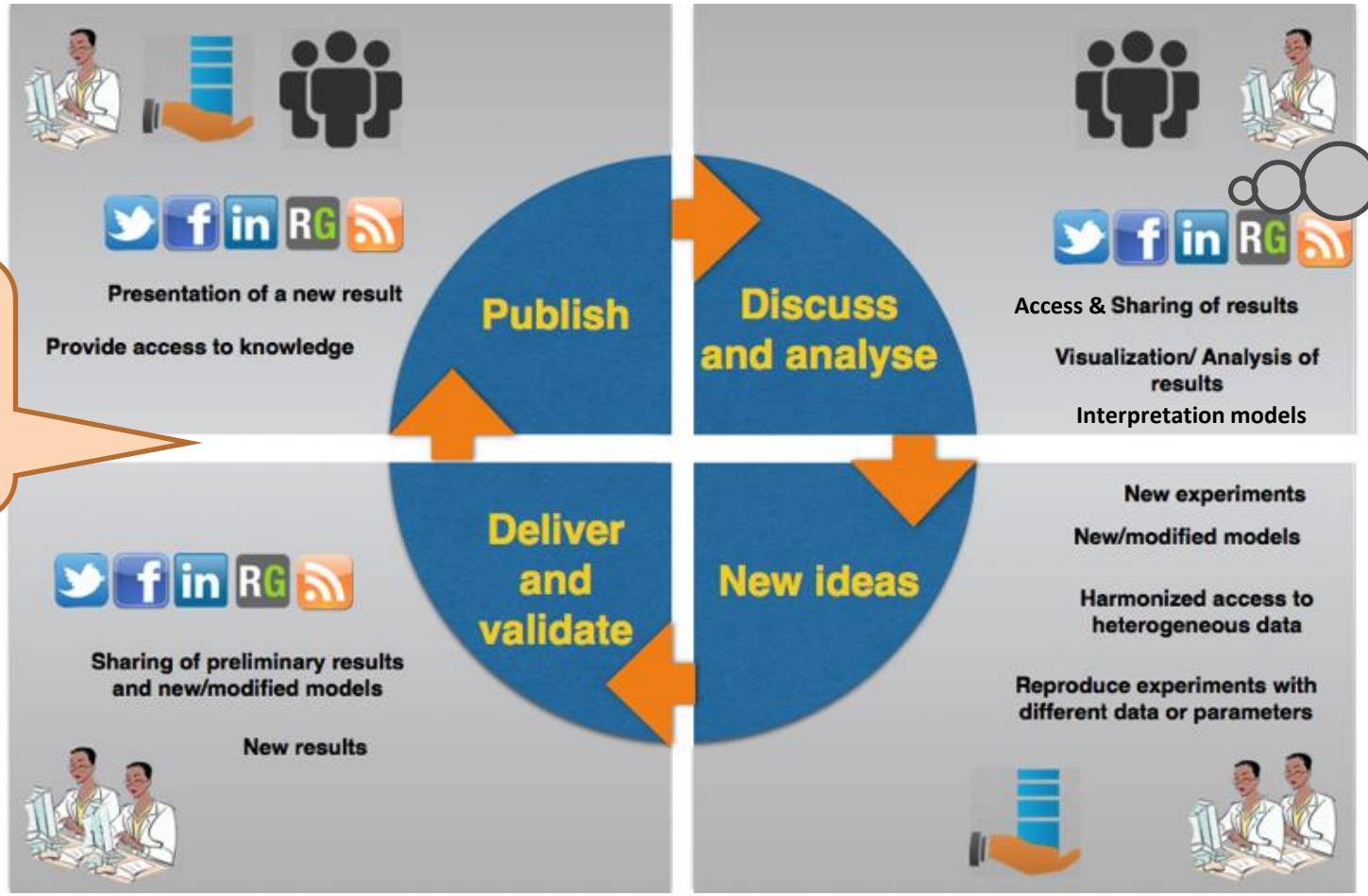

# Managing, Preserving and Disseminating Research Objects in Earth Science with the ROHub ScienceGateway

**Raul Palma**, Jose M. Gomez Perez, Andres Garcia, Cezary Mazurek

This project is co-funded by the European Union

# The Research and Information Lifecycle in Earth Science

A continuous, iterative and dynamic process followed by scientists for conducting, validating and disseminating scientific knowledge




