WWU MÜNSTER

# Infrastructures and Practices for Reproducible Research in Geography, Geosciences, and GIScience

PhD Thesis Defense | 2022-02-14 | **Daniel Nüst**, Dipl-Geoinf.

This document: **http://go.wwu.de/wklef**
Full reference: Nüst, Daniel. 2021. Infrastructures and Practices for Reproducible Research in Geography, Geosciences, and GIScience. Doctoral dissertation, University of Münster, Germany. https://doi.org/10.5281/zenodo.4768096

ifgi
Institute for Geoinformatics
University of Münster

living.knowledge

Welcome to my PhD thesis defense. Let’s start with a quick introduction to the topic. You can find a typable link to the presentation at the bottom of each slide.

# Scope & motivation



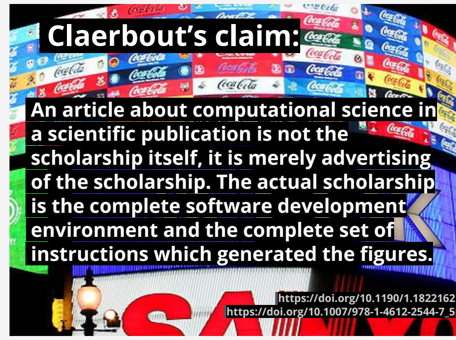
First off, what is the scope of the work, and why did I pursue it?

## A problematic situation

https://giphy.com/gifs/usnationalarchives-nasa-scientist-scientists-1F1JGyGZhiSAA8Vuhn


https://giphy.com/gifs/with-computers-fascination-PxSFAnuubLkSA

https://giphy.com/gifs/david-hasselhoff-M3o3fL9nnxG4o




CC-BY 3.0, Sebastian Bertalan, Wikimedia Commons

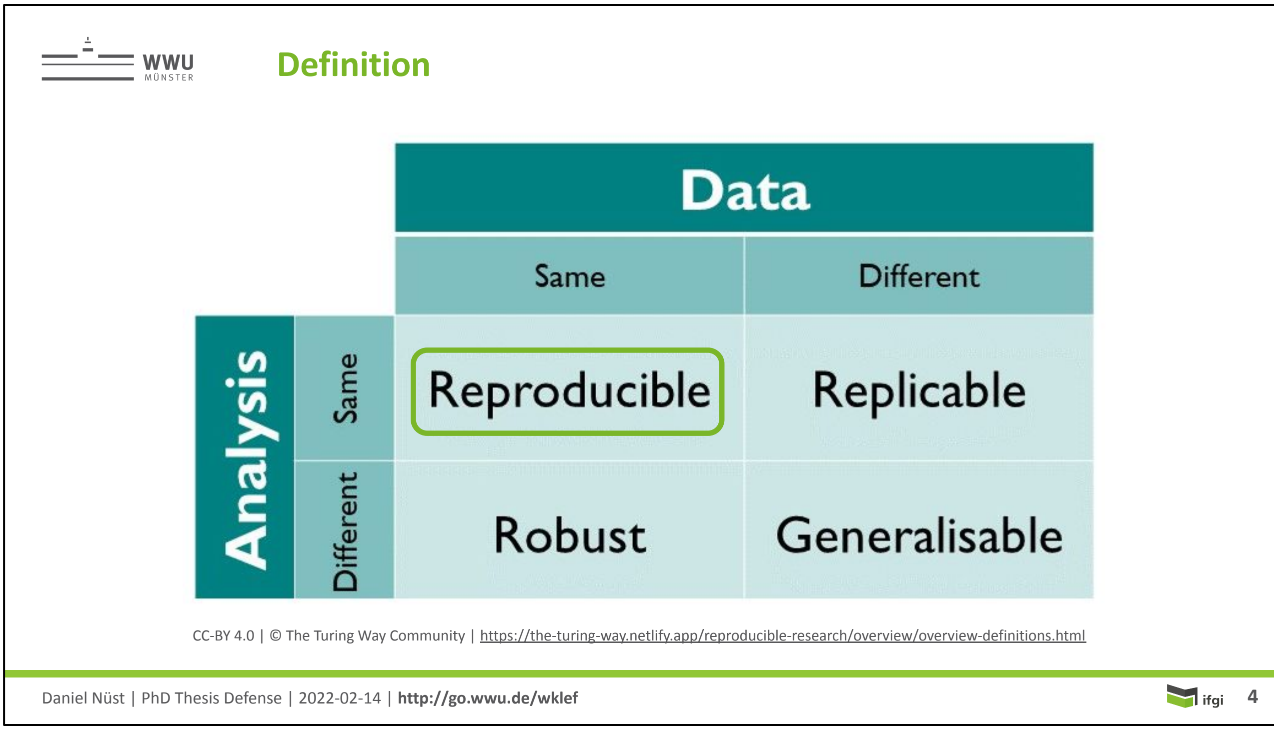


The root of all problems is clear:

🖱️ **Researchers use computers** because computers can help us answer questions, which would otherwise be impossible to investigate.
However, we do not have an equivalent to "lab procedures" with our computers. There are no established practices for note taking and wearing safety-gear, we simply turn on the computer and expect it to work.

🖱️ Because we are a bit reckless, the collaboration with our digital tools is **rarely frictionless**. It feels like software will always continue to work, when in reality it is a house of cards.

And when it comes to publishing our results, 🖱️ for too long we simply said: **look, this is what we did**. We basically showed fellow scientists our screens, so they can look AT our work, but not look INTO it.
The problem is that the means of communication stayed the same although research became complicated and complex when we started use computers. To share results we use a PDF, although 🖱️ **Claerbout's claim** on computational research already described the reality much better*: An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*




My work focuses on computational reproducibility, not on reproducing lab experiments or field studies.

When working with computers, 🖱️ **"reproducible research" means** that an independent party can execute the original code with the same dataset and gets the same results. In case of reproducibility, this should be the same actual numbers, within limitations for example in case of floating point arithmetic.

Reproducible research is distinct from replicable, robust, and generalisable research. These terms describe different combinations of same or different data or code. Naturally, these are not about acquiring the actual same results in terms of numerical values, but about a scientific interpretation of the results that is, for example, robust towards previous findings.

In my view, reproducibility is the most basic of the four and a prerequisite to effectively reach the other goals.

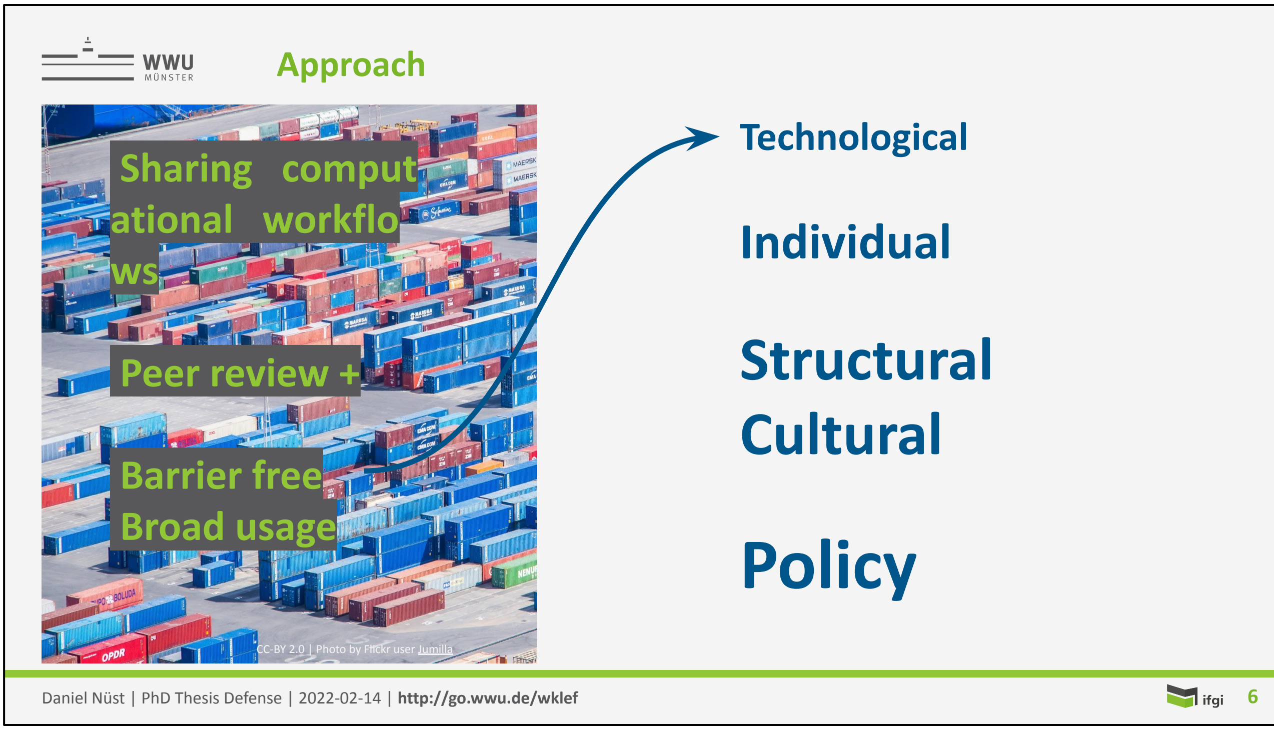

Without being complete, let’s remind ourselves of some of the biggest problems with **closed and irreproducible research**:

It cannot be fully understood nor **verified**.

It obstructs effective **reuse and extension**, therefore it is unsustainable.

It prevents usage in **education** and citizen science.

And it slows **innovation** through closeness and repetition.

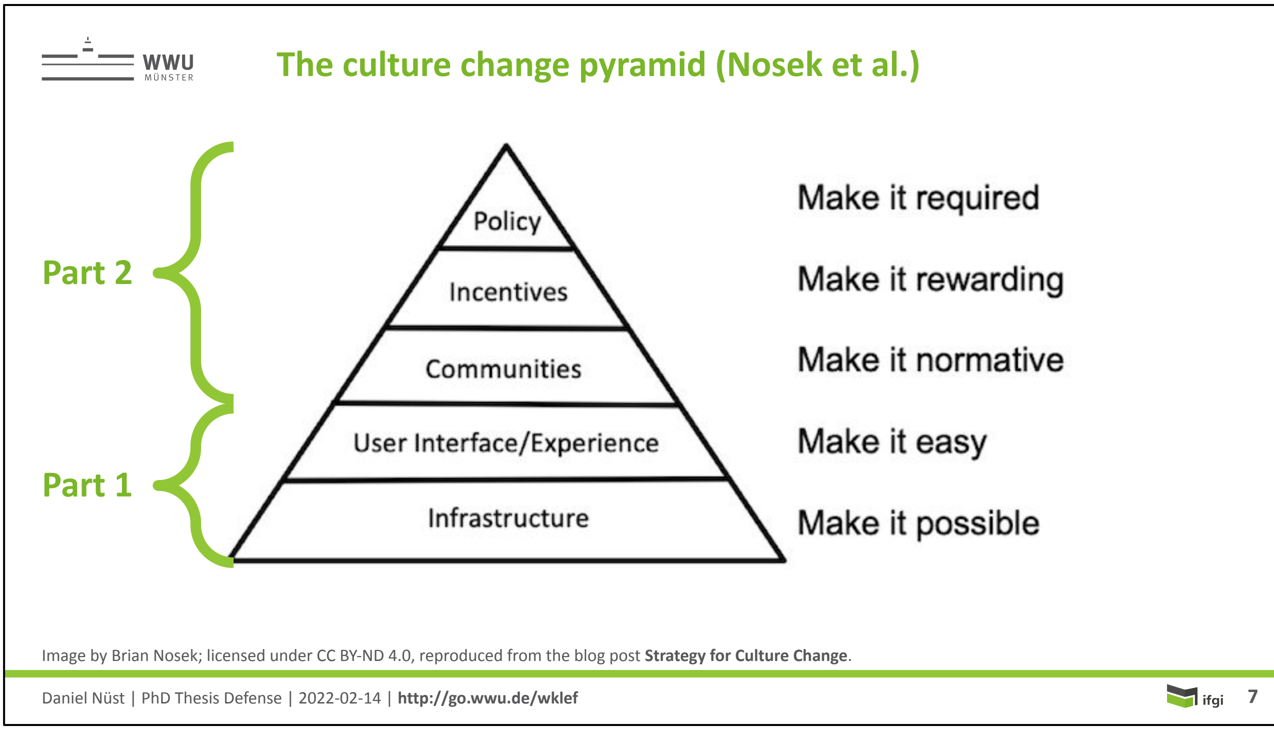



The work's core idea is to a method from mainstream IT, namely **containerisation**, to capture and control computing environments. The goal is to make the **sharing, evaluating, and extending of computational workflows** become a regular part of the **peer review** process.

As these technologies are complex, an important objective is to make their **use** for reproducible research **easy and understandable for the broader communities of researchers in geography and geosciences** even without a lot of programming experience.

The expected **technical challenge** quickly turned into a question of how to serve **individual needs** of all participants in scholarly communication and subsequently into the **structural and cultural contexts and eventually science policy**.

Because reproducibility needs to be approached considering all of these perspectives, I use the **culture change pyramid** to guide us through this talk.




The culture change pyramid has five layers that, going from bottom to the top, show how cultural change can be realised, in this case the cultural change towards higher reproducibility of research.

Each layer has a specific field of activity given in the pyramid, and a goal, shown on the right hand side next to the pyramid.

🖱️ **Part 1** of my thesis covers the lower two layers, the infrastructure and user experience.
They make reproducible research possible and easy.

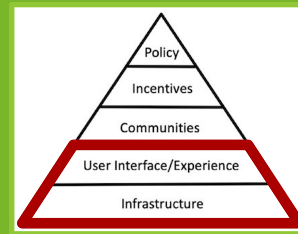
🖱️ **Part 2** covers the upper three layers on how to make reproducible research normative, rewarding, and required.

🥤

⏱️ **5**

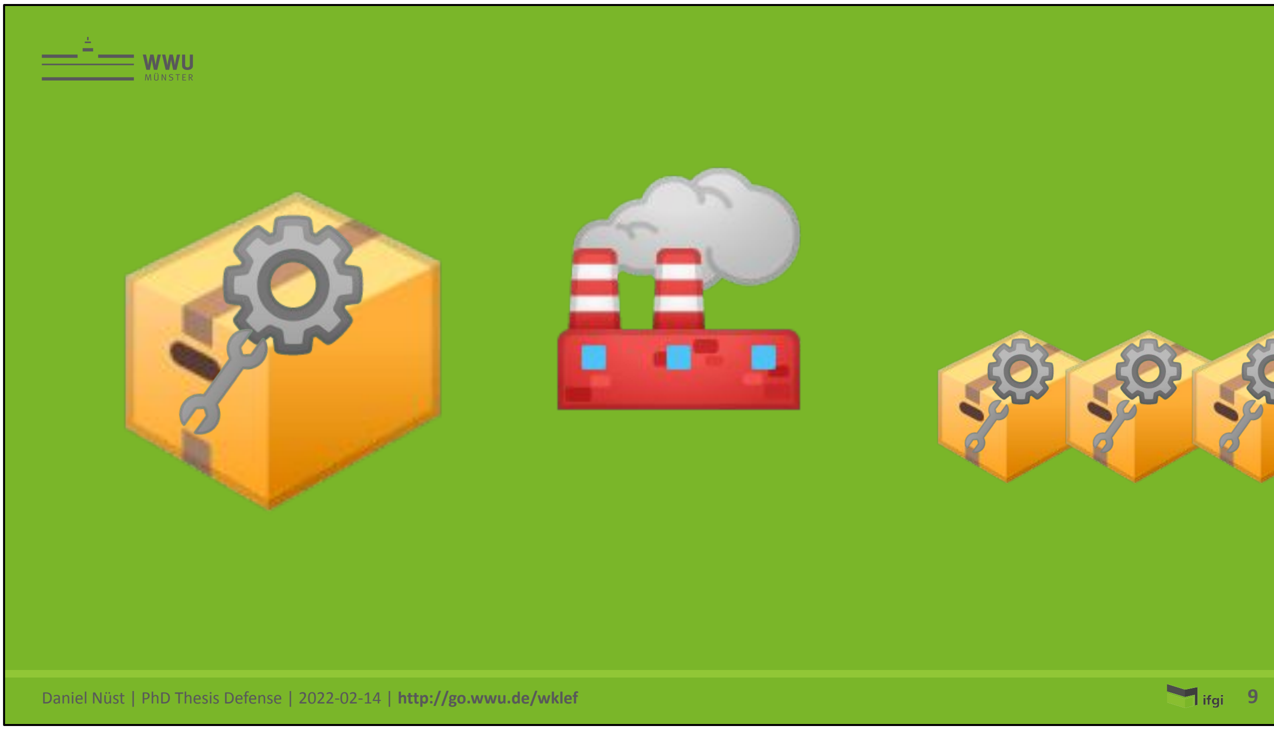
WWU MÜNSTER

# Part 1: Infrastructure & user experience






Let's start with part 1…



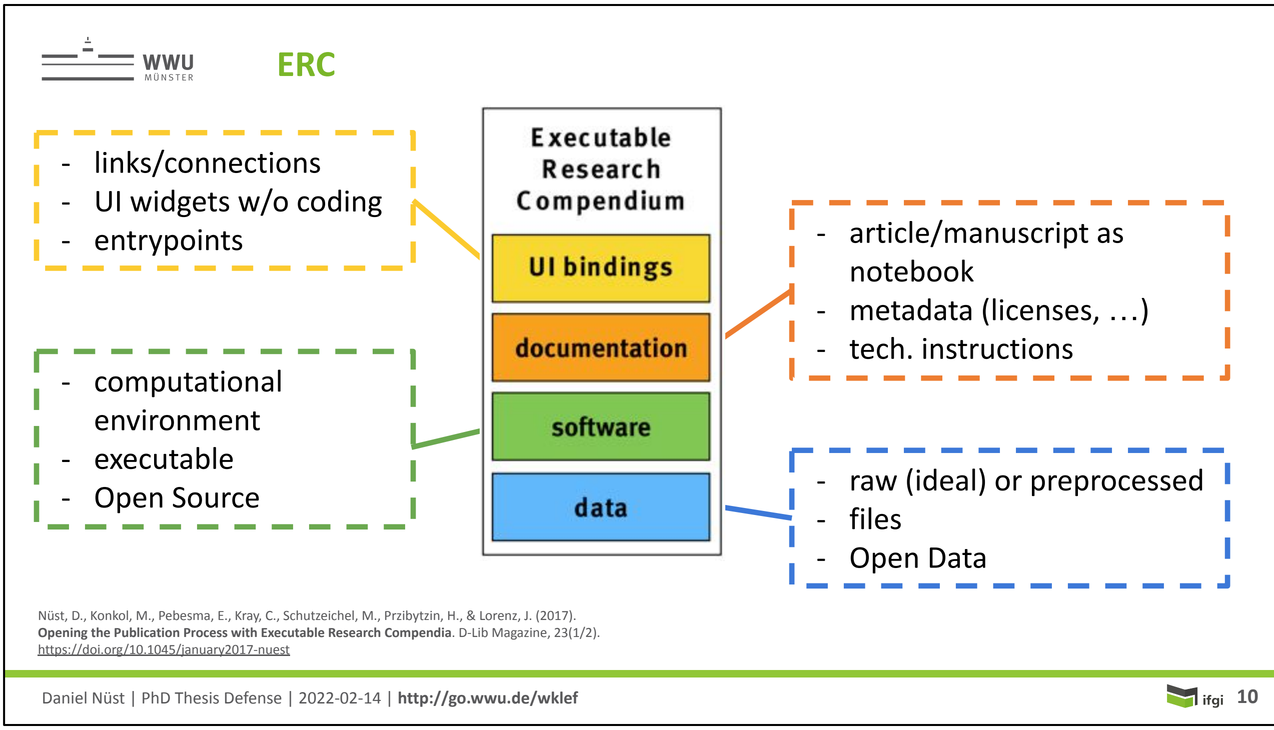

The questions that guide this part of the research are concerned with the **packaging** of computational research so that it is useful and usable, how this packaging can include an **executable runtime environment**, and how we can connect the executable packages with the existing **infrastructure and practices for publishing research**.

----------------

How can packaging of computational analyses serve the needs of authors, publishers, readers, and preservationists?

To what extent can the process of capturing the runtime, software, data, and metadata of reproducible research packages be automated in geoscientific analyses?

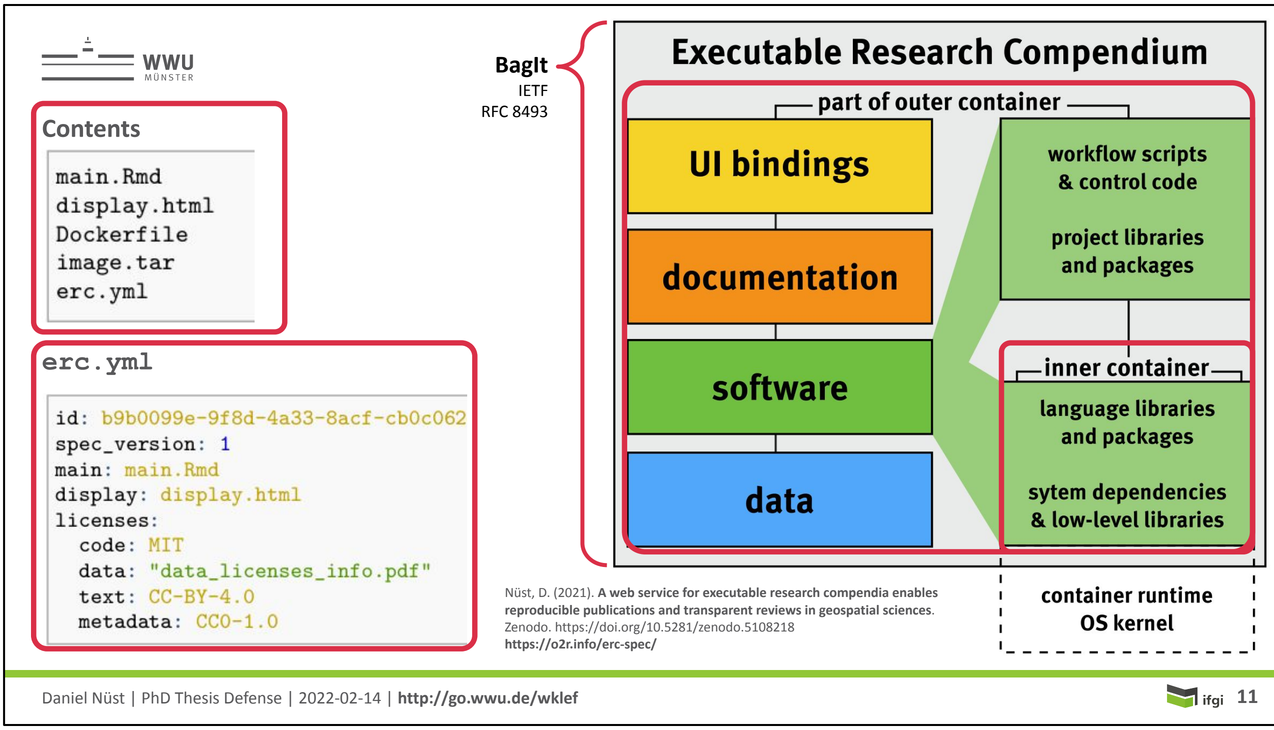
How can the ERC fit into the existing practices and infrastructure for research and publishing in geography, geosciences, and GIScience?




At the beginning of the o2r project the executable research compendium (short: the ERC) was designed by the o2r team to make this packaging and connection possible. During this dissertation, the ERC was implemented.

The ERC has four core parts:

🖱 **data**, e.g., input data, pre-processed data, result data, all open
🖱 **software**, e.g., scripts or special libraries, and always a containerised runtime environment,
🖱 **documentation**, e.g., developer docs, user docs, but also the actual scientific article, and
🖱 **UI bindings** as entry points for user interaction.




Let's look at the ERC and it's technical specification

The ERC uses the idea of **nested containers**. There is an **inner** container based on Docker for capturing the computing environment, except the operating system kernel. This allows to quickly start and execute the workflow and provides a human and computer readable description of the computing environment.

The **outer container** contains all files of a research project, an archive of the actual Docker image, the recipe for the image, the Dockerfile, and other metadata files. The outer packaging follows the **BagIt specification** from the archival domain, and enables deposition in data repositories and file preservation.

The ERC **contains** a main document, which manages the whole workflow.
A display document, which needs to be an HTML file, points users and systems to the file that should be displayed to users when the want to read an ERC.
These two entry points into the captured workspace are accompanied by even more way to explore the ERC based on UI bindings. UI bindings are a key contribution by my fellow o2r PhD student Markus Konkol and are not on the agenda today. I can only recommend to take a look at his thesis to learn more about bindings.

Finally, there is the **ERC configuration** file shown at the lower left. It only points to the two main documents, as standard names for the environment description and container file are used, and lists the licenses for the four parts. The ERC tries to follow the paradigm of “convention over configuration”. So you could even omit the display and main file here.

🖱️ That’s all.

The ERC enables reproducible research by building on the **literate programming paradigm** for the main document to capture a full computational workflow.
It also enables new kinds of in-depth examination, including manipulation and substitution of parts and the reuse of results.
The only requirement for researchers is to share their work in a computational notebook.

With the goal of a barrier-free creation in mind, we tried to **automate** the process of ERC creation, especially the inner container tailored to the ERC’s software.

🥤
⏱️ **10**

# containerit

Nüst, D., & Hinz, M. (2019). **containerit: Generating Dockerfiles for reproducible research with R**. Journal of Open Source Software, 4(40), 1603. https://doi.org/10.21105/joss.01603

**Capturing** an R session (script, Rmd) in a Dockerfile

User only uses **R functions**
System **dependency** resolving

Always **executes** script (`callr`), hard to fool (unlike static progr. anal.)

**github.com/o2r-project/containerit**

```
> suppressPackageStartupMessages(library("containerit"))
> my_dockerfile <- containerit::dockerfile(from = utils::sessionInfo())
> print(my_dockerfile)
FROM rocker/r-ver:3.5.2
LABEL maintainer="daniel"
RUN export DEBIAN_FRONTEND=noninteractive; apt-get -y update \
  && apt-get install -y git-core \
    libcurl4-openssl-dev \
    libssl-dev \
    pandoc \
    pandoc-citeproc
RUN ["install2.r", "curl", "digest", "evaluate", "formatR", \
  "futile.logger", "futile.options", "htmltools", "jsonlite", \
  "knitr", "lambda.r", "magrittr", "Rcpp", "rjson", \
  "rmarkdown", "rsconnect", "semver", "stevedore", "stringi", \
  "stringr", "xfun", "yaml"]
WORKDIR /payload/
CMD ["R"]
```

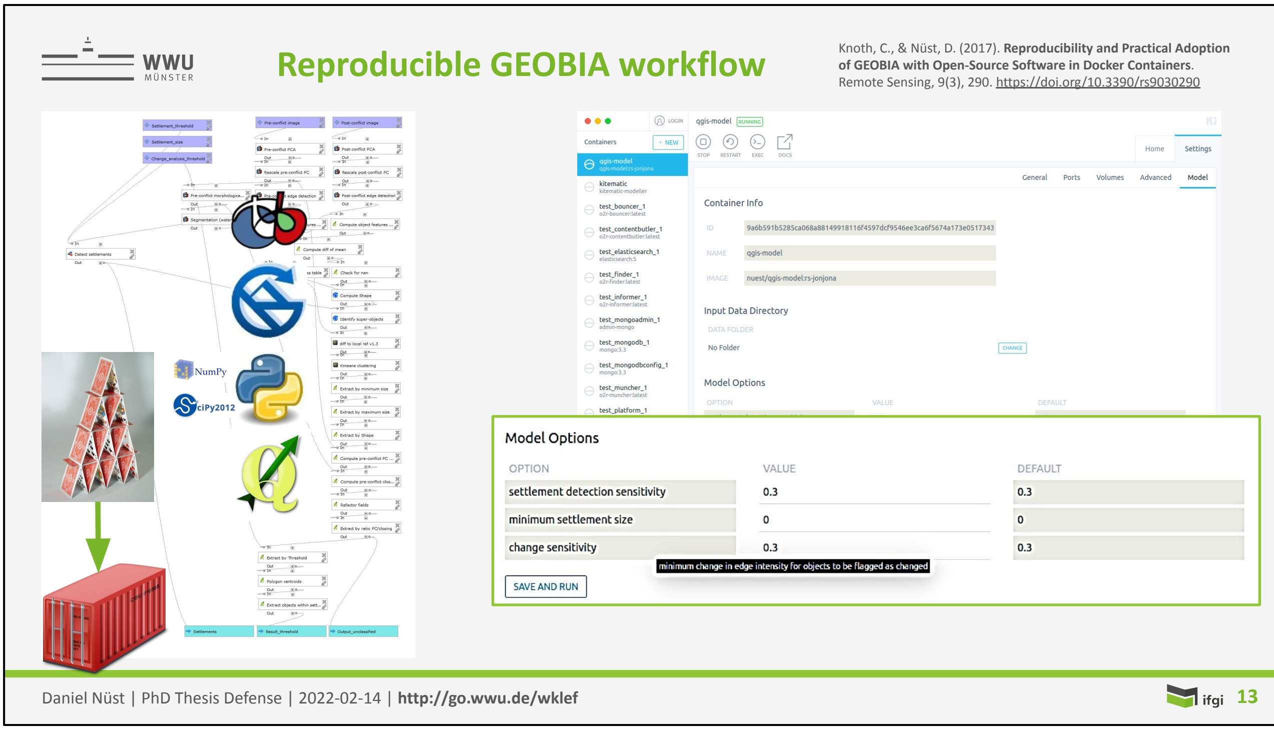


The tool we came up with for the automated creation of the inner container is the R package **containerit**.

containerit helps you to **capture** the R session needed to run a specific R script or R Markdown document in a Dockerfile using **only plain R** commands. The code on the right hand side shows you how quickly this can be. The created Dockerfile includes R packages and their system dependencies.

A core idea is that containerit **always executes** the script and inspects the actual R session, instead of parsing scripts or documents to find out which packages are needed.
containerit is an important tool in the o2r platform and the core building block to show that it is possible to capture everything needed to create an ERC without any user actions.

**So automation is possible. What about the experiences from manual research compendia creation, and what about non-R workflows?**



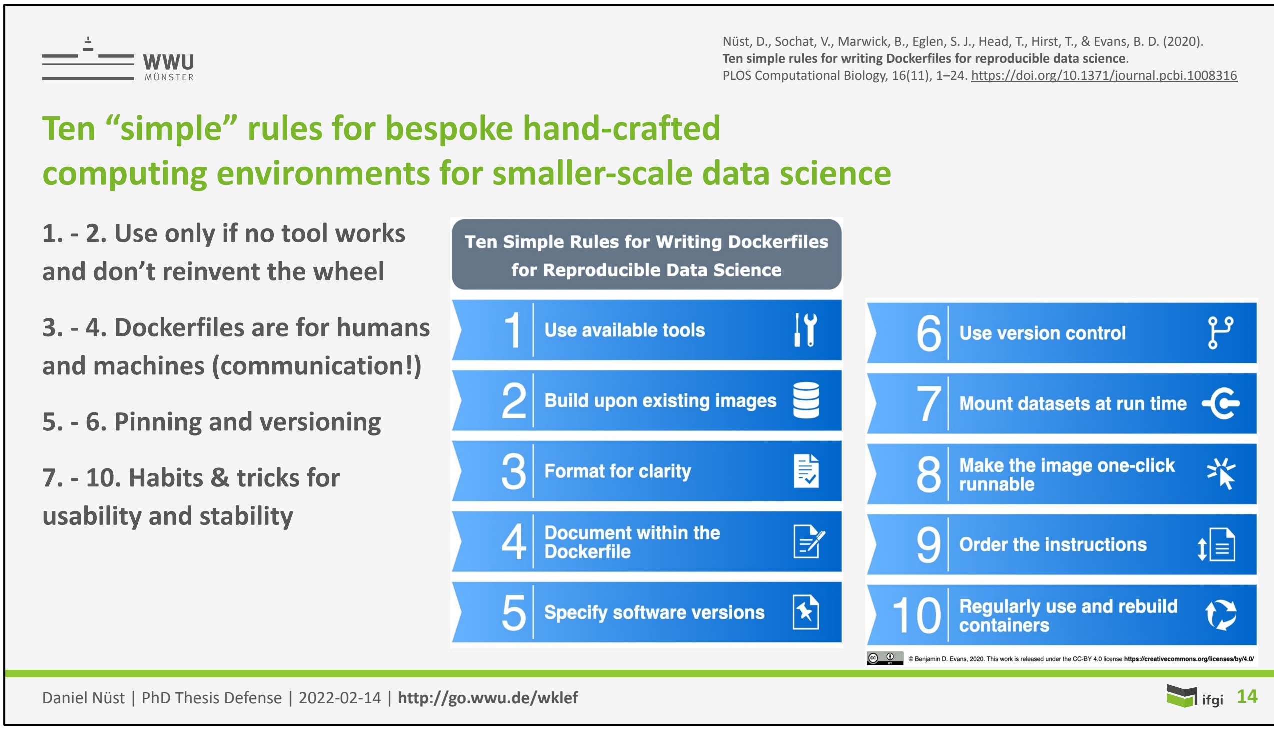

An practical and extensive research compendium was created in a collaboration with Christian Knoth, who implemented a geographic object-based image analysis workflow with open source tools. I helped to put together the complex manual steps of the pipeline using multiple tools and packaged the **pipeline in Docker container**. Besides the packaging, the challenge was to transfer a complex workflow from the dominating closed commercial software into a more open alternative. The context of Christian's works in the area of human rights NGOs meant that there are severe budget restrictions and that transparency is important, so a UI was needed and technical demands should be low.

The work involved quite a lot of fiddling, because specific versions needed to be installed from various sources to achieve compatibility.
**User can even manipulate parameters** of the the pipeline and exchange the input data with a simple form-based user interface.

--------------------

**Main benefits of the approach are** that it is low cost, easy to use, reproducible, and customizable, and thereby very well fits the use case.

**Main constraints of the approach** we found were the functionality of open source software for GEOBIA (no single tool to realise it all) and the creation of containers for a research compendium was therefore complex.




And this article would have been very helpful for the GEOBIA paper.

🖱️ **The work** presents ten rules for writing Dockerfiles for reproducible data science. It targets smaller-scale data science research (not HPC).

Reproducibility in research is often a question of putting in the best efforts, not about achieving the perfect solution, and what is best may even change over time.
The core ideas of the ten, to be honest not so simple, rules are
🖱️ using established tools and images over own reimplementations,
🖱️ to put documentation for human users into the Dockerfiles,
🖱️ to pin computing environments and use versioning,
🖱️ and to follow some useful tricks and habits to make sure that one actually works within the container every day, not just when something is "finished" or about to be published.

containerit is one of the tools you can use to follow rule 1. Even more rules are followed if you create an ERC with the o2r reproducibility service.
So let's take a look into that.

⏱️ **15**

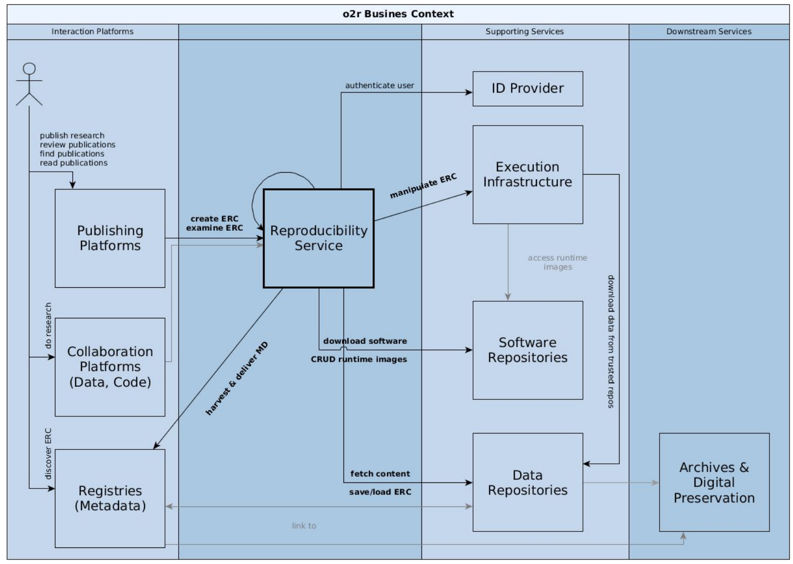
WWU MÜNSTER

# o2r web service for reproducibility

Nüst, D. (2021). **A web service for executable research compendia enables reproducible publications and transparent reviews in geospatial sciences**. Zenodo. https://doi.org/10.5281/zenodo.5108218

**ERC Reproducibility Service (ERS):**

- Context
- Architecture
- Specification
- Web API



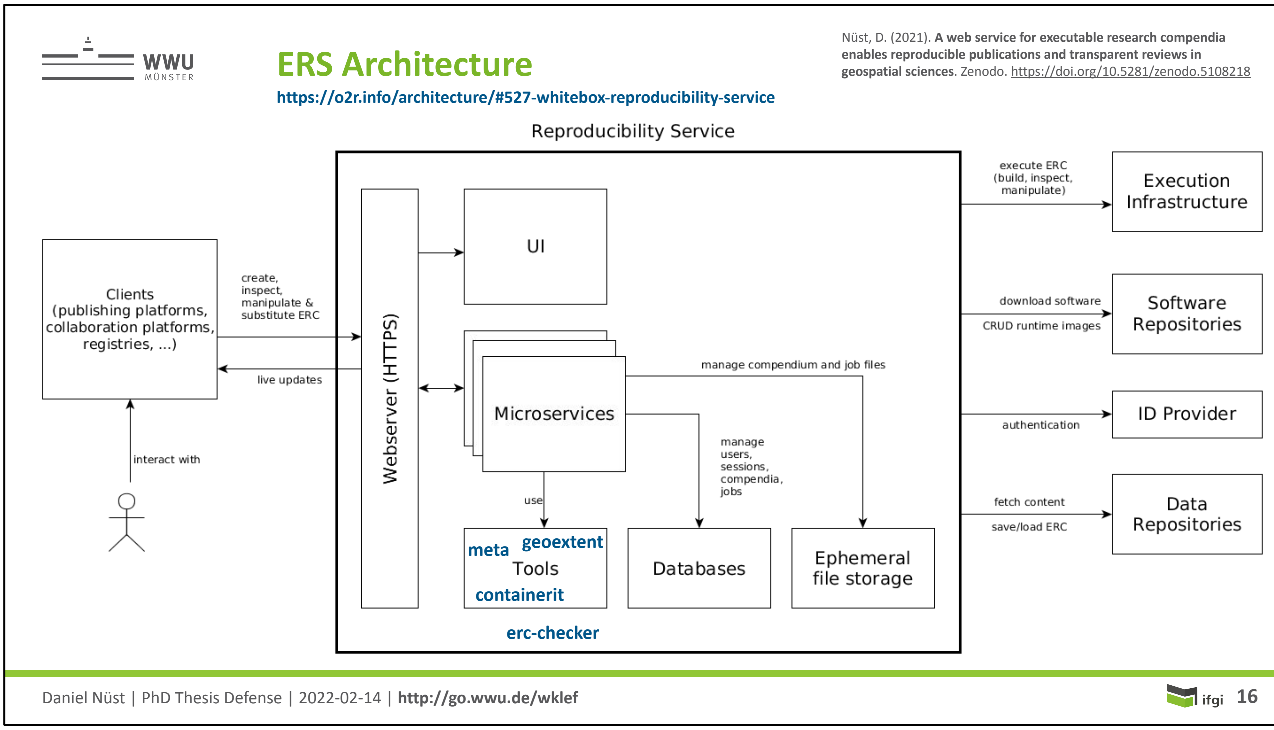



To connect executable research compendia with the platforms and practices of scientific publishing, we designed the ERC reproducibility service (ERS).

