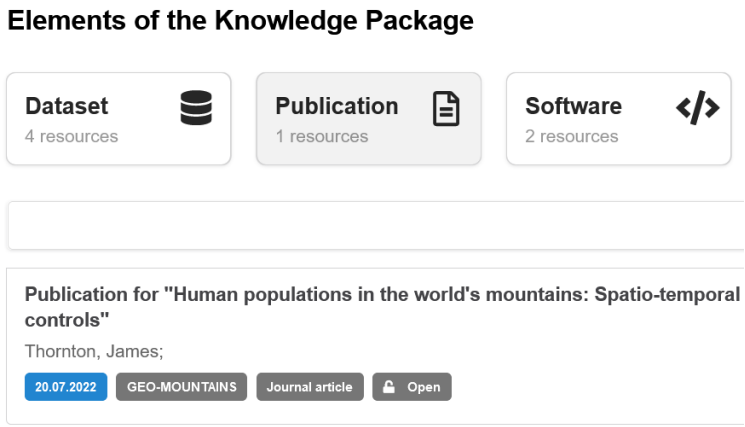
Supplementary Information (data, code, and figures) for "Human populations in the world's mountains: Spatio-temporal patterns and potential controls (Thornton et al. 2022; https://doi.org/10.1371/journal.pone.0271466). The project involved collaboration between GEO Mountains, GEO Human Planet, and other organisations. The code provided enables the results of the study to be replicated, and the workflow transferred or extended to other applications.

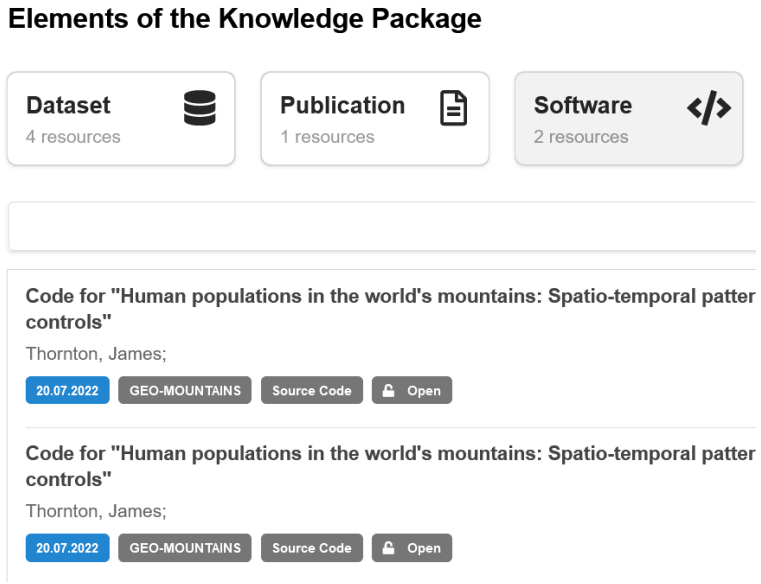
Further reading: GEO Knowledge Hub,

52north

GEO Knowledge Hub







Further reading: [GEO Knowledge Hub](),

Things to discuss and consider

Reproducing results can also mean reproducing errors.

Others might become discouraged to collect data for replication.

Is ORR the ultimate goal or just the basis?

Reproducible research is not necessarily of high quality.

Wrap up

- **Reproducible Research** refers to achieving the **same results** (e.g., tables, figures, numbers) as reported in the paper by using the **same publicly available source code and data**.
- Despite a number of ("selfish") reasons to do reproducible research, it is not common practice (due to missing time & skills).
- Accessible code and data is not necessarily reproducible (technical issues, design-related differences, deviating results).
- Reproducibility principles (e.g., Executable Research Compendium) and tools (Computational notebooks, Binder) can help to avoid issues.

19:00–20:00 Dinner

20:00–21:30 **Reproducibility Hackathon**

52north

THANK YOU!

"Openness is not all-or-nothing [...] Fully open research is a long-term goal, not a switch we should expect to flip overnight."
MCKIERNAN et al. (2016)

References

Last access of all URLs: 11th April 2021.