**(peer) Review:** ***"Are these novel findings? Was the method sound?"***

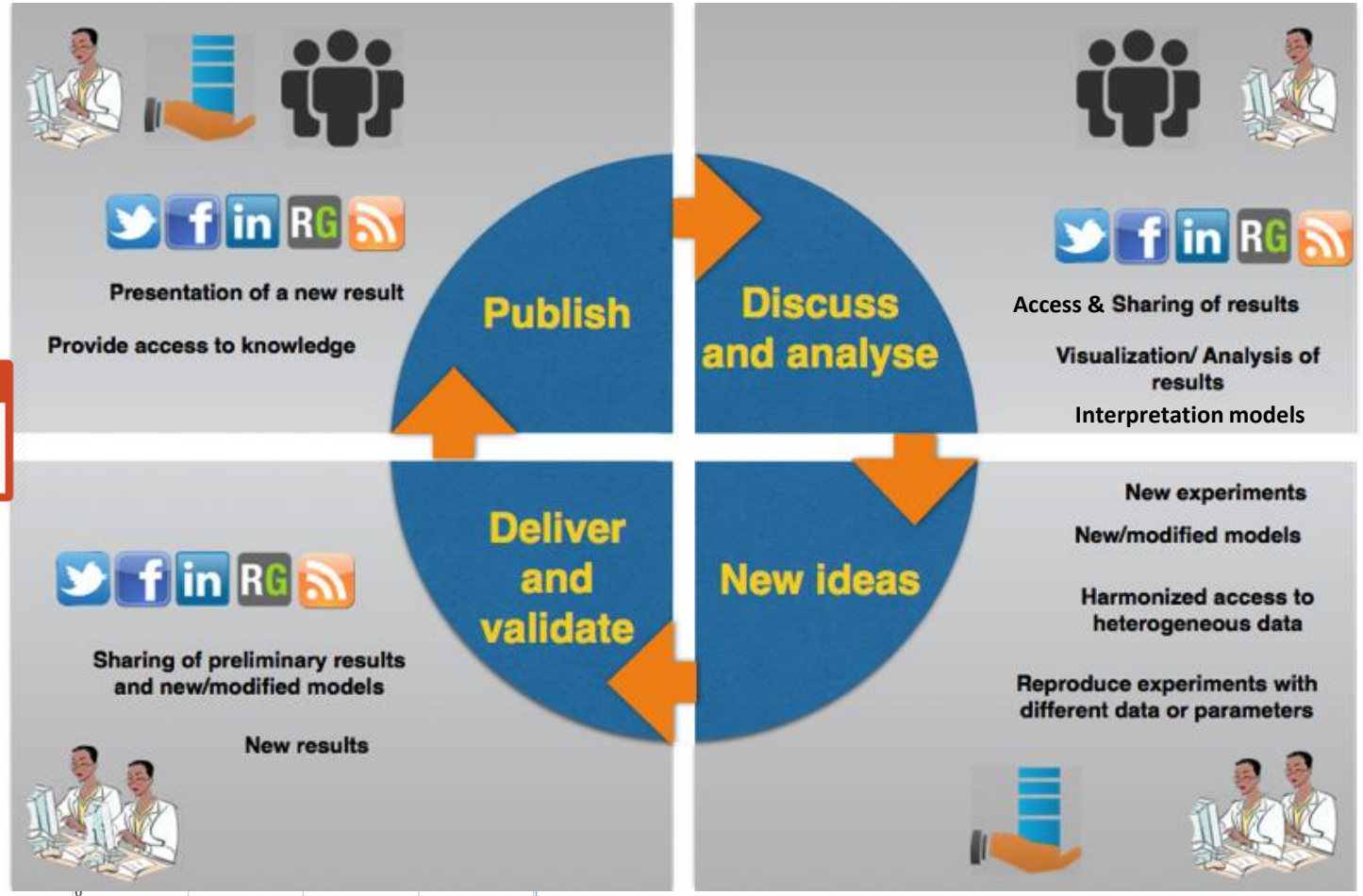
**Reader:** ***"I trust the method and data are sound."***

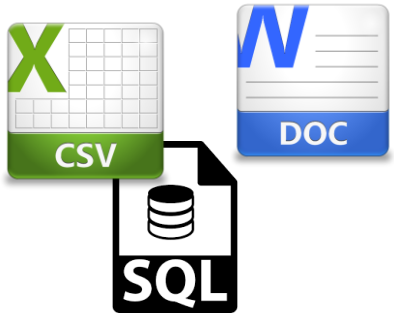
**Colleagues:** ***"Let's collaborate!"***