Left column: user interaction platforms

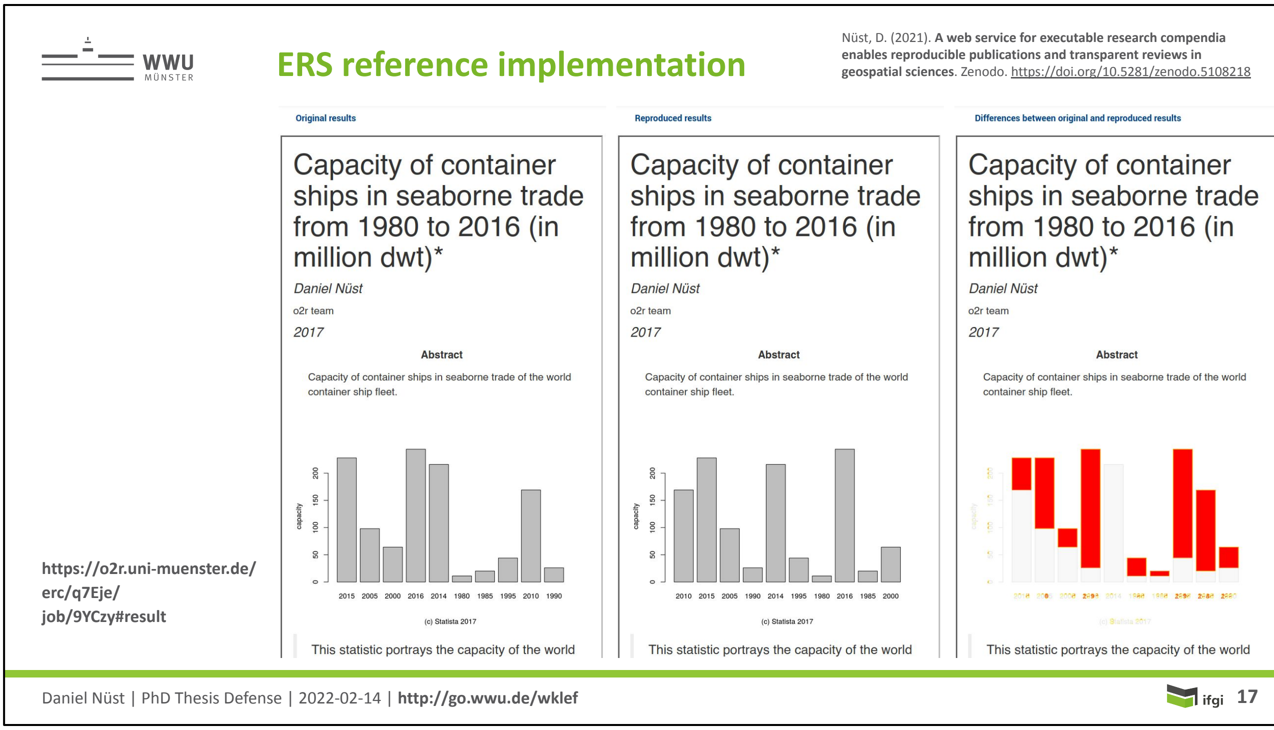
Third column: not-reinvented base services

[no further notes]




**Containerised microservices**

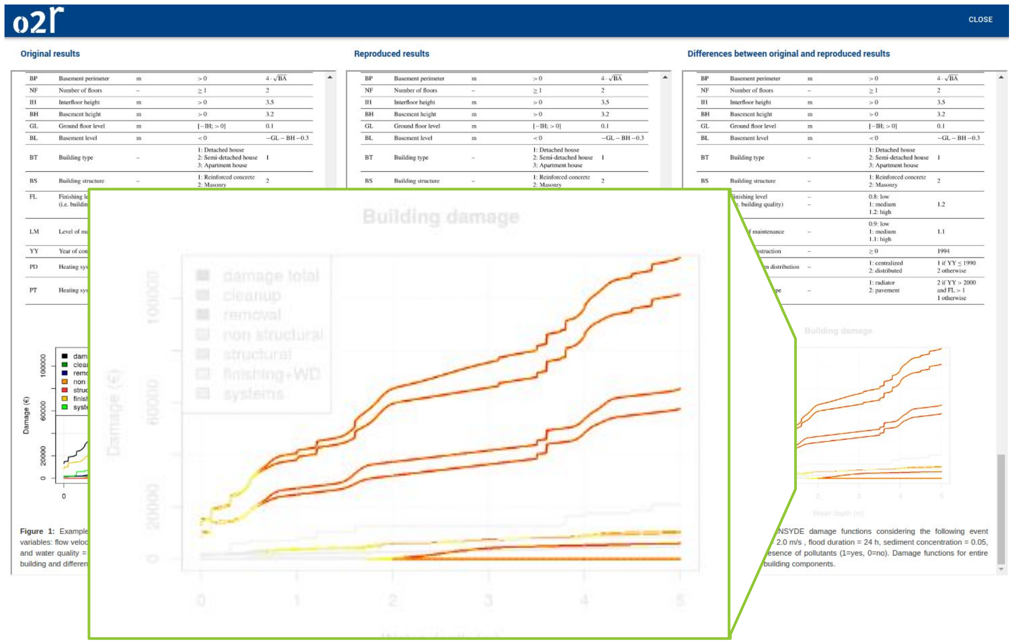
🖱️ **Independent Tools** > reusable, containerised, multiple languages




[…]

Now, imagine this view being embedded into a journal platform and a little badge on the article saying “Executable Paper”.

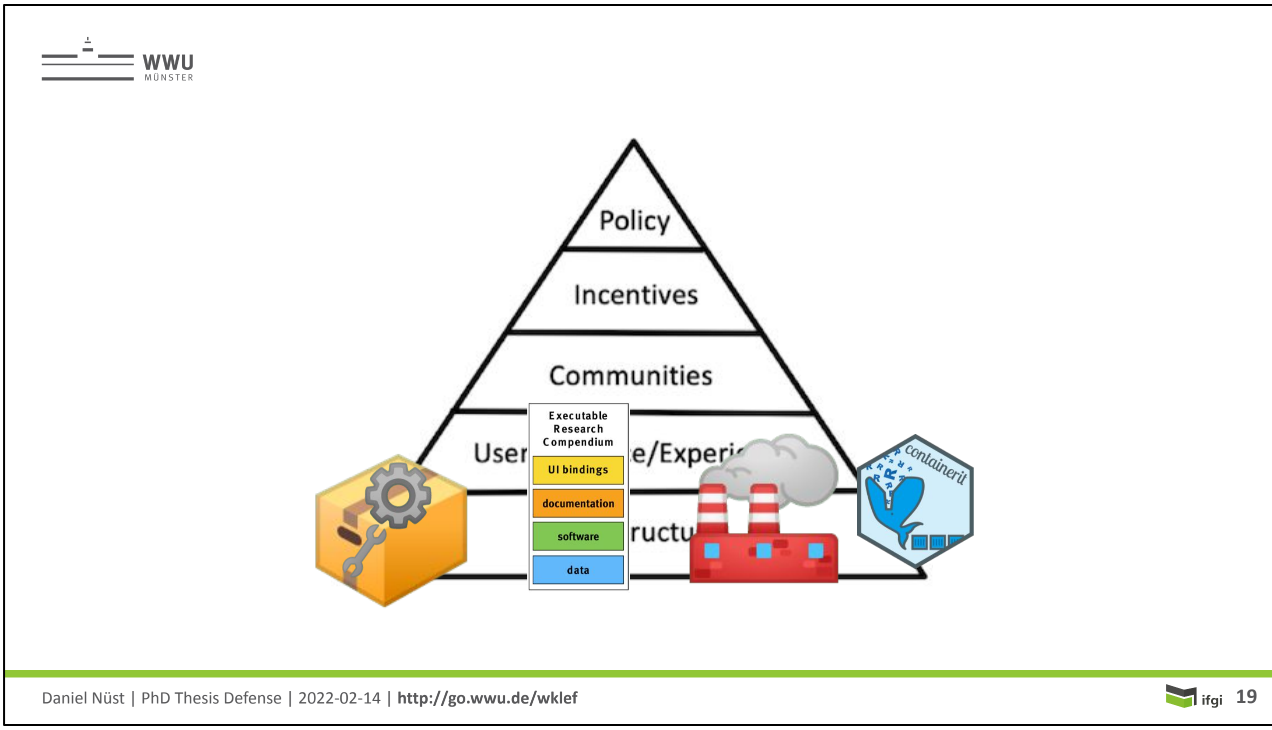
WWU MÜNSTER




Nüst, D. (2021). **A web service for executable research compendia enables reproducible publications and transparent reviews in geospatial sciences**. Zenodo. https://doi.org/10.5281/zenodo.5108218



ERS reference implementation




So 🖱️ **now** we have manually crafted research compendia and the ERC. We identified literate programming with computational notebooks and containerisation as the foundation for research compendia.

The ERC is also very well connected with publishing infrastructure. Even non IT experts can create the ERC.
We have made the change towards reproducibility possible and easy, or at least easier, with infrastructure and user experience.

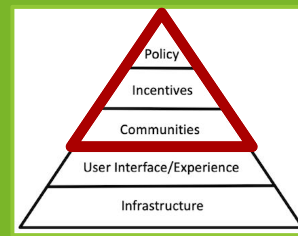
Of course, the ERC is **not the only option to package** computational research. And of course the **use cases for using Docker to create portable environments** are much more diverse than what containerit can capture automatically.
We looked into alternative packaging formats and platforms in a **review paper**, and we surveyed the landscape of applications using Docker with R and called it the **Rockerverse**.
I cannot cover these articles in detail in this talk today, but I want to mention that practically all other platforms also use containerisation technology.

It's time to move up in in pyramid!

⏱️ **18**

WWU MÜNSTER

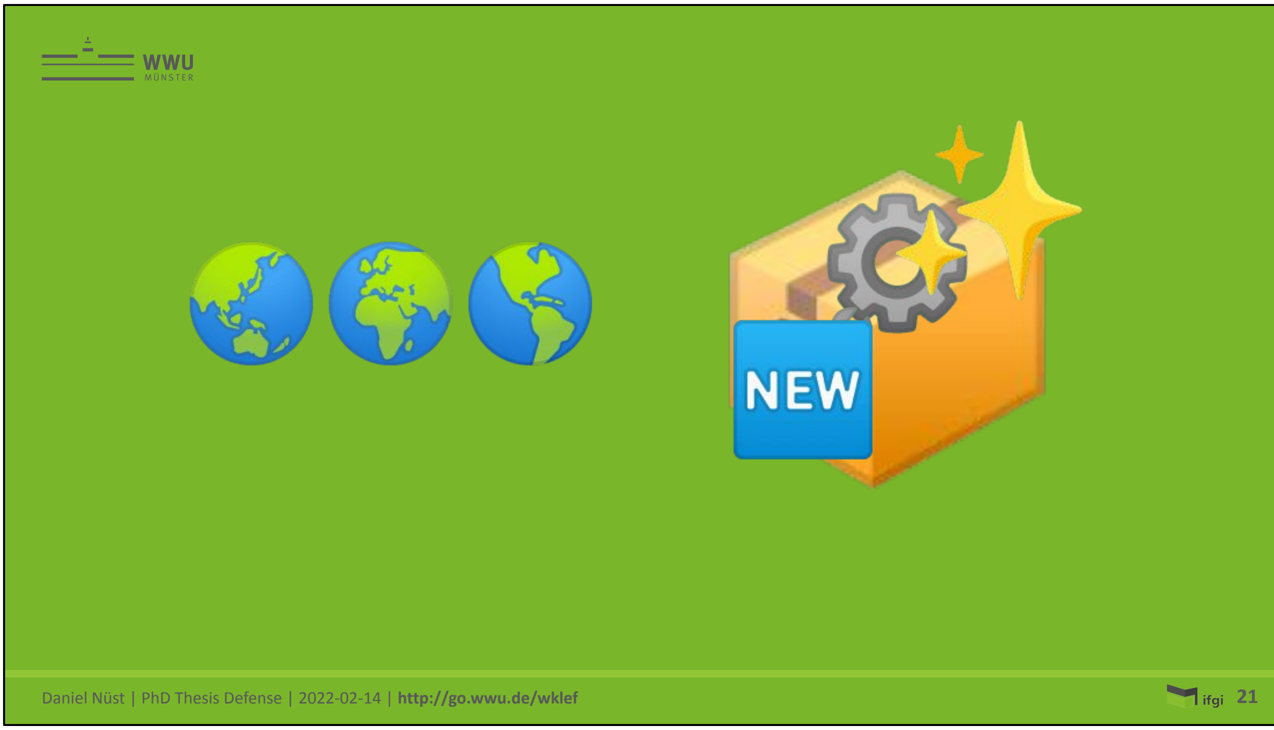
# Part 2: Communities, incentives & policy






Leaving technical challenges behind, let's think about why and how to put these solutions into practice by tackling the top three layers of the culture change pyramid, namely: communities, incentives, and policy.

**These three layers especially are very much intertwined.**

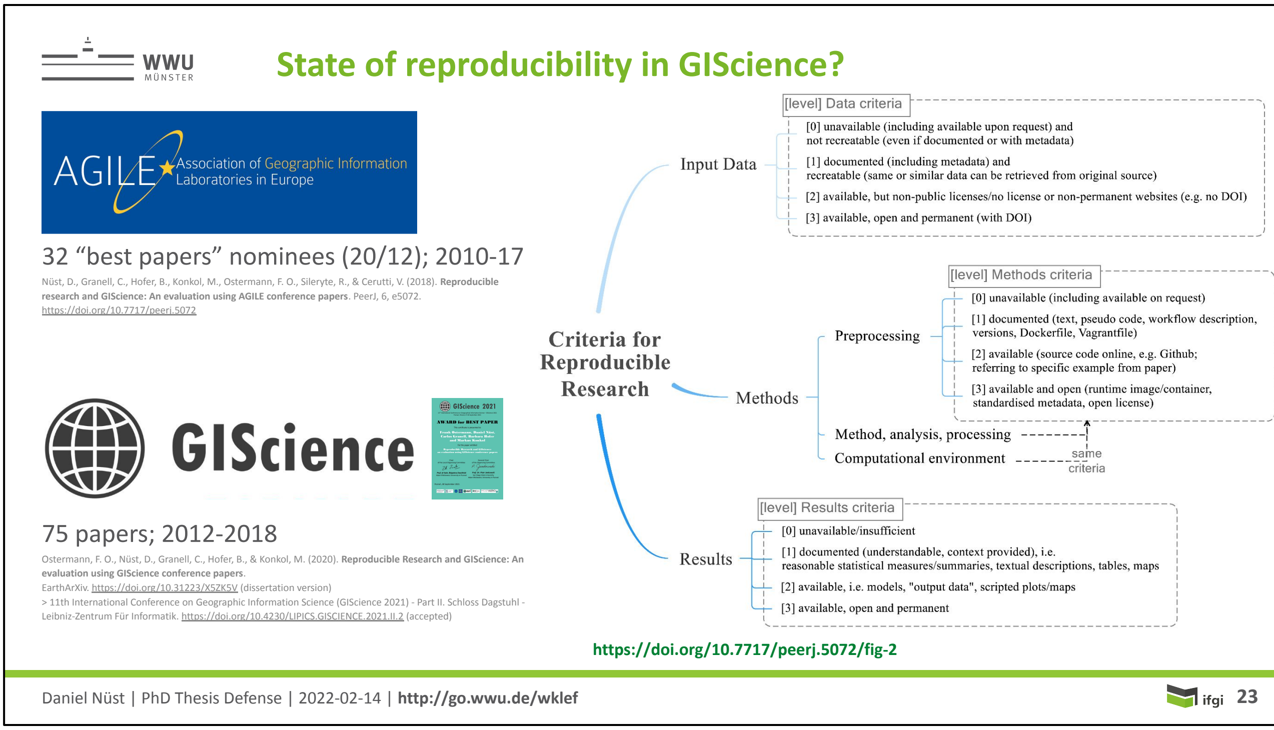



In this part, the guiding questions concern the 🖱️ **domain-specific challenges and solutions to reproducible research**, and what 🖱️ **new features** we can built upon packaged workflows, and how?

-----------------------------------------------

What are domain-specific challenges and solutions for the geography, geosciences, and GIScience domains in the context of reproducible publications?

What new services and features can be built upon reproducible workflows, e.g., when packaged as an ERC?




To effectively propose change, one must first understand the state of reproducibility in a discipline.

That is why we conducted two studies for two established community-led conferences in the field of GIScience: the yearly AGILE conference and the biannual GIScience conference.

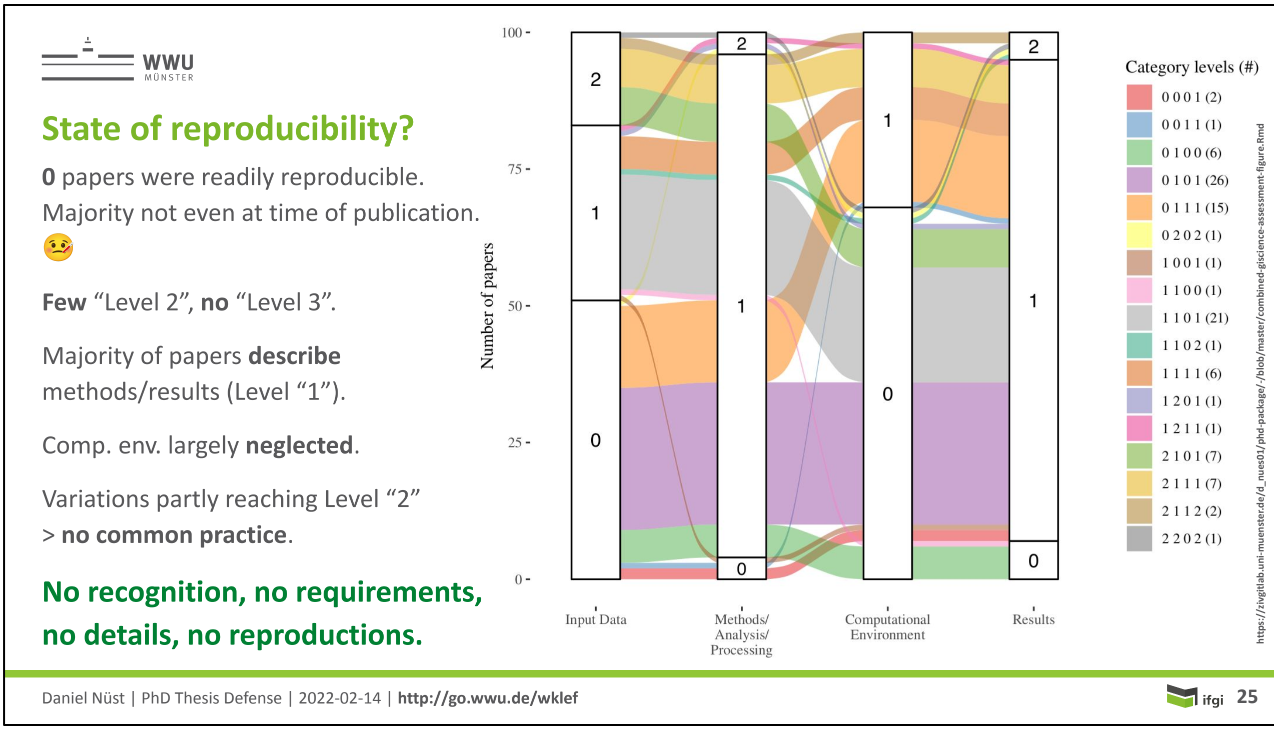
The team behind these works developed a **rubric** to classify papers. The rubric has **five categories** (input data, preprocessing, method, computational environment, results) and each category can reach **four levels** from 0 (unavailable) to 3 (resource is available, open, and permanent).

**------------------------------------------------**

Similar **time frames** were used. In the first study for AGILE only the nominees for "best paper" awards were evaluated, for the second study we decided to look at all available full papers.

**Assessment** was done by two assessors. If there was disagreement in the assessment of a paper, the whole group discussed the work. This approach was originally chosen because we did not expect to have the time to put in the extra effort to conduct actual reproductions, though in reality, it never would have come to that extra effort.

It should be noted that both works clearly **separate the validity of results and the quality of the work** from the reproducibility of workflows. These aspects of papers that have been published years ago should be measured by the standards applied then.




The studies to assess the state of reproducibility in GIScience show quite clear results which are summarised in this alluvial plot combining the data for both studies. It shows the classification for 100 papers.
**The plot** has **four columns** for the categories. I left out "preprocessing" because the category was difficult to apply clearly. Groups of papers with the same properties are shown as coloured bands flowing through the diagram. What is the result?

**Reproducibility is low**, even when looking back just a few years. Both studies show that, at the time of investigation, **none** of the assessed 107 papers provided sufficient details and references to deposits of data or code to allow a reproduction to be practical at the time of the assessment.
( i.e., without intensive interaction with the authors or without substantial efforts to recreate large parts of the workflow).
For the vast majority of papers, this is the case even at the time of publication!

The **main patterns** are:

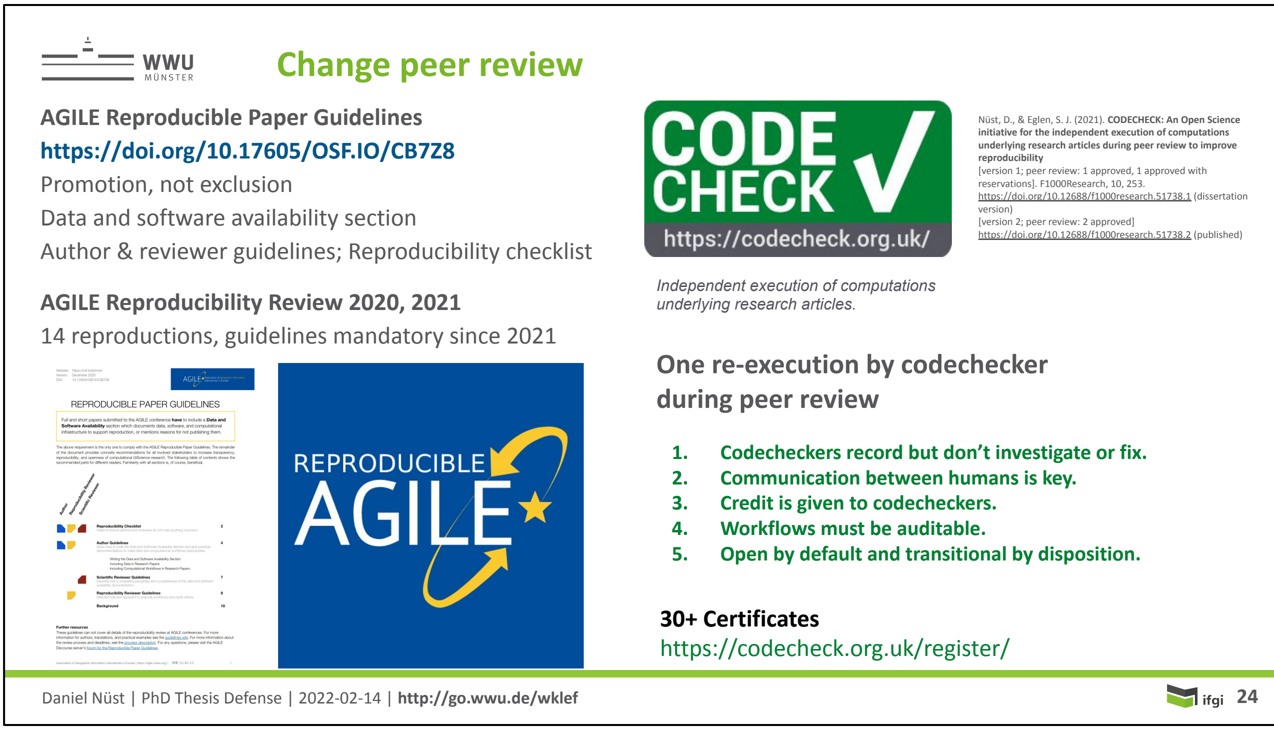
- The largest groups regarding the numbers of papers do not reach level 2 (i.e., access provided, but not permanently and open) in **any category**.
- Noteworthy is also the **complete lack of papers that achieve the highest level of 3** (open and published permanently) in any of the categories.
- The majority of papers describe methods and results sufficiently to make a **recreation in principle possible** (level 1).
- However, about half of the papers **do not give enough details to access**

- **data** and about **two thirds do not document the computational environment** at all (level 0). This shifts the major burden of reproduction to the reader.
- 🖱️ Within the smaller groups that achieve levels of 2, the figure shows that there's plenty of jumping between levels and most combinations seem to exist. This is a sign of the **lack of common practice** and the differing opinions and understanding of authors even when they try to work openly and reproducible.

🖱️ The assessments' core takeaway is that unavailability of data and code and incompleteness of methodological details are **barriers to reproductions**. We largely attribute this to a **lack of recognition** and requirements, so the articles give concrete **recommendations** for improvement to authors and conference organisers to **enable reproducibility within peer review**.

🥤

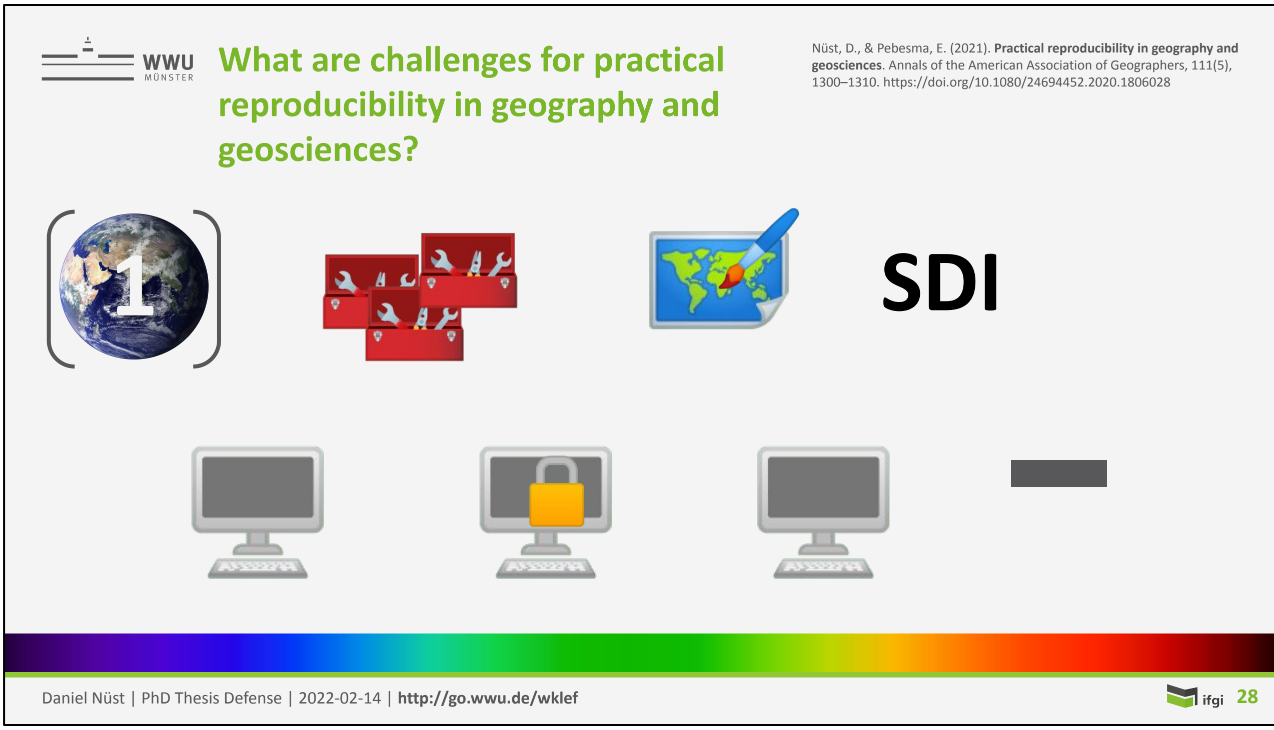
FIGURE: Combined alluvial diagram. Includes groups of papers across four categories for the merged AGILE (Chapter 11) and GIScience (Chapter 12) datasets; the category Preprocessing was dropped because of difficulties to clearly assess it; included are 100 papers without any "not applicable" value from 2010 to 2018




So, how can we change peer review practices?

One direct result of the recommendations of the AGILE assessment are the AGILE reproducible paper guidelines. Since 2020, a reproducibility review successfully reproduced 14 full papers at the AGILE conference. The fact that these 14 papers could be reproduced strongly suggests that **all** of them, unless data could only shared privately, would likely **reach at least a level of 2 across all categories**.

🖱️ **Furthermore, we developed CODECHECK.** CODECHECK is a collaboration with Stephen Eglen from the University of Cambridge. It's approach is to have one re-execution of a workflow by a human codechecker during peer review. It follows a set of principles to set the scope and ensure recognition. Up to today, we have created over 30 CODECHECK certificates, which include the AGILE reproductions.




So if we could come up with answers to the process, what about actual practical challenges for the disciplines in question?

First, the fact that we have **only one Earth** and **unique locations** is more an issue for replicability. The same applies for the **unique systems** that we rely on to observe Earth, such as remote sensing satellites or even social media networks. Since we cannot just come up with a second system or duplicate study areas, the more important is the transparency and openness of the research.

Another thing we have a lot more than just 1 of are the **tools** we use for our research. This multiplicity is indeed a challenge in the disciplines in question. Data collection, analysis, and presentation are often not just done in one software. If you want to use latest methods, you may rely on R for data wrangling. Yet for cartography and data collection you might rely on a proprietary GIS.

**Making maps** is also often a design process that is not entirely driven by code, out of habit, visual preference, or familiarity with specific tools, or due to limitations of analysis software.

**Spatial data infrastructures** are great, because we can rely on a lot of standardised interfaces to access data. We also need them for hosting and processing big data. But remote data sources often need authentication and the access is not scripted but manual, with reduces reproducibility.

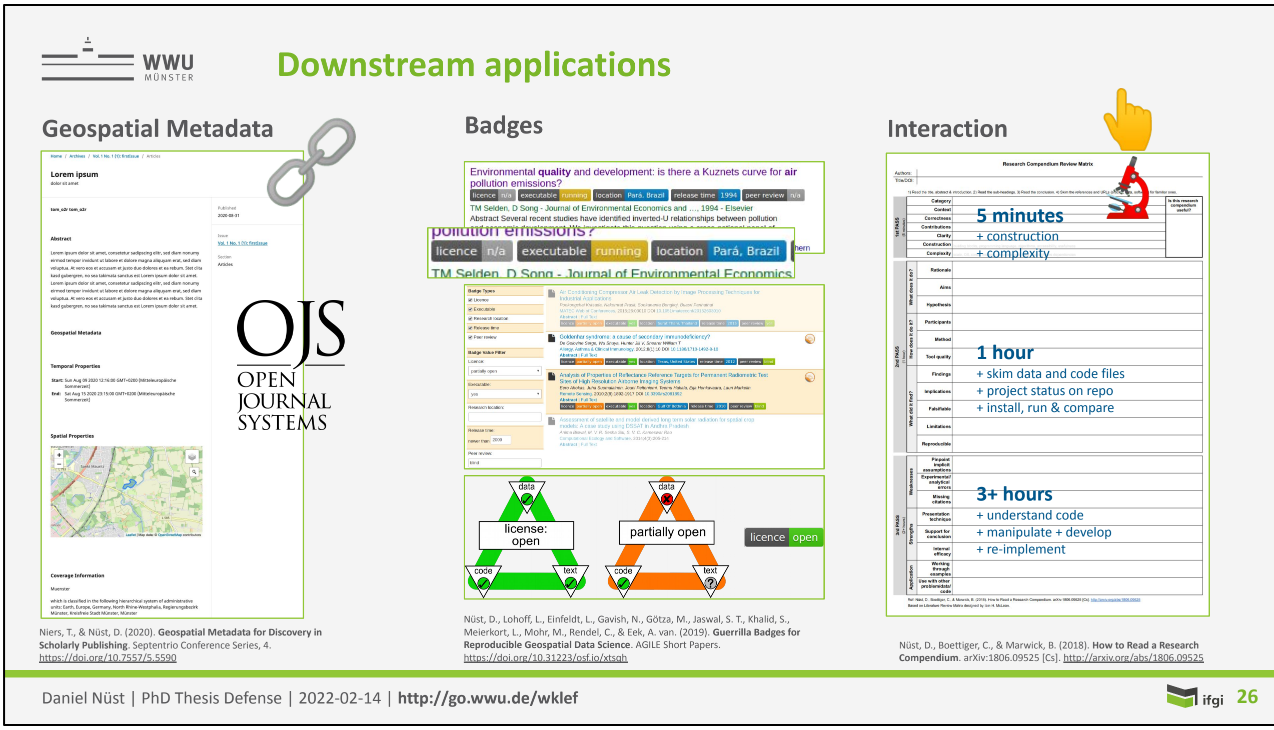
🖱️ **Remote data processing on “free” platforms** such as Google Earth Engine made large remote sensing datasets readily available for many users, but we rely on closed systems here again, although alternatives now exist. The same issues apply to the very widespread 🖱️ **proprietary GIS**. Open tools today may require a higher literacy, but they are also much easier to combine with reproducibility practices.

Some fields, such as 🖱️ **qualitative GIS** have been deemed non-reproducible. That may indeed be the case in principle, but it should be an excuse not to make every part of the research process that can be made reproducible, even if it is just the data visualisation. Similarly, we can make workflows with 🖱️ **sensitive data** (typical for geospatial data!) reproducible, that is not a reason not to try as hard as we can.

And finally, the research that happens under the umbrella of geography and geoscience is distributed across a very 🖱️ **diverse spectrum**, from chemical experiments in labs to human-subjects studies in the field to high performance computing of satellite data. This complicates education and guidelines, as researches will often feel like they are not in the target audience, that their work is different. This diversity makes the decision which **degree of reproducibility is “good enough,”** even harder, but every community of practice needs to find their own answer to that. We believe that in most cases “very, very close to the original” is feasible and practical despite all challenges.

-----------------------------------------------------

https://en.wikipedia.org/wiki/Talk:Visible_spectrum#/media/File:Linear_visible_spectrum.svg

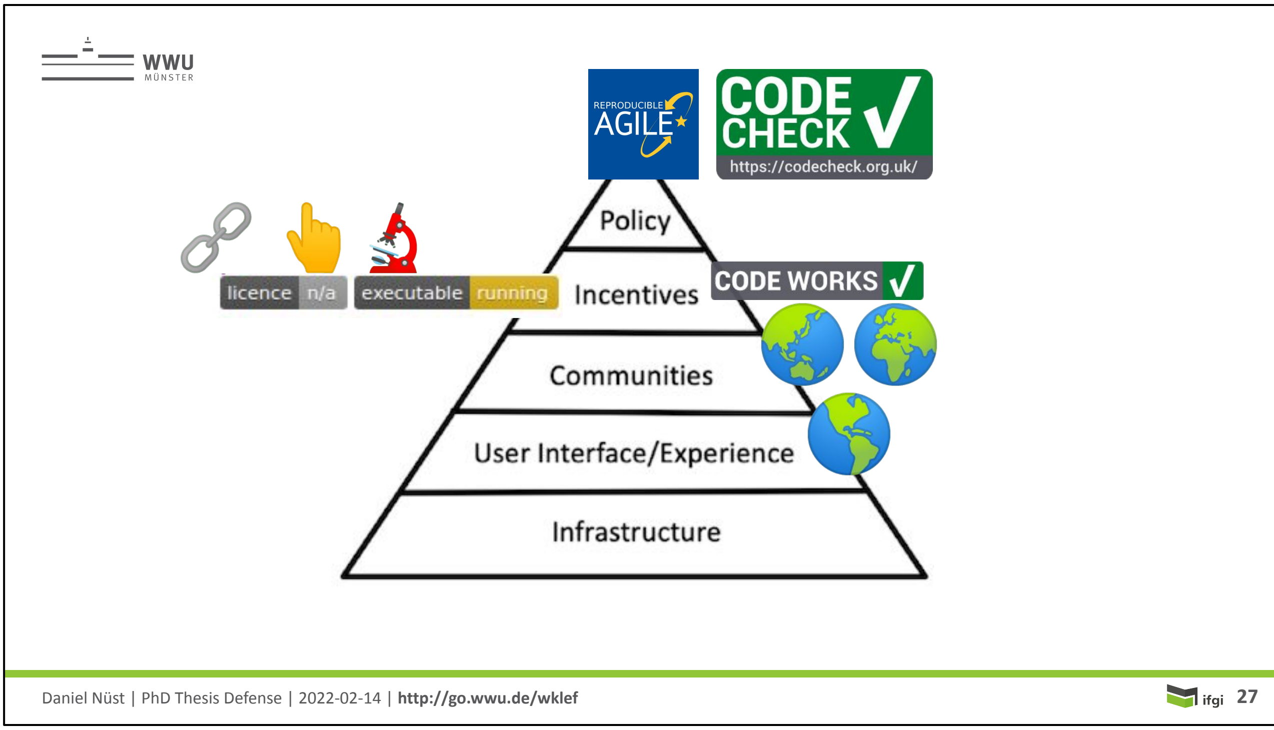



Besides making Other improvements that can be build upon ERCs were explored as well.