Alston, J. & Rick, J. (2020). A Beginner's Guide to Conducting Reproducible Research. http://dx.doi.org/10.32942/osf.io/h5r6n

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

Binder (2017). User guide – Get started. https://mybinder.readthedocs.io/en/latest/index.html#get-started

Brinckman, A. et al. (2019). Computing environments for reproducibility: Capturing the "Whole Tale." *Future Generation Computer Systems*, 94, 854–867. http://dx.doi.org/10.1016/j.future.2017.12.029.

Buckheit, J.B. & Donoho, D.L. (1995). WaveLab and Reproducible Research. *Lecture Notes in Statistics*, 55–81. http://dx.doi.org/10.1007/978-1-4612-2544-7_5

Chawla, D. S. (2020) Critiqued coronavirus simulation gets thumbs up from code-checking efforts *Nature* 582, 323-324. https://doi.org/10.1038/d41586-020-01685-y

Culina, A., van den Berg, I., Evans, S., Sánchez-Tójar, A. (2020) Low availability of code in ecology: A call for urgent action. PLOS Biology 18(7): e3000763. https://doi.org/10.1371/journal.pbio.3000763

Davies, N. G., Kucharski, A. J., Eggo, R. M., Gimma, A., Edmunds, W. J., Jombart, T., ... & Liu, Y. (2020). Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health*, *5*(7), e375-e385. https://doi.org/10.1016/S2468-2667(20)30133-X

References

Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., … & Yu, X. (2016). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science*, 3(10), 388–415. https://doi.org/10.1002/2015ea000136

GitHub. Choose an open source license. https://choosealicense.com/

Eglen, S. J. (2020). CODECHECK certificate 2020-008. *Zenodo*. https://doi.org/10.5281/zenodo.3746024

Goodman, S. N., Fanelli, D. & Ioannidis, J.P.A. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8(341). http://dx.doi.org/10.1126/scitranslmed.aaf5027

Hasselbring, W., Carr, L., Hettrick, S., Packer, H. & Tiropanis, T. (2020). Open Source Research Software. *Computer*, 53, 8, 84-88. https://doi.org/10.1109/MC.2020.2998235

Hettne, K., Proppert, R., Nab, L., Rojas-Saunero, L. P., & Gawehns, D. (2020). ReprohackNL 2019: how libraries can promote research reproducibility through community engagement. *IASSIST Quarterly*, *44*(1-2), 1–10. https://doi.org/10.29173/iq977

Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science*, 359(6377), 725–726. http://dx.doi.org/10.1126/science.359.6377.725

Katz, D. S., Gruenpeter, M. & Honeyman, T. (2021). Taking a Fresh Look at FAIR for Research Software. *Patterns* 2, 3. https://doi.org/10.1016/j.patter.2021.100222

Konkol, M., Kray, C., & Pfeiffer, M. (2018). Computational reproducibility in geoscientific papers: Insights from a series of studies geoscientists and a reproduction study. *International Journal of Geographical Information Science*, 33:2, 408-429. https://doi.org/10.1080/13658816.2018.1508687

52north

References

Konkol, M., Nüst, D. & Goulier, L. (2020). Publishing computational research - a review of infrastructures for reproducible and transparent scholarly communication. *Res Integr Peer Rev* 5**,** 10. https://doi.org/10.1186/s41073-020-00095-y

Krishnamurthi, S. & Vitek, J. (2015). The real software crisis: repeatability as a core value. *Commun. ACM* 58, 3, 34–36. https://doi.org/10.1145/2658987

Lasser, J. (2020). Creating an executable paper is a journey through Open Science. *Commun Phys* **3,** 143. https://doi.org/10.1038/s42005-020-00403-4

Lewis, L.M. et al., 2018. Replication Study: Transcriptional amplification in tumor cells with elevated c-Myc. eLife, 7. Available at: http://dx.doi.org/10.7554/elife.30274

Makowski, D. (2018). How to Cite Packages. *R-bloggers.* https://www.r-bloggers.com/2018/08/how-to-cite-packages/

Markowetz, F. (2015). Five selfish reasons to work reproducibly. *Genome Biol* **16,** 274. https://doi.org/10.1186/s13059-015-0850-7

Marwick, B. (2015). How computers broke science – and what we can do to fix it. *The Conversation*. https://theconversation.com/how-computers-broke-science-and-what-we-can-do-to-fix-it-49938