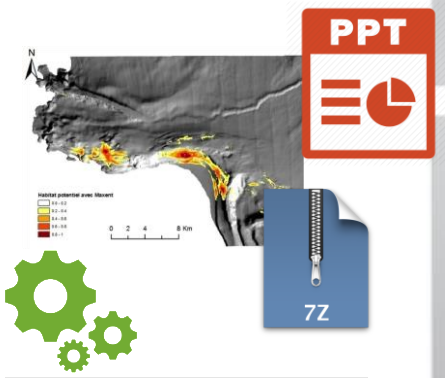
This project is co-funded by the European Union

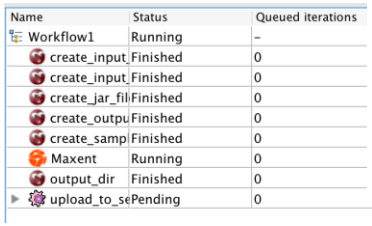
# The Research and Information Lifecycle in Earth Science

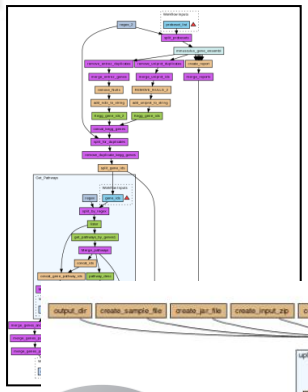
Data and process intensive: increasingly consuming and generating a variety of digital resources **->** need to i) capture the lifecycle and ii) provide single-entry point to resources involved