🖱 **First**, we showed how to enhance scholarly articles with geospatial metadata to improve discoverability without manual efforts.
[If authors submit ERCs to a journal, the geospatial metadata could be extracted automatically from the included data to reducing errors and manual labour.]

🖱 **Second**, we demonstrated how more meaningful metadata can improve search and discovery in academic search engines.
[We inserted badges as visualisations of complex article metadata, more precisely the kind of metadata that would be readily available with ERCs. Information such as executability or human-readable location information can be retrieved or derived from the o2r API and helps researchers to decide if something is relevant for them.]

🖱 **Third**, we took the reader's perspective at research compendia and extended Keshav's method for evaluating, reading, and understanding a scientific article in three passes with additional steps to explore research compendia.
[This work is partly based on experience and partly a theoretical exercise to contribute to the discourse about benefits and challenges of sharing more than just a PDF before novel practices are set in stone.]




**In part 1, manually crafted research compendia as well as the ERC created with the reproducibility service gave us the infrastructure and user experience for cultural change towards more reproducible research.**

Now after part 2, we have moved all the way up in the pyramid.

We answered the question of the 🖱️ **state of reproducibility** in GIScience, investigated 🖱️ **domain-specific challenges**, gave examples how research compendia enable 🖱️ **new features** in scholarly communication, and showed how the 🖱️ **practice of peer review can be improved**.

What we learned can help to turn reproducible research into a **normative, rewarding, and eventually required practice**.

**------------------------------------------------**

What are domain-specific challenges and solutions for the geography, geosciences, and GIScience domains in the context of reproducible publications?
What new services and features can be built upon reproducible workflows, e.g., when packaged as an ERC?

# Key contributions



**Thank you for your attention so far - to wrap up the talk, I'd like to take a look at the key contributions of this dissertation.**


This work demonstrates a working 🖱 **infrastructure** for more reproducible scholarly communication based on the Executable Research Compendium.

We offer innovations in the application of containerisation for reproducible research, including 🖱 **handcrafted and automatically captured computing environments**.

The dissertation describes the 🖱 **state of reproducibility in GIScience**. Based on the dire state, we develop a path for communities to adopt reproducible research practices. We could even show that this path is very promising and …

🖱 **present approaches that introduce reproducibility in peer review and publishing practices.** The current disruptions in scholarly publishing, such as questioning big deals with publishers and increasing adoption of preprints, should be used as an opportunity to rethink the way we share and collaborate.

If journals, conferences, and researchers acknowledge these contributions and adopt them, I am convinced we would get a big step closer towards 🖱 **verifiable,** 🖱 **reusable and extendable research products that would be useful for** 🖱 **education** and enable 🖱 **innovation** through openness and transparency.

With these main contributions in mind, one must also consider the question of 🖱

**what is the geoinformatics in all this? I will leave that to the discussion :-).**

**-------------**

https://cyclingsolutions.info/cost-benefit-of-cycling-infrastructure/
https://www.pxfuel.com/en/free-photo-xlwxm
https://www.publicdomainpictures.net/en/view-image.php?image=95827&picture=blue-containers

**https://www.maxpixel.net/Helping-Each-Other-Together-Teamwork-Winning-2643652**
**https://visibleearth.nasa.gov/images/57723/the-blue-marble (public domain)**

WWU MÜNSTER

# Citation & Publications

**Nüst, Daniel. 2021. Infrastructures and Practices for Reproducible Research in Geography, Geosciences, and GIScience. Doctoral dissertation, University of Münster, Germany. https://doi.org/10.5281/zenodo.4768096**


**Infrastructure & use experience**

Knoth, C., & Nüst, D. (2017). Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers. Remote Sensing, 9(3), 290. https://doi.org/10.3390/rs9030290

Konkol, M., Nüst, D., & Goulier, L. (2020). Publishing computational research - a review of infrastructures for reproducible and transparent scholarly communication. Research Integrity and Peer Review, 5(1), 10. https://doi.org/10.1186/s41073-020-00095-y

Nüst, D. (2021). A web service for executable research compendia enables reproducible publications and transparent reviews in geospatial sciences. Zenodo. https://doi.org/10.5281/zenodo.5108218

Nüst, D., Eddelbuettel, D., Bennett, D., Cannoodt, R., Clark, D., Daróczi, G., Edmondson, M., Fay, C., Hughes, E., Kjeldgaard, L., Lopp, S., Marwick, B., Nolis, H., Nolis, J., Ooi, H., Ram, K., Ross, N., Shepherd, L., Sólymos, P., Swetnam, T. L., Turaga, N., Petegem, C. V., Williams, J., Willis, C., & Xiao, N. (2020). The Rockerverse: Packages and Applications for Containerisation with R. The R Journal, 12(1). https://doi.org/10.32614/RJ-2020-007

Nüst, D., & Hinz, M. (2019). containerit: Generating Dockerfiles for reproducible research with R. Journal of Open Source Software, 4(40), 1603. https://doi.org/10.21105/joss.01603

Nüst, D., Konkol, M., Pebesma, E., Kray, C., Schutzeichel, M., Przibytzin, H., & Lorenz, J. (2017). Opening the Publication Process with Executable Research Compendia. D-Lib Magazine, 23(1/2). https://doi.org/10.1045/january2017-nuest

Nüst, D., & Pebesma, E. (2021). Practical reproducibility in geography and geosciences. Annals of the American Association of Geographers, 111(5), 1300–1310. https://doi.org/10.1080/24694452.2020.1806028

Nüst, D., Sochat, V., Marwick, B., Eglen, S. J., Head, T., Hirst, T., & Evans, B. D. (2020). Ten simple rules for writing Dockerfiles for reproducible data science. PLOS Computational Biology, 16(11), 1–24. https://doi.org/10.1371/journal.pcbi.1008316

**Communities, incentives & policy**

Niers, T., & Nüst, D. (2020). Geospatial Metadata for Discovery in Scholarly Publishing. Septentrio Conference Series, 4. https://doi.org/10.7557/5.5590

Nüst, D., Boettiger, C., & Marwick, B. (2018). How to Read a Research Compendium. arXiv:1806.09525 [Cs]. http://arxiv.org/abs/1806.09525

Nüst, D., & Eglen, S. J. (2021). CODECHECK: An Open Science initiative for the independent execution of computations underlying research articles during peer review to improve reproducibility. F1000Research, 10, 253. https://doi.org/10.12688/f1000research.51738.1

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). Reproducible research and GIScience: An evaluation using AGILE conference papers. PeerJ, 6, e5072. https://doi.org/10.7717/peerj.5072

Nüst, D., Lohoff, L., Einfeldt, L., Gavish, N., Götza, M., Jaswal, S. T., Khalid, S., Meierkort, L., Mohr, M., Rendel, C., & Eek, A. van. (2019). Guerrilla Badges for Reproducible Geospatial Data Science. AGILE Short Papers. https://doi.org/10.31223/osf.io/xtsqh

Ostermann, F. O., Nüst, D., Granell, C., Hofer, B., & Konkol, M. (2020). Reproducible Research and GIScience: An evaluation using GIScience conference papers. EarthArXiv. https://doi.org/10.31223/X5ZK5V




Here is a full list of the publications that comprise the thesis.

WWU MÜNSTER

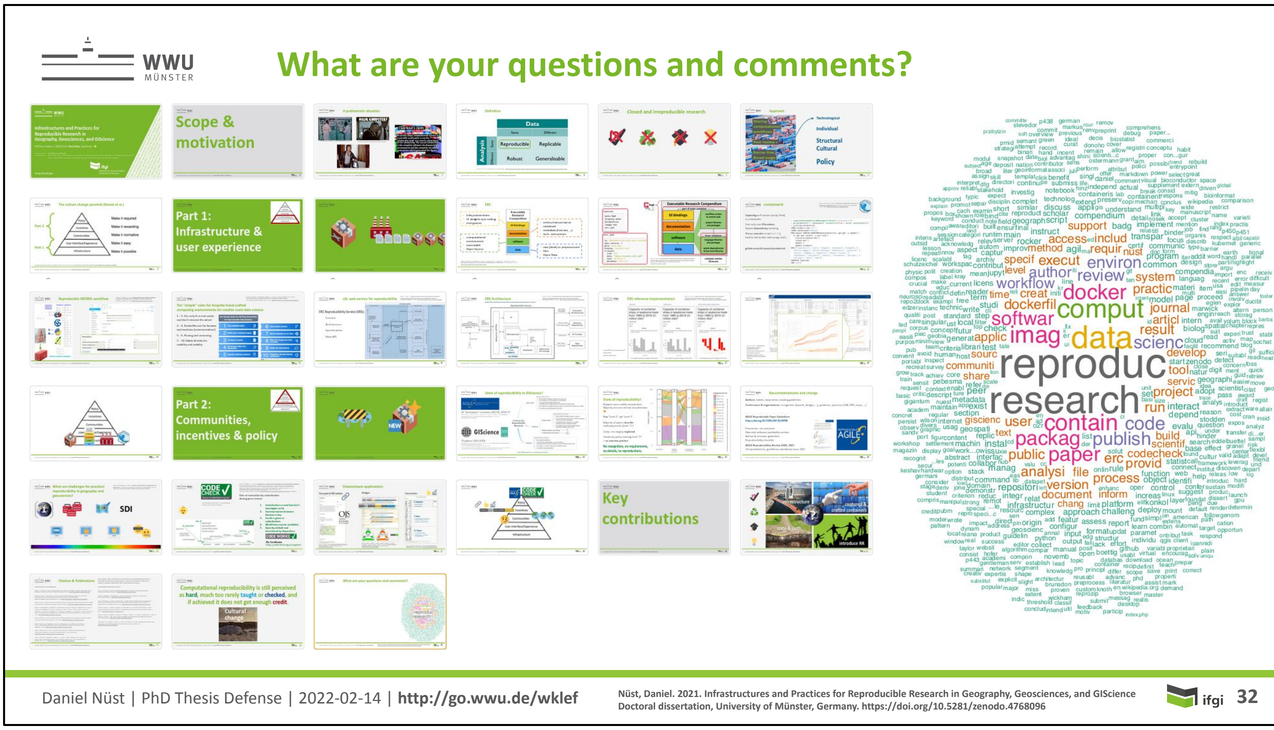
# Computational reproducibility is still perceived as hard, much too rarely taught or checked, and if achieved it does not get enough credit.




Of course, these contributions do not put an end to the fact that computational reproducibility is still perceived as hard, much too rarely taught or checked, and if achieved it does not get enough credit. 🖱 **We need to continue to work on this, but acknowledge that cultural change is slow.**

I personally hope platforms such as presented today will be adopted by researchers and journals within the next **10 years?**

One small step at a time, the way how we **share research can catch up with the methods of conducting research**.




I want to **thank** the o2r project team, especially my supervisor **Edzer Pebesma** and my colleague **Markus Konkol**, for making this research possible and fun.

In addition, I would like to **thank Ben Marwick**, who quickly agreed to serve as second assessor of my work, and **Professors Dan Sui, Chris Kray, and Norbert Hölzel** for serving on the evaluation committee, and many **friends and family** who supported me on this journey.

**What are your questions?**

--------------------------------------------

Figure 3: Word stem cloud. Contains all terms used in this document, including references, excluding select stop words (based on tidytext::stop_words and selected abbreviations, publisher names, code, markup, domains, etc.); **figure is based on 98626 word stems, minimum occurrence is manually set to 21 times; the top word stem "reproduc" occurs 1797 times [= 1,82 % of words ~ every 50th word]**.

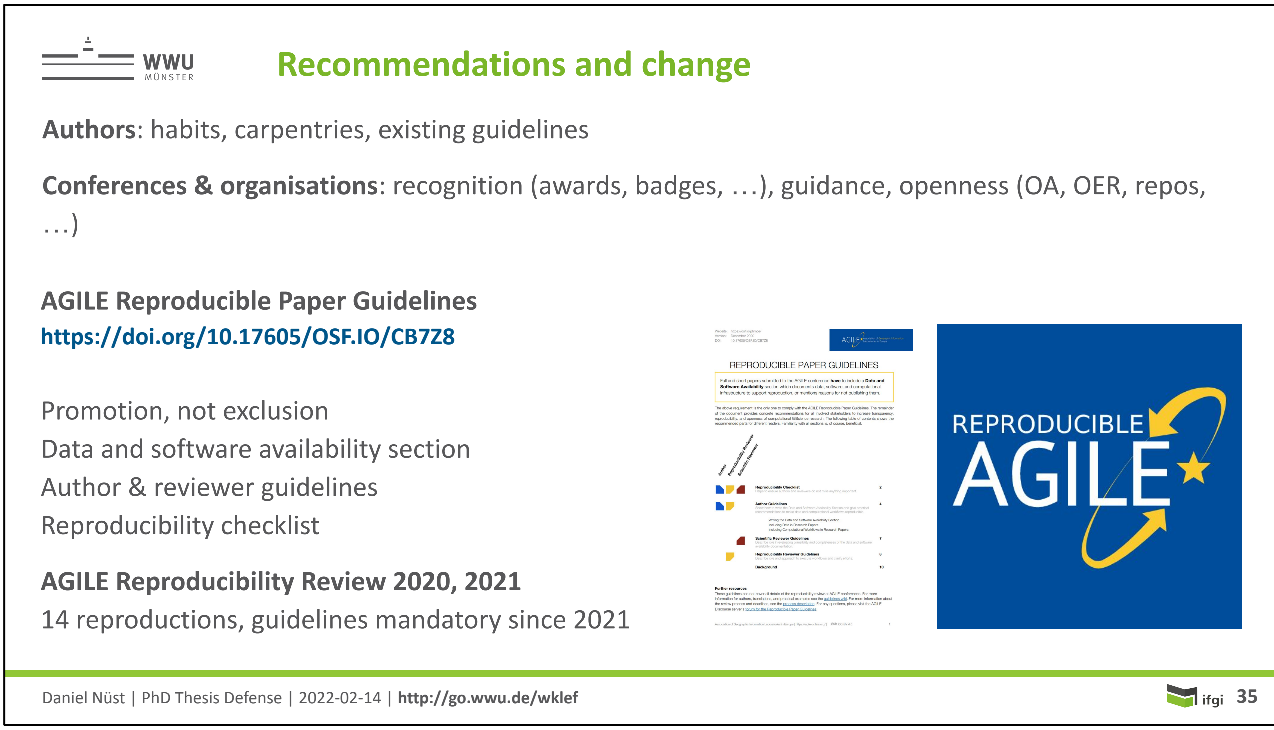
# Encore

WWU MÜNSTER

1. GI solves geospatial problems with IT, the problem: **computational reproducibility**
2. Adaptation and transfer of mainstream IT to domain: **containerisation** (little needed, future work re. maps?!)
3. ERC infrastructure & metascience are **transferable**, the direct addressing (examples, community membership) is **not** - without domain focus just a **theoretical exercise**
4. Geoinformatitians **translate and interpret** between geo-scientists and developers



In my understanding, as a field of “hyphenated informatics”, geoinformatics researches solutions to geospatial problems using information technology (IT). Geospatial problems often are the questions and issues that geographers and geoscientists face in their work. In the case of the research presented in this dissertation, the issue is reproducibility. Domain-specific solutions using IT regularly require the adaptation and transfer of novel concepts and tools from computer sciences and general IT. In this case, containerisation was applied and made available for a broader community of geospatial disciplines. Of course, the designed infrastructure could just as well be applied to other natural sciences, and the methods to understand and shift community practice are transferable metascience, too. However, directly addressing geographers, geoscientists, and GIScience researchers is needed to communicate the challenges and approaches of reproducible research successfully. Activities like research on reproducibility have a very generic component, nevertheless they are important to carry out in a specific domain so they do not remain a theoretical exercise. The communication with researchers and the need for a practical evaluation require the bridging between informatics and geo-disciplines and the consideration of barriers and opportunities across all levels of the culture change pyramid. **That requirement is met by these key contributions.**

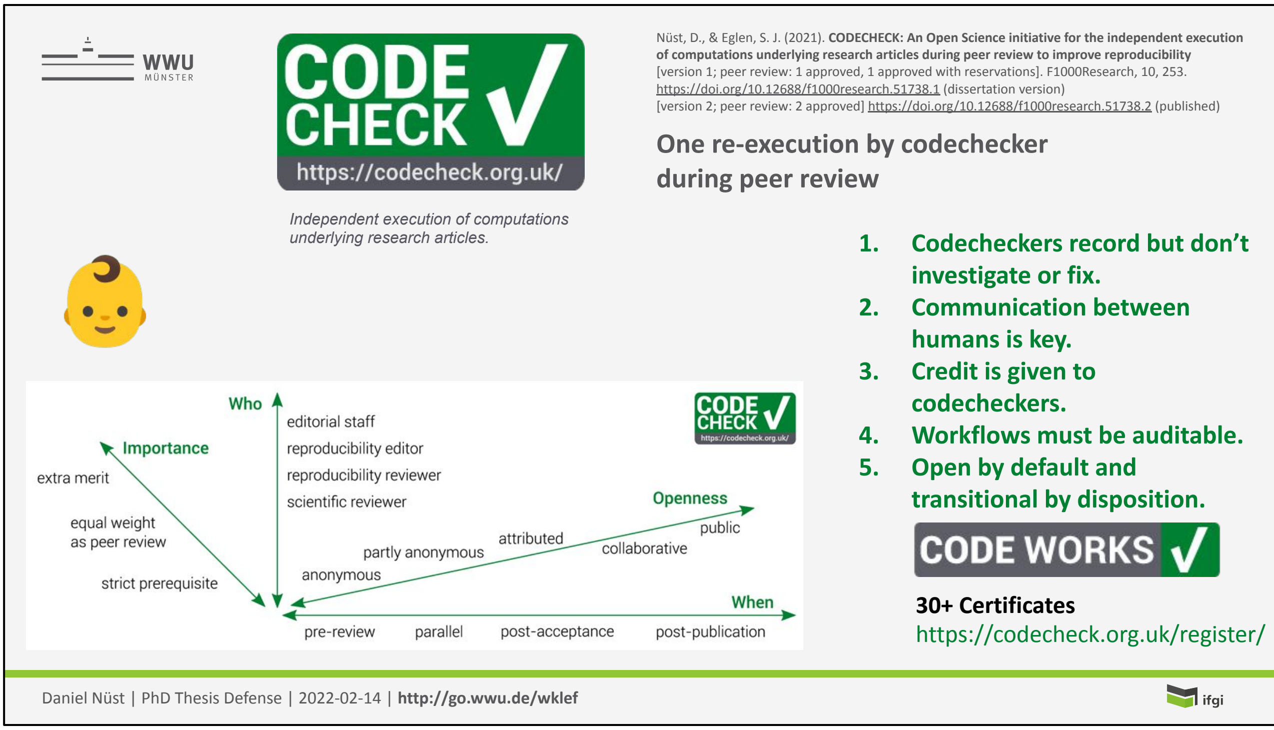



For authors, we recommend to improve habits and self-educate using existing resources.
However, the majority of the responsibility to change the dire situation lies with the **communities and their organisations**. They need to promote reproducibility and provide clear guidance on what is expected.

As a direct result from the assessment, **The Reproducible AGILE initiative**, with support from the AGILE organisation, developed the **AGILE Reproducible Paper Guidelines**. These guidelines aim to recognise challenges and promote excellence, not to exclude. All past papers that were assessed in the studies above could stay the same and still conform to the guidelines by adding a **Data and Software Availability section** that makes the (non-)availability of data and software transparent.

The initiative shows shows how quickly **practices can change**, provided that the topic finds champions in the community as well as institutional support.
Based on the guidelines, a **reproducibility review committee completed 14 successful reproductions.**

The fact that these 14 papers could be reproduced strongly suggests that **all** of them, unless data could only shared privately, would likely **reach at least a level of 2 across all categories**.
This is a clear improvement over the earlier assessment of publications.




**One suggestion for what is "good enough" comes with the CODECHECK initiative.** This chapter of my thesis is a collaboration with Stephen Eglen from the University of cambridge.

With CODECHECK, the idea is to have a special role, the codechecker, in peer review who re-executes the workflow underlying a scientific publication once as part of peer review.

The CODECHECK workflow follows **five principles** and if a reproduction is successful, the CODECHECK certificate is published, more than 30 of these certificates were created, and about half a dozen as part of actual peer reviews where the now published articles link to the certificate.

**There are many variants** how to integrate a codecheck into a publication and review process, each giving more or less relevance to the reproducibility.

**All** variants have in common that codecheck is particularly interesting for **early career researchers**. They have the skills and an interest to be introduced into peer reviewing.

**OF COURSE, codechecking becomes much more simple when researchers create an ERC!**

WWU MÜNSTER

Nüst, D., & Pebesma, E. (2021). **Practical reproducibility in geography and geosciences**. Annals of the American Association of Geographers, 111(5), 1300–1310. https://doi.org/10.1080/24694452.2020.1806028

## Practical Reproducibility in Geography and Geosciences

Review of **common** guidance for RR

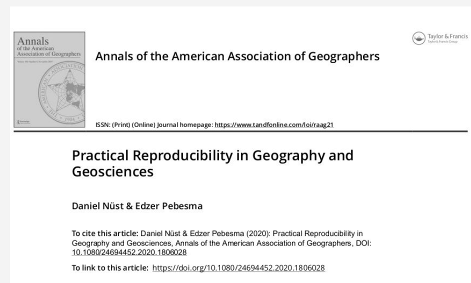
Core idea: consciously control & share computing environment

physical,

logical, and

cultural components

*Use scripts and notebooks, create research compendia.*


Annals of the American Association of Geographers

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Practical Reproducibility in Geography and Geosciences

Daniel Nüst & Edzer Pebesma

To cite this article: Daniel Nüst & Edzer Pebesma (2020): Practical Reproducibility in Geography and Geosciences, Annals of the American Association of Geographers, DOI: 10.1080/24694452.2020.1806028

To link to this article: https://doi.org/10.1080/24694452.2020.1806028



We have mentioned three core concepts for practical reproducibility: literate programming with computational notebooks, containerisation of the runtime environment, and research compendia.
In the chapter “Practical Reproducibility in Geography and Geosciences”, we extend these three logical components further components to present a generic guidance for how reproducible research can be conducted. Besides the logical components, we added documentation of the things that are not easy to share, such as physical hardware, and the cultural component of a common practice. This way, one can control and share the *full computing environment*.

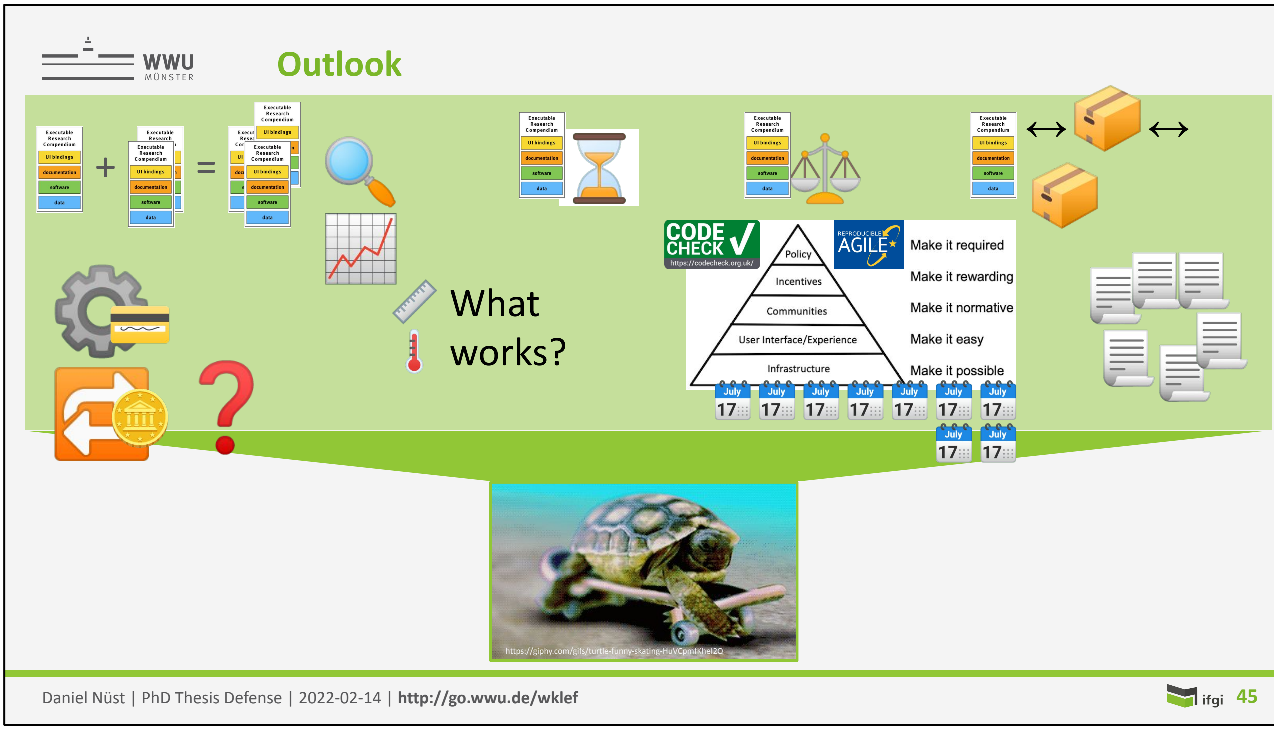
**More concretely, if you use scripts or notebooks and share research compendia you can follow the common good practices for reproducible research, also a geographer, geoscientist, or GIScience researcher.**
In terms of this foundational technology, there are no specialities to these disciplines.

**However, the creation and interaction with research compendia can of course be tedious. And just by using containers and notebooks, we can package computational research, but we don’t connect it well to the publishing infrastructure yet.**

# Outlook



So let's take a brief look into what I see as useful next steps to pursue.




ERC-based research is better from an individual's perspective, but it also facilitates strategic research building to accumulate evidence, which can more effectively deal with reproducibility challenges. 🖱 **With ERCs**, we don't just publish results but collaborate on the advancement of science. The idea of the ERC as a self-contained publishable piece of research output has not lost any appeal, and beyond the downstream applications mentioned above, one should mention adapted representations depending on the target users, e.g., for public communication, discovery solutions that take advantage of the ERC's building blocks. Furthermore, there are of course engineering challenges for the o2r platform and tools, regarding scalability for example, and what the costs of ERC-based publishing are.

Challenges around the ERC that remain are first, 🖱 **preservation**. The ERC is a complex compound. How can they be maintained, and how archived?
What are the 🖱 **legal** implications of having so many different pieces of software, text, and data in one package?

It should be a worthwhile exercise to explore 🖱 **conversions** between approaches to package computational workflows. Maybe there is a minimal common ground, so that no "winner" has to be declared, but instead the most suitable of compatible tools is used. The shared foundation of containers is a promising start for connecting, e.g., Whole Tales, ERAs, ERCs, and ReproZip packages.

Furthermore, the 🖱 **cost of reproducibility** remains unclear. Should there be direct payments by readers, or should costs be factored into publication charges? With limited resources, one might also ask how to determine what should be reproduced and what should not.

With the recent awareness on reproducibility, it is unsurprising that more and more journals and conferences consider evaluating computational workflows. Hard requirements are still far away, but publicity is given to the leading works in context. The next generation of researchers will find the sharing and evaluation of data and code much more natural. To support scientific venues in the uptake of reproducibility reviews, there needs to be a survey on 🖱 **code execution during peer review** in journals and conferences so that we learn what works, and how, and what does not work.

These technical approaches form the basis for change, education, and community interaction. But to be successful, they will have to be embedded in a broader 🖱 **cultural shift to change policies**. The CODECHECK and Reproducible AGILE initiatives have just started here but show how it can work. Longitudinal studies to observe the impact of these guidelines will provide further evidence for those that are still sceptical.

And if you need even more convincing, I’d be really interested to see 🖱 **large scale reproductions of classic foundational works across geospatial sciences, or possibly even replications**.

The topic of reproducibility gets a lot of attention at the moment, even considering I’m pretty deep in the reproducible research bubble. This is a great opportunity as individuals and community can relate to the ideals and benefits of reproducibility, possibly more easily than the even bigger challenges the academia as a whole faces. My personal 🖱 **optimistic outlook** is that conquering reproducibility will help to transform science and academia in a much broader sense and adoption of open reproducible research methods will be among the quicker shifts that will happen in the coming years.

(equity/diversity/inclusion, openness, research assessment & evaluation, metrics & incentives, predatory publishing, misinformation, involving the Global South, healthy work environments, career opportunities for software experts, publication pressure & bias, and valuing reuse over piecewise publishing)


The technological innovations and means to influence norms which were identified in this dissertation nevertheless contribute to a wider-scale adoption of reproducible research practices. While the fundamental ideas for reproducibility have been around for a long time, notably with the work on electronic documents by Claerbout & Karrenbach almost 30 years ago and with research compendia first presented in 2007 by Gentleman and Temple Lang, the modern technologies transferred to reproducibility tools and scholarly practices in this work provide a new level of accessibility and pathway for reproducible methods. However, the age of the landmark papers also shows that adoption is too slow and a technology-driven approach does not suffice, despite all individual and collective benefits, and despite the progress on Openness in academia in general.

# One thing

## Have a README: all else is details.

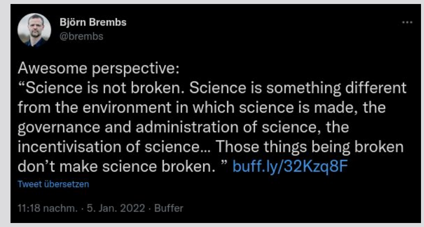
Inspired by Greg Wilson's Teching Tech Together (http://teachtogether.tech/en/index.html) Rule 1.

# Research Compendia

**research-compendium.science**

# Is there a *reproducibility crisis* in

# geography?
# geosciences?
# GIScience?




https://www.incaseofpeace.org/currencies-in-academia/

Crisis narrative useful?

**Unclear!**

A lot of work to proof on the **technical/practical** side, even not separable from general **academic crisis**?

How likely is it that these disciplines are so different from others (psychology) that there really is nothing?

## Generalisability

Yes, CODECHECK, Reproducible AGILE, and the o2r ERS should work across all disciplines.

“Geo” is already very mixed discipline, manifold methods and specialisations.
E.g., qualitative research? Make reproducible what is based on code.

Other documents? Atlases, books (Jupyter Book), …

## Limitations

**ERCs**

Big data, long workflows

User evaluation of ERC/ERS
> proof benefits

Small # workflows

Long term study

**Prototype**

Bindings environment != ERC image

Real world deployment missing
- BagIt profile + real archive

# Research Questions

## All Research Questions

### Infrastructure & user experience

1) How can packaging of computational analyses serve the needs of authors, publishers, readers, and preservationists?
2) To what extent can the process of capturing the runtime, software, data, and metadata of reproducible research packages be automated in geoscientific analyses?
3) How can the ERC fit into the existing practices and infrastructure for research and publishing in geography, geosciences, and GIScience?

### Communities, incentives & policy

1) What are domain-specific challenges and solutions for the geography, geosciences, and GIScience domains in the context of reproducible publications?
2) What new services and features can be built upon reproducible workflows, e.g., when packaged as an ERC?

WWU MÜNSTER

**How can packaging of computational analyses serve the needs of authors, publishers, readers, and preservationists?**

**To what extent can the process of capturing the runtime, software, data, and metadata of reproducible research packages be automated in geoscientific analyses?**

- ✍️ efficiency in (self-)collaboration, understandability = persuading/accessibility
- 🏛️ Innovative leadership position, more interesting product, costs unclear; ERC: adaptable and flexible
- 📖 Reduced barriers to understanding/evaluation, extend workflows, become collaborator
- 🧑‍💻 Assume completeness, ERC (plain text, meaningful links & entrypoints; snapshot with consistent packaging = one preservation strategy)

**Automation** of ERC creation for large majority of workflows *starting from a notebook or fully scripted workflow* is possible **with containerisation**

Capturing **large data** and **HPC** environments challenging

**Manual alternative** important for researcher freedom

**Manual checks** for crucial metadata needed



First of all, we could stop worrying about automation if instead we could establish the required skills and change education. But that’s even harder and more long term than the technical solutions I just presented.

In short, automated packaging using containers is possible, though large datasets and non-Desktop environments are open challenges.

Manual alternatives will remain, not the least to allow innovation and ensure research freedom.

Manual checks by humans, for example for metadata, are a reasonable complement to the automation.

## How can the ERC fit into the existing practices and infrastructure for research and publishing in geography, geosciences, and GIScience?

**ERS** can make the ERC the **unit of publication**, interacting with existing open or even closed services, **no duplication of services**

ERC can alleviate the issues of **procedural and cultural shift** in publication practices

**Notebooks** are established practice for reproducibility

**Missing** (for ERC): private data (solutions exist),
huge data (under development),
commercial support
uptake/investment by publishers

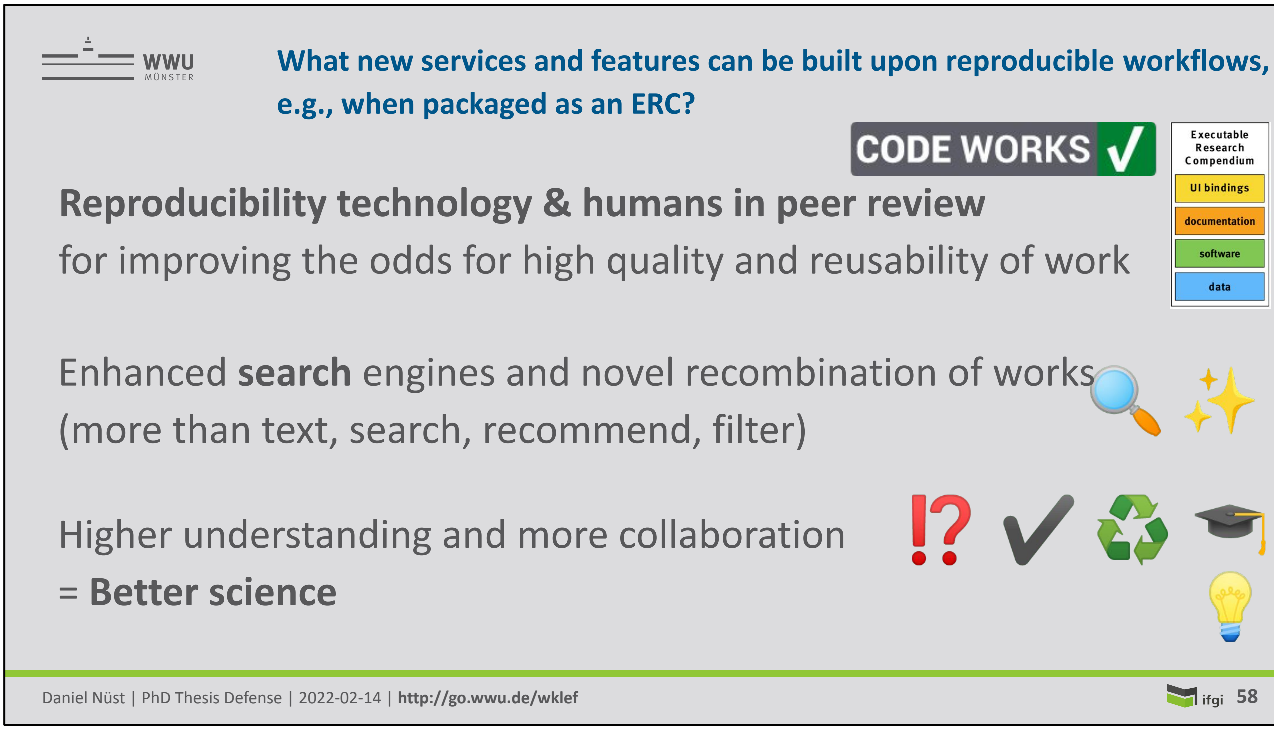
**Unique ERS**: full openness (spec, impl), substitutions, bindings

**What are domain-specific *challenges* and *solutions* for the geography, geosciences, and GIScience domains in the context of reproducible publications?**

👨‍💻 👩‍💻 Technical: generally none, some at concrete level
All recommendations draw from other disciplines

👍 🪧 Incentives & policy: same as academia at a whole?
CAN change within our communities

👨‍👨‍👧‍👧 👩‍👩‍👧‍👧 Communities: have a lot of technical literacy




First, we presented the ERC itself and a **platform for ERC creation and inspection**. The interactivity and connection to other services in scholarly communication are only possible with ERCs, and the means for authors to convince reviewers, and reviewers to evaluate submissions, are considerable. The balance between technological solutions and **adopting new ways to communicate** is deeply embedded in the culture change pyramid. This balance is reflected by the inclusion of both the innovative, technology-driven approach for scientific commons based on ERCs and the "less is more" approach that highlights the **human interaction** and community practices with CODECHECK and Reproducible AGILE, which are enabled by reproducible workflows and embrace reproducible workflows.

We also looked into downstream applications that demonstrate advantages of applications based on sharing research works in the form of (E)RCs. As just briefly summarised in the previous slide, there are **enhancements for search and discovery of scientific outputs**. With ERCs, we can look deeper into these products and provide more relevant well-defined information, such as data about executability of workflows. We can unearth relevant information otherwise hidden in the text.

And finally, one "feature" would be **the novel ways that scientists as**

**readers can engage with a piece of published research**. As research compendia gain more traction, readers have more ways to interact, use, or extend other people's works. This will lead to higher understanding and hopefully less reinventing of wheels but more collaboration. That provides the service of "better science", and can fix the issues with closed and irreproducible work mention at the beginning of this presentation.

