

# Packaging geoscience research for transparency and reproducibility: approaches and examples

**Daniel Nüst**  
Opening Reproducible Research (o2r)  
Institute for Geoinformatics, University of Münster  
Münster, Germany  
d.n@wwu.de

@nordholmen



https://o2r.info

doi:10.5281/zenodo.2630483



## Advantages


### Encourages good habits

- Structure helps to work effectively, ideally **community/peer's conventions** are used
- Tools & workflows for Openness often increase **efficiency** at the same time
- **Tidy** well-documented **virtual laboratory** (keep input, methods, output separate)
- Packaging process can even simulate and **independent reproduction** (starting from scratch in container)
- **Quality** of work improves with good programming practices, (perceived) overhead pays off quickly
- **Prepare** for requirements by funders and journals (**Open Science** will become Science)

### Improves transparency

- Easier to **understand** (readers, students, self)
- More **convincing** and **inspectable** (reviewers)
- Higher trust and **confidence** in results because of independent re-execution

### Enables collaboration

- Future **you** is your best collaborator! 
- Data, methods, and software are demonstrated to work, which improves **reuse** and **discoverability**
- Extensibility through good practices

## Examples

[Images from left to right.]

### Geosimulation model with PCRaster/Python

https://github.com/JudithVerstegen/PLUC\_Mozambique/

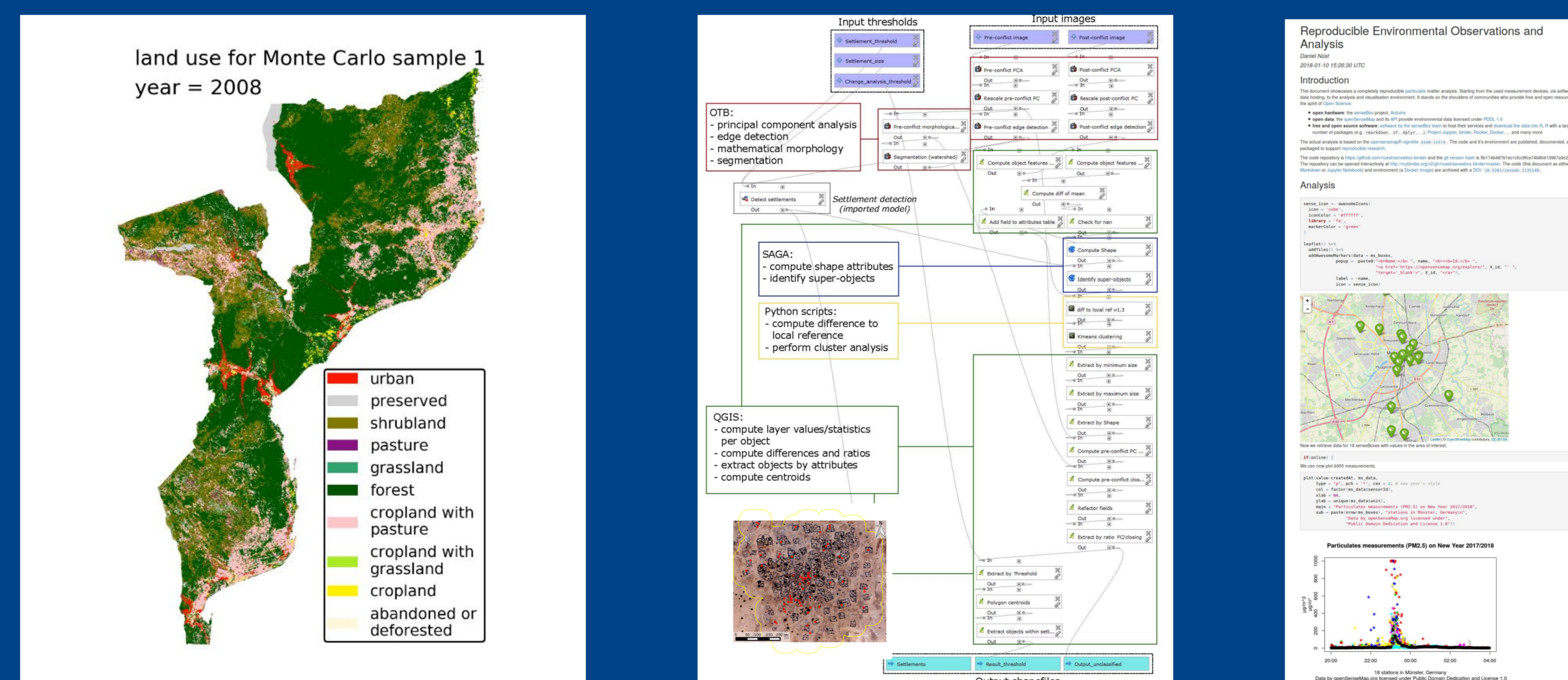
### GEOBIA workflow

doi:10.3390/rs9030290

### Environmental citizen science

doi:10.5281/zenodo.1135139

# Packaging research\* encourages good habits, improves transparency, and enables collaboration

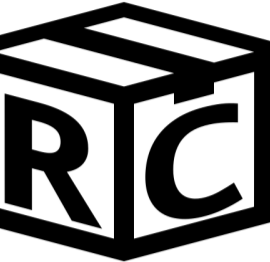


\* As research compendium, ERC, Docker/Singularity image/container, VM, Binder, ReProZip package, Tale, Compute Capsule, ...

## Containerisation/VMs & UI

- x11docker makes sharing of display between host and container simple and secure (Linux)
- Web-based UI (incl. API) best
- Special UI encourages **deliberate communication**
- Containers scripted (Dockerfile) and better for reuse and "making"
- VMs larger, good for dissemination, but for **use** only... when did the Kernel ever break?

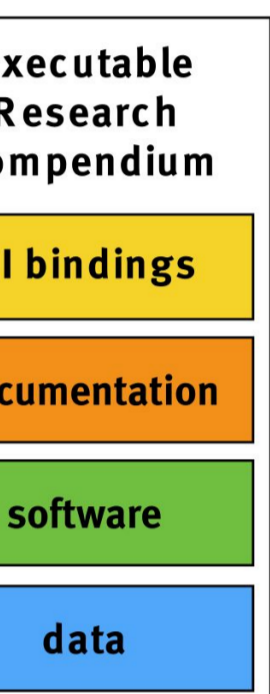
## Research Compendium



https://research-compendium.science/

- Accompanies/enhances/is a scientific publication
- Container for data, code, docs, notebooks; under clear licenses; distribution & managing
- "package-pendium": based on programming language modules, e.g. PyPI, CRAN, npm.

