Publishing Reproducible Geoscientific Papers: Status quo, benefits, and opportunities
Markus Konkol

https://o2r.info/
Twitter: @o2r_project, @MarkusKonkol
https://github.com/o2r-project
Open reproducible research

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

Replicable research refers to coming to similar conclusions based on newly collected data or a newly implemented analysis.

Replicability & reproducibility are essential for scientific work.

Reproducible research refers to achieving exactly the same results (e.g. tables, figures, numbers) as reported in the paper by using the same source code and data. In Open reproducible research, these materials are publicly accessible.
Why is unreproducible research a problem?

- Difficult to find errors in the analysis
- Reviewers cannot verify but need to trust the results
  - Extra effort from authors and reviewers required
- Analysis not fully understandable
- Materials not reusable (sustainable)

Computational geosciences: papers based on code/data
What is the status quo of open reproducible research in the computational geosciences?
What is the status quo of open reproducible research in the computational geosciences?

Which incentives are provided by open reproducible research beyond re-computable results?
➢ What is the status quo of open reproducible research in the computational geosciences?
➢ Which incentives are provided by open reproducible research beyond re-computable results?
➢ How can we assist geoscientists in publishing open reproducible research?
What is the status quo of open reproducible research in the computational geosciences?

Which incentives are provided by open reproducible research beyond re-computable results?

How can we assist geoscientists in publishing open reproducible research?

How can we create interactive geoscientific papers to realize the incentives provided by ORR?
What is the status quo of open reproducible research in the computational geosciences?

Which incentives are provided by open reproducible research beyond re-computable results?

How can we assist geoscientists in publishing open reproducible research?

How can we create interactive geoscientific papers to realize the incentives provided by ORR?

Outlook

Open questions
How can we assist geoscientists in publishing open reproducible research?

- What is the status quo of open reproducible research in the computational geosciences?
- Which incentives are provided by open reproducible research beyond re-computable results?
- How can we assist geoscientists in publishing open reproducible research?
- How can we create interactive geoscientific papers to realize the incentives provided by ORR?

Outlook

Open questions
What is the status quo of ORR in the computational geosciences?
The status quo of ORR in the comp. geosciences

- 49% say they publish reproducible research results often or always
The status quo of ORR in the comp. geosciences

- 49% say they publish reproducible research results often or always
- 33% say their papers include links to datasets often or always

The status quo of ORR in the comp. geosciences

- 49% say they publish reproducible research results often or always
- 33% say their papers include links to datasets often or always
- 12% say their papers include links to source code often or always


The status quo of ORR in the comp. geosciences

- 49% say they publish reproducible research results often or always
- 33% say their papers include links to datasets often or always
- 12% say their papers include links to source code often or always
- 7% say they try to reproduce others’ results often or always

The status quo of ORR in the comp. geosciences

Cultural issues

Obstacles while authors publish reproducible research: e.g. no incentives for making code/data accessible, missing tools

Obstacles while readers reproduce others' work: e.g. missing materials, too much effort

Required changes in the scientific culture

- Better citation systems/acknowledgement for code/data
- Handling errors in papers: should not end a career
  - But: ERCs help to detect errors earlier (during review)
- “Leaving the comfort zone”: Just publishing PDF files without code and data was accepted but it is not enough
The status quo of ORR in the comp. geosciences

Technical issues

- Minor, e.g. when a library was not installed
- Substantial, e.g. a wrong file directory
- Severe, e.g. a flawed functionality
- Sys.-dependent, e.g. different software versions

The status quo of ORR in the comp. geosciences

The status quo of ORR in the comp. geosciences

Status quo?
Which incentives are provided by open reproducible research beyond re-computable results?
Incentives for publishing ORR

ORR can help you to...

- ...make your research better findable

Incentives for publishing ORR

ORR can help you to...

- ...make your research better findable
- ...show reviewers how specific results were achieved

Incentives for publishing ORR

ORR can help you to...

- ...make your research better findable
- ...show reviewers how specific results were achieved
- ...create interactive figures for changing parameter values

Incentives for publishing ORR

ORR can help you to...