# Platforms

# Reproducibility Platform Landscape

Konkol, M., Nüst, D., & Goulier, L. (2020). **Publishing computational research - a review of infrastructures for reproducible and transparent scholarly communication**. Research Integrity and Peer Review, 5(1), 10. https://doi.org/10.1186/s41073-020-00095-y

**Table 1** Overview of applications we included in the analysis

| Application | Description |
|---|---|
| Authorea | In Authorea, authors can create executable papers collaboratively. They can attach code and data to figures to make them reproducible. Authors can also directly submit to a journal and, at the same time, publish a preprint. |
| Binder | Binder creates a containerized executable environment based on a repository (e.g., on GitHub/Lab, Zenodo) including a Jupyter Notebook [24]. Readers can launch the analysis and inspect the workflow in a browser. |
| Code Ocean | Code Ocean creates "capsules" containing code, data, and the computational environment. While reading, users can execute and inspect the analysis in a separate window below the article or on Code Ocean's website [25]. |
| eLife Reproducible Document Stack (RDS) | RDS originates from the life sciences. Authors can publish executable documents based on Stencila (https://stenci.la/), an open-source editor for articles. The executable document, which contains the whole narrative and executable code snippets, is not only a supplement but the actual scientific article. |
| Galaxy | Galaxy [26] provides features tailored to use cases in the life sciences. It is a web app for developing comput. Analyses without programming expertise. Scientists can upload and analyze data using Jupyter Notebooks [27]. |
| Gigantum | Gigantum packages code, data, the computational environment, and the work history into a Git repository. Gigantum is composed of a client app for creating as well as executing analyses locally and a cloud-based infrastructure for sharing computations and collaborating with peers. |
| Manuscripts | Manuscripts is an online tool for writing executable documents collaboratively based on the concept of literate programming, but featuring a "What you see is what you get" user interface. The runtime environment of the author is, however, not considered. |
| o2r | o2r [22] originates from the geosciences and addresses publishers who want to extend their infrastructure via a reproducibility service during the process of paper submission [28]. Authors can create interactive figures, allowing readers to change model parameters using a slider [29]. |
| REANA | REANA [4, 30] originates from particle physics and provides a specification for capturing data, code, and the comput. Environment. Based on this structure and manually created configuration files, REANA provides command line interface (CLI) commands to run large analyses on a remote REANA cloud. |
| ReproZip | ReproZip [31, 32] provides a set of CLI commands for encapsulating data, code, and the computational environment. Users can execute the resulting bundle on a server provided by ReproZip [33] or locally on different systems. |
| Whole Tale | With Whole Tale [34], authors can create so-called "Tales" that combine narrative, data, code, and the computational environment. Readers can inspect the materials and execute the analysis in the original environment. |

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

| | 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 | + | + | + | + | + | + | + | + | – | + | – |



We analysed the existing platforms and tools in a review papers. The gist of is that almost everybody uses literate programming and containerisation, and the ERC remains unique in his abilities to define specific manipulation options with bindings and with a mechanism for substituting specific parts of one compendium for re-combination and re-use.

Also, the level of complexity and automation between these platforms varies greatly, and the o2r ERC service is certainly on the side of high automation ease of use for domain scientists with limited programming experience.

But automation only goes so far - and at the beginning of my work, I collaborated with colleagues to get first hand experiences in manually packaging a workflow, which was crucial to design and implement the ERC and the o2r reproducibility service. Let's take a brief look at this early paper.

# Rockerverse

WWU MÜNSTER

# Rockerverse

**I) Packaging research reproducibly benefits from other use cases applying containerisation**

**II) Usability vs. Diversity vs. Stability vs. Uptake vs. Innovation vs. Funding**

Interfaces for Docker in R
Image stacks for communities of practice
Capture and create environments
Development, debugging, and testing
Processing
Deployment and continuous delivery
Using R to power enterprise software in production environments
Common or public work environments
Teaching
Packaging research reproducibly

| Functionality | AzureContainers | babelwhale | dockermachine | dockyard | googleCloudRunner | harbor | stevedore |
|---|---|---|---|---|---|---|---|
| Generate a Dockerfile | | | | ✓ | | | |
| Build an image | ✓ | | | ✓ | ✓ | | |
| Execute a container locally or remotely | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Deploy or manage instances in the cloud | ✓ | | ✓ | | ✓ | ✓ | ✓ |
| Interact with an instance (e.g., file transfer) | | ✓ | ✓ | | | | ✓ |
| Manage storage of images | | | | | | ✓ | ✓ |
| Supports Docker and Singularity | | ✓ | | | | | |
| Direct access to Docker API instead of using the CLI | | | | | | | ✓ |
| Installing Docker software | | | ✓ | | | | |

Nüst, D., Eddelbuettel, D., Bennett, D., Cannoodt, R., Clark, D., Daróczi, G., Edmondson, M., Fay, C., Hughes, E., Kjeldgaard, L., Lopp, S., Marwick, B., Nolis, H., Nolis, J., Ooi, H., Ram, K., Ross, N., Shepherd, L., Sólymos, P., Swetnam, T. L., Turaga, N., Petegem, C. V., Williams, J., Willis, C., & Xiao, N. (2020). **The Rockerverse: Packages and Applications for Containerisation with R**. The R Journal, 12(1). https://doi.org/10.32614/RJ-2020-007



We coined this the “Rockerverse”.

I won’t go into any detail here, but just point out two findings:

1. This overview article shows that we need to note is that reproducible environments is **the** common goal of numerous and very diverse applications. The use case to share computing environments for research purposes can benefit from the more remote applications.
2. The multi-dimensional requirements that applications face lead to multitude of packages and tools, which develop and disappear over time. To find resources for consolidation, for a shared foundation, and for maintenance remains the challenge.

# GEOBIA

# Reproducible GEOBIA workflow

Knoth, C., & Nüst, D. (2017). **Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers**. Remote Sensing, 9(3), 290. https://doi.org/10.3390/rs9030290

1) `docker run` starts a container and executes the entry point script `/qgis/model.sh` using a Bash shell
2) `/qgis/model.sh` ...
   a) copies model and script files
      from `/workspace/models/*` to `/root/.qgis2/processing/models`
      from `/workspace/scripts/*` to `/root/.qgis2/processing/sripts`
   b) executes `model.py` as a Python file with a virtual frame buffer
3) `/workspace/model.py` ...
   a) initiates QGIS application
   b) loads manipulation parameters and construct input and output paths
   c) runs the model `example_analysis_linux_v3.1.model` using the QGIS Python API passing configuration parameters
4) `/root/.qgis/processing/models/example_analysis_linux_v3.1.model` ...
   a) executes the model steps, using user scripts from `/root/.qgis/processing/scripts`
   b) saves the files to the result directory
5) `/results` holds the output files for user access

WWU MÜNSTER

# GEOBIA

Knoth, C., & Nüst, D. (2017). **Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers**. Remote Sensing, 9(3), 290. https://doi.org/10.3390/rs9030290

Listing 1: Excerpt of workspace directory tree; the full workspace is available on GitHub [62] and in the reproducibility package, see Section 3.4.

```
/workspace
|-- data
| |-- COPYRIGHT
| |-- jonjona_pos_conflict_proj.tif
| `-- jonjona_pre_conflict_proj.tif
|-- model.py
|-- models
| |-- detect_settlements_on_edgelayer.model
| `-- example_analysis_linux_v3.1.model
`-- scripts
  |-- diff_to_local_ref_v1.3.py
  `-- kmeans_clustering_v2.3.py
```

Listing 6: Full reproduction commands: run the container from Docker Hub and extract the result.

```
docker run -it --name repro nuest/qgis-model:rs-jonjona
docker cp repro:/workspace/results /tmp/repro_results
```

Listing 7: Result directory tree after execution, supplementary shapefile files, i.e., .dbf, .prj, .qpj, and .shx, and workspace files (see previous Listing 1) not shown.

```
|/result
|`-- 20161212-172947
|  |-- result_threshold.shp
|  |-- result_unclassified.shp
|  |-- settlements.shp
```

Listing 8: Analysis control and data switching examples. From top to bottom: (a) mounting another workspace; (b) mounting only input files; (c) changing model options via environment variables.

```
# (a)
docker run -it -v /my/analysis:/workspace nuest/qgis-model:rs-jonjona

# (b)
docker run -it -v mypreconflict.tif:/workspace/data/pre_conflict.tif
  -v mypostconflict.tif:/workspace/data/pos_conflict.tif nuest/qgis-model:rs-jonjona

# (c)
docker run -it -e change_analysis_threshold=0.28 nuest/qgis-model:rs-jonjona
```

## GEOBIA

Knoth, C., & Nüst, D. (2017). **Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers**. Remote Sensing, 9(3), 290. https://doi.org/10.3390/rs9030290

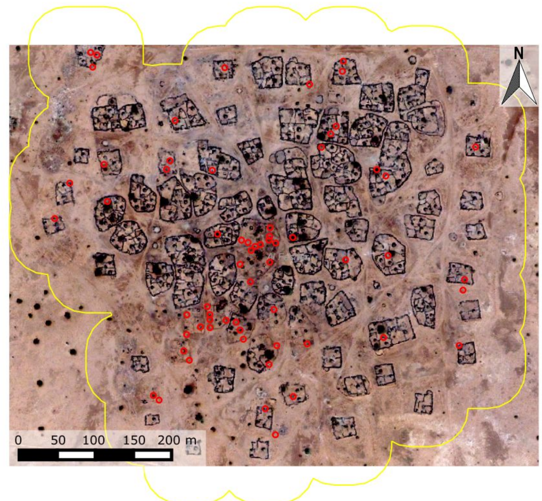



**Figure 4.** Post-conflict image (location: 13.686°N, 24.979°E) with two results of the example analysis. The detected settlement area is the yellow polygon. The results of the damage assessment, i.e., the disappeared dwellings, are the red circles (image © 2016 DigitalGlobe).

# AGILE conference

# AGILE paper corpus

**Table 1 Reproducibility-related keywords in the corpus, ordered by sum of matches per paper.** For full references of the corpus papers see Supplemental Material.

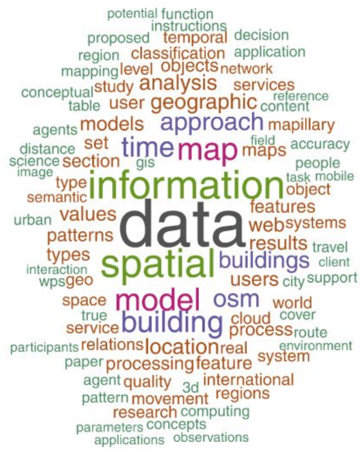
| Citation | Reproduc. | Replic. | Repeatab. | Code | Software | Algorithm(s) | (pre)process. | Data | Result(s) | All |
|---|---|---|---|---|---|---|---|---|---|---|
| Foerster et al. (2012) | 0 | 0 | 0 | 2 | 3 | 11 | 140 | 129 | 41 | 326 |
| Wiemann & Bernard (2014) | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 98 | 3 | 123 |
| Mazimpaka & Timpf (2015) | 0 | 0 | 0 | 3 | 0 | 4 | 4 | 97 | 10 | 118 |
| Steuer et al. (2015) | 0 | 0 | 0 | 0 | 0 | 25 | 12 | 64 | 17 | 118 |
| Schäffer et al. (2010) | 0 | 0 | 0 | 0 | 10 | 1 | 26 | 65 | 6 | 108 |
| Rosser et al. (2016) | 0 | 0 | 0 | 0 | 2 | 1 | 42 | 51 | 6 | 105 |
| Gröchening et al. (2014) | 0 | 0 | 0 | 0 | 0 | 3 | 2 | 69 | 27 | 101 |
| Almer et al. (2016) | 0 | 0 | 0 | 1 | 1 | 1 | 22 | 53 | 22 | 100 |
| Magalhães et al. (2012) | 0 | 0 | 0 | 2 | 1 | 20 | 52 | 9 | 1 | 85 |
| Juhász & Hochmair (2016) | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 55 | 11 | 70 |
| Wiemann (2016) | 0 | 0 | 0 | 0 | 3 | 0 | 8 | 55 | 1 | 69 |
| Fan et al. (2014) | 0 | 0 | 0 | 0 | 0 | 3 | 8 | 44 | 12 | 67 |
| Merki & Laube (2012) | 0 | 0 | 0 | 0 | 0 | 9 | 6 | 40 | 6 | 62 |
| Zhu et al. (2017) | 2 | 2 | 0 | 2 | 0 | 10 | 7 | 32 | 6 | 61 |
| Kuhn & Ballatore (2015) | 0 | 0 | 1 | 2 | 14 | 1 | 5 | 26 | 8 | 58 |
| Soleymani et al. (2014) | 1 | 0 | 0 | 0 | 0 | 0 | 4 | 39 | 9 | 56 |
| Fogliaroni & Hobel (2015) | 0 | 0 | 0 | 0 | 0 | 3 | 14 | 30 | 5 | 52 |
| Osaragi & Hoshino (2012) | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 36 | 7 | 48 |
| Stein & Schlieder (2013) | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 42 | 3 | 48 |
| Körner et al. (2010) | 0 | 0 | 0 | 0 | 0 | 6 | 5 | 30 | 4 | 45 |
| Knoth et al. (2017) | 0 | 0 | 0 | 3 | 2 | 1 | 6 | 25 | 7 | 44 |
| Raubal & Winter (2010) | 0 | 0 | 0 | 1 | 1 | 1 | 18 | 0 | 13 | 34 |
| Konkol et al. (2017) | 1 | 0 | 0 | 3 | 1 | 1 | 2 | 4 | 19 | 31 |
| Kiefer et al. (2012) | 1 | 0 | 0 | 0 | 2 | 1 | 9 | 10 | 8 | 31 |
| Haumann et al. (2017) | 0 | 0 | 0 | 0 | 0 | 6 | 8 | 10 | 2 | 26 |
| Josselin et al. (2016) | 0 | 0 | 0 | 0 | 2 | 1 | 9 | 5 | 8 | 25 |
| Heinz & Schlieder (2015) | 1 | 0 | 0 | 2 | 1 | 3 | 2 | 14 | 2 | 25 |
| Osaragi & Tsuda (2013) | 0 | 0 | 0 | 1 | 1 | 0 | 3 | 16 | 2 | 23 |
| Baglatzi & Kuhn (2013) | 1 | 0 | 0 | 0 | 0 | 0 | 6 | 12 | 3 | 22 |
| Scheider et al. (2014) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 13 | 4 | 19 |
| Brinkhoff (2017) | 0 | 0 | 0 | 0 | 1 | 9 | 2 | 3 | 2 | 17 |
| Schwering et al. (2013) | 0 | 0 | 0 | 0 | 0 | 4 | 2 | 3 | 5 | 14 |
| **Total** | 7 | 2 | 1 | 22 | 47 | 126 | 454 | 1,179 | 280 | 2,131 |

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). **Reproducible research and GIScience: An evaluation using AGILE conference papers**. PeerJ, 6, e5072. https://doi.org/10.7717/peerj.5072

WWU MÜNSTER

# AGILE paper corpus

A




B

| place | word | n | # papers |
|---|---|---|---|
| 1 | data | 1058 | 31 |
| 2 | information | 589 | 32 |
| 3 | spatial | 577 | 30 |
| 4 | map | 411 | 25 |
| 5 | model | 411 | 25 |
| 6 | building | 381 | 24 |
| 7 | time | 378 | 30 |
| 8 | approach | 297 | 32 |
| 9 | osm | 292 | 8 |
| 10 | buildings | 266 | 15 |
| 11 | geographic | 249 | 28 |
| 12 | location | 239 | 26 |
| 13 | analysis | 229 | 28 |
| 14 | users | 225 | 19 |
| 15 | results | 207 | 30 |
| 16 | web | 206 | 21 |
| 17 | models | 202 | 20 |
| 18 | values | 202 | 23 |
| 19 | patterns | 196 | 16 |
| 20 | maps | 189 | 20 |

**Figure 1** Two illustrations of the test corpus papers: word cloud, scaled and coloured by number of occurrence of words with at least 100 occurrences (96 unique words) (A); top words sorted by overall occurrence and number of papers including the word at least once (B).

Full-size DOI: 10.7717/peerj.5072/fig-1

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). **Reproducible research and GIScience: An evaluation using AGILE conference papers**. PeerJ, 6, e5072. https://doi.org/10.7717/peerj.5072

WWU MÜNSTER

# AGILE paper corpus levels

Table 3 **Reproducibility levels for paper corpus; '-' is category not available.** For full references of the corpus papers see Supplemental Material.

| Author | Short paper | Input data | Preprocessing | Method/analysis/ processing | Computational environment | Results |
|---|---|---|---|---|---|---|
| Zhu et al. (2017) | | 0 | 1 | 1 | 1 | 1 |
| Knoth et al. (2017) | | 0 | – | 0 | 1 | 1 |
| Konkol et al. (2017) | | 2 | 2 | 1 | 1 | 1 |
| Haumann et al. (2017) | X | 0 | 1 | 1 | 0 | 1 |
| Brinkhoff (2017) | X | 0 | – | 1 | 0 | 0 |
| Almer et al. (2016) | | 0 | – | 1 | 1 | 1 |
| Wiemann (2016) | | 2 | – | 1 | 1 | 1 |
| Juhász & Hochmair (2016) | | 0 | 1 | 1 | 0 | 0 |
| Josselin et al. (2016) | X | 1 | – | 0 | 0 | 1 |
| Rosser et al. (2016) | X | 0 | – | 1 | 0 | 0 |
| Kuhn & Ballatore (2015) | | – | – | – | – | – |
| Mazimpaka & Timpf (2015) | | 2 | 1 | 1 | 1 | 1 |
| Steuer et al. (2015) | | 2 | 0 | 1 | 1 | 1 |
| Fogliaroni & Hobel (2015) | X | – | – | – | – | – |
| Heinz & Schlieder (2015) | X | 0 | 0 | 1 | 1 | 1 |
| Scheider et al. (2014) | | 1 | 1 | 2 | 1 | 1 |
| Gröchening et al. (2014) | | 2 | 0 | 1 | 0 | 1 |
| Fan et al. (2014) | | 0 | 1 | 1 | 0 | 1 |
| Soleymani et al. (2014) | X | 0 | 0 | 1 | 0 | 0 |
| Wiemann & Bernard (2014) | X | 0 | 0 | 1 | 0 | 0 |
| Osaragi & Tsuda (2013) | | 0 | 1 | 1 | 0 | 1 |
| Baglatzi & Kuhn (2013) | | – | – | – | – | – |
| Li et al. (2013) | X | 0 | 0 | 1 | – | 1 |
| Stein & Schlieder (2013) | X | 0 | – | 1 | 0 | 1 |
| Osaragi & Hoshino (2012) | | 0 | 0 | 1 | 0 | 1 |
| Magalhães et al. (2012) | | 0 | 0 | 1 | 0 | 0 |
| Foerster et al. (2012) | | 1 | – | 1 | 1 | 1 |
| Merki & Laube (2012) | X | 0 | – | 1 | 1 | 1 |
| Kiefer et al. (2012) | X | 0 | 1 | 1 | 0 | 1 |
| Raubal & Winter (2010) | | – | – | – | – | – |
| Schäffer et al. (2010) | | 0 | 0 | 1 | 1 | 1 |
| Körner et al. (2010) | | – | – | – | – | – |

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). **Reproducible research and GIScience: An evaluation using AGILE conference papers**. PeerJ, 6, e5072. https://doi.org/10.7717/peerj.5072

## AGILE assessment results

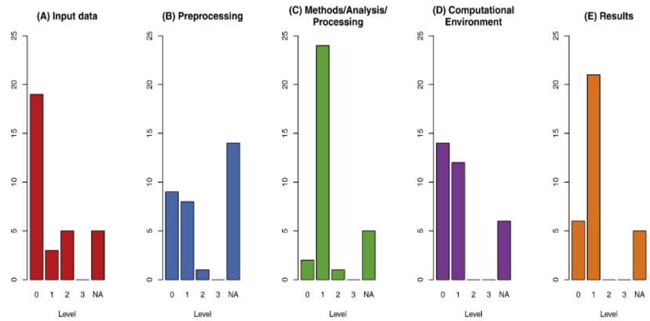



**Figure 3** **Results of reproducibility assessment across all categories for the assessment of reproducibility: *Data* (A), *Methods* with sub-categories preprocessing (B), method/analysis/processing (C) and computational environment (D), and *Results* (E).** The level of reproducibility ranges from 0 (not reproducible) to 3 (fully reproducible); NAs include 5 conceptual papers (all categories are NA).
Full-size DOI: 10.7717/peerj.5072/fig-3

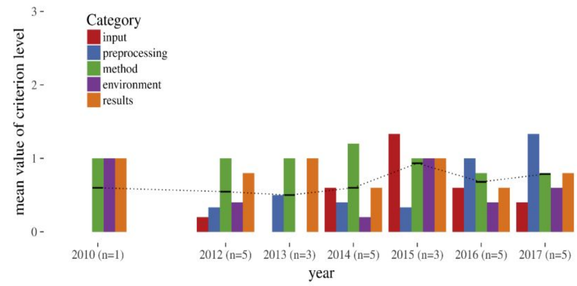



**Figure 4** **Mean reproducibility levels per category over time; black dotted line connects the mean per year over all categories (in 2010 only one of three papers could be assessed, reaching level 1 for methods).**
Full-size DOI: 10.7717/peerj.5072/fig-4

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). **Reproducible research and GIScience: An evaluation using AGILE conference papers**. PeerJ, 6, e5072. https://doi.org/10.7717/peerj.5072

WWU MÜNSTER

# AGILE survey results

**Table 6** Hindering circumstances for reproducibility for each survey response ($n = 17$) sorted by barrier type for the category with most "Main reason" occurences; each line is one response and background colour corresponds to cell text.

| Legal restrictions | Lack of time | Lack of tools | Lack of knowledge | Lack of incentive |
|---|---|---|---|---|
| Main reason | Strongly hindered | Not at all | Not at all | Strongly hindered |
| Main reason | Not at all | Not at all | Not at all | Moderately hindered |
| Main reason | Slightly hindered | Strongly hindered | Moderately hindered | Strongly hindered |
| Main reason | Not at all | Slightly hindered | Not at all | Not at all |
| Strongly hindered | Strongly hindered | Strongly hindered | Moderately hindered | Strongly hindered |
| Moderately hindered | Main reason | Not at all | Not at all | Not at all |
| Slightly hindered | Moderately hindered | Slightly hindered | Slightly hindered | Moderately hindered |
| Slightly hindered | Not at all | Main reason | Strongly hindered | Not at all |
| Not at all | Moderately hindered | Not at all | Moderately hindered | Not at all |
| Not at all | Strongly hindered | Strongly hindered | Strongly hindered | Slightly hindered |
| Not at all | Moderately hindered | Not at all | Not at all | Not at all |
| Not at all | Slightly hindered | Main reason | Not at all | Strongly hindered |
| Not at all | Main reason | Not at all | Not at all | Not at all |
| Not at all | Main reason | Not at all | Not at all | Not at all |
| Not at all | Moderately hindered | Moderately hindered | Not at all | Strongly hindered |
| Not at all | Not at all | Not at all | Not at all | Not at all |
| Not at all | Slightly hindered | Not at all | Slightly hindered | Not at all |

Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018).
**Reproducible research and GIScience: An evaluation using AGILE conference papers**.
PeerJ, 6, e5072. https://doi.org/10.7717/peerj.5072



Legal restrictions has most "main reasons", but lack of time has most "not at all", and the latter is the strong enough to overshadow all others

# GIScience conference

# GIScience assessment results

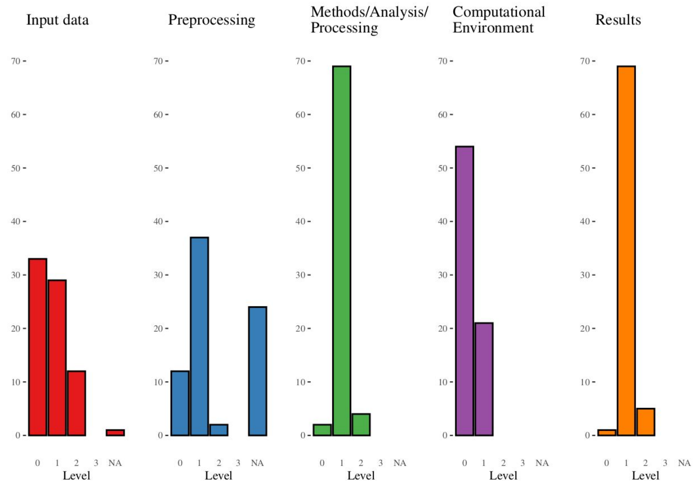



**Figure 1** Barplots of reproducibility assessment results; levels range from 0 (leftmost bar) to 'not applicable' (rightmost bar).

Ostermann, F. O., Nüst, D., Granell, C., Hofer, B., & Konkol, M. (2020). **Reproducible Research and GIScience: An evaluation using GIScience conference papers**. EarthArXiv. https://doi.org/10.31223/X5ZK5V (dissertation version)
> 11th International Conference on Geographic Information Science (GIScience 2021) - Part II. Schloss Dagstuhl - Leibniz-Zentrum Für Informatik. https://doi.org/10.4230/LIPICS.GISCIENCE.2021.II.2 (accepted)

# GIScience assessment results

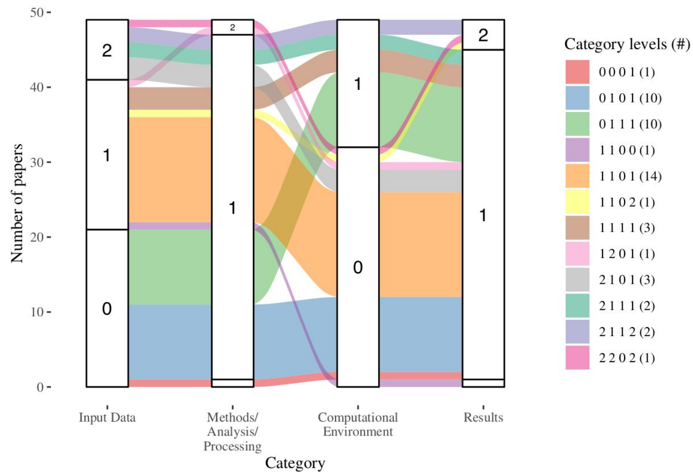
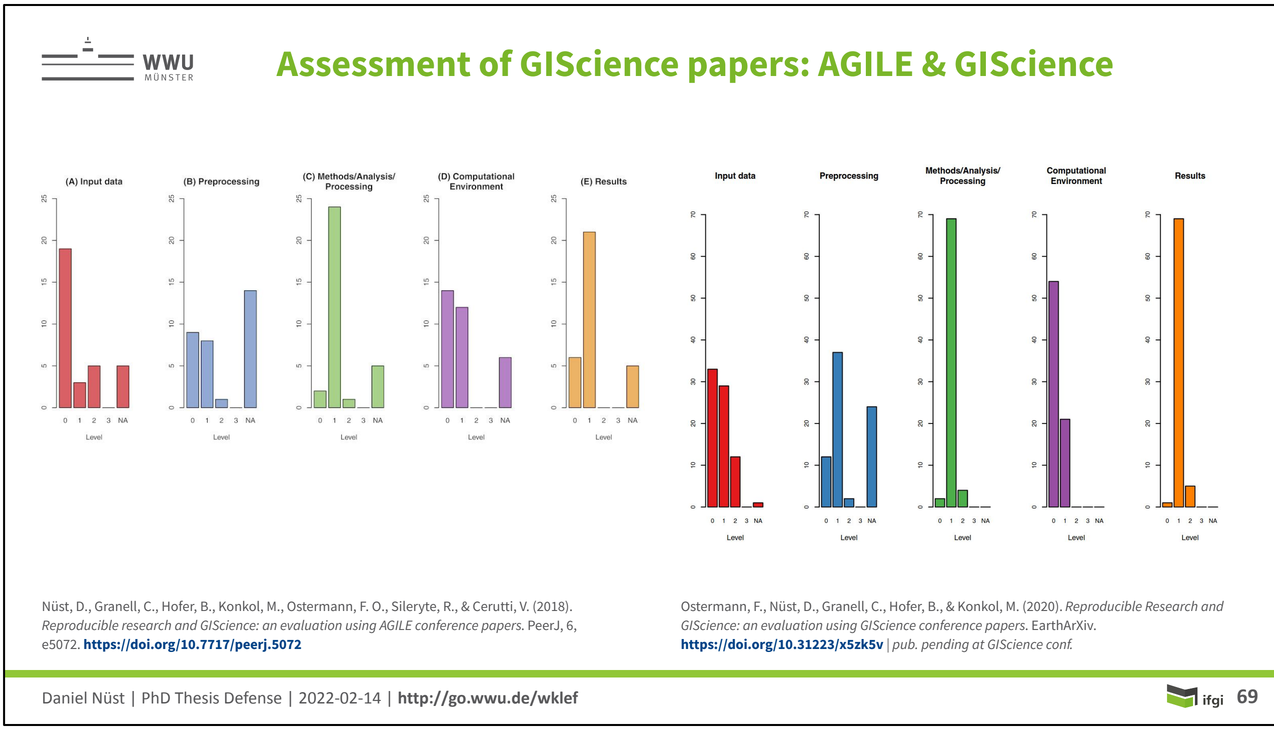



**Figure 2** Alluvial diagram of common groups of papers throughout 4 of 5 categories including only papers without any "not applicable" *(Level NA)* value; category *Preprocessing* was dropped because difficulty to clearly assess it lead to many "not applicable" values.


Ostermann, F. O., Nüst, D., Granell, C., Hofer, B., & Konkol, M. (2020). **Reproducible Research and GIScience: An evaluation using GIScience conference papers**. EarthArXiv. https://doi.org/10.31223/X5ZK5V (dissertation version)
> 11th International Conference on Geographic Information Science (GIScience 2021) - Part II. Schloss Dagstuhl - Leibniz-Zentrum Für Informatik. https://doi.org/10.4230/LIPICS.GISCIENCE.2021.II.2 (accepted)




We assessed over 100 papers of the two major community led conferences in the field of GIScience using a rubric of five categories and with four levels. None of the papers, even just looking at the information provided, would have been reproducible at the point of publication (without extended communication with the authors). In case of the AGILE conference, we also surveyed authors for the reasons, and the major barriers were legal restrictions and lack of time, but also lack of tools, knowledge and incentives.

We used the same rubric again on GIScience conference papers, with similarly dire results. Again, not a single paper reached level three in any category: available and deposited openly and permanently. In the data you see spikes at level 1 for methods and results, which basically means that we found the articles to be understandable and reasonable - the common bar to pass peer review across all disciplines, journals, and conferences.

*So how can we get to levels of 2 and higher, which means that data and methods are actually available, even if not permanently with a DOI?*

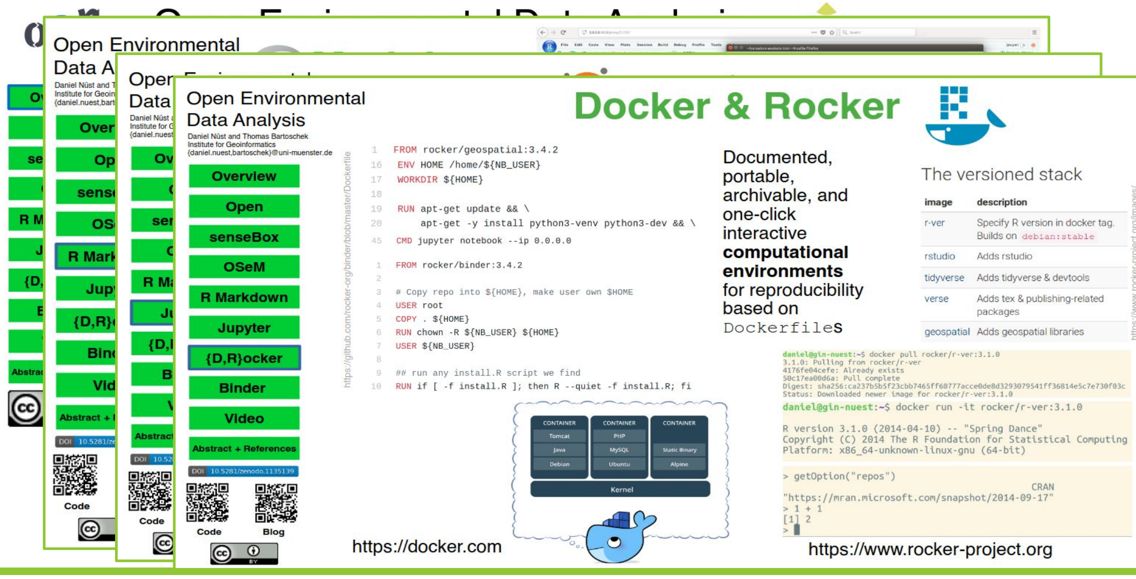
# Posters

WWU MÜNSTER

Nüst, D., & Schutzeichel, M. (2017). **An Architecture for Reproducible Computational Geosciences**. 20th AGILE Conference for Geoinformation Science poster session. **https://doi.org/10.5281/zenodo.1478542**

# EGU 2018 Poster

Nüst, D., & Bartoschek, T. (2018). Open Environmental Data Analysis. Geophysical Research Abstracts, Vol. 20; archived on Zenodo; ERC: https://o2r.uni-muenster.de/erc/Phbla. https://doi.org/10.5281/zenodo.1217912



# BSc & Msc theses

https://o2r.info/theses/

# Developing and Evaluating Infrastructure for ERC to Communicate with Data Repositories and Computing Services

Niklas George, 2021, MSc

Open up ERCs to allow controlled access to specific computing or data services; expert interviews and prototyping approaches (proxy, **DNS**, firewall) for Docker container/networks.

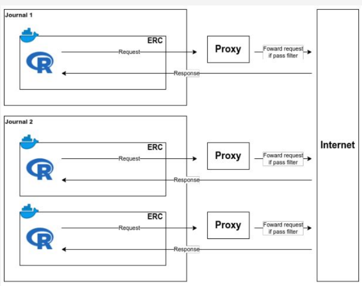


Figure 3.2: Concept for using a proxy server. Important: Every ERC uses the same proxy server.

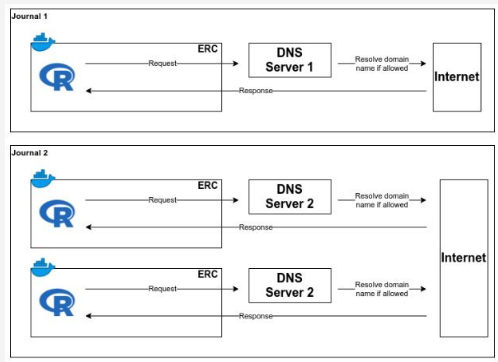


Figure 3.3: Concept for using a DNS server.

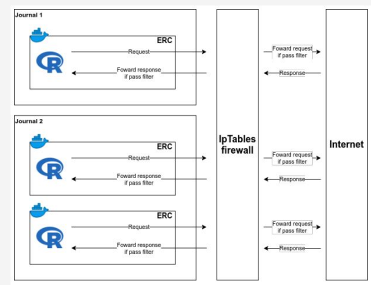


Figure 3.4: Concept for using IpTables.

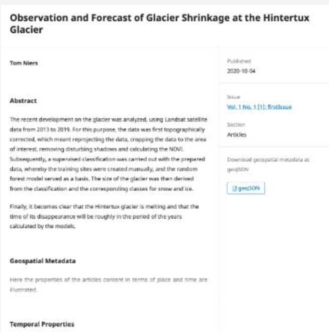
# Geospatial Metadata for Discovery in Scholarly Publishing

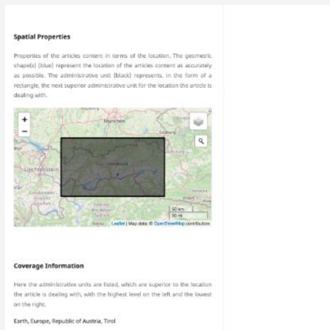
Tom Niers, 2020, BSc

**http://nbn-resolving.de/urn:nbn:de:hbz:6-69029469735**

**https://github.com/tnier01/geoOJS**

Geospatial metadata for articles as part of OJS; innovative matching of text and coordinate metadata.




# Testing Geospatial R Packages on Implementations of the R language and Platforms

Ismail Sunni, 2020, MSc GeoTech

**http://hdl.handle.net/10362/95140 | https://github.com/ismailsunni/altRnative/**

Using containers for R implementations (GNU R, MRO, Renjin, FastR, pqR, TERR) across several distributions (Debian, Fedora, Ubuntu, Arch), including benchmarking; challenging installations!