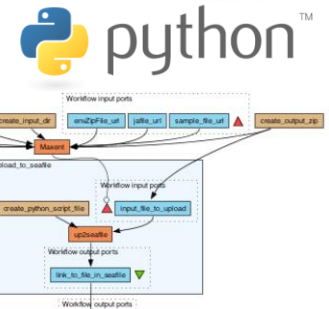



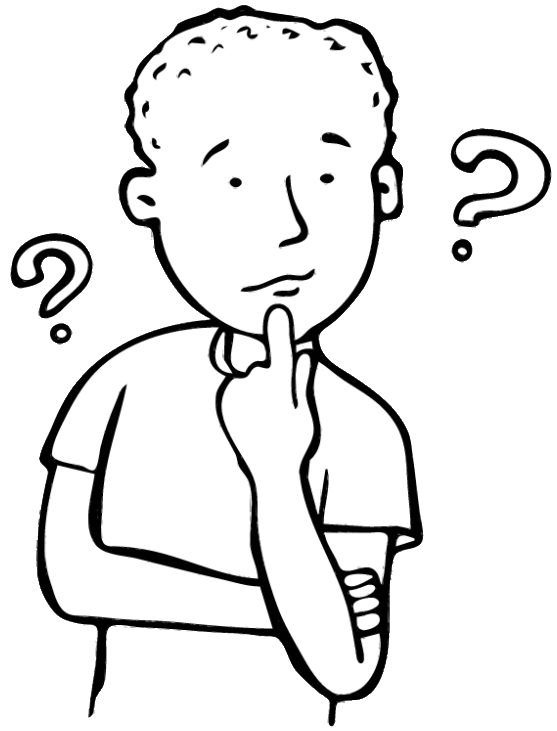




This project is co-funded by the European Union

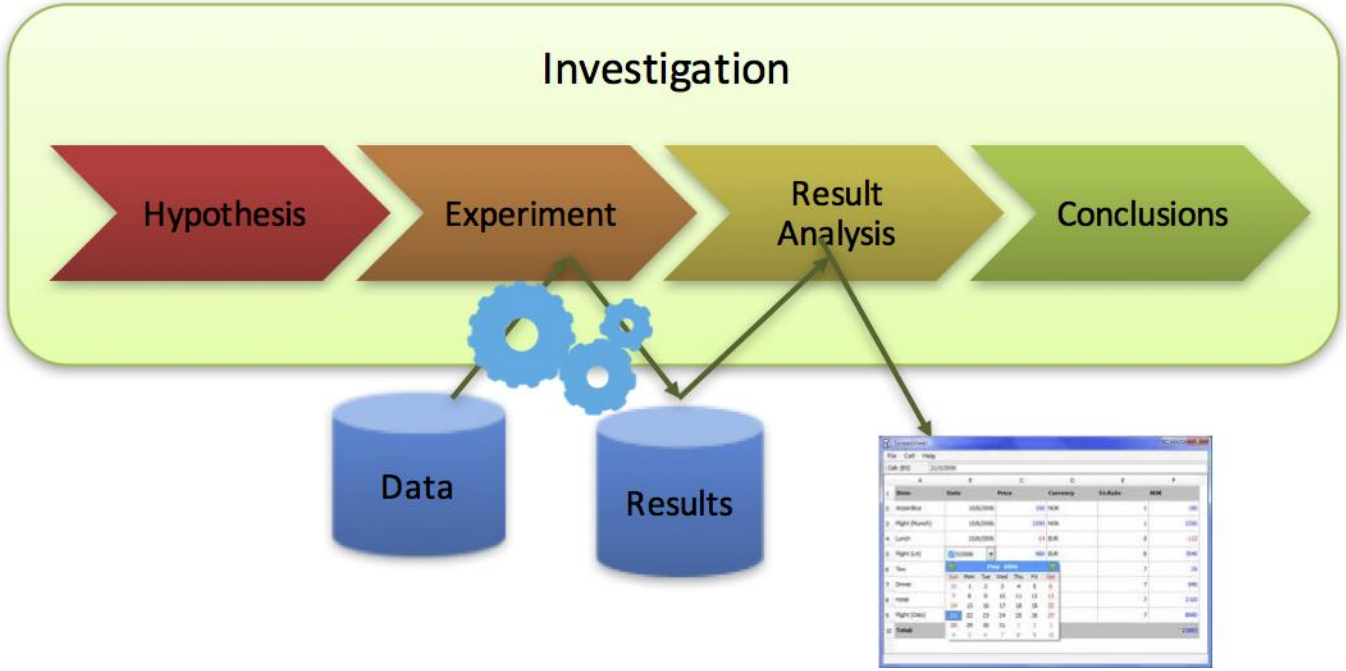
# Research in data and process intensive disciplines

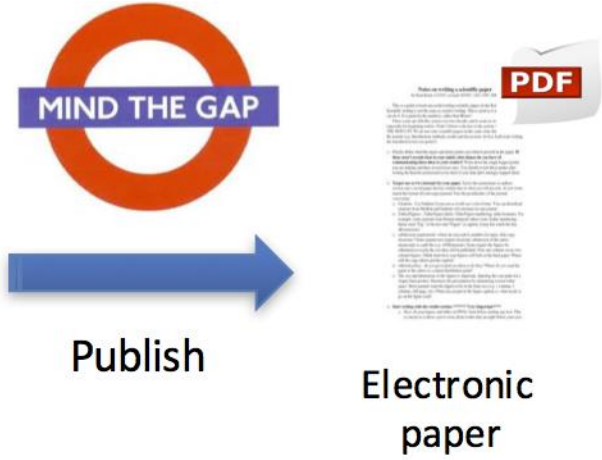
- How to provide a single-entry point to all the related resources ?
- How to capture the lifecycle of scientific investigations ?
- How to validate and disseminate scientific knowledge ?
- How to preserve the scientific knowledge for their reuse?



