

FAIR Research Objects for realizing Open Science with RELIANCE EOSC project

Anne Fouilloux[‡], Federica Foglini[§], Elisa Trasatti[|]

[‡] University of Oslo, Oslo, Norway

[§] Institute of Marine Sciences, National Research Council, Venice, Italy

[|] National Institute of Geophysics and Volcanology, Rome, Italy

Corresponding author: Anne Fouilloux (anne.fouilloux@gmail.com)

Abstract

The H2020 [Reliance](#) project delivers a suite of innovative and interconnected services that extend European Open Science Cloud (EOSC)'s capabilities to support the management of the research lifecycle within Earth Science Communities and Copernicus Users. The project has delivered 3 complementary technologies: Research Objects (ROs), Data Cubes and AI-based Text Mining.

[RoHub](#) is a Research Object management platform that implements these 3 technologies and enables researchers to collaboratively manage, share and preserve their research work.

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.

The development of RoHub is co-designed and validated through multidisciplinary and thematic real life use cases led by three different Earth Science communities: Geohazards, Sea Monitoring and Climate Change communities.

A RO commonly starts its life as an empty Live RO. ROs aggregate new objects through their whole lifecycle. This means, a RO is filled incrementally by aggregating new relevant resources such as workflows, datasets, documents according to its typology that are being created, reused or repurposed. These resources can be modified at any point in time.

We can copy and keep ROs in time through snapshots which reflect their status at a given point in time. Snapshots can have their own identifiers (DOIs) which facilitates tracking the evolution of a research. At some point in time, a RO can be published and archived (so called Archived RO) with a permanent identifier (DOI). New Live ROs can be derived based on an existing Archived RO, for instance by forking it.

To guide researchers, different types of Research Objects can be created:

- Bibliography-centric: includes manuals, anonymous interviews, publications, multimedia (video, songs) and/or other material that support research;
- Data-centric: refers to datasets which can be indexed, discovered and manipulated;
- Executable: includes the code, data and computational environment along with a description of the research object and in some cases a workflow. This type of ROs can be executed and is often used for scripts and/or Jupyter Notebooks;
- Software-centric: also known as “Code as a Research Object”. Software-centric ROs include source codes and associated documentation. They often include sample datasets for running tests.
- Workflow-centric: contains workflow specifications, provenance logs generated when executing the workflows, information about the evolution of the workflow (version) and its components elements, and additional annotations for the workflow as a whole.
- Basic: can contain anything and is used when the other types do not cover the need.

To ease the understanding and the reuse of the ROs, each type of RO (except Basic RO) has a template folder structure that we recommend researchers to select. For instance an executable RO has 4 folders:

- *'biblio'* where researchers can aggregate documentations, scientific papers that led to the development of the software/tool that is aggregated in the tool folder;
- *'input'* where all the input datasets required for executing the RO are aggregated;
- *'output'* where some or all the results generated by executing the RO are aggregated;
- *'tool'* where the executable tool is aggregated. Typically, we aggregate Jupyter Notebook and/or executable workflows (Galaxy or snakemake workflows).

In addition to the different types of ROs and associated template structures, researchers can select the type of resources that constitutes the main entity of a RO: for instance, a Jupyter Notebook can be selected as the main entity of an executable RO. As shown on Fig. 1, this additional metadata is then visible to everyone (and machine readable) to ease reuse. Examples of Bibliography-centric and Data-centric Research Objects are shown on Fig. 2: the overall overview of any types of Research Object is always the same with mandatory metadata information such as the title, description, authors & collaborators, sketch (featured plots/images), the content of the RO (with different structures depending on the type of ROs). Additional information is displayed on the right panel such as number of downloads, additional discovered metadata (automatically

discovered from the Reliance text enrichment service), free keywords (added by end-users) and citation. The 'toolbox' and 'share' sections allows end-users to download, snapshot and archive the RO and/or share it.

Any Research Object in RoHub is a FAIR digital object that is for instance findable in OpenAire, including Live ROs.

In our presentation, we will showcase different types of ROs for the 3 Earth Science communities represented in Reliance to highlight how the scientists in our respective disciplines changed their working methodology towards Open Science.

Keywords

Geohazard, Climate Change, Sea Monitoring, open science, research life cycle

Presenting author

Anne Fouilloux

Presented at

First International Conference on FAIR Digital Objects, presentation

Grant title

RELIANCE (REsearch Lifecycle mAnagement for Earth Science Communities and CopErnicus users in EOSC).

The RELIANCE project has received funding from the European Union's Horizon 2020 INFRAEOSC programme under grant agreement No 101017501.

Author contributions

Author list is given by alphabetical order. Each author has contributed equally to the work.

Conflicts of interest

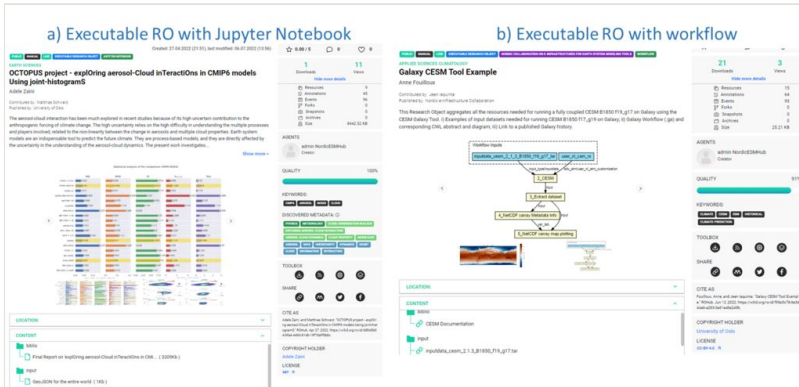


Figure 1.

Examples of Executable ROs: a) [Live Research Object in RoHub with the OCTOPUS project](#) where the main resource is a Jupyter notebook; b) [Live Research Object in RoHub for the Galaxy Community Earth System Modelling \(CESM\) Tools](#) where the main resource is a workflow.

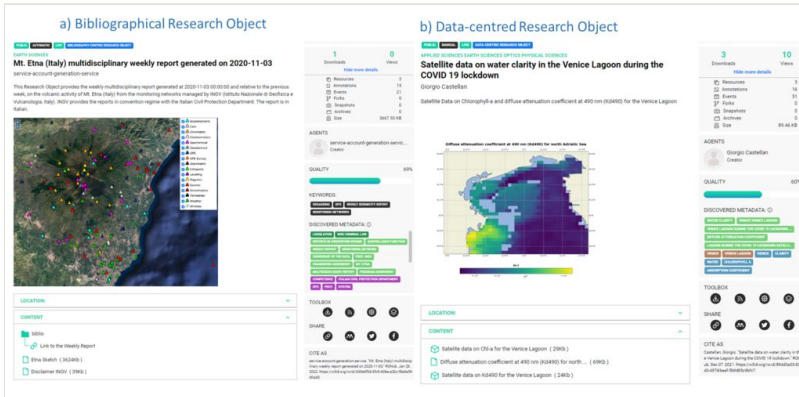


Figure 2. Examples of a) Bibliographical Research Object on [Mt. Etna \(Italy\) multidisciplinary weekly report generated on 2020-11-03](#) and b) Data-centred Research Object on [Satellite data on water clarity in the Venice Lagoon during the COVID 19 lockdown](#).