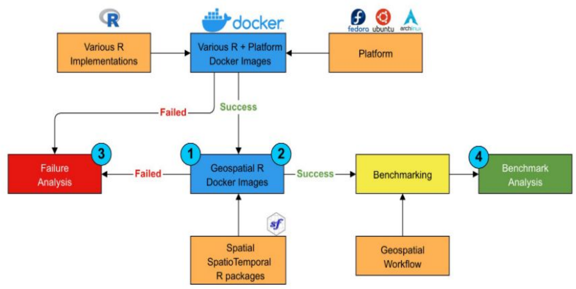



Figure 3.1: Methodology

Table 4.1: Vanilla R Docker Images

| | GNU R | MRO | Renjin | FastR | pqR | TERR |
|---|---|---|---|---|---|---|
| **Debian** | Yes | Yes | Yes | Yes | Yes | Yes |
| **Fedora** | Yes | Yes | Stop | Yes | Stop | Stop |
| **Arch Linux** | Yes | Yes | Stop | Stop | Stop | Stop |

Table 4.2: Geospatial R Docker Images

| | GNU R | MRO | Renjin | FastR | pqR | TERR |
|---|---|---|---|---|---|---|
| **Debian** | Yes | Yes | No | No | No | No |
| **Fedora** | Yes | Yes | No | No | No | No |
| **Arch Linux** | Yes | ? | No | No | No | No |

# Testing Geospatial R Packages on Implementations of the R language and Platforms (cont.)

Ismail Sunni, 2020, MSc GeoTech

**http://hdl.handle.net/10362/95140 | https://github.com/ismailsunni/altRnative/**

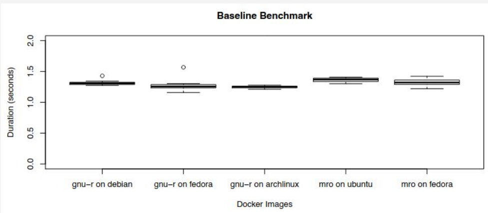



Figure 4.1: Baseline Benchmark

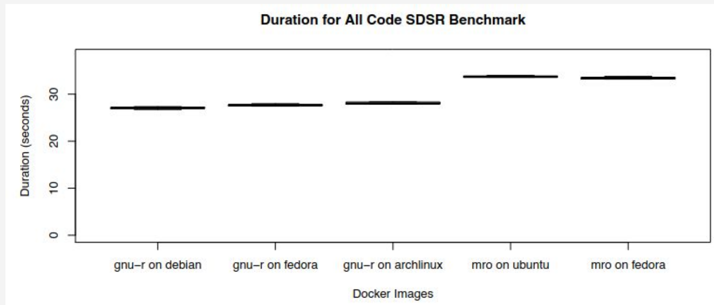



Figure 4.2: Duration for All Code SDSR Benchmark

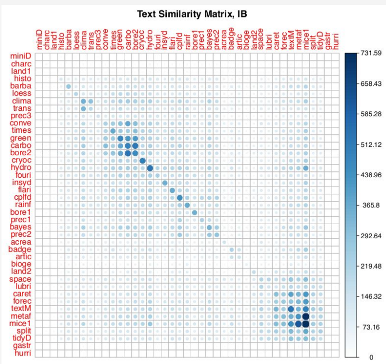


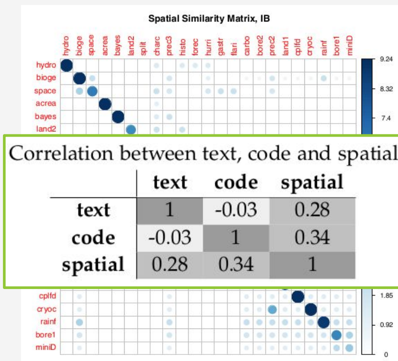
Baseline benchmark = run R and calculate 1+1

# Similarity Measurements for Executable Research Compendia

Lukas Lohoff, 2018, MSc

**Use components of an ERC, code and geospatial metadata, to enhance search, i.e., find spatially (Geohash + Text similarity) and computationally similar works (e.g., loaded libraries)**




Correlation between text, code and spatial similarity

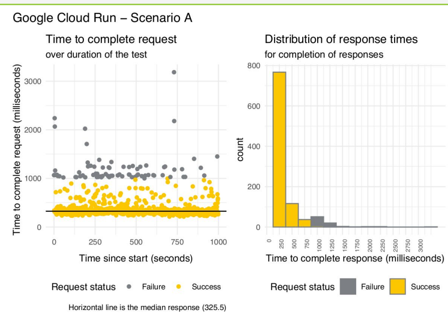
| | text | code | spatial |
|---|---|---|---|
| **text** | 1 | -0.03 | 0.28 |
| **code** | -0.03 | 1 | 0.34 |
| **spatial** | 0.28 | 0.34 | 1 |

# Serverless GEO Labels for the Semantic Sensor Web

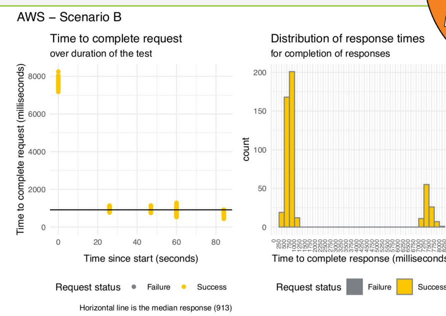
Graupner, A., & Nüst, D. (2020). **Serverless GEO Labels for the Semantic Sensor Web**.
Schloss Dagstuhl - Leibniz-Zentrum Für Informatik. https://doi.org/10.4230/LIPICS.GISCIENCE.2021.I.4

Scalable generation of meaningful and rich metadata visualisations with labels (GIScience '20)


**> Adaptable to ERC badges!**

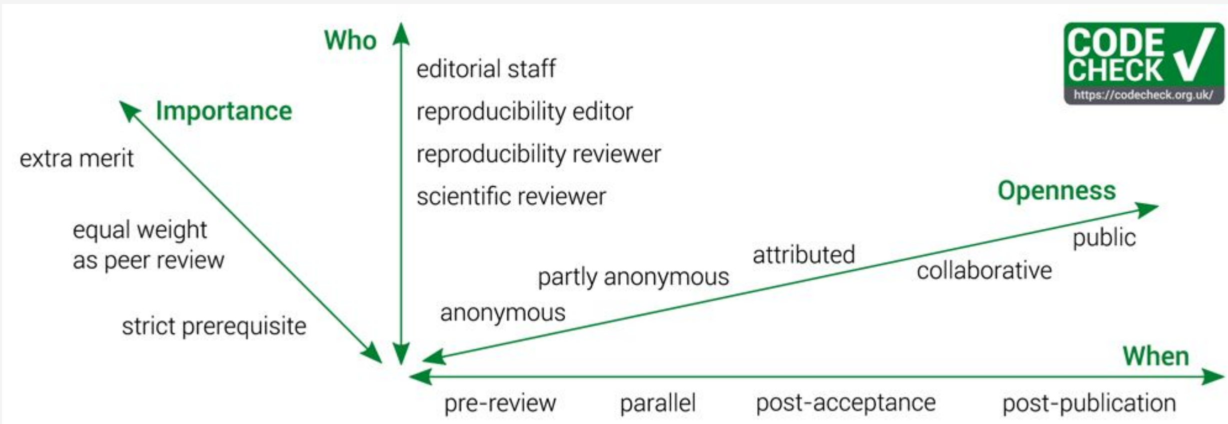



■ **Figure 2** Plots of response time for GCR deployment under scenario A: elapsed time to complete requests (left); histrogram with distribution of response times (right); result data file: `GCR_Scenario_2_V1`.




■ **Figure 5** Plots of response time for AWS deployment in scenario A: elapsed time to complete requests (left); histrogram with distribution of response times (right); result data file: `AWS_Scenario_3_V2`.

# CODECHECK


Who
editorial staff
reproducibility editor
reproducibility reviewer
scientific reviewer
Importance
extra merit
equal weight
as peer review
strict prerequisite
Openness
anonymous
partly anonymous
attributed
collaborative
public
When
pre-review
parallel
post-acceptance
post-publication
CODE CHECK
https://codecheck.org.uk/

# CODECHECK Experiences

*Independent execution of computations underlying research articles.*

30+ certificates

Several journals, one conference > ongoing contacts

25+ codecheckers signed up, 1 check from not core team member

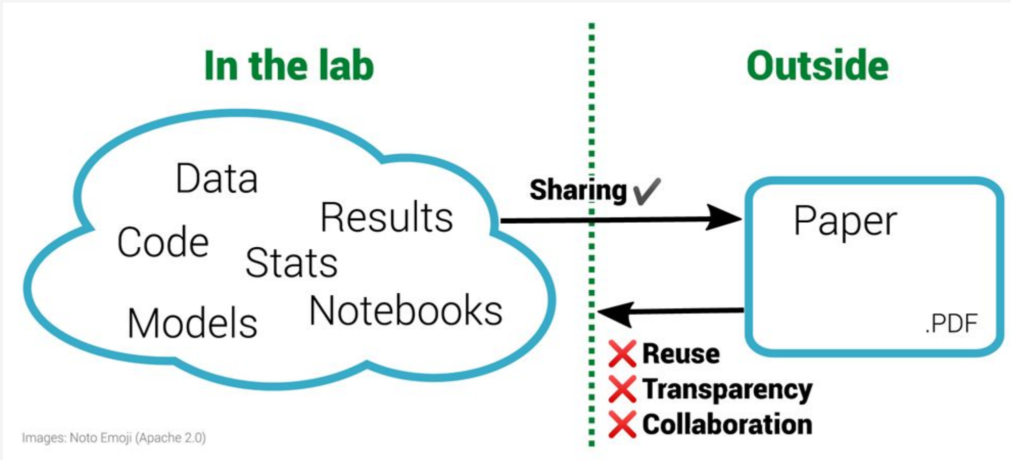
**Next**

Mentoring + practical experiences (ReproHack collaboration?), funding! (codechecks for diamond OA journals?)

CODECHECK + R2S2 @ ITC: https://www.itc.nl/research/open-science/codecheck/

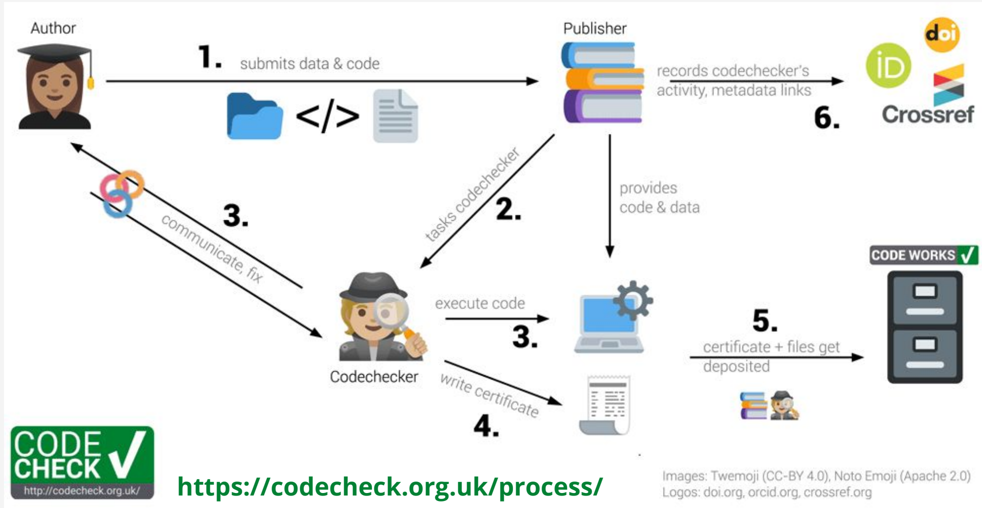
WWU MÜNSTER

## CODECHECK: The inverse problem in research




The inverse problem in reproducible research. Figure 1 of https://doi.org/10.12688/f1000research.51738.1

The left half of the diagram shows a diverse range of materials used within a laboratory. These materials are often then condensed for sharing with the outside world via the research paper, a static PDF document. Working backwards from the PDF to the underlying materials is impossible. This prohibits reuse and is not only non-transparent for a specific paper but is also ineffective for science as a whole. By sharing the materials on the left, others outside the lab can enhance this work.




**The CODECHECK example process implementation.** Figure 2 of https://doi.org/10.12688/f1000research.51738.1

The left half of the diagram shows a diverse range of materials used within a laboratory. These materials are often then condensed for sharing with the outside world via the research paper, a static PDF document. Working backwards from the PDF to the underlying materials is impossible. This prohibits reuse and is not only non-transparent for a specific paper but is also ineffective for science as a whole. By sharing the materials on the left, others outside the lab can enhance this work.

# Reproducible AGILE

REPRODUCIBLE AGILE

https://reproducible-agile.github.io/

2017, ‘18 & ‘19: Workshops on reproducibility

2019: Reproducible publications at AGILE conferences (initiative)

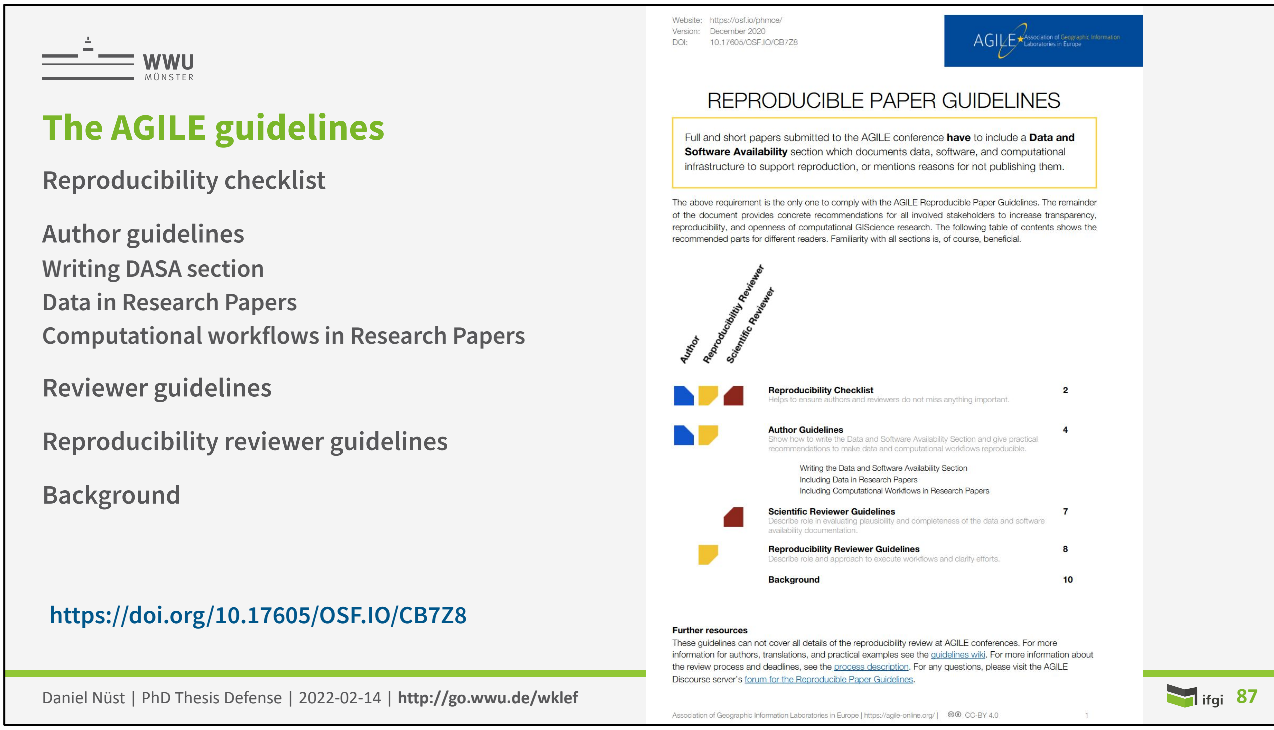
2020: AGILE Reproducible Paper Guidelines v1

2020: First AGILE reproducibility review



Two workshops in 2017 and 2018 and a piece of meta-research into the reproducibility of papers published at the AGILE conference were the basis for a small group from the AGILE community to apply for an initiative to develop guidelines for reproducible publications. The AGILE council granted support for the initiative, and the guidelines were published shortly before the conference in 2019.
In 2020, the guidelines were for the first time recommended in the call for papers, and for the first time a reproducibility committee was put in place to evaluate the reproducibility of accepted papers. Find out more at the Reproducible AGILE website.

Today, I will introduce the Reproducible Paper Guidelines. Then, I’ll talk about the reproducibility review process at the last conference in detail and try to summarise the results and lessons learned in a way that makes them useful to other communities.




These are the contents of the guidelines: Primarily, they include AUTHOR GUIDELINES. These give advice how to incorporate DATA and COMPUTATIONAL WORKFLOWS IN RESEARCH PAPERS, and how to WRITE THE DATA AND SOFTWARE AVAILABILITY SECTION section. There also is a pre-submission checklist for authors.
The document also includes some background on the rationale and motivation, as well as reviewer guidelines and, brand new in the current draft, *reproducibility reviewer guidelines*.

# Checklist & Writing the DASA section

REPRODUCIBLE AGILE

## REPRODUCIBILITY CHECKLIST

For all **datasets** included/produced in the paper, check if data:

- ❏ Is provided in a non-proprietary format
- ❏ Is documented for third parties to reuse
- ❏ Is accessible in a public repository and has an open data licence

For all **software tools/libraries/packages** and **computational workflows** included/produced, check if:

- ❏ Reproduction steps are explained in a README (plain text file), flowchart, or script
- ❏ Computational environments (including hardware) are documented or provided
- ❏ Versions of relevant software components (libraries, packages) are provided
- ❏ All parameters and expected execution times for the computational workflow are provided
- ❏ Software developed by the authors is available in a public repository and has an open licence
- ❏ There is a clear connection between **tables, figures, maps, and statistical values** and the data and code that they are based on, e.g., using file names or documentation in the README

In the **Data and Software Availability section**, check if you include:

- ❏ Data and software statements (see examples below)
- ❏ The reasons, if any, for not being able to share (parts of) data or code

For all **data and software** check that:

- ❏ All datasets and code (used or mentioned) are assigned DOIs
- ❏ Datasets and code are cited throughout the paper

After acceptance in the **camera-ready paper** check that:

- ❏ If data has been shared privately or anonymously for peer review, they are updated with all metadata and accessible via a DOI and referenced from the paper
- ❏ If a reproducibility review report will be published for your paper, a DOI URL in the Data and Software Availability section is included using the following template:
  *A reproducibility report for this paper is available confirming that [considerable parts of the computational workflow / all results / Figures 1 and 4] could be independently reproduced, see https://doi.org/link_to_report.*

## WRITING THE DATA AND SOFTWARE AVAILABILITY SECTION

The DASA section provides references to where data, software and documentation is available (e.g., paper section or README file) and under what conditions (e.g., copyright, licenses or access procedures for protected data). It should be concise and contain persistent links to repositories using Digital Object Identifiers[7] (DOI). You may remove links for anonymity during peer review ("xxx"), or share anonymized links[8] if your repository supports them. Data, software and (third-party) tools should be cited following recommended citation or standard citation guidelines. Possible statements for the DASA section are provided below. You may include one of these statements or draft your own.

***Statements for non-computational or conceptual work***

*No data or code was collected, developed, or used in this work.*

*The full list of reviewed literature is available at* [link to attachment or citable deposit of bibliography].

*The full concept maps are available at* [link] *and the ideas were first sketched in a blog post at* [link].

***Research data/code supporting this publication ...***

*… is available in* [name of the repository(-ies)] *and is accessible via the following DOI* [DOI link(s)]

*… was accessed on* [date of dataset access/download] *with the following* [query parameters, if applicable] *under the license* [dataset license].

*… was downloaded manually using the services at* [name of organisation] *(using a departmental subscription for costs) and* [name of organisation]. *The compiled dataset cannot be redistributed due to licensing restrictions.*

*...is not available due to* [indicate reasons, e.g., licenses, sensitive data on human subjects, privacy statements; if there are processes to obtain the data, describe them].

***The computational workflow supporting this publication …***

*… is executed via* [choose, e.g., a single command/file, a workflow management software, a set of numbered scripts] *published under license* [the license] *at* [DOI of repository].

*… is published in a* [language] *module/package at* [link of software project]. *The used version is archived at* [DOI of repository].

*… is provided as a* [container/VM] *published at* [DOI of repository] *with instructions included in the file README.md in the repository.*

https://doi.org/10.17605/OSF.IO/CB7Z8



The bare minimum for authors to look at and know!

WWU MÜNSTER

## The guidelines for data

"What if…"

Examples

INCLUDING DATA IN RESEARCH PAPERS

| | Minimum requirements | Recommended practices |
|---|---|---|
| **What?** | • All input data and configuration<br>• Data description/documentation, including provenance, field or column types, etc.<br>• If data is retrieved from an external source, documentation on collection queries and download steps | • Standardised, discipline-specific metadata[8] and ontologies to describe your data<br>• Data download scripts |
| **Where?** | • Publish data in a public repository providing a DOI<br>• Cite data (including date and version) in the paper | • Discipline- or data type-specific repository[9]<br>• Include recommended citation in dataset description (unless already provided by repository)<br>• Create a registration for OSF projects[10] and use the DOI to cite it |
| **How?** | • Use open data formats; export from proprietary format for publication<br>• Specify the license | • Use plain text-based file formats |

REPRODUCIBLE AGILE

https://doi.org/10.17605/OSF.IO/CB7Z8



Here are the guidelines for data in a nutshell: Data must be (a) published (b) in a data repository (DOI) and (c) using open data formats.
Ideally the provided metadata and the used repository are discipline specific (e.g., PANGAEA for measurement data) and FAIR guidelines are followed.
The guidelines also include a few "What if…" statements to answer commonly faces concerns and questions of authors, and some examples, with more examles in the Wiki.

WWU MÜNSTER

# The guidelines for computational workflows

REPRODUCIBLE AGILE

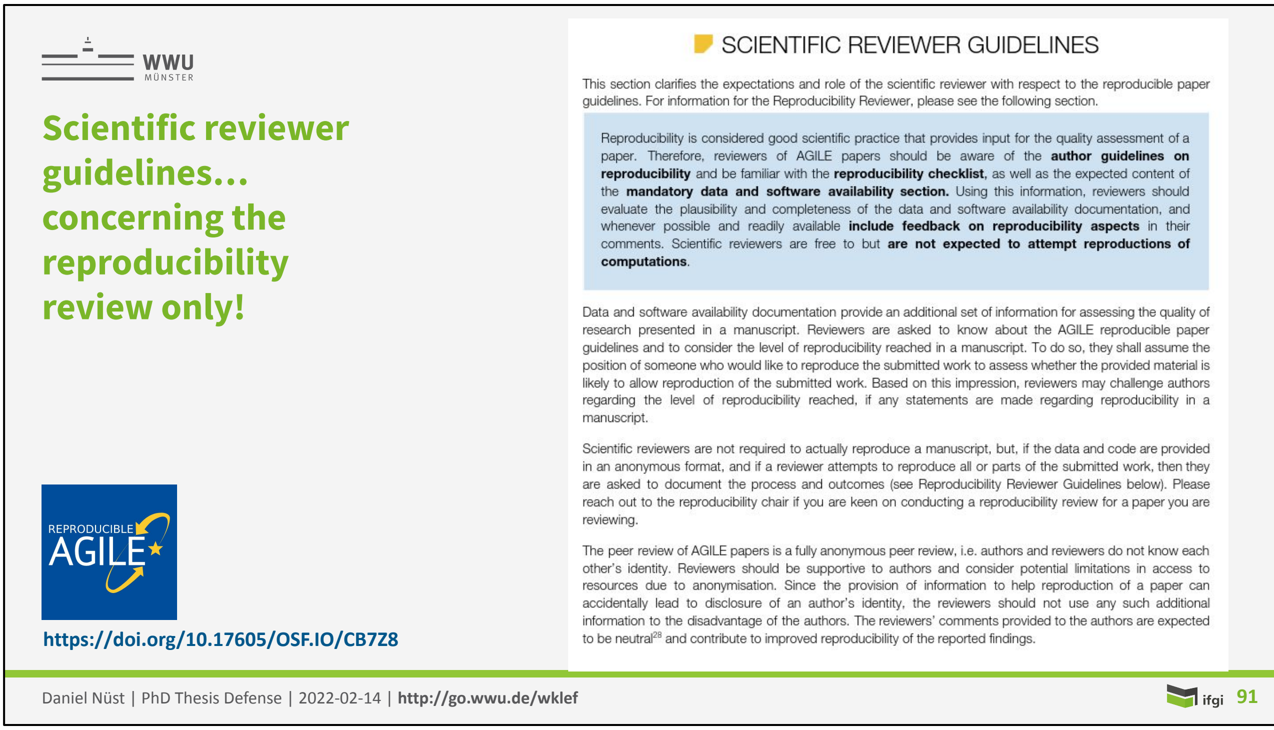
https://doi.org/10.17605/OSF.IO/CB7Z8

INCLUDING COMPUTATIONAL WORKFLOWS IN RESEARCH PAPERS

| | | Minimum requirements | Recommended practices |
|---|---|---|---|
| **What?** | Computational environment | • Describe the used environment and computational infrastructure, e.g., hardware specs, operating system<br>• List software versions<br>• Cite used software[14] | • Provide the actual environment, e.g., a Dockerfile + container[15] or a Virtual Machine (e.g., using OSGeo-Live)<br>• Provide a pinned freeze of your dependencies (structured configuration files with dependency information)<br>• Add a colophon or "reproducibility receipt"[16] to your notebooks<br>• Installation and execution instructions for different operating systems |
| | Computation steps | • Document the detailed steps in a text file and/or flowchart (every action/click)<br>• Document expected execution times given computing power unless negligible<br>• Ask a colleague to try out the instructions | • Scripts/models and a README file that explains their use<br>• All figures are fully scripted and a peer has read your README's instructions (incl. interactive visualisations and interactive adjustments<br>• Multi-panel plots are composited with scripts[17]<br>• Software package with structured metadata[18], tests/CI[19], and a pipeline framework[20] or workflow language[21]<br>• Live documents for analyses, e.g., Binder[22]<br>• Live demo of APIs/online applications (e.g., anonymous cloud resources, such as Google Cloud Run or AWS)<br>• Subset or a synthetic dataset for quick evaluation |
| **Where?** | | • Repository providing a persistent identifier, e.g., a DOI or SWHID[23] | • Versioned code repository, such as GitHub or GitLab, and ongoing open development |
| **How?** | Tools used | • Use generally available tools (avoid proprietary tools that are not available to reviewers and other researchers) | • Use and create Open Source tools<br>• Cite core modules/tools/language used |
| | Development practices | • Use clear licenses[24] that fit your environment<br>• Follow one of "Good enough practices in scientific computing"[25] | • Follow all "Good enough practices.." Use development guidelines for your environment / language of choice (e.g., for R[26]) |



Here are the guidelines for workflows:
**Minimum** boils down to having a README that describes the generally available software and hardware used, with files published with a DOI.
**Intermediate** adds some structure and requires a scripted workflows.
**Ideal** increases the requirements on the environment (e.g., container, VM), structured metadata, code repository (open development, not just snapshot), open source tools, proper software citations, and good software practices.
The guidelines also include a examples, with more examples in the Wiki.




Here are the guidelines for workflows:
**Minimum** boils down to having a README that describes the generally available software and hardware used, with files published with a DOI.
**Intermediate** adds some structure and requires a scripted workflows.
**Ideal** increases the requirements on the environment (e.g., container, VM), structured metadata, code repository (open development, not just snapshot), open source tools, proper software citations, and good software practices.
The guidelines also include a examples, with more examples in the Wiki.

![WWU Münster logo]

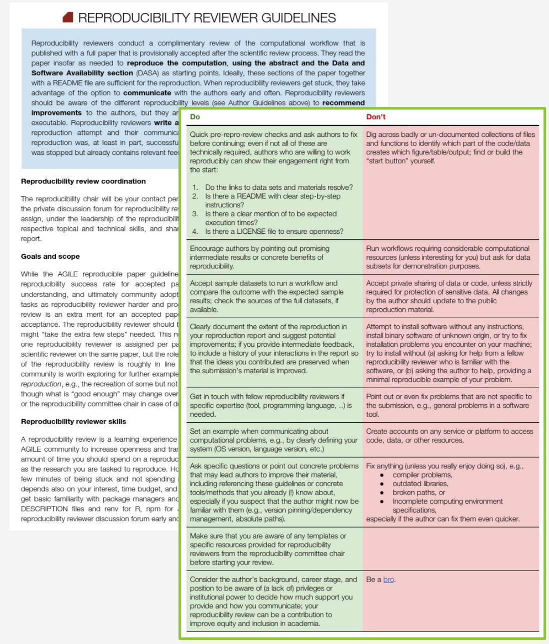
## The guidelines for reproducibility reviewers

Ideal vs. realistic

Role & skills

Examples for “Do’s and Don’ts”:

- Do shift burden to author
- Do encourage and s
- Private data/code sharing last resort
- Document your work in report (impact)
- Be kind (career stage, knowledge, privileges)
- No rummaging



REPRODUCIBILITY REVIEWER GUIDELINES

Reproducibility reviewers conduct a complimentary review of the computational workflow that is published with a full paper that is provisionally accepted after the scientific review process. They read the paper insofar as needed to **reproduce the computation**, **using the abstract and the Data and Software Availability section** (DASA) as starting points. Ideally, these sections of the paper together with a README file are sufficient for the reproduction. When reproducibility reviewers get stuck, they take advantage of the option to **communicate** with the authors early and often. Reproducibility reviewers should be aware of the different reproducibility levels (see Author Guidelines above) to **recommend improvements** to the authors, but they ar

**Reproducibility review coordination**

**Goals and scope**

**Reproducibility reviewer skills**

| Do | Don’t |
|---|---|
| Quick pre-repro-review checks and ask authors to fix before continuing; even if not all of these are technically required, authors who are willing to work reproducibly can show their engagement right from the start:<br>1. Do the links to data sets and materials resolve?<br>2. Is there a README with clear step-by-step instructions?<br>3. Is there a clear mention of to be expected execution times?<br>4. Is there a LICENSE file to ensure openness? | Dig across badly or un-documented collections of files and functions to identify which part of the code/data creates which figure/table/output; find or build the “start button” yourself. |
| Encourage authors by pointing out promising intermediate results or concrete benefits of reproducibility. | Run workflows requiring considerable computational resources (unless interesting for you) but ask for data subsets for demonstration purposes. |
| Accept sample datasets to run a workflow and compare the outcome with the expected sample results; check the sources of the full datasets, if available. | Accept private sharing of data or code, unless strictly required for protection of sensitive data. All changes by the author should update to the public reproduction material. |
| Clearly document the extent of the reproduction in your reproduction report and suggest potential improvements; if you provide intermediate feedback, to include a history of your interactions in the report so that the ideas you contributed are preserved when the submission’s material is improved. | Attempt to install software without any instructions, install binary software of unknown origin, or try to fix installation problems you encounter on your machine; try to install without (a) asking for help from a fellow reproducibility reviewer who is familiar with the software, or (b) asking the author to help, providing a minimal reproducible example of your problem. |
| Get in touch with fellow reproducibility reviewers if specific expertise (tool, programming language, ..) is needed. | Point out or even fix problems that are not specific to the submission, e.g., general problems in a software tool. |
| Set an example when communicating about computational problems, e.g., by clearly defining your system (OS version, language version, etc.) | Create accounts on any service or platform to access code, data, or other resources. |
| Ask specific questions or point out concrete problems that may lead authors to improve their material, including referencing these guidelines or concrete tools/methods that you already (!) know about, especially if you suspect that the author might now be familiar with them (e.g., version pinning/dependency management, absolute paths). | Fix anything (unless you really enjoy doing so), e.g.,<br>• compiler problems,<br>• outdated libraries,<br>• broken paths, or<br>• Incomplete computing environment specifications,<br>especially if the author can fix them even quicker. |
| Make sure that you are aware of any templates or specific resources provided for reproducibility reviewers from the reproducibility committee chair before starting your review. | |
| Consider the author’s background, career stage, and position to be aware of (a lack of) privileges or institutional power to decide how much support you provide and how you communicate; your reproducibility review can be a contribution to improve equity and inclusion in academia. | Be a bro. |



Here are some **examples** for things to do and not to do.

We want to shift the main burden to the author;

Reviewers are very welcome to encourage and share their knowledge;

Reviewers should not accept private sharing of data without very good reason;

The reproducibility review is a contribution to science, that is why the recommendations and feedback should be documented and is published in a document with a DOI.

The reviewers are encouraged to acknowledge the challenges that individual authors might face, for example disadvantages they have, and *be kind*.
So far, that was not a problem at all, of course, but it hopefully encourages more authors to be open and transparent.

The hardest task in my experience is **”no rummaging”**: it is really hard not to get excited and dig deeper all the time feeling “I am so close” !

WWU MÜNSTER

## Review process

Proceedings:
https://www.agile-giscience-series.net/review_process.html

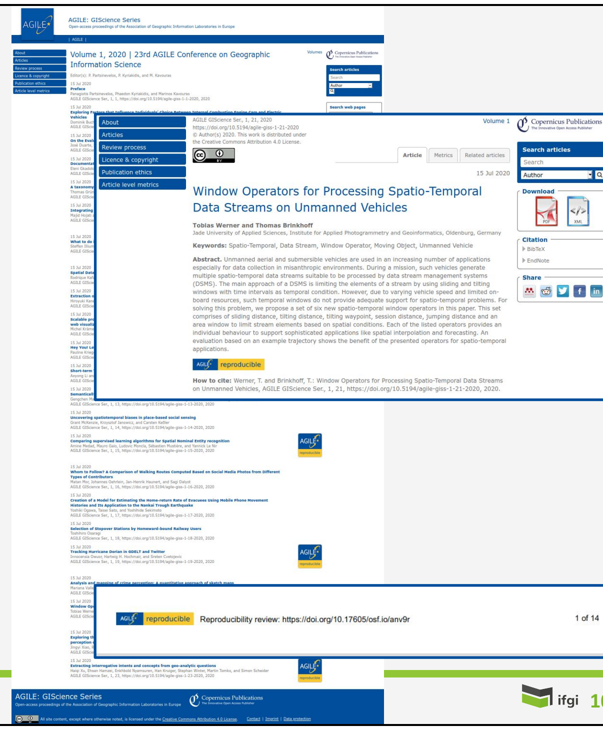
Process documentation:
https://osf.io/7rjpe/

Reproducibility review *after* accept/reject decisions

Reproducibility review & communication

Community conference & volunteers

Badges on proceedings website, article website with link, and first article page
(💖 Copernicus!)






First a few words about the AGILE reproducibility review process. You can find the details in the linked document.

Let’s start with the **order**: the reproducibility review happened after the authors were notified about the acceptance or rejection. Ideally, we wanted to review only papers that were identified by the scientific reviewers, but we found that that was not reliable so we briefly checked all papers ourselves. Also, none of the regular reviewers could be intrigued to attempt a reproduction.

This reduced the number of papers and allowed us to communicate directly with the authors during the reproducibility reviews.

The process was not smooth, because AGILE is a community-led conference managed by volunteers, and the authors are the ones who prepare the camera-ready copies of the articles. So there is a lot of moving parts with no central control. On top of that, the conference switched the publisher and the Coronavirus destroyed all schedules.

Eventually, the papers were published. The publisher added badges on the landing pages of the proceedings, which link to the reproducibility reports, as you can see on the right hand side. The papers themselves do not reference the reports.

We originally planned to have short presentations of selected reproducible papers in the final conference session, where candidates for the the best paper award are presented and voted live by the audience. Unfortunately, the in-person event was cancelled, so we could *not* have a showcase of the first batch of reproducible papers as we would have liked to.

**Besides publishing the reports on OSF and linking to them from the proceedings website, as shown on a previous slide, the AGILE conference's reproducibility reports are published in the CODECHECK register. The CODECHECK initiative tries to establish independent executions of computations that underly research articles as part of peer review. CODECHECK defines a set of principles which we at AGILE try to embrace.**

# Reproducibility Reports

Published on OSF with a DOI
Title page, cites the paper

Paper links to report via URL (no citation)

Automatically added to ORCID profile

WWU MÜNSTER

# Reproducibility review results 2021

## 9 reproducibility reports published (2020: 6)

- no starting point in the paper
- documentation insufficient for third party

8 not reproducible:

- conceptual papers
- data not shared (choice, licence)
- code not shared (choice) or proprietary software (repro reviewer matching failed)

Reproducibility review of: Building Change Detection of Airborne Laser Scanning and Dense Image Matching Point Clouds using Height and Class Information
Friese
Reproduction report and material.

Reproducibility review of: Investigating drivers' geospatial abilities in unfamiliar environments
Friese
Reproduction report and material.

Reproducibility review of: Extraction of linear structures from digital terrain models using deep learning
Nüst & Graser

Reproducibility review of: A Comparative Study of Typing and Speech For Map Metadata Creation
Ostermann & Nüst

Reproducibility review of: A Socially Aware Huff Model for Destination Choice in Nature-based Tourism
Krukar

Reproducibility review of: Automated Extraction of Labels from Large-Scale Historical Maps
Nüst

Reproducibility review of: Flood Impact Assessment on Road Network and Healthcare Access – at the example of Jakarta, Indonesia
Graser

Reproducibility review of: H-TFIDF: What makes areas specific over time in the massive flow of tweets related to the covid pandemic?
Nüst

Reproducibility review of: An Approach to Assess the Effect of Currentness of Spatial Data on Routing Quality
Nüst & Kmoch

REPRODUCIBLE AGILE

https://osf.io/h64sd/



**9 is more than 8!**


In just three years, we were able to plant the seed of cultural and structural change (if I myself may say so). So here is how to put your community on a path towards more reproducibility in 5 easy steps:

1. Build a team of enthusiasts, for example by organising a workshop
2. Assess the current state and talk about it - this is how people will realise you are serious; in our case it was pre-conference workshops and a paper with a strong message
3. Try to get institutional support - in our case the support of the AGILE Council and the conference committee chairs made everything possible
4. Try to get the community on your side by staying positive and inclusive - irreproducibility is usually not a sign for bad science, but for a lack of knowledge and incentives, which you can try to create
5. Keep at it, cultural change takes time - putting reproducibility on the map is an achievement in itself, and every little step counts; providing feedback as part of the reproducibility reviews will, one author at a time, educate your community.

WWU MÜNSTER

## Next steps for reproducible AGILE

AGILE reproducible

Do it again in 2022 🎉

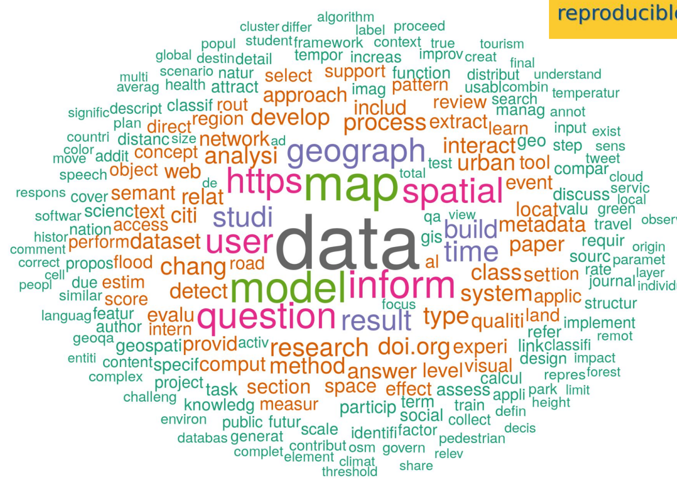
Grow reproducibility reviewer team
Opportunity ECRs (mentoring/workshops/…)

Continue discourse (meaning of rprdcblty)

Re-assess new papers > impact?

Towards opening scholarship
Scope, requirements, acceptance condition?
Open review if tenured? Format-free first submission
CRediT

*Phase out when standard practice...*




Word-stem cloud of all AGILE 2021 submissions (full/short/poster & accepted/rejected)



I'd like finish with a short outlook on the next steps.

Most importantly, we are doing it again next year and try to learn from our experiences.

