FAIR Notebooks: opportunities and challenges for the geoscience community

Alejandro Coca-Castro[1], Anne Fouilloux[2], J. Scott Hosking[1,3], and Environmental Data Science Book community[4]

[1]The Alan Turing Institute, Data Science for Science and Humanities Research Programme, London, United Kingdom of Great Britain – England, Scotland, Wales (acoca@turing.ac.uk)

[2]Simula Research Laboratory, Oslo, Norway (annef@simula.no)

[3]AI Lab, British Antarctic Survey, Cambridge, United Kingdom of Great Britain – England, Scotland, Wales (jask@bas.ac.uk)

[4]Open Science Community, Global (environmental.ds.book@gmail.com)

Making assets in scientific research **Findable, Accessible, Interoperable and Reusable (FAIR)** is still overwhelming for many scientists. When considered as an afterthought, FAIR research is indeed challenging, and we argue that its implementation is by far much easier when considered at an early stage and focusing on improving the researchers' day to day work practices. One key aspect is to bundle all the research artefacts in a FAIR Research Object (RO) using RoHub (https://reliance.rohub.org/), a Research Object management platform that enables researchers to collaboratively manage, share and preserve their research work (data, software, workflows, models, presentations, videos, articles, etc.). RoHub implements the full RO model and paradigm: resources associated to a particular research work are aggregated into a single FAIR digital object, and metadata relevant for understanding and interpreting the content is represented as semantic metadata that are user and machine readable. This approach provides the technical basis for implementing FAIR executable notebooks: the data and the computational environment can be "linked" to one or several FAIR notebooks that can then be executed via EGI Binder Service with scalable compute and storage capabilities. However, the need for defining clear practises for writing and publishing FAIR notebooks that can be reused to build upon new research has quickly arised. This is where a community of practice is required. The **Environmental Data Science Book (or EDS Book)** is a pan-european community-driven resource hosted on GitHub and powered by Jupyter Book. EDS Book provides practical guidelines and templates that help to translate research outputs into curated, interactive, shareable and reproducible executable notebooks. The quality of the FAIR notebooks is ensured by a collaborative and transparent reviewing process supported by GitHub related technologies. This approach provides immediate benefits for those who adopt it and can feed fruitful discussions to better define a reward system that would benefit Science and scientific communities. All the resources needed for understanding and executing the notebook are gathered into an executable Research Object in RoHub. To date, the community has successfully published ten FAIR notebooks covering a wide range of topics in environmental data science. The notebooks consume open-source python libraries e.g. $intake$, $iris$, $xarray$, $hvplot$ for fetching, processing and interactively visualising environmental research. While these notebooks

are currently python-based, EDS Book supports other programming languages such as R and Julia, and we are aiming at engaging with computational notebooks communities alike towards improving the research practices in environmental science.