This project is co-funded by the European Union

# Electronic paper is not enough!



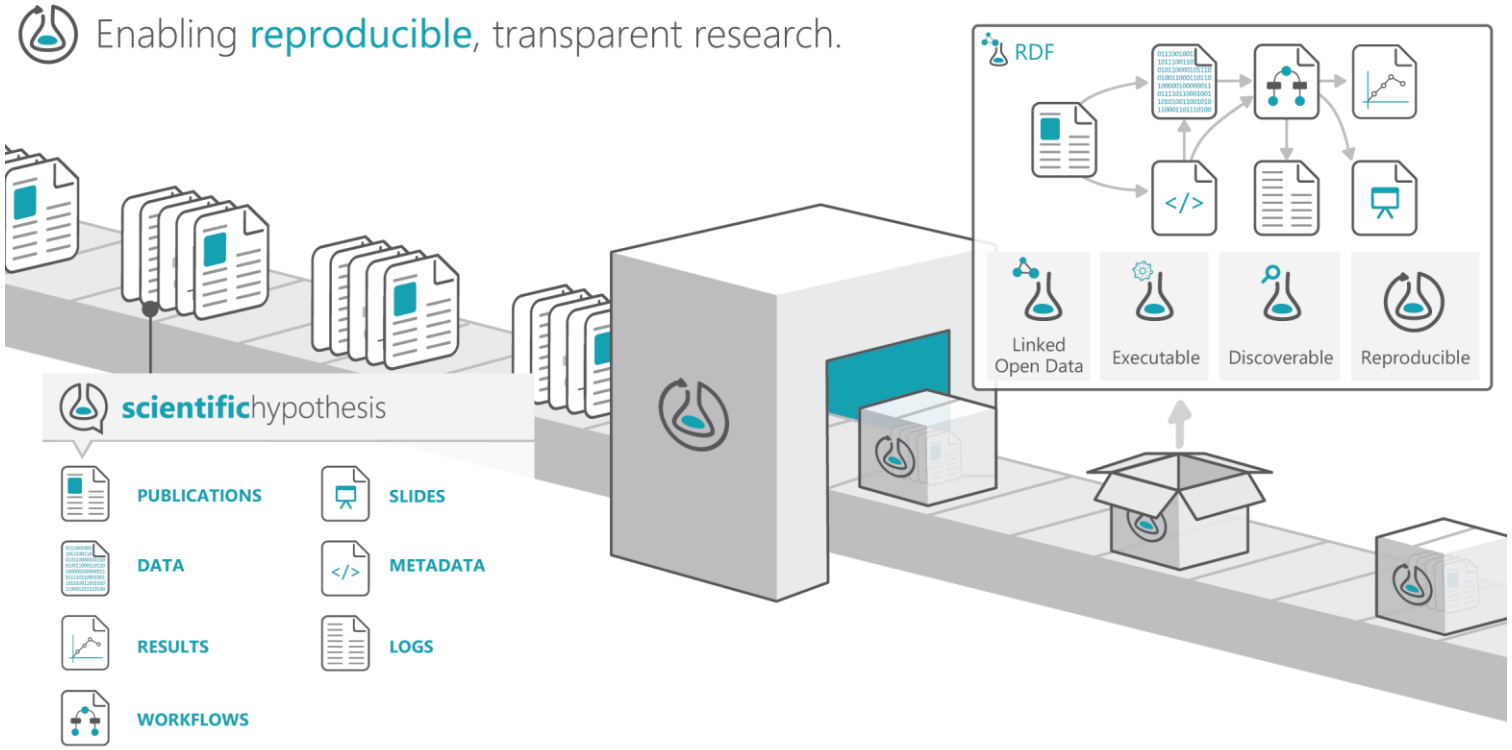

**not sufficient** to **communicate and share** information about research **investigations**, as **recognized** by researchers and evidenced by **OR**.

**Open Research** movement: Openly share the data of your experiments

This project is co-funded by the European Union

# Research objects

**Research objects goal**: Account, describe and share *everything* about your research, including how those things are related




**Unique identifier**, e.g. DOI
**Hypotheses**
**Data** used and results produced
**Methods** employed to produce and analyse data
**Scientific workflows** implementing such methods
**Provenance** of their executions
**Versioning** information
**People** involved in the investigation
**Annotations** about these resources

http://www.researchobject.org

Workflow 4Ever

ROHUB

myexperiment

This project is co-funded by the European Union

# Why research objects

i. To **organize and describe** the resources, materials, and methods of an investigation

ii. To **share** your research materials with other scientists at **discrete milestones of your investigation**. Uniquely identified by an URI, pref. as a DOI
*(RO as a social object)*