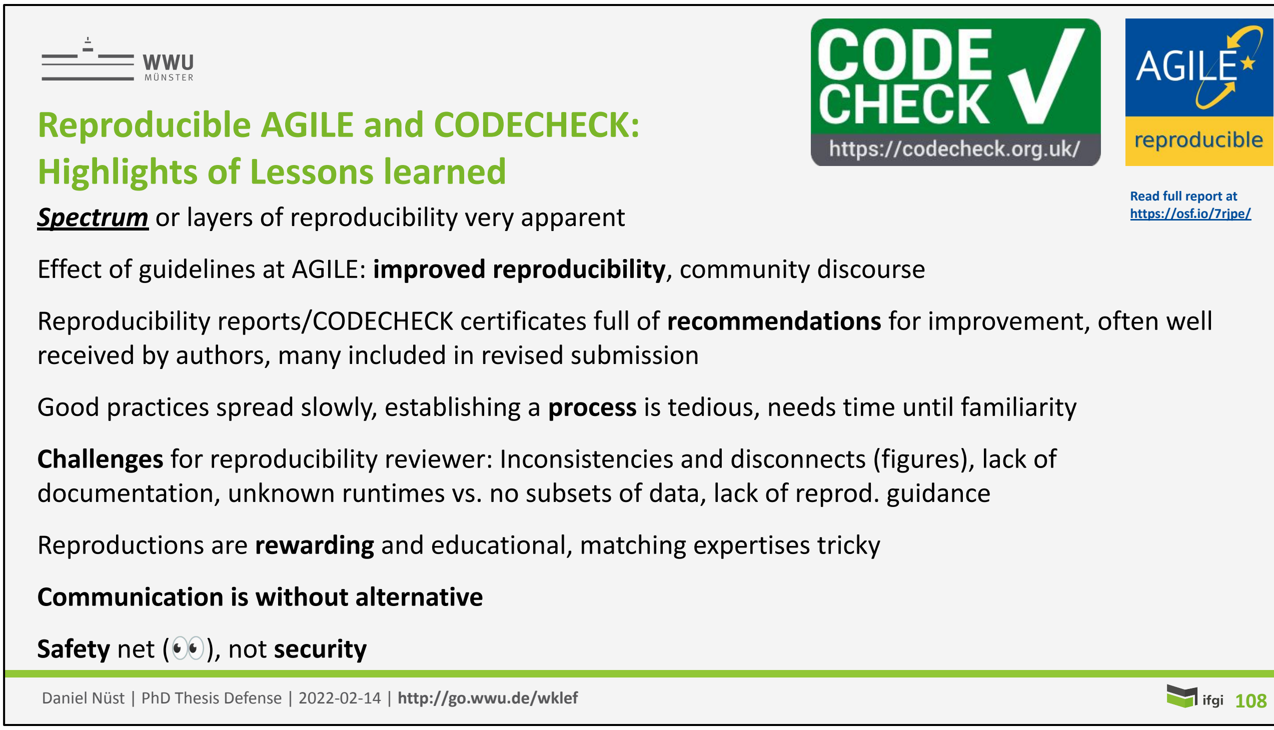
Furthermore, the guidelines will be improved. Most likely we will simplify them and move details to other places. I'm keen on having the guidelines translated to make them more accessible, and hope they find adoption beyond the conference.

We need to grow reproducibility reviewer team, which I think is a great opportunity to involve early career researchers in peer review. More people means more diverse skill sets to better match the given manuscripts.
I hope that we will be able to give reproducibility (and regular) reviewers public credit soon, for example by putting reviewer activity into ORCID profiles.

We continue to research the state of reproducibility in GIScience and have just published a preprint on another conference series, the GIScience conference. The results are equally sobering as for the AGILE conference three years ago.

And, of course, the continued interaction with the community is key to increase reproducibility further. We will debate the scope of the review and when to increase requirements, maybe even require reproductions for acceptance?
In the long term, I personally hope the awareness on reproducibility is a starting point for more open practices in general, for example requiring tenured reviewers to disclose their identity, accept first submissions without formatting, or using the CRediT taxonomy for author contributions.
*And most importantly, I'd be completely fine to phase out the special reproducibility*

*review, when reproducible papers have become standard practice.*




Highlights of things that we learned during the first reproducibility review at an AGILE conference and the codechecks we did so far.

We saw the full spectrum of reproducibility, which encourages us to continue with our current approach of positive encouragement and education.

At AGIIE, we saw that compared to previous years' submissions, the guidelines did effectively increase community awareness and improved reproducibility. Of course we also looked harder than before, but we are still very happy with the overall outcome of the reproductions.

The reproducibility reports have a lot of details with recommendations how to improve workflows. Some authors picked these up pretty quickly and even improved their material in time with the publication of the articles, which I think is a great reward for the reproducibility reviewers. I hope the reports were as friendly as the ReproHackers' feedback!

Also, good practices spread slowly because people need to change their (even daily) habits; at AGILE this led to reviews being less strict than originally anticipated. I had to leard that *slow change can be a good thing* because you can keep everybody on board. Organising the whole process is hard, and needs improvements over time - so I try to worry less about that.

The challenges for reviewers respectively codecheckers are manifold, but mostly easy to solve in a bilateral conversation.

While everyone *should* be able to reproduce any workflow, matching reviewer expertise can be tricky yet important for effective reviews - in any case I personally found the reproductions to be always educational and a rewarding task.

There is not better way to make a reproduction a success and a good experience, and *effective*, than direct open communication.

And finally, there are of course limitations. The AGILE reproducibility review and CODECHECK provide safety nets, a layer to make sure we are not making any avoidable mistakes such as incomplete datasets. They do not provide security against fraud or malicious activities, though they can ensure that future investigators have all they need to dig deeper.

----------

https://www.quora.com/What-is-the-difference-between-safety-and-security
**Safety** is the prevention of accidents (accidents which may or may not involve human agents, but are in any case not intentional).
**Security** is the prevention of malicious activities by people (mugging, burglary, robbery, terrorist activities, etc.).

Avoid inherent dangers (driving sober with a regularly inspected car), not protect against external factors (tree falling onto the road, someone tampering with the breaks).

## What can communities and institutions do?

Introduce reproducibility reviews - CODECHECK (or not) - at your journals, labs, collaborations!

Workshops on RCR, ReproHacks

Provide support (**R2S2**, PhD edu.)

Rewards and incentives

Community discourse

**Awareness > Change**



Introduce reproducibility reviews, top down if you're an editor, or buttom up if your're author or reviewer. Just do it, it will shift practice over time.

Become and advocate and educator! You can organise a workshop or a ReproHack, and help more and more community members build up the expertise to work more reproducibly. A lack of reproducibility is rarely intentional or supported by a strong opinion, but a lack of knowledge and rewards.

Create incentives and rewards.

Awareness leads to change. In the AGILE community, the team behind Reproducible AGILE was able to get from zero to six reproducible papers in three years, which I'm really proud of. However, I must admit this was only possible because we had institutional support.
We're still working hard to find journals who want to establish CODECHECKs.

One tiny step at a time is still progress, and as this dancer shows us, you can still be very powerful and graceful.

WWU MÜNSTER

# Concepts, metaphors, memes

WWU MÜNSTER

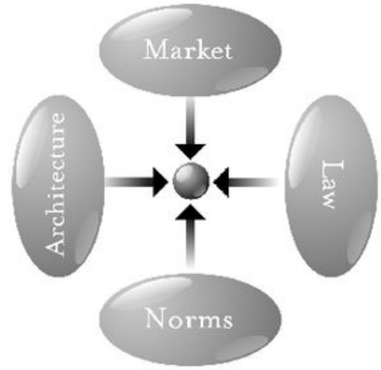
## Lessig’s pathetic dot theory

**Law**
policies, sanctions

**Social norms**
community enforcement

**Markets**
supply & demand > price of items & behaviours

**(Social) Architecture**
Made or found features & properties & infrastructure (biology, physics, major social/cultural forces); constraints




Theory of regulation, applied to internet *but also fitting scholarly communication* > unlike real world, architecture (= code) is created and controlled by humans resp. *scientists*, yet still are a force on our behaviour.

https://en.wikipedia.org/wiki/Pathetic_dot_theory



A similarly helpful concept was introduced by Lawrence Lessig’s pathetic dot theory, where he describes four forces that regulate our lives as individuals (= the pathetic dots).

These forces can also help to describe the professional live of scientists.

**Law** provides the policies and incentives as well as possibly threatening sanctions.

**Social norms** and **markets** are controlled by and provide value for the community.

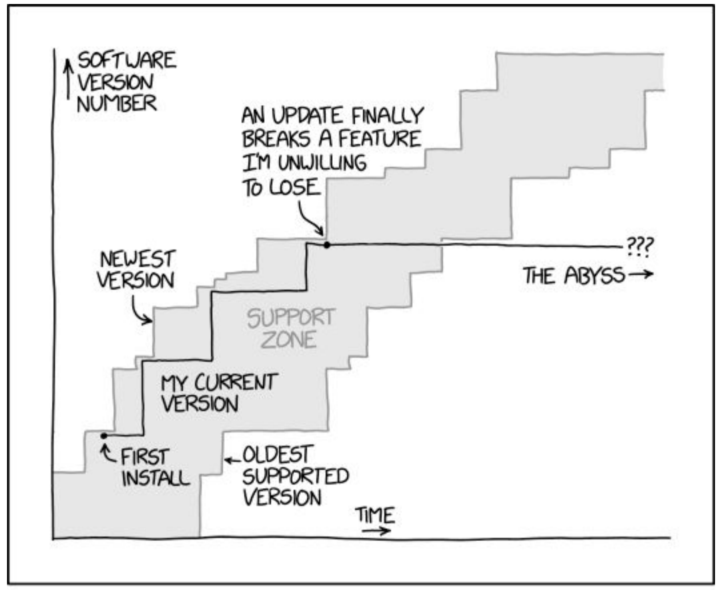
**Architecture** comprises the made or found constraining facts and infrastructure.

If the forces are applied to scholarly communication and scientific progress, the important point is that all forces can be shaped by scientists, who can influence or should control other stakeholders when they work together, just as the layers of the culture change pyramid.

# *Digital information lasts forever, or five years - whichever comes first.*

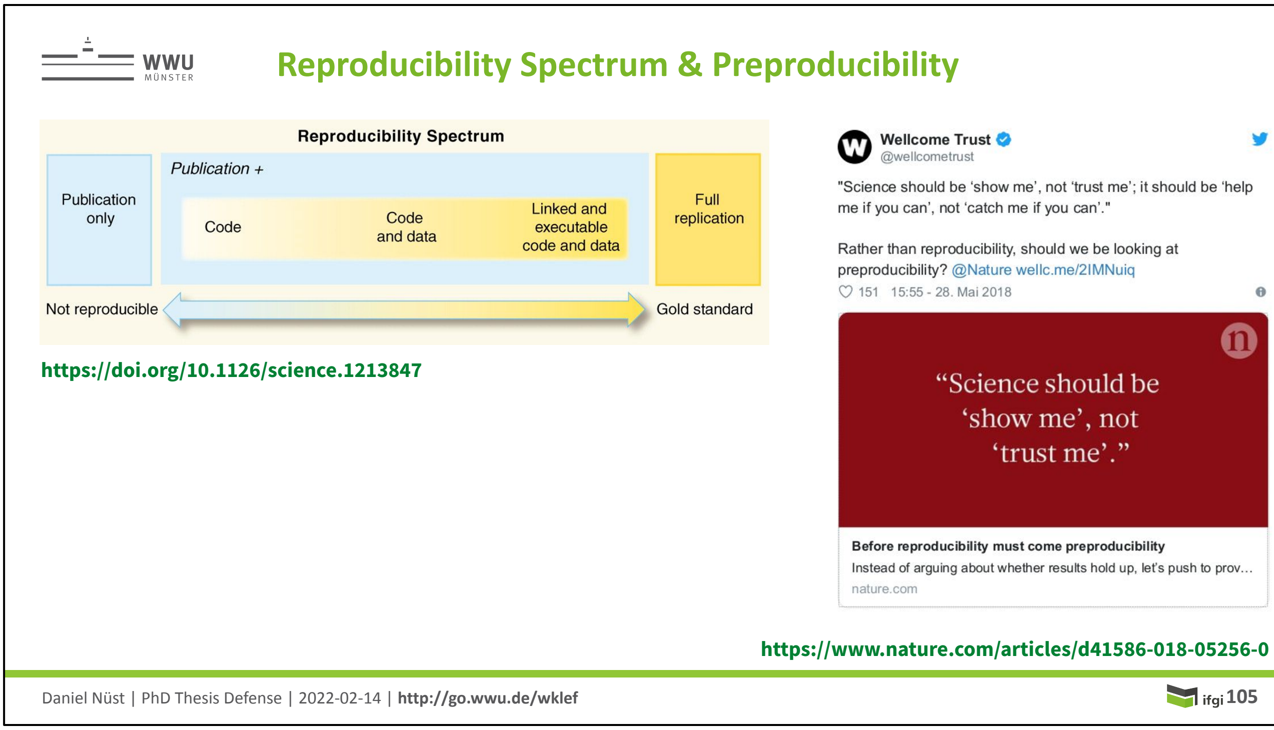
Rothenberg, Jeff. 1995. "Ensuring the Longevity of Digital Documents." *Scientific American* 272 (1): 42–47.
via https://twitter.com/snet_jklump/status/1141934045820887040?s=09


http://phdcomics.com/comics.php?f=1689



ALL SOFTWARE IS SOFTWARE AS A SERVICE.

https://xkcd.com/2224/




Reproducibility Spectrum by Roger Peng

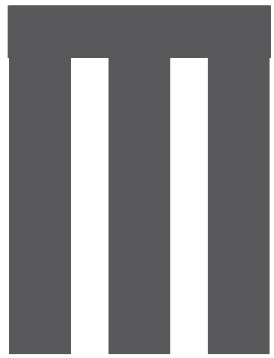
------

**Preproducibility** ist ein Neologismus von Philip B. Stark als als Alternative zu den verwirrenden oder unterschiedlich interpretierten existierenden Begriffen, nämlich: *"An experiment or analysis is preproducible if it has been described in adequate detail for others to undertake it. Preproducibility is a prerequisite for reproducibility, and the idea makes sense across disciplines. [...] Science should be 'show me', not 'trust me'; it should be 'help me if you can', not 'catch me if you can'."* Lesenswert: 10.1038/d41586-018-05256-0 !

# Traditional and modern scientists

Broad knowledge: across disciplines
collaborate with other experts, apply outside of own field

**Deep knowledge:** expertise and skills
within a single field



**Computer & method skills**
statistics, reproducibility, programming,
data science

https://en.wikipedia.org/wiki/T-shaped_skills

https://doi.org/10.1007/s10816-015-9272-9

https://jakevdp.github.io/blog/2014/08/22/hacking-academia/
https://www.sciencemag.org/careers/2013/05/when-all-science-becomes-data-science
https://escience.washington.edu/community-level-data-science-and-its-spheres-of-influence-beyond-novelty-squared/

US RSE
US RESEARCH SOFTWARE ENGINEER ASSOCIATION

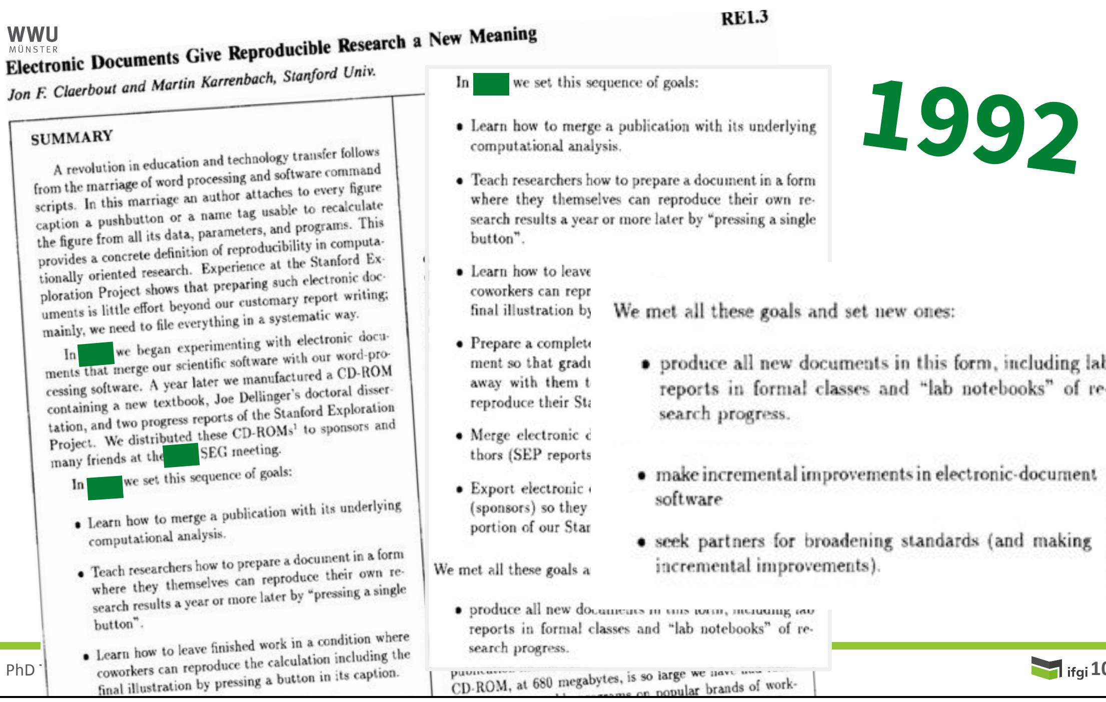
de RSE

RSE
RESEARCH SOFTWARE ENGINEERS ASSOCIATION

<ReSA>
Research Software Alliance

1992

http://dx.doi.org/10.1190/1.1822162



RE1.3

**Electronic Documents Give Reproducible Research a New Meaning**

*Jon F. Claerbout and Martin Karrenbach, Stanford Univ.*


SUMMARY

A revolution in education and technology transfer follows from the marriage of word processing and software command scripts. In this marriage an author attaches to every figure caption a pushbutton or a name tag usable to recalculate the figure from all its data, parameters, and programs. This provides a concrete definition of reproducibility in computationally oriented research. Experience at the Stanford Exploration Project shows that preparing such electronic documents is little effort beyond our customary report writing; mainly, we need to file everything in a systematic way.


In [redacted] we began experimenting with electronic documents that merge our scientific software with our word-processing software. A year later we manufactured a CD-ROM containing a new textbook, Joe Dellinger's doctoral dissertation, and two progress reports of the Stanford Exploration Project. We distributed these CD-ROMs[1] to sponsors and many friends at the [redacted] SEG meeting.

In [redacted] we set this sequence of goals:

- Learn how to merge a publication with its underlying computational analysis.
- Teach researchers how to prepare a document in a form where they themselves can reproduce their own research results a year or more later by "pressing a single button".
- Learn how to leave finished work in a condition where coworkers can reproduce the calculation including the final illustration by pressing a button in its caption.
- Prepare a complet... ment so that gradu... away with them t... reproduce their Sta...
- Merge electronic ... thors (SEP reports...
- Export electronic ... (sponsors) so they ... portion of our Star...

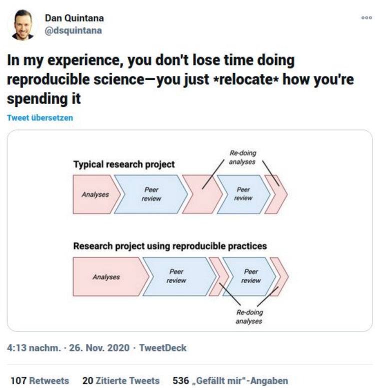
We met all these goals and set new ones:

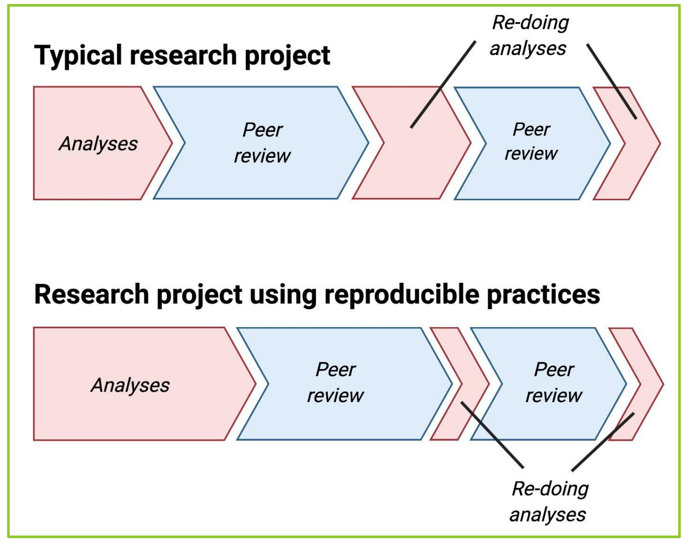
- produce all new documents in this form, including lab reports in formal classes and "lab notebooks" of research progress.
- make incremental improvements in electronic-document software
- seek partners for broadening standards (and making incremental improvements).

... CD-ROM, at 680 megabytes, is so large we have ... popular brands of work-

WWU MÜNSTER

## Reproducibility is "more work"?




Quintana, D. S. (2020, November 28). Five things about open and reproducible science that every early career researcher should know. **https://doi.org/10.17605/OSF.IO/DZTVQ**



And while better reproducibility to answer journal requirements might be the only way to introduce real change, let's be clear that using reproducible practices is not something that you only do for you, but something that you can apply just for yourself! This is a recent tweet by Dan Quintana, who makes it clear that the extra time you might feel you have to spend in the beginning, really pays off later in the research process when you have to adjust you analysis, and even further down the road when you want to write your next paper.

*Reproducibility is the basis for reusability and collaboration, and future you is your best collaborator.*

**RSEng = create research software**
**RSEs = people behind research softw**
**RSEs ≠ IT !!!**

**Reproducibility guru** dives deeply into manifold software and tools to make his research reproducible and develops his own software in a sustainable way.

**Researcher** uses scripts for data analysis and needs working stable software for her work. She learns what is necessary to achieve her research goals.

**Person for tough problems** knows how to solve all kinds of computer-related issues; he was not hired for that, but enjoys to help and spends time to get to the bottom of other people's challenges.

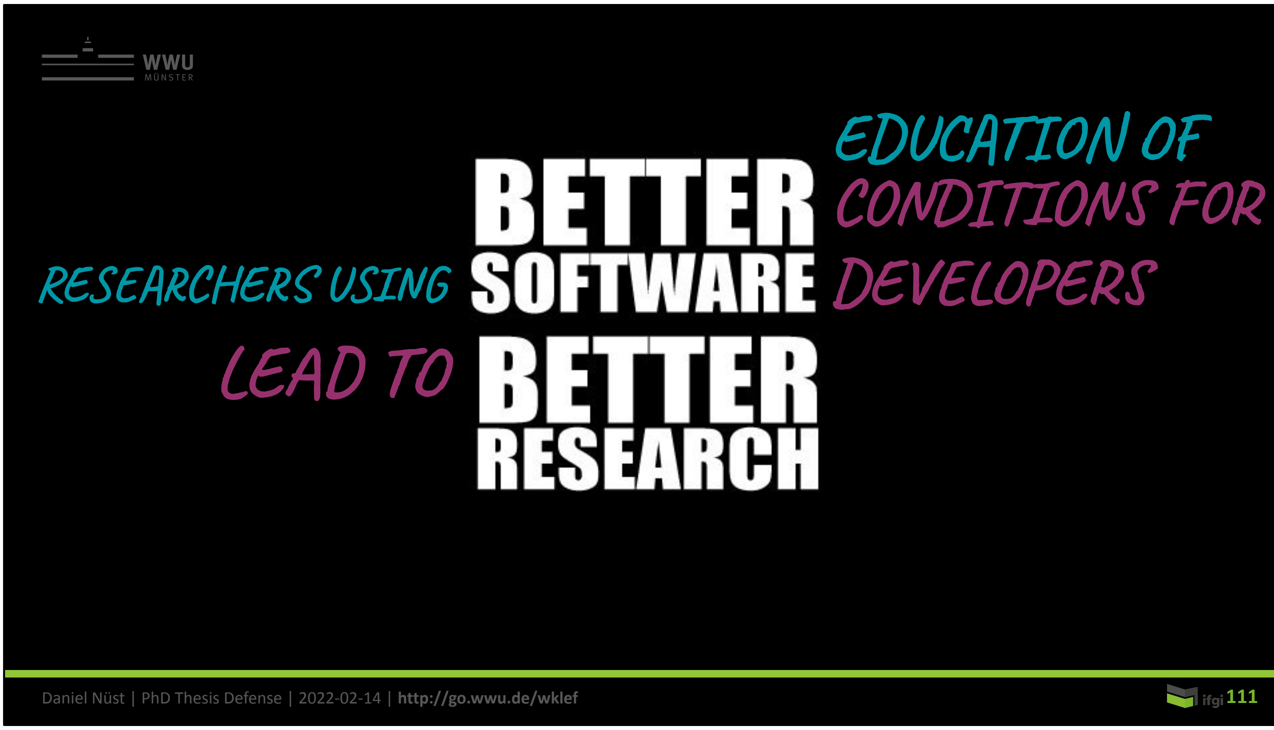
# "Software is 95% human and only 5% code" *

**Software developer** was hired to implement software for a research project and contributes to large collaborative software projects to realise the next generation of digital infrastructure for science.

**Geek** writes software as part of her research project and would like to code more, but must keep an eye on her career in science and needs to write papers.

* Eric Albers, CCC2019, https://media.ccc.de/v/thms-49-ber-die-nachhaltigkeit-von-software | Bilder © H. Seibold, S. Janosch, OSD2019


Here's an image that I find useful that explains why professionalisation matters:
We can all prepare food so that we do not starve. For some that means that they pour boiling water into a cup of ramen, for some that means to fire up the barbeque with friends, for others it means to cook fresh and healthy food every day.
*However*, when we want to eat very good, or we want to serve many people at once (think about the Mensa), we go to the professionals.
Or is here someone, who cooks themselves when they throw a party or take out their own cooker in the office?
***Without professional-grade software development, the progress in science will be slower. And these developers must, to some extent, understand the science and speak the language of the researchers.***




Software development is an essential, integral part of research activity. Research software increasingly supports the acquisition, processing and analysis of empirical data, but also the modeling and the simulation of complex processes. Thus, software has a significant influence on the quality of research results. The British Software Sustainability Institute (SSI) has coined the slogan “Better Software - Better Research”.


Home
About
Search Twitter
Have an account? Log in
Sylvain Deville
Freezing
Observat
Principl
and Use
Sylvain
@DevilleSy
Follow
When you try to replicate a paper using the methods section
9:56 AM - 31 Jan 2018
2,605 Retweets 5,877 Likes
54
2.6K
5.9K
Sylvain
@DevilleSy
Freezing stuff since 1876. Will science for chocolate. ORCID Id 0000-0002-3363-3184. Author of "Freezing Colloids"
springer.com/fr/book/978331…
France
sylvaindeville.net
© 2018 Twitter About Help Center Terms
Privacy policy Imprint Cookies Ads info
https://twitter.com/DevilleSy/status/958761021421903872
112

# Code review, journals, crisis

WWU MÜNSTER

## Code Review != Reproduction/Reproducibility Review

rOpenSci

Boettiger, C., Chamberlain, S., Hart, E., & Ram, K. (2015). Building Software, Building Community: Lessons from the rOpenSci Project. Journal of Open Research Software, 3(1), e8. doi:10.5334/jors.bu

pyOpenSci

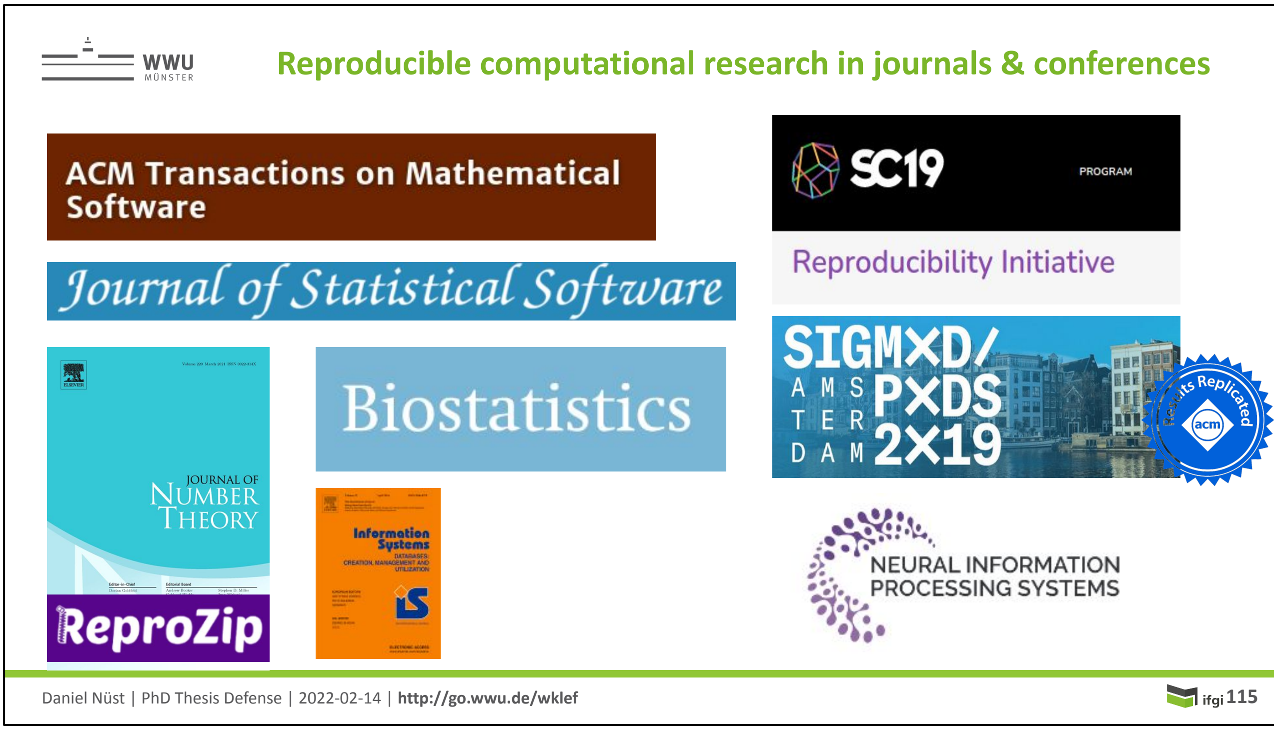
Journal of open research software

The Journal of Open Source Software

**Code Review Community Working Group**



I won't cover actual code review today. I think it is very important, but I am not convinced peer review of *publications* is when code review should happen. During publication peer review, the whole workflow combined with the scientific method should be under scrutiny.
If you're interested in software review and software publications, I invite you to check out any of these organisations and journals if you don't know them yet, and to join the Code Review Community Working Group.




Actually, there are quite a few journals and conferences that have **reproducibility editors**, so what I present here today is not unique, just not yet common enough. Here are some examples, maybe one from your discipline?
We're currently trying to compile an exhaustive list and learn more about all the existing variants. Some of these use **badges** to highlight reproducible works. The journal Information Systems takes a unique approach: invited reproducibility papers are published after the original paper and the reproducer become coauthors. *Take that, impact factor!*
**However, I can better talk about solutions to connect reproducibility with peer review that I'm involved with myself.**
**And these are the AGILE conference's reproducibility review and CODECHECK.**

NATURE | NEWS FEATURE

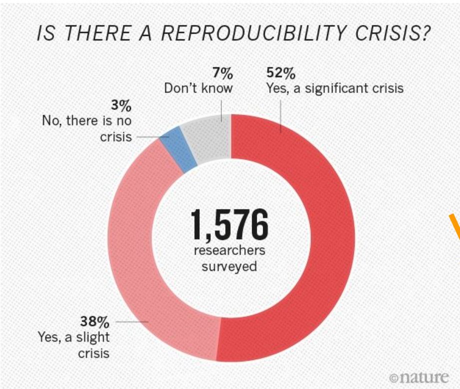
https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

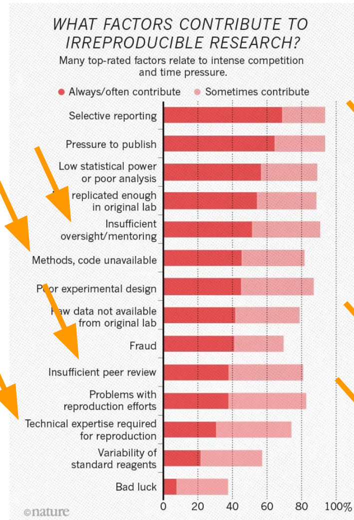
# 1,500 scientists lift the lid on reproducibility

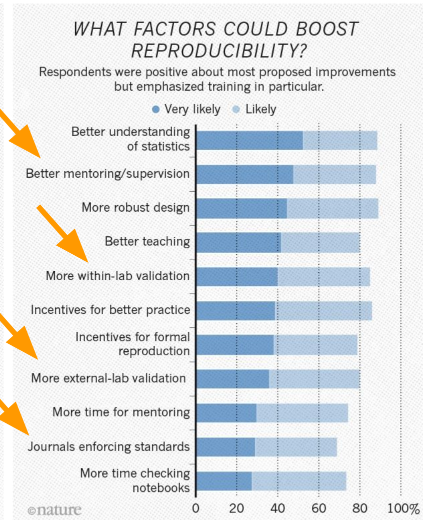
**Survey sheds light on the 'crisis' rocking research.**

Monya Baker

https://www.youtube.com/watch?v=j7K3s_vi_1Y

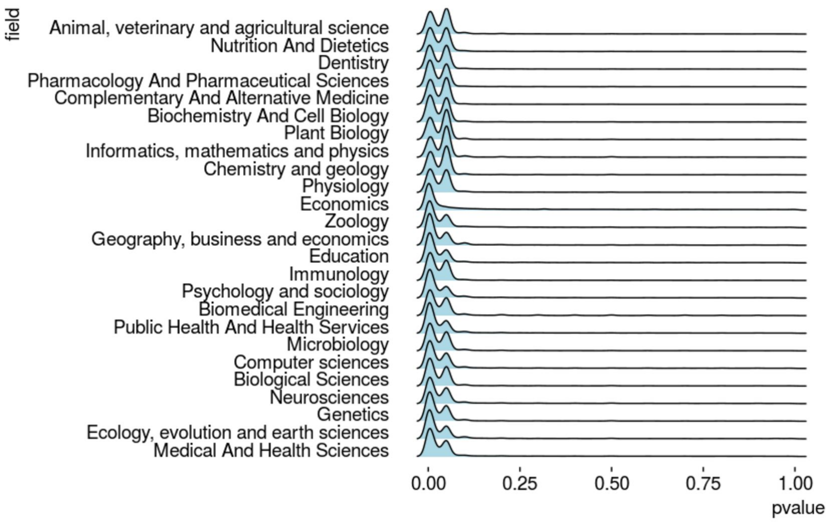


## J. Leek's tidypvals

The tidypvals package is an effort to find previous collections of published p-values, synthesize them, and tidy them into one analyzable data set. The currently available p-value data sets in this package are:




"Notice Anything funny?"

https://simplystatistics.org/2017/07/26/announcing-the-tidypvals-package/

WWU MÜNSTER

Comment | Open Access | Published: 08 December 2015

# Five selfish reasons to work reproducibly

Florian Markowetz

Genome Biology 16, Article number: 274 (2015) | Cite this article



https://doi.org/10.1186/s13059-015-0850-7

1. reproducibility helps to avoid disaster
2. reproducibility makes it easier to write papers
3. **reproducibility helps reviewers see it your way**
4. reproducibility enables continuity of your work
5. reproducibility helps to build your reputation

WWU MÜNSTER

nature International weekly journal of science

## Publish your computer code: it is good enough

**Freely provided working code — whatever its quality — improves programming and enables others to engage with your research, says Nick Barnes.**

Nick Barnes

I am a professional software engineer and I want to share a trade secret with scientists: most professional computer software isn't very good. The code inside your laptop, television, phone or car is often badly documented, inconsistent and poorly tested.

Why does this matter to science? Because to turn raw data into published research papers often requires a little programming, which means that most scientists write software. And you scientists generally think the code you write is poor. It doesn't contain good comments, have sensible variable names or proper indentation. It breaks if you introduce badly formatted data, and you need to edit the output by hand to get the columns to line up. It includes a routine written by a graduate student which you never completely understood, and so on. Sound familiar? Well, those things don't matter.

**https://doi.org/10.1038/467753a**



"I am a professional software engineer and I want to share a trade secret with scientists: most professional computer software isn't very good. The code inside your laptop, television, phone or car is often badly documented, inconsistent and poorly tested."
"Why does this matter to science? Because to turn raw data into published research papers often requires a little programming, which means that most scientists write software. And you scientists generally think the code you write is poor. It doesn't contain good comments, have sensible variable names or proper indentation. It breaks if you introduce badly formatted data, and you need to edit the output by hand to get the columns to line up. It includes a routine written by a graduate student which you never completely understood, and so on. Sound familiar? Well, those things don't matter."

