Geophysical Research Abstracts
Vol. 20, EGU2018-17461, 2018
EGU General Assembly 2018

Open Environmental Data Analysis

Daniel Nüst and Thomas Bartoschek
University of Münster, Institute for Geoinformatics, Münster, Germany ({daniel.nuest,bartoschek}@uni-muenster.de)

Open Science is a major movement. It is deeply connected with other initiatives on openness, availability, and transparency: Open Access, Open Data, Reproducible Research, Free and Open Source Software (FOSS), and Open-source Hardware (OSH, https://en.wikipedia.org/wiki/Open-source_hardware). Free to use platforms supporting Open Science become more recognised and many are FOSS themselves. They range from large, publicly-funded players such as Zenodo (https://zenodo.org), to community projects with first grant funding such as Binder (Pérez et al., 2017). The senseBox community (https://sensebox.de/en) provides a Do-It-Yourself Toolkit for Citizen Science as Open Hardware and an Open Data API (https://opensensemap.org) for collecting, sharing, and visualising observations, including particulates (Pfeil et. al., 2015). Both approaches were designed following usability engineering methods to enable broad use, i.e. in education or by non-professional users (Wirwahn Bartoschek, 2015).

We present an open environmental data analysis building on all these projects. A central R Markdown (http://rmarkdown.rstudio.com/) document provides a recipe for data retrieval (with backup in plain-text files), analysis, and visualisation with `R` (https://r-project.org). The document was created in a containerised environment with Docker (https://docker.com), more specifically the Rocker (Boettiger, 2017) image `rocker/binder` (https://hub.docker.com/r/rocker/binder/) which provides Jupyter Notebook (https://jupyter.org/) and RStudio (http://rstudio.com/) user interfaces. The container's advantages are (a) portability, (b) completeness using configuration files and scripts, (c) at least limited archivability, (d) simple integration of browser-based interactive user interfaces, and (e) one click online execution with Binder. All resources are published on GitHub (https://github.com/nuest/sensebox-binder) and the Zenodo research data repository (Nüst, 2017).

We demonstrate an open analysis workflow for environmental data utilising numerous Open X projects. It showcases the potential of latest technology and may serve as a template towards development of best practices, but also provides insight into remaining stumbling blocks. Connecting all these parts requires some technical proficiency. In the future reproducible computational workflows should become part of researcher's education, making them aware of simple yet effective practices (e.g. text-based formats, scripted analyses). In the meantime templates and semi-automatic assistants (cf. Nüst et al., 2017) can mitigate challenges.

References

Boettiger, C. and D. Eddelbuettel, 2017. An Introduction to Rocker: Docker Containers for R. The R Journal, RJ-2017-065. https://rjournal.github.io/archive/2017/RJ-2017-065/index.html

Nüst, D., 2018. Open environmental data analysis with senseBox, openSenseMap, Jupyter Notebook, RStudio, and BinderHub (Version v2). Zenodo. doi: 10.5281/zenodo.1139929

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

Pérez, F., Granger, B., and Ragan-Kelley, B., 2017. Binder: enabling sharing and publication of reproducible computational research. https://figshare.com/s/e9d0ad7bdc4e405cccfa | project website: https://mybinder.org

Pfeil, M., Bartoschek, T., and Wirwahn, J.,2015. OpenSenseMap – A Citizen Science Platform for Publishing and Exploring Sensor Data as Open Data. FOSS4G Conference Proceedings, pp. 122-139, Seoul, 2015. http://www.meci.co.kr/societyevent/FOSS4G2015/download/FOSS4G%20PROCEEDINGS%20ver2.pdf

Wirwahn, J. and Bartoschek, T., 2015. Usability Engineering for Successful Open Citizen Science. FOSS4G Conference Proceedings, pp. 68-78, Seoul, 2015. http://www.meci.co.kr/societyevent/FOSS4G2015/download/FOSS4G%20PROCEEDINGS%20ver2.pdf