iii. To enable **reproducibility** and **reuse** of scientific methods

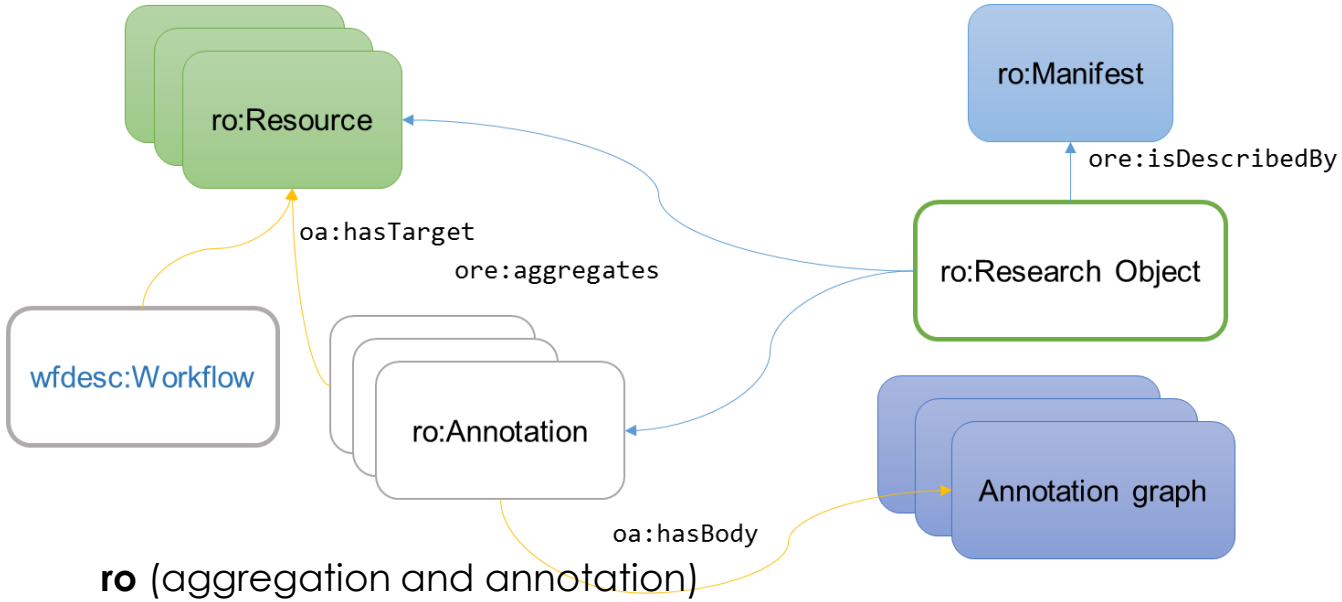
iv. To be **recognized** and **cited** *(even constituent parts)*

v. To **preserve** results and **prevent decay**
*(curation of workflow definition)*

vi. To provide **evidence** to findings claimed in **scholarly articles**

This project is co-funded by the European Union

# Representing research objects: The RO Model




**ro** (aggregation and annotation)
**wfdesc** (workflow description)
**wfprov** (workflow provenance)
**roevo** (evolution model)
**minim** (minimum information model)

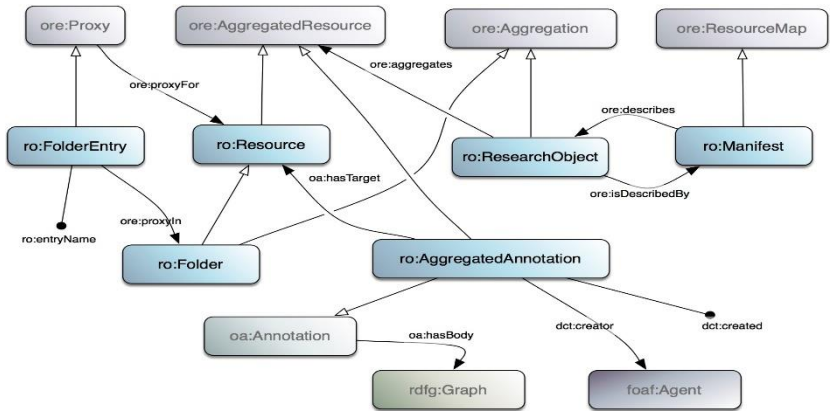
**RO primer:** http://wf4ever.github.com/ro-primer
**RO specification:** http://wf4ever.github.com/ro

- **RO = ORE + AO + vocabularies**
- Object Re-use and Exchange **(OAI-ORE)**
  - Describes aggregations of resources
  - data, metadata, papers, etc.
- Annotation Ontology **(AO)**
  - Associates RDF metadata descriptions with resources
- Generic and domain-specific **vocabularies**
  - Used in annotation bodies to provide information about resources
  - Involve types, dependencies, descriptions