**No excuse is good enough!**

*What can we do to make the software better?*

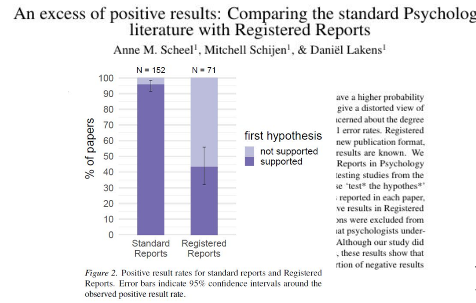
# Structural challenges

**Metrics** for acknowledging/measuring impact in science **are broken** (impact factor, ..) and they lead to publication bias, HARKing, p-Hacking, intransparency and lack of reproducibility

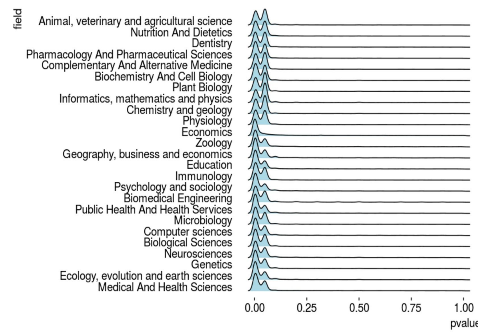
Leiden Manifesto: **http://www.leidenmanifesto.org**
DORA: **https://sfdora.org**
Vienna Principles: **https://viennaprinciples.org**

***Acknowledging data and software as valuable products of research (instead of shoehorning software into papers)***


http://doi.org/10.5334/dsj-2020-003




https://doi.org/10.31234/osf.io/p6e9c



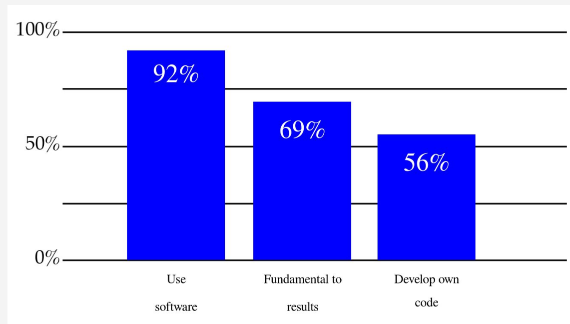

https://simplystatistics.org/2017/07/26/announcing-the-tidypvals-package/

WWU MÜNSTER

## Motivation for RSE

Back to 2010 The Software Sustainability Institute (SSI, UK) run a study (1000 randomly chosen researchers) …

**“It's impossible to conduct research without software, say 7 out of 10 UK researchers”**



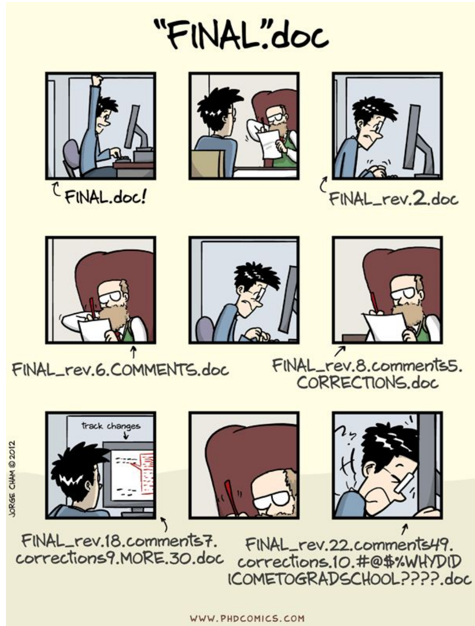

**https://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers**

## Motivation for RSEng

A study of Nature papers from Jan-March 2016 reveals that

**“32 of the 40 papers examined mention software, and the 32 papers contain 211 mentions of distinct pieces of software, for an average of 6.5 mentions per paper.”**

[2] Nangia, Udit; Katz, Daniel S. (2017): Understanding Software in Research: Initial Results from Examining Nature and a Call for Collaboration. doi:**10.1109/eScience.2017.78**




http://phdcomics.com/comics/archive.php?comicid=1531

WWU MÜNSTER

# Learn more about code execution practices at journals and conferences

**osf.io/x32nc**

## Code Execution and Peer Review

### Idea

Research outputs are more than just PDF papers, but include data and software. With an increasing number of journals and conferences giving guidance on sharing data and code, the actual execution of workflows underlying research papers is still relatively rare. To better understand the different approaches to realise code execution (limitations, roles) and the different levels these reviews can take, we want to run a survey/series of interviews. Based on the experiences made, we hope to derive guidelines and a common language for integrating workflow execution into peer review.

### Project status

The idea was conceived by Daniel Nüst, Stephen Eglen, and Heidi Seibold. A survey was designed, with help from Lea Schulz-Vanheyden. A list of journals and contact points is ready to be used to start either interviews or send out the survey. See the tasks document for the completed steps and the original ideas how to continue.

All material is published on OSF at https://doi.org/10.17605/osf.io/x32nc. The main documents are Google Docs shared at https://drive.google.com/drive/folders/1ageeYBIFGDL82Pn55u30BsjUsP0YuSz4?usp=sharing.

**Daniel Nüst** @nordholmen

Sometimes the idea & the people are right, but the timing is bad. No question about the people: @HeidiBaya & @StephenEglen are awesome collaborators and great supporters of #OpenScience & #ReproducibleResearch. Here is our idea that we cannot pursue. You tell us if it is good: 🧵

Tweet übersetzen

4:49 nachm. · 1. Juni 2021 · Typefully

23 Retweets 3 Zitierte Tweets 31 „Gefällt mir"-Angaben

Twittere deine Antwort — Antworten

**Daniel Nüst** @nordholmen · 1. Juni
Antwort an @nordholmen
Data and code are first class research outputs. No doubt about it! Is there a long way to go to broad recognition? Sure. Do we need more rewards for open practices and transparency? Definitely. Do we need more pressure? Possibly.

What about simply trying out stuff during review?

1 1

**Daniel Nüst** @nordholmen · 1. Juni
It's great that peer review practices are challenged in academia (#preprints, #OpenPeerReview, post publication review). We need to improve them.

Sadly, one thing is often missing from reviews: the attempt to actually execute the workflow underlying a research paper.

2 2 7

**Daniel Nüst** @nordholmen · 1. Juni
There are reasons for code & data review not commonly being part of peer review, but this thread is not them. Based on the simple #CODECHECK perspective (codecheck.org.uk) we want to learn more about journals and conferences who simply try to execute a paper's code.

CODECHECK
CODECHECK is a process for independent execution of computations underlying scholarly ...
codecheck.org.uk

1 1 4

**Daniel Nüst** @nordholmen · 1. Juni
So we came up with a survey and collected a long list of journals and events that include workflow evaluation and code execution. We realised the survey might not be a great fit to learn what we want to know: to understand what works, who must be involved, and how to sustain.

1 2

**Daniel Nüst** @nordholmen · 1. Juni
To understand code execution practices during peer review we decided the best approach is to conduct semi-structured interviews with associated editors for reproducibility, reproducibility chairs... But we don't have the time atm. Do you think it is a worthwhile project?

# Code execution in peer review

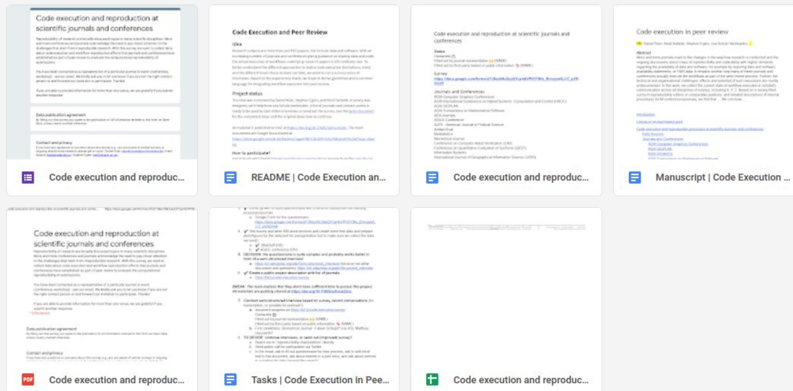
https://osf.io/x32nc/

**Daniel Nüst, Heidi Seibold, Stephen Eglen, Lea Schulz-Vanheyden, Limor Peer, Josef Spillner**

Survey practices of code execution as part of peer review

Text survey design ✔️
Manuscript outline ✔️
List of journals and events ✔️

Surveying ❌ (interviews?)



# Metadata

# The role of metadata in reproducible computational research

Jeremy Leipzig, Daniel Nüst, Charles Tapley Hoyt, Karthik Ram, Jane Greenberg
Patterns (N Y). 2021 Sep 10;2(9):100322. **https://doi.org/10.1016/j.patter.2021.100322**


| Type of standard | Purpose |
|---|---|
| Reporting standards | Ensure adequate metadata for reproduction |
| Terminology artifacts or semantics | Concept disambiguation and semantic relationships |
| Models and formats | Interoperability |
| Identifier schemata | Discovery |

| Metadata level | Description | Examples of metacontent | Examples of standards | Projects and organizations |
|---|---|---|---|---|
| 1. Input | metadata related to raw data and intermediates | sequencing parameters, instrumentation, spatiotemporal extent | MIAME, • EML, • DICOM • , GBIF CIF ThermoML, CellML, DATS, FAANG, ISO/TC 276, NetCDF, OGC, GO | OBO, NCBO, FAIRsharing, Allotrope |
| 2.Tools | metadata related to executable and script tools | version, dependencies, license, scientific domain | CRAN DESCRIPTION file, • Conda • meta.yaml/environment.yml, pip requirements.txt, • pipenv Pipfile/Pipfile.lock, Poetry pyproject.toml/poetry.lock, EDAM, • CodeMeta, • Biotoolsxsd, DOAP, ontosoft, SWO | Dockstore, Biocontainers |
| 3.Statistical reports and notebooks | literate statistical analysis documents in Jupyter or knitr, overall statistical approach or rationale | session variables, ML parameters, inline statistical concepts | OBCS, STATO • SDMX DDI, MEX, • MLSchema, MLFlow, • Rmd YAML • | Neural Information Processing Systems Foundation |
| 4.Pipelines, preservation, and binding | dependencies and deliverables of the pipeline, provenance | file intermediates, tool versions, deliverables | CWL, • CWLProv, • RO-Crate, • RO, WICUS, OPM, PROV-O, ReproZip Config, ProvOne, WES, BagIt, BCO, ERC | GA4GH, ResearchObjects, WholeTale, ReproZip |
| 5.Publication | research domain, keywords, attribution | bibliographic, scientific field, scientific approach (e.g., "GWAS") | BEL, • Dublin Core, JATS, ONIX, MeSH, LCSH, MP, Open PHACTS, SWAN, SPAR, PWO, PAV | NeuroLibre, JOSS, ReScience, Manubot |

Metadata standards, including MIAME,[40] EML,[50] DICOM,[51] GBIF,[52] CIF,[53] ThermoML,[54] CellML,[55] DATS,[56] FAANG,[57] ISO/TC 276,[58] GO,[41] Biotoolsxsd,[59] meta.yaml,[60] DOAP,[61] ontosoft,[62] EDAM,[63] SWO,[64] OBCS,[65] STATO,[66] SDMX,[67] DDI),[68] MEX,[69] MLSchema,[70] CWL,[71] WICUS,[72] OPM,[73] PROV-O,[74] CWLProv,[75] ProvOne,[76] PAV,[77] BagIt,[78] RO,[47] RO-Crate (abstract by Sefton et al., 2019), BCO,[79] Dublin Core,[80] JATS,[81] ONIX,[82] MeSH,[83] LCSH,[84] MP,[85] Open PHACTS,[86] BEL,[87] SWAN,[88] SPAR,[89] PWO.[90] • Standards that are featured within this article. Examples of all standards can be found at https://github.com/leipzig/metadata-in-rcr.

# Knowledge Exchange

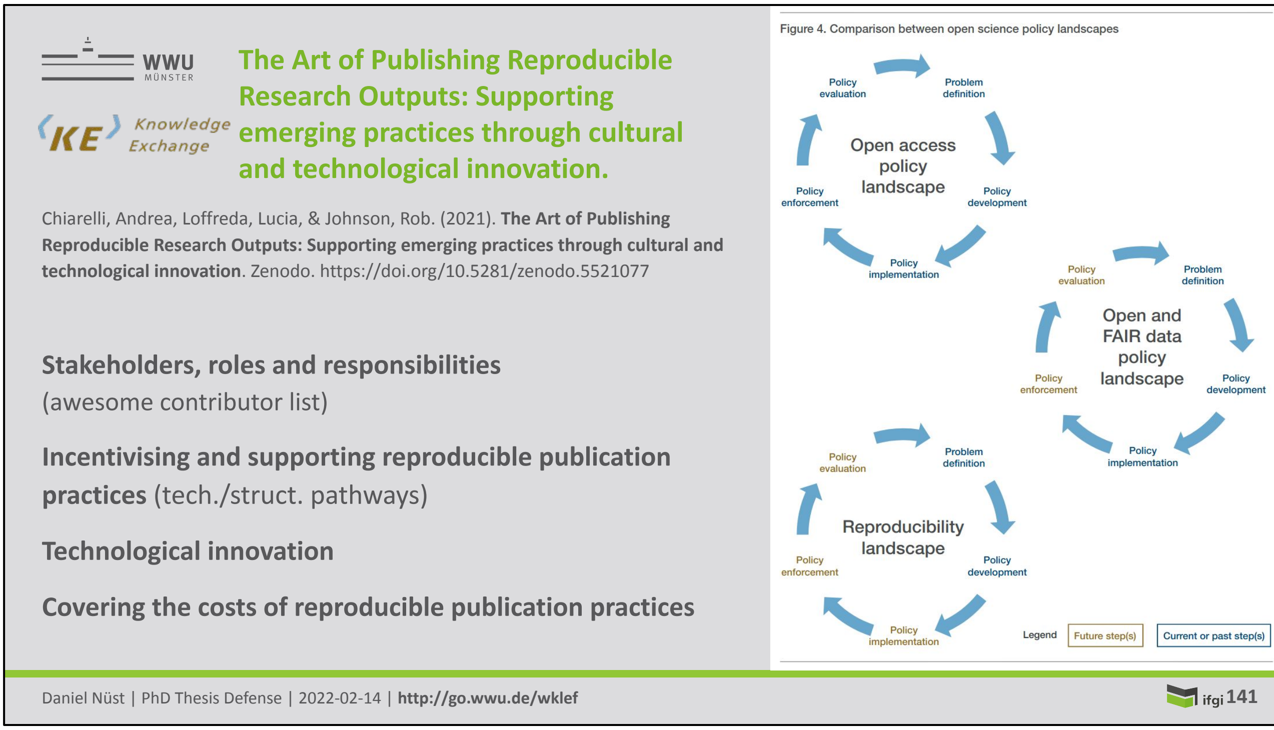

**Five take-away messages**

1. Reproducibility is part of the vision for open science, alongside concepts such as replication, robustness and the generalisation of research findings. It is difficult to pursue **culture change** with regard to reproducibility without considering this **broader context**.
2. Stakeholder collaboration is needed to continue developing reproducible publication **practices**. All players from the individual researcher to national and international bodies have a role to play, including in the context of policy development and implementation.
3. **Incentives** for reproducible publication practices are currently limited. Research performing organisations are beginning to support researchers in meeting their **growing reproducibility expectations**, and there is increasing demand for new **training** and support **pathways** in this area.
4. The management, curation and **sharing of research data and methods** are necessary conditions for reproducible publication. It is essential for these **practices to become the norm** to push the reproducibility agenda forward, and some dedicated institutional roles such as data stewards may be required to keep up with the demand for support.
5. Reproducible publication practices require a range of **technological solutions**, but most contributors agreed that these are already **available** in today's research landscape. The key technical gap appears to be the **interoperability** between available tools and workflows; however, we also note that technological solutions for reproducibility are not currently covered as part of training curricula.

The purpose of this activity of Knowledge Exchange was to explore current practices and barriers in the area of research reproducibility, with a focus on the publication and dissemination stage. We wanted to determine how technical and social infrastructures can support future developments in this area. In this work, we defined research reproducibility **as cases where data and procedures shared by the authors of a study are used to obtain the same results as in their original work**.

We captured the views of **research funding organisations, research performing organisations, learned societies, researchers, academic publishers and infrastructure and service providers**. We did a comprehensive literature review and a series of interviews and focus groups with a total of 51 contributors. The results of our activity give answers to the following questions:

What are the main benefits and barriers of publishing reproducible research outputs?
What are the roles of the different stakeholders involved?
How expensive are reproducibility checks?
What kind of digital tools and infrastructure are needed to publish reproducible research output?




**Five take-away messages**

1. Reproducibility is part of the vision for open science, alongside concepts such as replication, robustness and the generalisation of research findings. It is difficult to pursue **culture change** with regard to reproducibility without considering this **broader context**.
2. Stakeholder collaboration is needed to continue developing reproducible publication **practices**. All players from the individual researcher to national and international bodies have a role to play, including in the context of policy development and implementation.
3. **Incentives** for reproducible publication practices are currently limited. Research performing organisations are beginning to support researchers in meeting their **growing reproducibility expectations**, and there is increasing demand for new **training** and support **pathways** in this area.
4. The management, curation and **sharing of research data and methods** are necessary conditions for reproducible publication. It is essential for these **practices to become the norm** to push the reproducibility agenda forward, and some dedicated institutional roles such as data stewards may be required to keep up with the demand for support.
5. Reproducible publication practices require a range of **technological solutions**, but most contributors agreed that these are already **available** in today's research landscape. The key technical gap appears to be the **interoperability** between available tools and workflows; however, we also note that technological solutions for reproducibility are not currently covered as part of training curricula.

The purpose of this activity of Knowledge Exchange was to explore current practices and barriers in the area of research reproducibility, with a focus on the publication and dissemination stage. We wanted to determine how technical and social infrastructures can support future developments in this area. In this work, we defined research reproducibility **as cases where data and procedures shared by the authors of a study are used to obtain the same results as in their original work**.

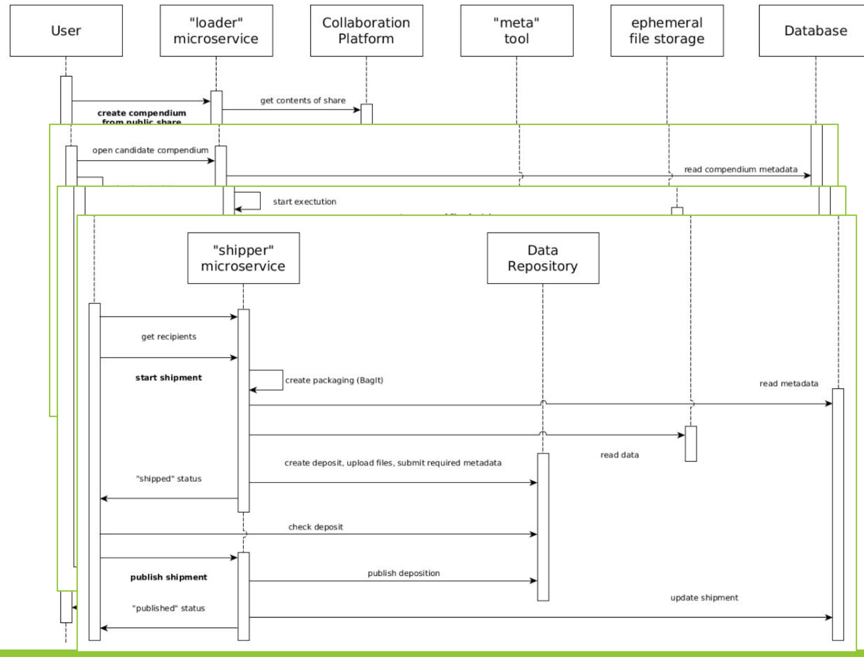
We captured the views of **research funding organisations, research performing organisations, learned societies, researchers, academic publishers and infrastructure and service providers**. We did a comprehensive literature review and a series of interviews and focus groups with a total of 51 contributors. The results of our activity give answers to the following questions:

What are the main benefits and barriers of publishing reproducible research outputs?
What are the roles of the different stakeholders involved?
How expensive are reproducibility checks?
What kind of digital tools and infrastructure are needed to publish reproducible research output?

# o2rX

# ERC creation sequence

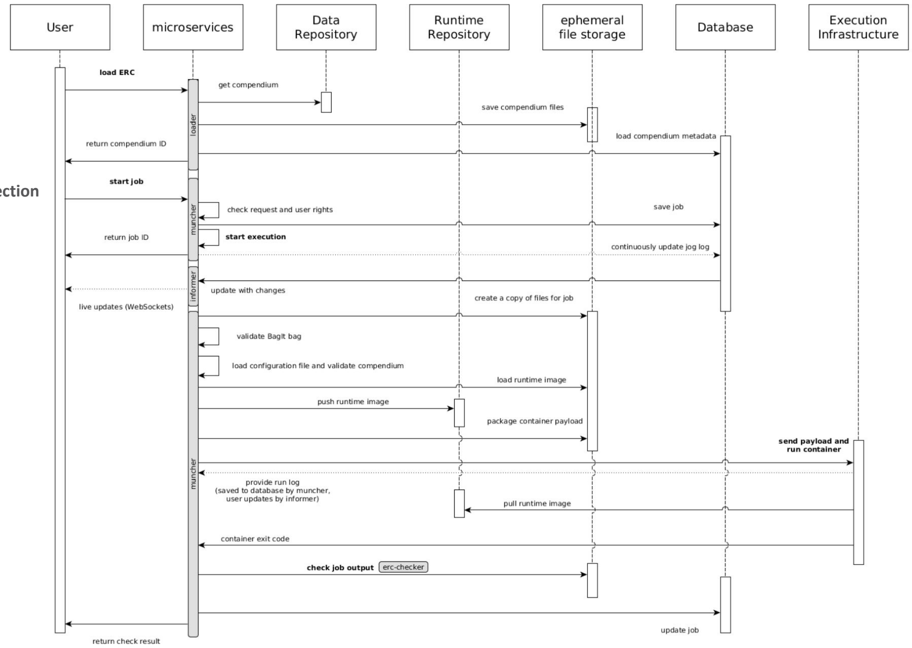
**https://o2r.info/architecture/#61-erc-creation**




Nüst, D. (2021). **A web service for executable research compendia enables reproducible publications and transparent reviews in geospatial sciences**. Zenodo. https://doi.org/10.5281/zenodo.5108218

# ERC examination sequence

**https://o2r.info/architecture/#62-erc-inspection**




Nüst, D. (2021). **A web service for executable research compendia enables reproducible publications and transparent reviews in geospatial sciences**. Zenodo. https://doi.org/10.5281/zenodo.5108218

# ERS Web API

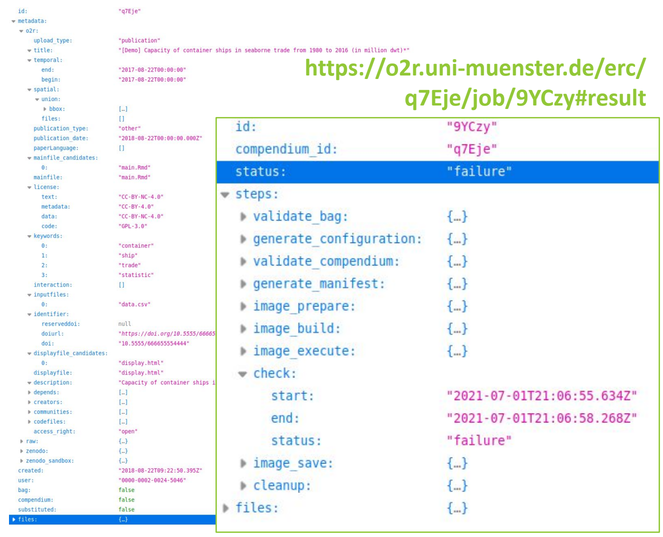
Nüst, D. (2021). **A web service for executable research compendia enables reproducible publications and transparent reviews in geospatial sciences**. Zenodo. https://doi.org/10.5281/zenodo.5108218

OpenAPI Spec: **https://o2r.info/api/**

Demo: **https://o2r.uni-muenster.de/api/v1/**

```
{
  "auth": "/api/v1/auth",
  "compendia": "/api/v1/compendium",
  "jobs": "/api/v1/job",
  "users": "/api/v1/user",
  "search": "/api/v1/search",
  "shipments": "/api/v1/shipment",
  "recipients": "/api/v1/recipient",
  "substitutions": "/api/v1/substitution",
  "links": "/api/v1/link"
}
```

**https://o2r.uni-muenster.de/api/v1/compendium/q7Eje** (/jobs)




- Authentication
- Compendium
- Metadata
- Execution
- Shipment
- Bindings
- Users
- API Info

# Secondary metadata in the ERC & preservation

**More formats, higher chance of long-term meaningful access**

https://o2r.info/erc-spec/spec/#preservation-of-erc

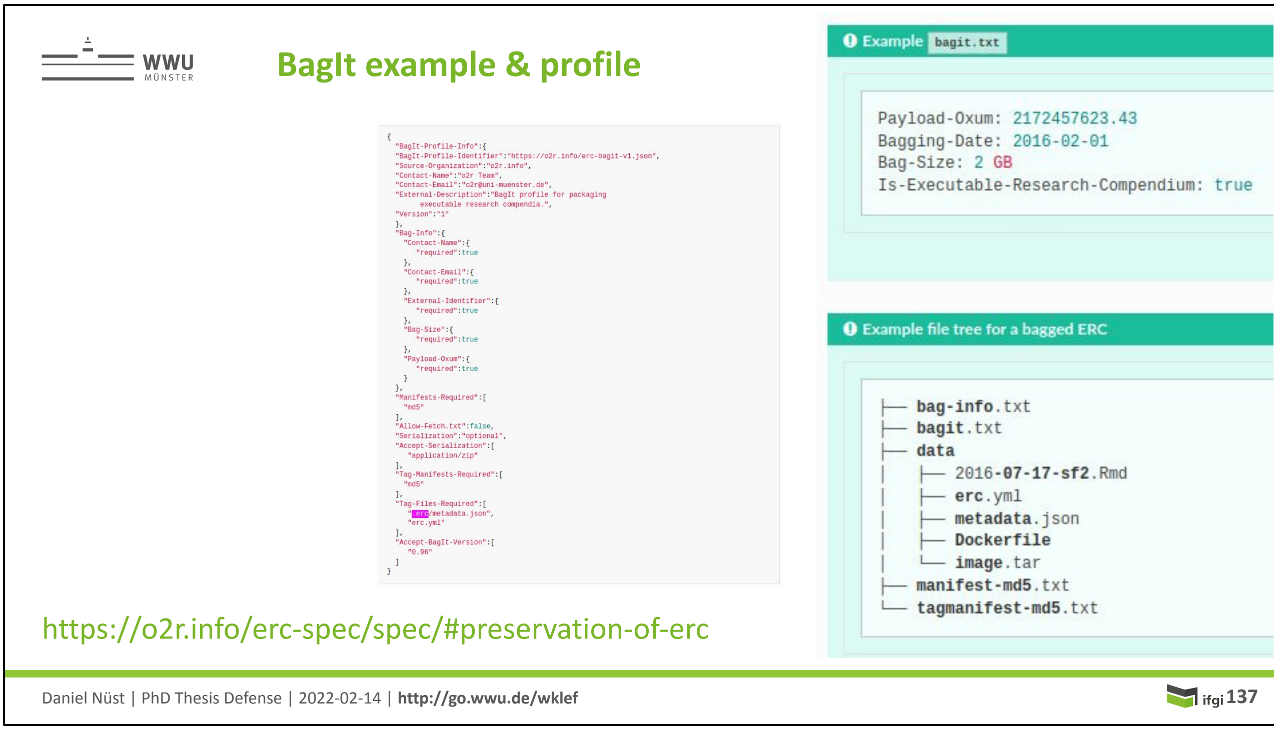
Leaflet

**.erc** folder

- DateCite
- Zenodo
- o2r (extraction, options)

Example package leaflet


```
{
    "standards_used": [
        {
            "o2r": {
                "map_description": "maps raw extracted metadata to
                    o2r schema compliant metadata",
                "mode": "json",
                "name": "o2r",
                "outputfile": "metadata_o2r.json",
                "root": ""
            }
        },
        {
            "zenodo_sandbox": {
                "map_description": "maps o2r schema compliant MD to
                    Zenodo Sandbox for deposition creation",
                "mode": "json",
                "name": "zenodo_sandbox",
                "outputfile": "metadata_zenodo_sandbox.json",
                "root": "metadata"
            }
        }
    ]
}
```




The elements of the o2r Bagit Profile is yet to be specified. This section is under development. Current BagIt tools do not include an option to add a BagIt Profile automatically.

A BagIt Profile as outlined below would make the requirements more explicit. The BagIt Profiles Specification Draft allows users of BagIt bags to coordinate additional information, attached to bags.

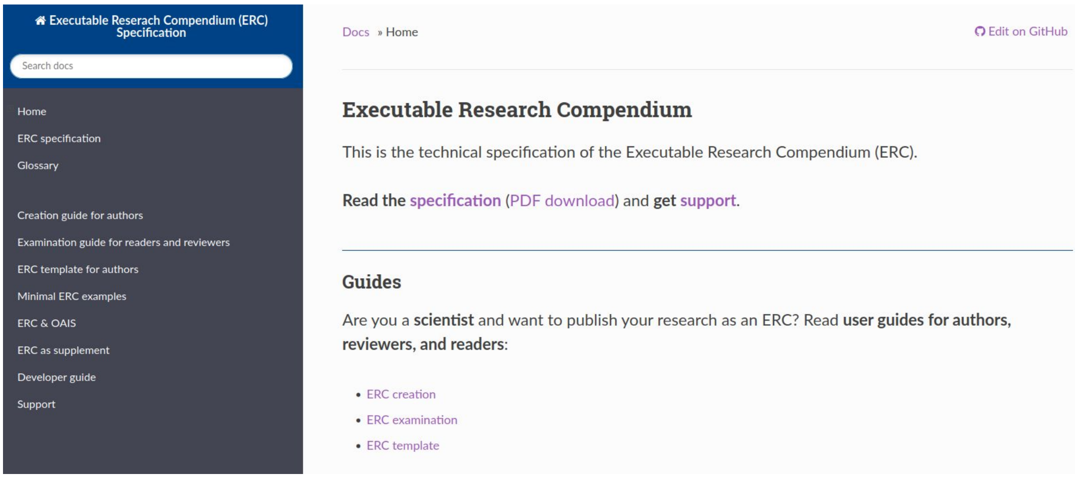
WWU MÜNSTER

# erc.yml

```yaml
id: b9b0099e-9f8d-4a33-8acf-cb0c062efaec
spec_version: 1
main: paper.rmd
display: paper.html
execution:
  bind_mounts: ...
licenses:
  code: MIT
  data: ODbL-1.0
  text: "data_licenses_info.pdf"
  metadata: CC0-1.0
convention: https://github.com/ropensci/rrrpkg
ui_bindings:
  interactive: true
  bindings:
    - purpose: http://.../data-inspection
      widget: http://.../tabular-browser
      code: [...]
      data: [...]
      text: [...]
    - purpose: http://.../parameter-manipulation
      widget: http://.../dropdown
```

# ERC specification

https://o2r.info/erc-spec/

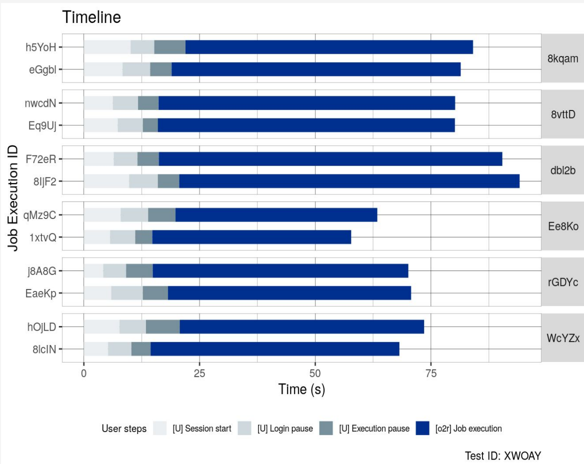


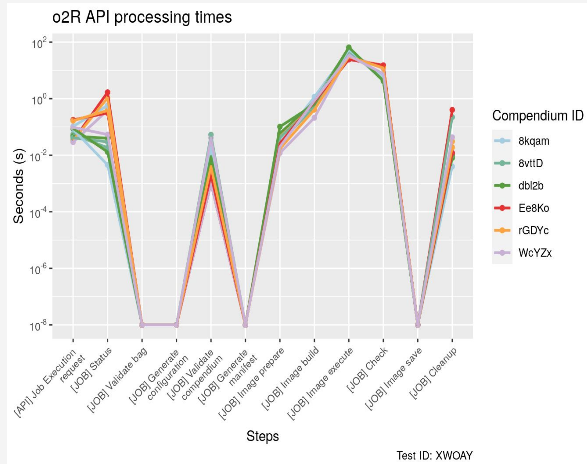


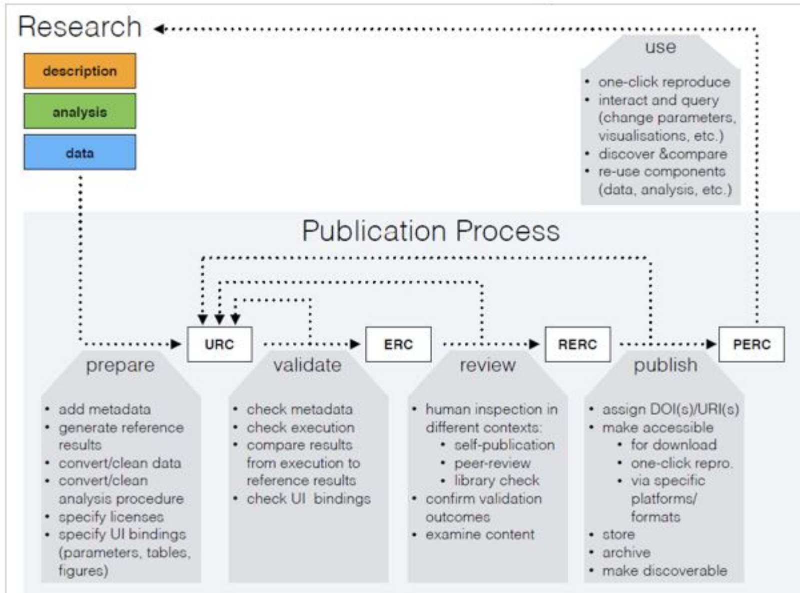

Glossary
User guides
Creation guide for authors
Examination guide for readers and reviewers
ERC template for authors
Minimal ERC examples
ERC & OAIS
ERC as supplement
Developer guide
Support

# Load test of o2r ERS

## https://github.com/o2r-project/api/pull/84






Nüst, D., Konkol, M., Pebesma, E., Kray, C., Schutzeichel, M., Przibytzin, H., & Lorenz, J. (2017). **Opening the Publication Process with Executable Research Compendia**. D-Lib Magazine, 23(1/2). https://doi.org/10.1045/january2017-nuest

## ERC Vision

opening reproducible research https://o2r.info

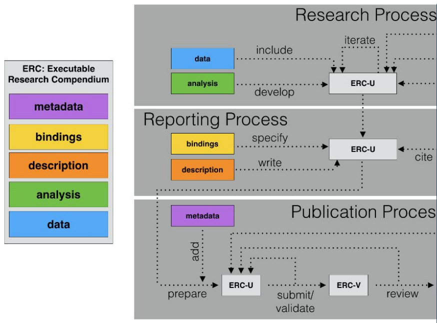



**Figure 1** Executable Research Compendium (ERC) with its five l
ERCs can be integrated into the research, reporting and publicati
stands for an unvalidated ERC, ERC-V for a validated one, ERC-R fo
for a published one. Processes are sequentialised to make the figure

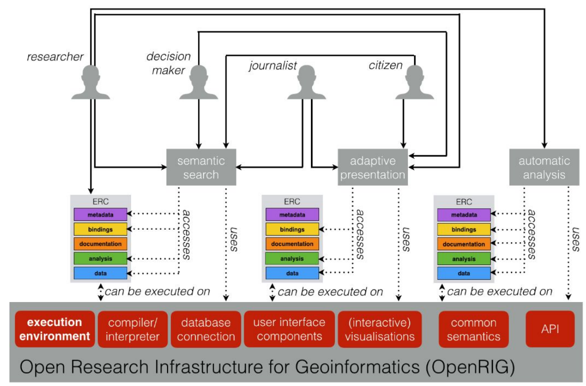



**Figure 2** Open Research Infrastructure for Geoinformatics (OpenRIG): key components (red), essential functionalities enabled by it (grey boxes) and different stakeholders wanting to access them.

Reproducible Research in Geoinformatics: Concepts, Challenges and Benefits (Vision Paper) *Kray C, Pebesma E, Konkol M, Nüst D.* doi:10.4230/LIPIcs.COSIT.2019.8
GenR blog: https://genr.eu/wp/a-vision-for-reproducible-research-in-geoinformatics-geography-and-geosciences/



The ERC

The OpenRIG
perform reasoning

We are part way there to this Open Research Infrastructure for Geoinformatics with the project Opening Reproducible Research and our reproducibility service.