## Executable Research Compendium (ERC)



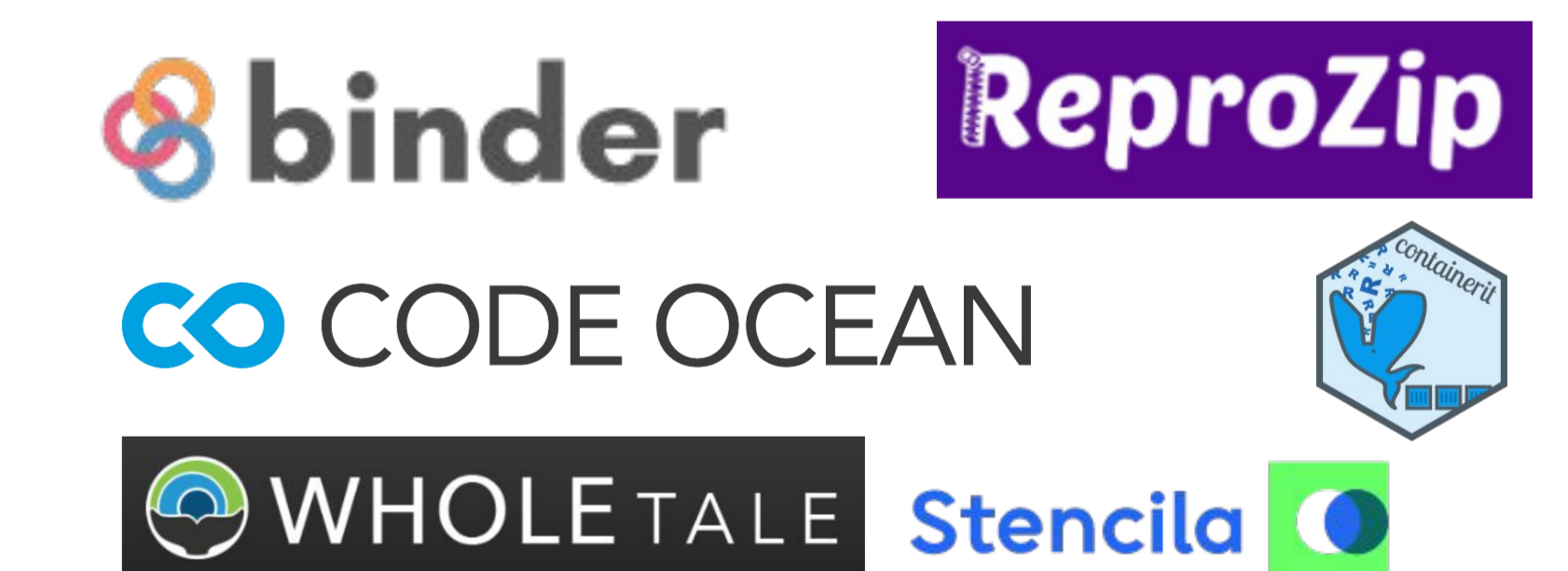
https://o2r.info/erc-spec/

- **Notebook** serves as control file and input for display file (dual entry points)
- Software includes container for **runtime environment**
- UI bindings provide **interactivity & linking** of parts (Konkol et al.)

*"An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship."*

Buckheit & Donoho (1995) doi:10.1007/978-1-4612-2544-7\_5

## Assisted Containerisation



## More & References

- Brown, C. T., 2012. "Virtual machines considered harmful for reproducibility". <http://ivory.idyll.org/blog/vms-considered-harmful.html>
- Chirigati, F., Rampin, R., Shasha D., Freire, J., 2016. "ReProZip: Computational Reproducibility With Ease". Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 2085-2088, 2016.
- container: <https://github.com/o2r-project/container/>
- Editorial, 2018. "Does your code stand up to scrutiny?". Nature, doi:10.1038/d41586-018-02741-4
- Jupyter et al., 2018. "Binder 2.0 - Reproducible, Interactive, Sharable Environments for Science at Scale." Proceedings of the 17th Python in Science Conference. doi:10.25080/Majora-4af11417-011
- Knoth, C., Nüst, D., 2017. "Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers". Remote Sens. 9, no. 3: 290. doi:10.3390/rs903029
- Konkol, M., Kray, C., Suleiman, J., 2018. "Creating interactive scientific publications using bindings". EarthArXiv. doi:10.31223/osf.io/jtd
- Marcowertz, F., 2015. "Five selfish reasons to work reproducibly". doi:10.1186/s13059-015-0850-7
- Marwick, B., Boettiger, C., Mullen L., 2018. "Packaging data analytical work reproducibly using R (and friends)". PeerJ Preprints 8:e3192v2. doi:10.7287/peerj.preprints.3192v2
- Nüst, D., Konkol, M., Schutzzeichel, M., Pöbesma, E., Kray, C., Przybytzin, H. and Lorenz, J., 2017a. "Opening the Publication Process with Executable Research Compendia". D-Lib Mag. doi:10.1045/january2017-nuest
- Nüst, D., Schutzzeichel, M., 2017b. "An Architecture for Reproducible Computational Geosciences". Zenodo. doi:10.5281/zenodo.1478543
- Nüst, D., Boettiger, C., & Eddelbuettel, D., 2018a. "rocker/geospatial: a flexible runtime environment for geoscientific data analysis". Zenodo. doi:10.5281/zenodo.1216751
- Nüst, D., 2018. "Open environmental data analysis with senseBox, openSenseMap, Jupyter Notebook, RStudio, and BinderHub". Zenodo. doi:10.5281/zenodo.1139929
- Nüst, D., Granell C, Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., Cerutti, V., 2018b "Reproducible research and GIScience: an evaluation using AGILE conference papers". PeerJ 8:e5072. doi:10.7717/peerj.5072
- Ram, K., 2019. "Research compendium principles". <http://mndata.org/talks/rstd19/#/>
- x11docker: <https://github.com/mviereck/x11docker/>



License: Create Commons Attribution 4.0 International (CC BY 4.0)  
<https://creativecommons.org/licenses/by/4.0/>