McKiernan, E.C. et al. (2016). Author response: How open science helps researchers succeed. http://dx.doi.org/10.7554/elife.16800.008

Nature Human Behavior. Supporting computational reproducibility through code review. *Nat Hum Behav* 5**,** 965–966 (2021). https://doi.org/10.1038/s41562-021-01190-w

Nüst, D., Ostermann, F. O., Sileryte, R., Hofer, B., Granell, C., Teperek, M., … Wang, Y. (2021). AGILE Reproducible Paper Guidelin
https://doi.org/10.17605/OSF.IO/CB7Z8

52north

References

Nüst et al. (2017). Opening the Publication Process with Executable Research Compendia. *D-Lib Magazine*. http://www.dlib.org/dlib/january17/nuest/01nuest.html

Nüst, D. & Eglen S. J. CODECHECK: an Open Science initiative for the independent execution of computations underlying research articles during peer review to improve reproducibility. *F1000Research* 2021, 10:253. https://doi.org/10.12688/f1000research.51738.2

Open Source Initiative. https://opensource.org/

Peng, R.D. (2011). Reproducible Research in Computational Science. *Science*, 334(6060), 1226–1227. http://dx.doi.org/10.1126/science.1213847

Quintana, D. S. (2020, December 5). Five things about open and reproducible science that every early career researcher should know. https://doi.org/10.17605/OSF.IO/DZTVQ

RStudio (2020). R Markdown Introduction. URL: https://rmarkdown.rstudio.com/lesson-1.html

Sandve, G. K., Nekrutenko, A., Taylor, J., Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. *PLoS Comput Biol* 9(10): e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Stark, P. B. (2018). Before reproducibility must come preproducibility. *Nature*, 557(7707), 613–613. https://doi.org/10.1038/d41586-018-05256-0

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., … & Taufer, M. (2016). Enhancing reproducibility for computational methods. Science, 354(6317), 1240–1241. https://doi.org/10.1126/science.aah6168

The Carpentries (2022). Version Control with Git. https://swcarpentry.github.io/git-novice/index.html

52north

References

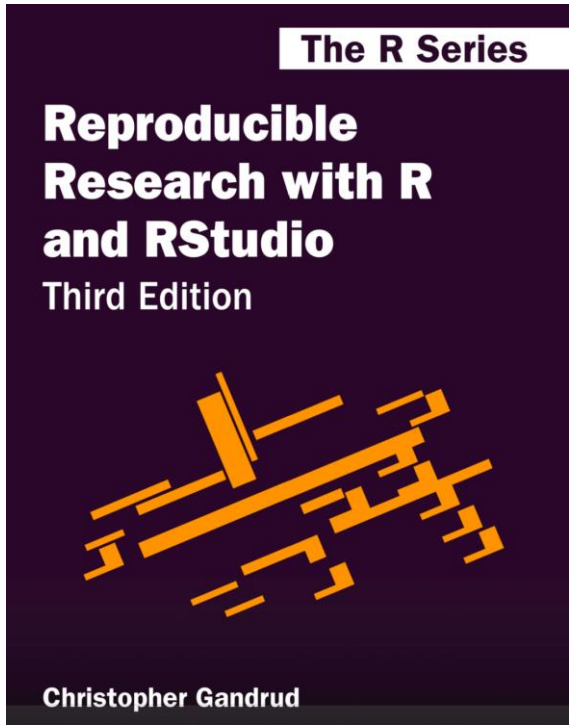
The Jupyter Book Community. UC Berkeley Data Science Modules. https://ds-modules.github.io/modules-textbook/introduction/ds-intro.html

The Royal Society. Philosophical Transactions: 350 years of publishing at the Royal Society (1665 – 2015). https://royalsociety.org/-/media/publishing350/publishing350-exhibition-catalogue.pdf

The Turing Way Community, Arnold, B., Bowler, L., Gibson, S., Herterich, P., Higman, R. … & Whitaker, K. (2019). The Turing Way: A Handbook for Reproducible Data Science. *Zenodo*. http://doi.org/10.5281/zenodo.3233986

52north

References


The R Series
Reproducible Research with R and RStudio
Third Edition
Christopher Gandrud


Frank Miedema
Open Science: the Very Idea
OPEN ACCESS
Springer