- ...make your research better findable
- ...show reviewers how specific results were achieved
- ...create interactive figures for changing parameter values
- ...make your materials reusable (also for yourself)

How can we assist geoscientists in publishing open reproducible research?
Assisting geoscientists in publishing ORR

Executable Research Compendium

Assisting geoscientists in publishing ORR

Executable Research Compendium

Data:
- Ideally as raw data
- Included as a file

Assisting geoscientists in publishing ORR

Assisting geoscientists in publishing ORR


Executable Research Compendium

Documentation:
- Article
- Metadata

Data:
- Ideally as raw data
- Included as a file

Software:
- Entire runtime environment
- Code scripts
- 1-click reproduce

documentation
software
data
Assisting geoscientists in publishing ORR

How can we create interactive geoscientific papers to realize the incentives provided by ORR?
Creating interactive scientific papers

Idea:
Connecting only those source code lines and data subsets that are needed to produce a specific computational result.


```r
velocity = 2.0 # m/s
ComputeDamage <- function (velocity) {
  # do some analysis
  printFigure1()
}
```
Creating interactive scientific papers

A binding describes which source code lines and data subsets were used to produce an individual computational result, such as a figure, table, or number in the text.

Creating interactive scientific papers

A **binding** describes which source code lines and data subsets were used to produce an individual computational result, such as a figure, table, or number in the text.

It explicitly refers to specific parameters in the code.
Creating interactive scientific papers

A **binding** describes which source code lines and data subsets were used to produce an individual computational result, such as a figure, table, or number in the text.

It explicitly refers to specific parameters in the code.

A **binding** can be used to create an interactive figure with the help of user interface widgets, e.g. a slider.

Konkol et al. (2019). Creating interactive scientific publications using bindings. [https://doi.org/10.1145/3331158](https://doi.org/10.1145/3331158)
Creating interactive scientific papers

A **binding** describes which **source code lines and data subsets** were used to produce an **individual computational result**, such as a figure, table, or number in the text.

It explicitly refers to specific **parameters** in the code.

A **binding** can be used to create an interactive figure with the help of **user interface widgets**, e.g. a slider.

Konkol et al. (2019). Creating interactive scientific publications using bindings. [https://doi.org/10.1145/3331158](https://doi.org/10.1145/3331158)
A binding describes which source code lines and data subsets were used to produce an individual computational result, such as a figure, table, or number in the text. It explicitly refers to specific parameters in the code.

A binding can be used to create an interactive figure with the help of user interface widgets, e.g. a slider.

Creating interactive scientific papers (author’s perspective)

Using interactive scientific papers
(reader’s perspective)

Konkol et al. (2019). Creating interactive scientific publications using bindings. [https://doi.org/10.1145/3331158]
Publication process

Outlook

Second phase:
- 2.5-year project, 2 RAs
- Collaboration between ULB, ifgi and publishers

Goal 1
- Pilot applications
  - collaboration with journals
  - implementation of UI and the repro. services
  - Self-hosted pilot
    - Open journal system plugin
    - Host OJS instance
    - ERC @ education
**Outlook**

**Goal 2:** Eliminate barriers  
- creating bindings  
- robust UI  
- Update specification and documentation

**Goal 3:** Evaluation  
- technology: stress tests, monitoring  
- user study about the understanding of ERCs

**Beyond:**  
- reproducible infrastructure @ WWU  
- “more geo” in schol. comm. platforms
Open Research Infrastructure for Geoinformatics

- Hub for ERCs
- Comparing ERCs
- Semantic search
- Harmonise figures across ERCs
- Adapt visualizations to needs of different stakeholders
- Combine parts of ERCs

5 Take-home messages in 30 seconds
Take-home messages

- Accessible code/data is not necessarily **reproducible**
- Incentives for publishing ORR outweigh the extra effort
- ERCs make research verifiable (one-click reproduce)
- Bindings help to understand how parameters affect results
- PDFs are not suitable for communicating (geo)scientific comp. research (particularly when published as ERCs)
Some take-home questions
Open questions

● How to handle biased data and code?
● How about privacy?
● Is more scientific output (papers) a desired goal?
● Does reproducibility lower the motivation for replicating data/code?