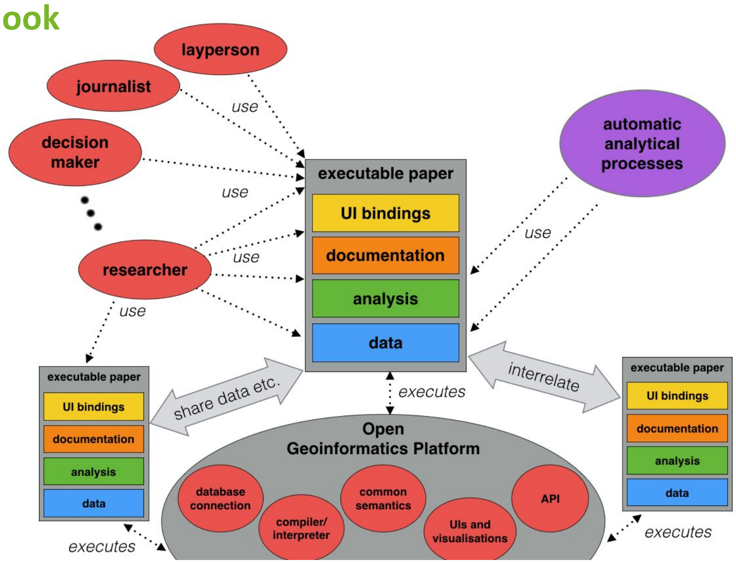
# ERC Vision: Outlook

Explore further options such as other tech interactions

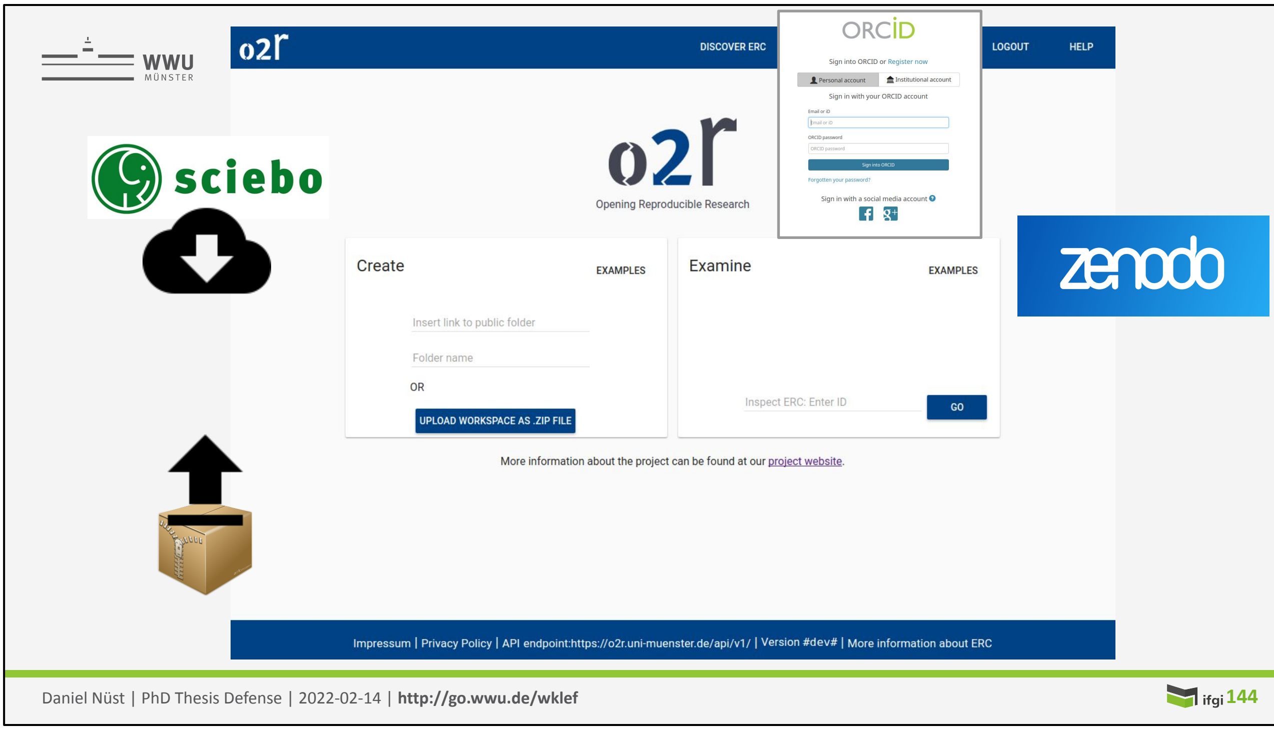
Deploy in practice

Use in teaching

**Towards Vision of Geoinformatics V2**

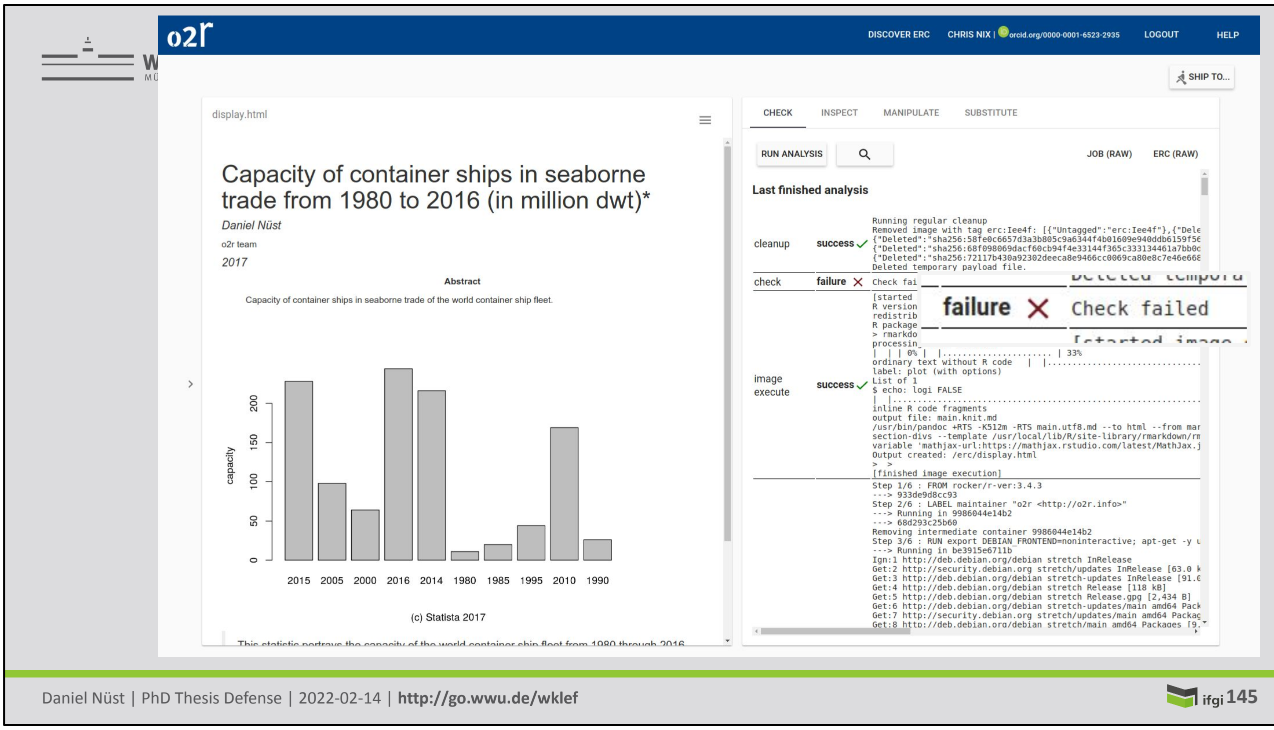



Reproducible Research in Geoinformatics: Concepts, Challenges and Benefits (Vision Paper) *Kray C, Pebesma E, Konkol M, Nüst D.* doi:10.4230/LIPIcs.COSIT.2019.8

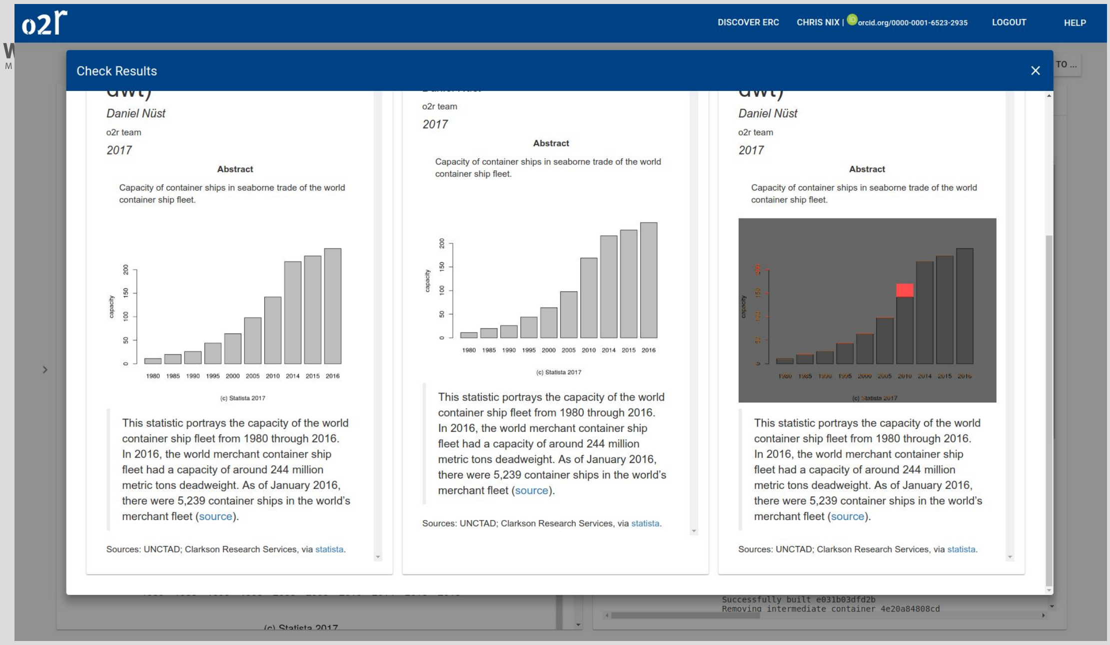



Load with **ZIP**, load from **Sciebo**
**https://uni-muenster.sciebo.de/index.php/s/hAh8AZLYvHNgNA9**

Icons: Material
https://upload.wikimedia.org/wikipedia/commons/thumb/2/20/Ark2_icon.svg/1000px-Ark2_icon.svg.png



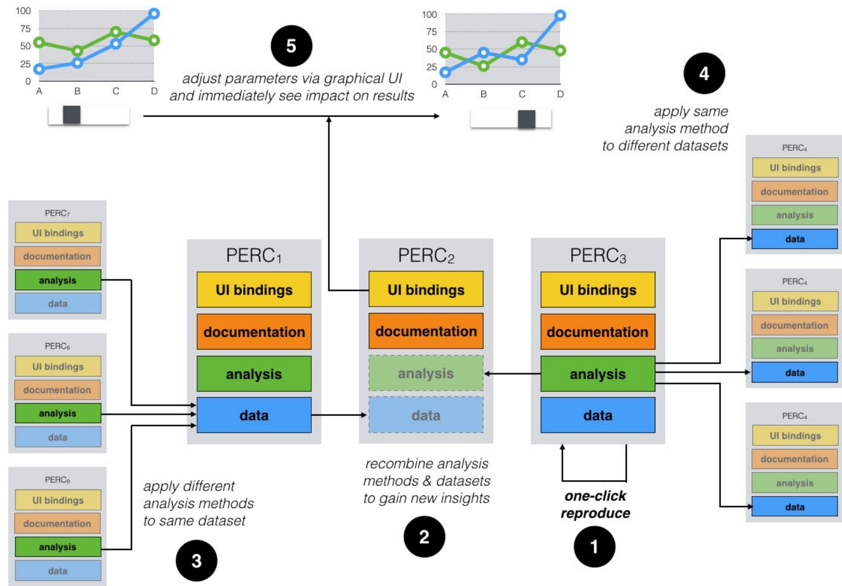

Random effect == check fails > **workspace-rmd-data-random.zip**






Submitted HTML differs from the one produced on the server, with **workspace-rmd-data_wrong-displayfile.zip**

# ERC benefits



# geoextent

**https://o2r.info/geoextent/**

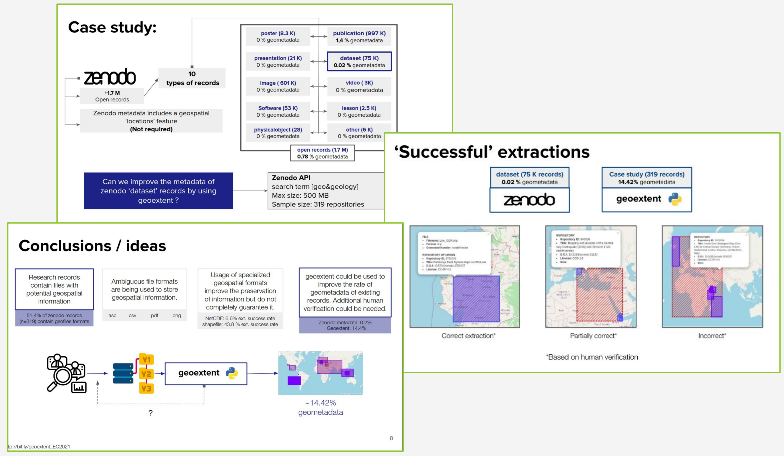
Sebastian Garzón and Nüst, Daniel. 2021. **Exploring Research Data Repositories with geoextent**. EarthCube annual meeting. https://doi.org/10.5281/zenodo.5496311

Extraction of geospatial metadata (spatial and temporal extent) from data files in workspaces submitted to the ERC reproducibility service. *Integrated in ERS as containerised CLI tool via o2r-meta.*

File formats (via GDAL):
GeoJSON, CSV, GeoTIFF, Shapefile, GeoPackage, GPX, GML, KML, (tbc)

```
geoextent -b -t muenster_ring_zeit.geojson
```

```
{'format': 'geojson',
 'geoextent_handler': 'handleVector',
 'tbox': ['2018-11-14', '2018-11-14'],
 'bbox': [7.6016807556152335,
  51.94881477206191,
  7.647256851196289,
  51.974624029877454],
 'crs': '4326'}
```



# o2r meta

https://github.com/o2r-project/o2r-meta

**Extraction** of metadata (publication, geospatial, code, licenses, …) from workspaces submitted to the ERC reproducibility service,

**Mapping** of metadata documents from one schema to another for target systems (Zenodo, archives, …),

**Validation** of metadata, and

**Harvesting** of catalogues for metadata completion (OAI-PMH).

*Integrated in ERC reproducibility service as a containerised CLI tool.*

# ERC in peer review in o2r pilots

**https://o2r.info/pilots/**

Copernicus journal **Earth System Science Data**

Deep-sea sediments of the global ocean by Markus Diesing (Data description paper)
https://essd.copernicus.org/articles/12/3367/2020/essd-12-3367-2020-discussion.html

**Referee comment**
https://essd.copernicus.org/preprints/essd-2020-22/essd-2020-22-RC1.pdf

introduction, data, methods, results, limitations of the approach, potential usage, data availability, and conclusion. These sections are streamlined towards the understanding of the algorithmic implementation and its results; they retain completeness while remaining pleasantly concise, "Limitations of the approach" being the only exception to this. All accompanying figures and tables are clear and understandable, both, in digital form and in paper.

The software was tested for reproducibility using the ERC tool under https://o2r.uni-muenster.de/#/erc/GWME2voTDb5oeaQFuTWMCEMveKS1MiXm, and performed positively in this aspect. Upon closer examination, the discrepancies that led to it being flagged with failed reproducibility multiple times, appear to be minor formatting changes. The data products found under https://doi.pangaea.de/10.1594/PANGAEA.911692 are accessible, complete, and use standard file types.

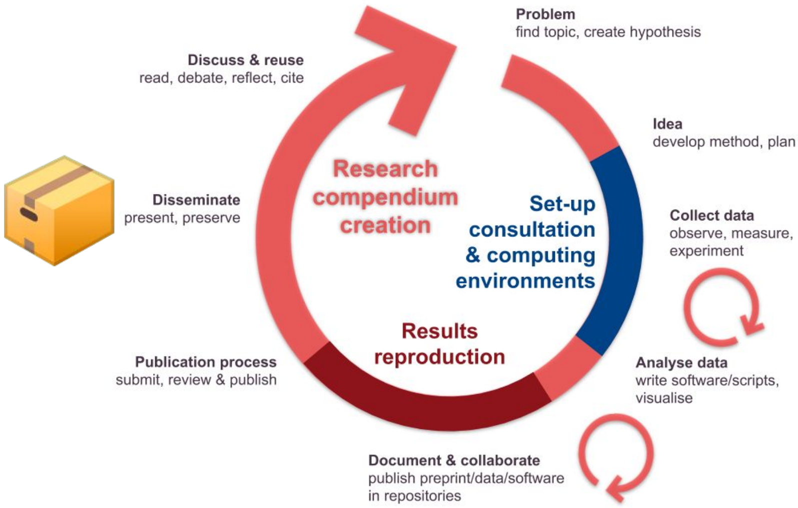
For the most part, the methodology was clearly explained, with enough references to the sources for the used techniques as well as a clear specification of the software

ESSDD

Interactive comment

# R2S2

## Reproducible Research Support Services in the Research Lifecycle

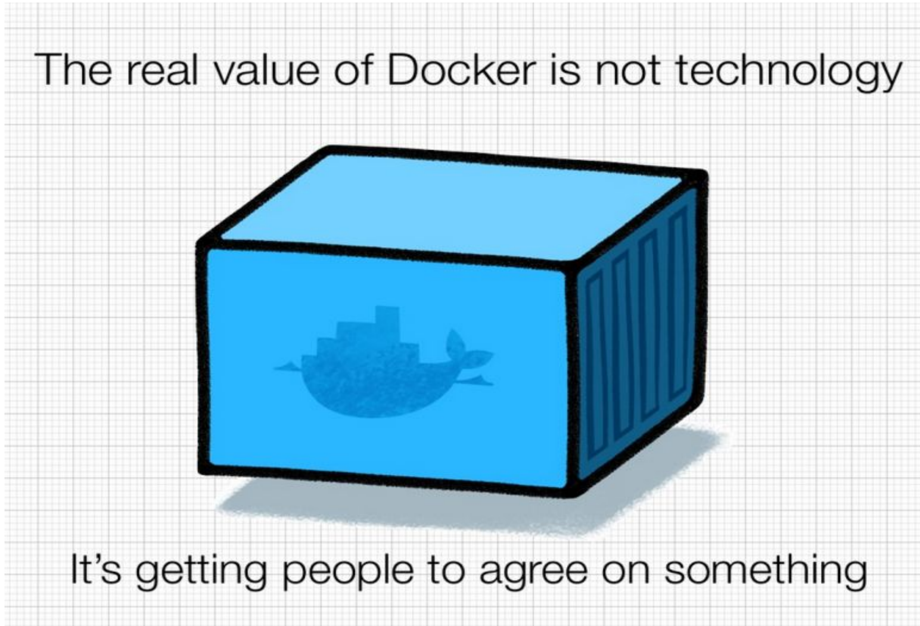



R2S2

**R**eproducible **R**esearch **S**upport **S**ervice

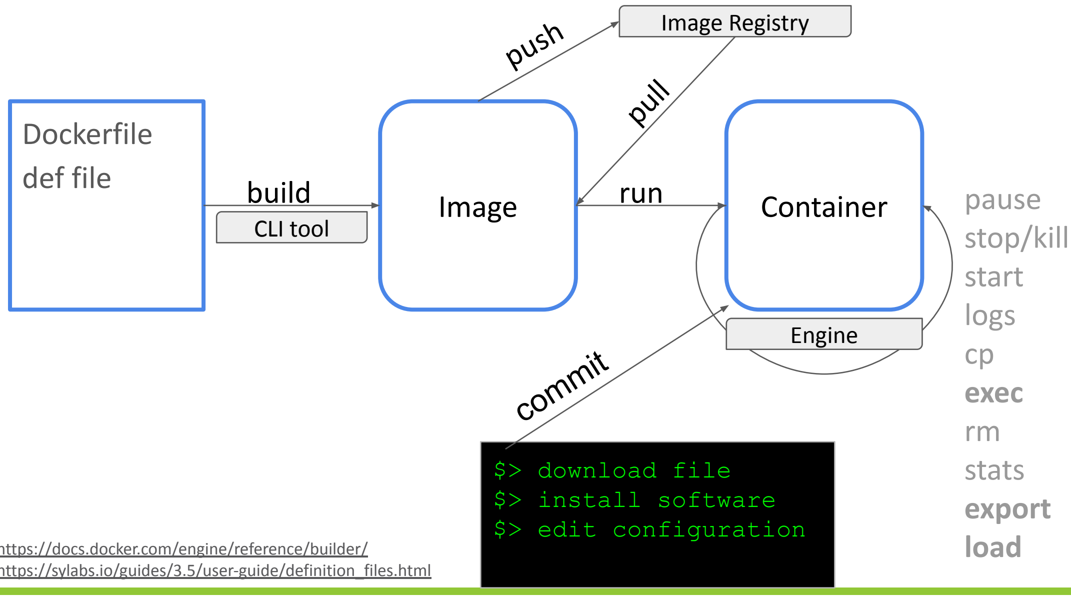
https://go.wwu.de/r2s2

# Containers

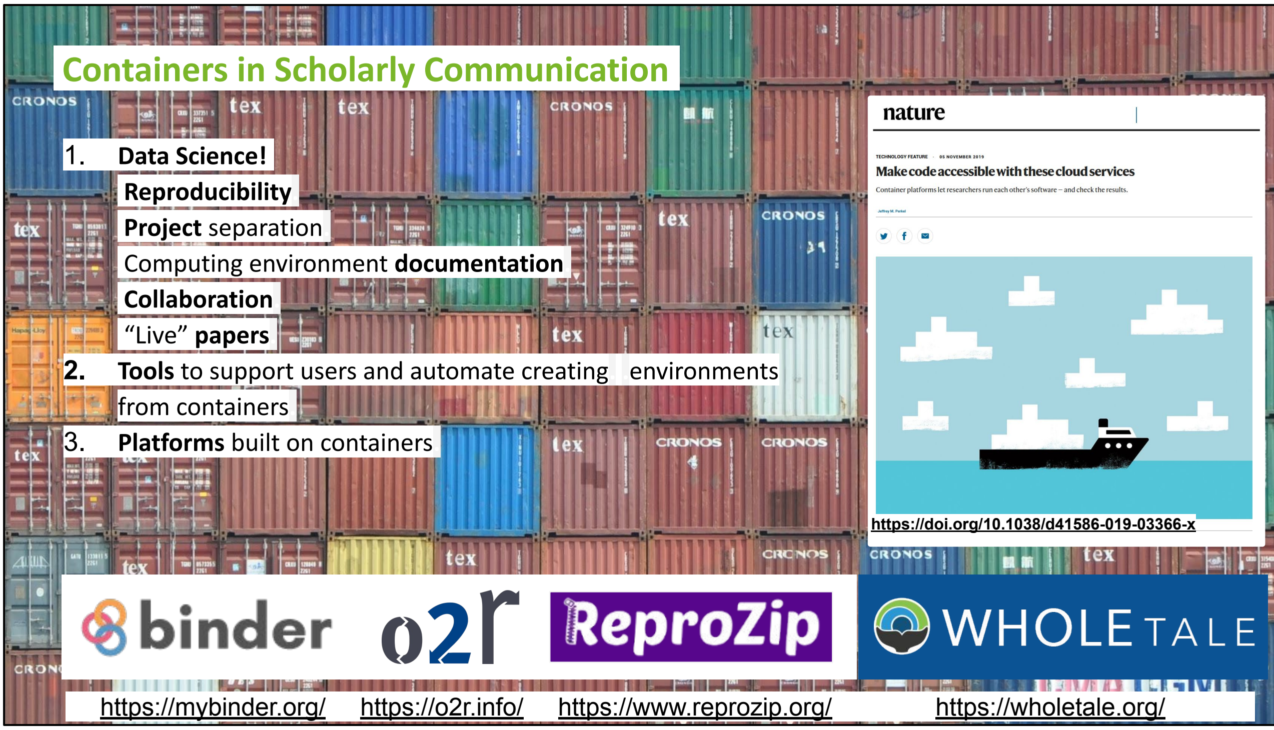



Slide by Docker inventor & Docker, Inc. CTO Solomon Hykes, DockerCon 2014

# Containerisation Basics




https://docs.docker.com/engine/reference/builder/
https://sylabs.io/guides/3.5/user-guide/definition_files.html

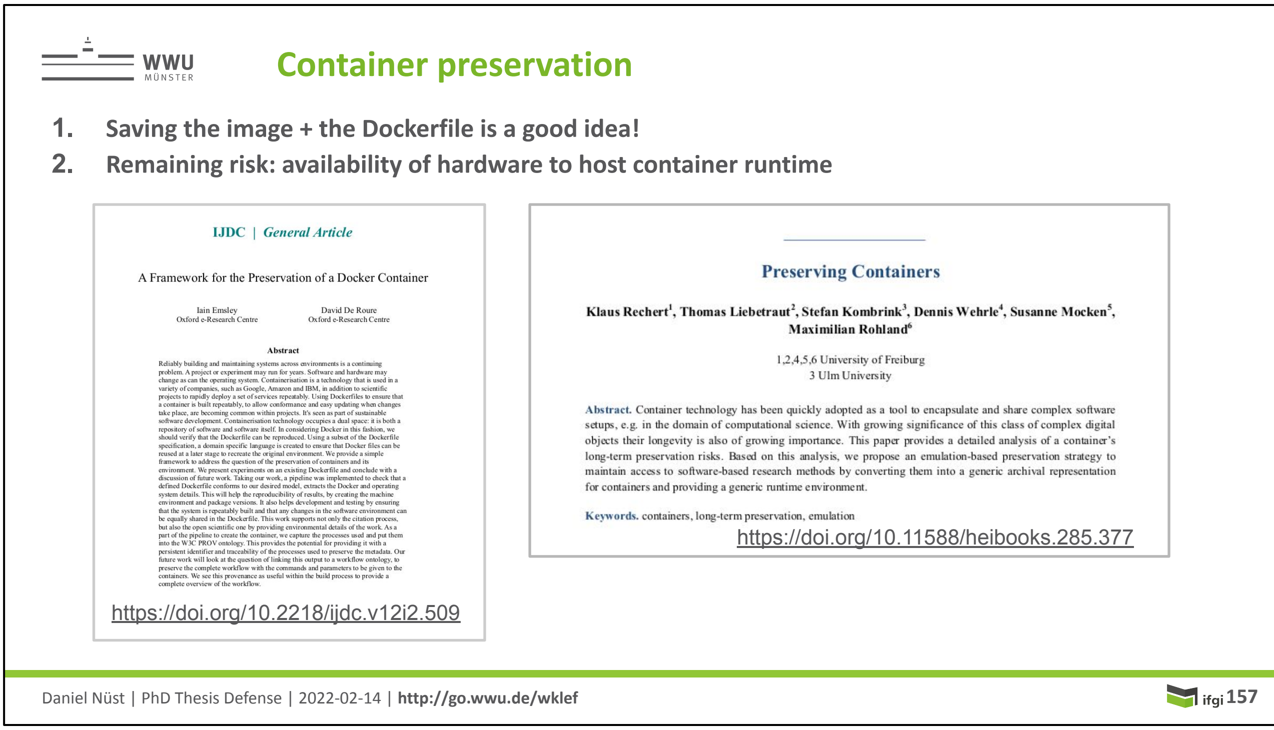



Based on containers, infrastructures are build for researchers to improve communication, collaboration, and reproducibility (cf. https://www.nature.com/articles/d41586-019-03366-x).

With the increasing use of containers they are likely to become a topic for research librarians, albeit their background of scalable cloud information technology being distintcly incompatible with traditional services and perspectives of libraries.

Nevertheless, libraries could provide guidelines for container usage and build infrastructures to leverage the advantages of a containerised workflow, but they might also have to handle container-related artefacts as products of research projects.

https://www.pexels.com/photo/cargo-cargo-container-cargo-containers-container-1563624/

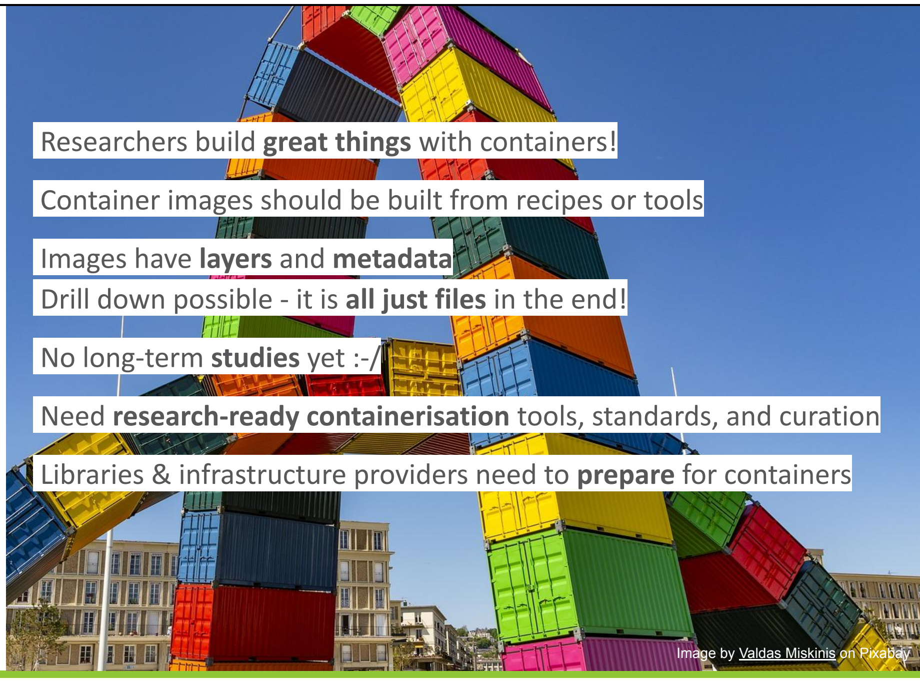



Further I'll give an overview of the state of the art in container preservation (other's work, e.g. https://doi.org/10.2218/ijdc.v12i2.509, Rechert et al. in https://doi.org/10.11588/heibooks.285.377 ).


**Solutions: researchers trust libraries!**
**Author guidelines are the only time when researchers are willing to do as others say!**

Based on my perspective as a research software engineer, I will speculate on the potential and the challenges for container archival and preservation, and how in ten years a container image might help the inspection of research published today.

WWU MÜNSTER



Researchers build **great things** with containers!

Container images should be built from recipes or tools

Images have **layers** and **metadata**

Drill down possible - it is **all just files** in the end!

No long-term **studies** yet :-/

Need **research-ready containerisation** tools, standards, and curation

Libraries & infrastructure providers need to **prepare** for containers

Image by Valdas Miskinis on Pixabay

# Abstracts

WWU MÜNSTER

# Abstract (for indexing and search)


Reproducibility of computational research, i.e., research based on code and data, poses enormous challenges to all branches of science. In this dissertation, technologies and practices are developed to increase reproducibility and to connect it better with the process of scholarly communication with a particular focus on geography, geosciences, and GIScience. Based on containerisation, this body of work creates a platform that connects existing academic infrastructures with a newly established executable research compendium (ERC). It is shown how the ERC can improve transparency, understandability, reproducibility, and reusability of research outcomes, e.g., for peer review, by capturing all parts of a workflow for computational research. The core part of the ERC platform is software that can automatically capture the computing environment, requiring authors only to create computational notebooks, which are digital documents that combine text and analysis code. The work further investigates how containerisation can be applied independent of ERCs to package complex workflows using the example of remote sensing, to support data science in general, and to facilitate diverse use cases within the R language community. Based on these technical foundations, the work concludes that functioning practical solutions exist for making reproducibility possible through infrastructure and making reproducibility easy through user experience. Several downstream applications built on top of ERCs provide novel ways to discover and inspect the next generation of publications.

To understand why reproducible research has not been widely adopted and to contribute to the propagation of reproducible research practices, the dissertation continues to investigate the state of reproducibility in GIScience and develops and demonstrates workflows that can better integrate the execution of computational analyses into peer review procedures.

We make recommendations for how to (re)introduce reproducible research into peer reviewing and how to make practices to achieve the highest possible reproducibility normative, rewarding, and, ultimately, required in science. These recommendations are rest upon over 100 GIScience papers which were assessed as irreproducible, the experiences from over 30 successful reproductions of workflows across diverse scientific fields, and the lessons learned from implementing the ERC.

Besides continuing the development of the contributed concepts and infrastructure, the dissertation points out broader topics of future work, such as surveying practices for code execution during peer review of manuscripts, or reproduction and replication studies of the fundamental works in the considered scientific disciplines. The technical and social barriers to higher reproducibility are strongly intertwined with other transformations in academia, and, therefore, improving reproducibility meets similar challenges around culture change and sustainability. However, we clearly show that reproducible research is achievable today using the newly developed infrastructures and practices. The transferability of cross-disciplinary lessons facilitates the establishment of reproducible research practices and, more than other transformations, the movement towards greater reproducibility can draw from accessible and convincing arguments both for individual researchers as well as for their communities.

# Zusammenfassung

Die Reproduzierbarkeit von rechnergestützter Forschung stellt alle Wissenschaftszweige vor enorme Herausforderungen. In dieser Dissertation werden Technologien und Praktiken entwickelt, um die Reproduzierbarkeit zu erhöhen und sie besser mit dem Prozess der wissenschaftlichen Kommunikation zu verbinden, mit besonderem Fokus auf Geographie, Geowissenschaften und GIScience. Basierend auf Containerisierung wird in dieser Arbeit eine Plattform geschaffen, die bestehende akademische Infrastrukturen mit einem neuartigen ausführbarem Forschungskompendium (Executable Research Compendium; ERC) verbindet. Es wird gezeigt, dass das ERC die Transparenz, Verständlichkeit, Reproduzierbarkeit und Wiederverwendbarkeit von Forschungsergebnissen, zum Beispiel für Peer-Reviews, verbessert, indem es alle Teile eines computergestützten Arbeitsablaufs erfasst. Das Kernstück der ERC-Plattform ist eine Software, welche die Rechenumgebung automatisch erfassen kann, so dass die Autoren nur noch sogenannte computational notebooks, digitale Notizbücher die Text und Analysecode verbinden, erstellen müssen. Die Arbeit untersucht weiter, wie Containerisierung unabhängig von ERCs angewendet wird und werden kann, unter anderem bei einer komplexen Analyse aus der Fernerkundung, für Datenwissenschaften im Allgemeinen sowie innerhalb der Anwenderschaft der Programmiersprache R. Basierend auf diesen technischen Grundlagen kommt die Arbeit zu dem Schluss, dass es funktionierende praktische Lösungen gibt, die Reproduzierbarkeit durch geeignete Infrastruktur möglich machen und die Benutzung deutlich vereinfachen. Mehrere nachgelagerte Anwendungen, die auf ERCs aufbauen, bieten neuartige Möglichkeiten, die nächste Generation von Publikationen besser suchen und inspizieren zu können.

Um zu verstehen, warum reproduzierbare Forschung nicht weit verbreitet ist, und um zur Verbreitung reproduzierbarer Forschungspraktiken beizutragen, untersucht die Dissertation weiterhin den Stand der Reproduzierbarkeit in der wissenschaftlichen Disziplin GIScience. Sie entwickelt und demonstriert Arbeitsabläufe, mit welchen die Durchführung von rechnerischen Analysen besser in Peer-Review-Verfahren integriert werden können. Es werden Empfehlungen gegeben, wie reproduzierbare Forschung in Peer-Review-Verfahren (wieder) eingeführt werden kann und wie Praktiken um die höchstmögliche Reproduzierbarkeit zu erreichen in der Wissenschaft normativ, lohnend und letztlich verpflichtend werden können. Diese Empfehlungen stützen sich auf über 100 als irreproduzierbar befundenen Artikeln aus der GIScience, auf die Erfahrungen aus über 30 erfolgreichen Reproduktionen von computerbasierten Arbeitsabläufen in verschiedenen Wissenschaftsbereichen und auf die Erkenntnisse von der Implementierung des ERC.

Neben der Weiterentwicklung der eingebrachten Konzepte und der Infrastruktur weist die Dissertation auf weitergehende Themen zukünftiger Arbeit hin, wie zum Beispiel die Untersuchung von Prozessen für Code-Ausführung als Teil von Begutachtungen von Manuskripten, oder Reproduktions- und Replikationsstudien für grundlegende Arbeiten in den betrachteten Wissenschaftsdisziplinen. Die technischen und sozialen Barrieren für höhere Reproduzierbarkeit sind stark mit anderen Transformationsprozessen in der Wissenschaft verwoben und daher trifft die Verbesserung der Reproduzierbarkeit auf ähnliche Herausforderungen rund um Kulturwandel und Nachhaltigkeit. Die Arbeit zeigt jedoch klar, dass reproduzierbare Forschung jedoch schon heute auf Basis der neu entwickelten Infrastrukturen und Praktiken realisierbar ist. Die Übertragbarkeit von disziplinübergreifenden Erkenntnissen begünstigt die Etablierung reproduzierbarer Forschungspraktiken, und mehr als andere Transformationen kann die Bewegung hin zu mehr Reproduzierbarkeit aus zugänglichen und überzeugenden Argumenten sowohl für einzelne Forscher als auch für ihre Gemeinschaften schöpfen.

# Acknowledgements & Thanks

## Thanks!

Throughout the writing of this dissertation I have received a great deal of support, met interesting people, enjoyed collaborating with many different minds, and even made new friends. This dissertation bears one name but captures the efforts of many. I would first like to thank my supervisor, **Edzer Pebesma**, who not only trusted me with an exciting project to work on, but also always helped me with insightful feedback, patience, and wisdom when I needed it, but mostly gave me more leeway than any young scholar would dare to ask for. Thank you Edzer, I'm very lucky to have you as my boss! I would like to acknowledge my colleagues at the **o2r project, ifgi, and WWU** whose expertise and personalities made our little attempt at changing how science is done both educational and fun, be they professor, librarian, researcher, or student. I will do some people injustice by omission, but nevertheless would like to thank a few people by name: **Markus Konkol**, who sharing an office with made the good times of getting a PhD enjoyable and the bad times bearable. Thank you dear **Mark Schutzeichel, Christian Kray, Christian Knoth, Marius Appel, Thomas Bartoschek, Holger Przibytzin, Jörg Lorenz, my fellow GSGI students of all years, Matthias Mohr, Fabian Fermazin, Juan Sebastian Garzón Alvarado, Laura Goulier, Matthias Hinz, Nick Jakuschona, Jan Koppe, Timm Kühnel, Torben Kraft, Lukas Lohoff, Tom Niers, Jan Suleiman, and Yousef Qamaz**. I am extremely lucky to not only have colleagues in the same institute, but that I can rely on excellent and enjoyable collaborators from all over the world. I was not good at setting up a straight line to follow, but in all detours and digressions I learned a lot and I am very appreciative of the opportunities and leaps of faith that were extended to me. The papers and events I could contribute to, make me proud. Big thanks for the great times go to **Stephen Eglen, Frank O. Ostermann, Barbara Hofer, Carlos Granell, Rusnė Šilerytė, Vicky & Remi Rampin, Vanessa Sochat, Heidi Seibold, Dirk Eddelbuettel, Ben Marwick, Marta Teperek, Anita Graser, Karl Broman, Niels Drost, David Topping, Lesley Wyborn, Xenia van Edig, Martin Rasmussen, Dirk Fleischer, Tim Head, Tony Hirst, Ben Evans, Bernadette Fritzsch, Martin Hammitzsch, Peter Kedron, Werner Kuhn, Alexis Comber, Jeremy Leipzig, Carl Boettiger, Alexander Kmoch, and the eLife sprints' teams and participants (Thanks, Naomi & Emmy)**. It is a testament to Open Science's power and its brilliant inspirational kind communities, such as the Binder team, that the list above is incomplete. Like-minded researchers and developers from many different disciplines, backgrounds, and locations contributed to big parts of this dissertation. I would also like to thank **Celeste R. Brennecka** for proofreading almost all manuscripts and helping me to find better words. I would not have had the energy to pursue a career in research to this point without the awesome local and national **Ultimate Frisbee** communities, the great teams I could play with (InDISCutabel, UMS, Monster Mix, Deine Mudder Bremen), and the many friends I found while throwing and catching discs. Ultimate delivered a way to rest my mind outside of my research. You should try it, too! In particular, I thank the **Braun family** and **Katja & Simon** for generously welcoming me in their homes in Münster whenever I needed it. Much closer to work, the RSE community is the best support network and inspirational crowd of nerds there is—we will change how things are done to the better! To all my colleagues, collaborators, professional contacts, who luckily not always stayed strictly professional, and those of you who I missed to recognise here (Sorry!), I would like to restate my thanks—it has been an honour and a privilege. Finally and most importantly, I would like to thank my beloved wife **Maria** for her unconditional support, sympathetic ear, and wise counsel. And on top of that, you even helped with this dissertation—I am grateful, now if not then, for every little and much needed push. I also want to thank my parents, my sister, my extended **family**, and my closest friends who are always there for me and are always understanding—I could not have completed a single bit of this research work without your backing. Danke! This work is dedicated to you.

Big thanks go also to the second assessor **Ben Marwick** and the committee members Professors **Daniel Sui, Christian Kray, and Norbert Hölzel**.