



Executable research compendia in geoscience research infrastructures

Daniel Nüst

University of Münster, Institute for Geoinformatics, Münster, Germany (daniel.nuest@uni-muenster.de)

From generation through analysis and collaboration to communication, scientific research requires the right tools. Scientists create their own software using third party libraries and platforms. Cloud computing, Open Science, public data infrastructures, and Open Source enable scientists with unprecedented opportunities, nowadays often in a field "Computational X" (e.g. computational seismology) or X-informatics (e.g. geoinformatics) [0]. This increases complexity and generates more innovation, e.g. *Environmental Research Infrastructures* (environmental RIs [1]).

Researchers in Computational X write their software relying on both source code (e.g. from <https://github.com>) and binary libraries (e.g. from package managers such as APT, <https://wiki.debian.org/Apt>, or CRAN, <https://cran.r-project.org/>). They download data from domain specific (cf. <https://re3data.org>) or generic (e.g. <https://zenodo.org>) data repositories, and deploy computations remotely (e.g. European Open Science Cloud). The results themselves are archived, given persistent identifiers, connected to other works (e.g. using <https://orcid.org/>), and listed in meta-data catalogues. A single researcher, intentionally or not, interacts with all sub-systems of RIs: data acquisition, data access, data processing, data curation, and community support [3].

To preserve computational research [3] proposes the *Executable Research Compendium* (ERC), a container format closing the gap of dependency preservation by encapsulating the runtime environment. ERCs and RIs can be integrated for different uses:

- (i) Coherence: ERC services validate completeness, integrity and results
- (ii) Metadata: ERCs connect the different parts of a piece of research and facilitate discovery
- (iii) Exchange and Preservation: ERC as usable building blocks are the shared and archived entity
- (iv) Self-consistency: ERCs remove dependence on ephemeral sources
- (v) Execution: ERC services create and execute a packaged analysis but integrate with existing platforms for display and control

These integrations are vital for capturing workflows in RIs and connect key stakeholders (scientists, publishers, librarians). They are demonstrated using developments by the DFG-funded project *Opening Reproducible Research* (<http://o2r.info>). Semi-automatic creation of ERCs based on research workflows is a core goal of the project.

References

- [0] Tony Hey, Stewart Tansley, Kristin Tolle (eds), 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.
- [1] P. Martin et al., Open Information Linking for Environmental Research Infrastructures, 2015 IEEE 11th International Conference on e-Science, Munich, 2015, pp. 513-520. doi: 10.1109/eScience.2015.66
- [2] Y. Chen et al., Analysis of Common Requirements for Environmental Science Research Infrastructures, The International Symposium on Grids and Clouds (ISGC) 2013, Taipei, 2013, <http://pos.sissa.it/archive/conferences/179/032/ISGC>
- [3] Opening Reproducible Research, Geophysical Research Abstracts Vol. 18, EGU2016-7396, 2016, <http://meetingorganizer.copernicus.org/EGU2016/EGU2016-7396.pdf>