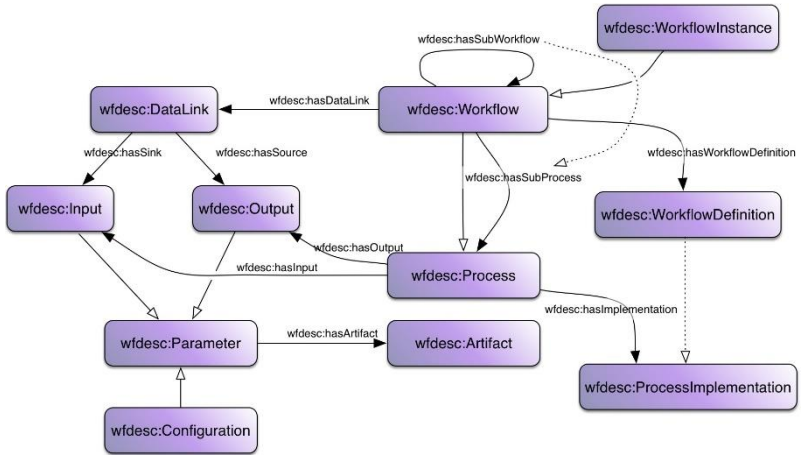
This project is co-funded by the European Union

# The RO ontology stack

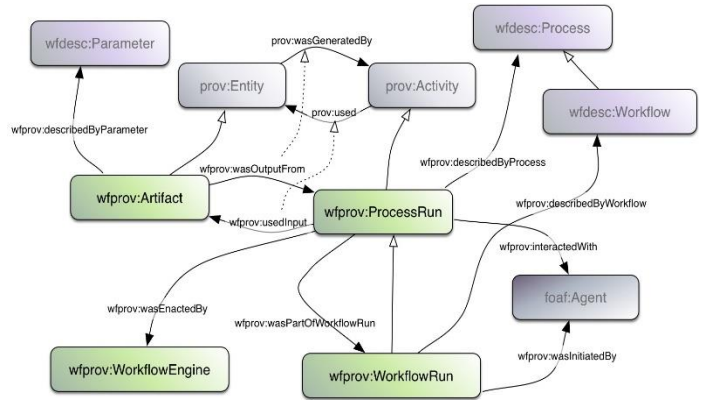
## Research object (ro)



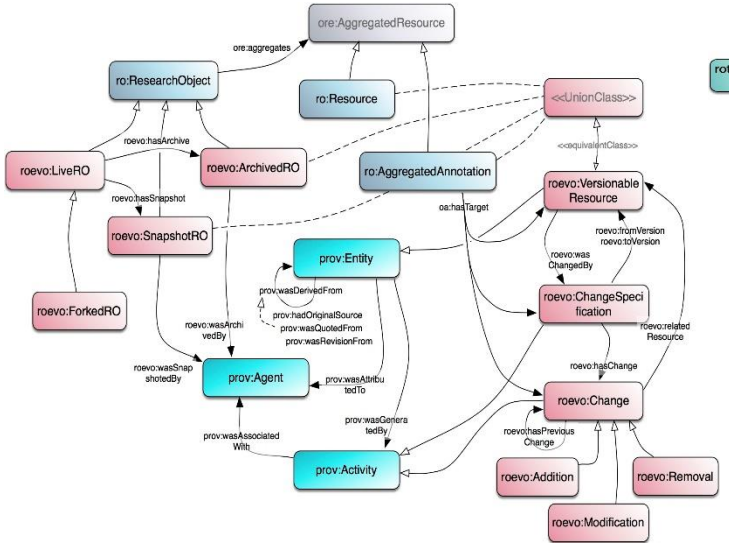

## Workflow description (wfdesc)



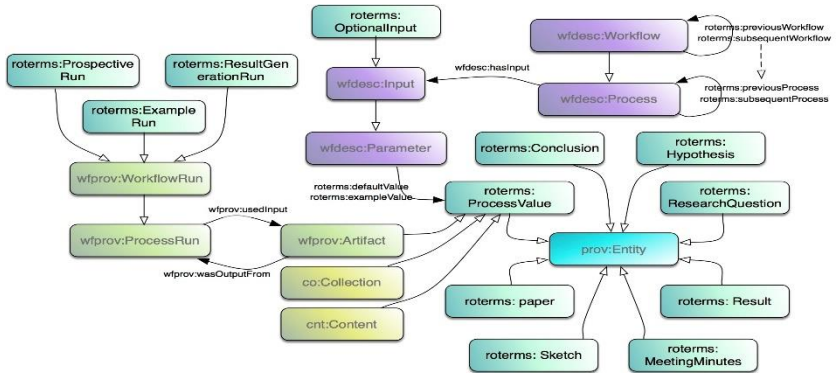

## Workflow provenance (wfprov)




## RO evolution (roevo)




## RO terms (roterms)



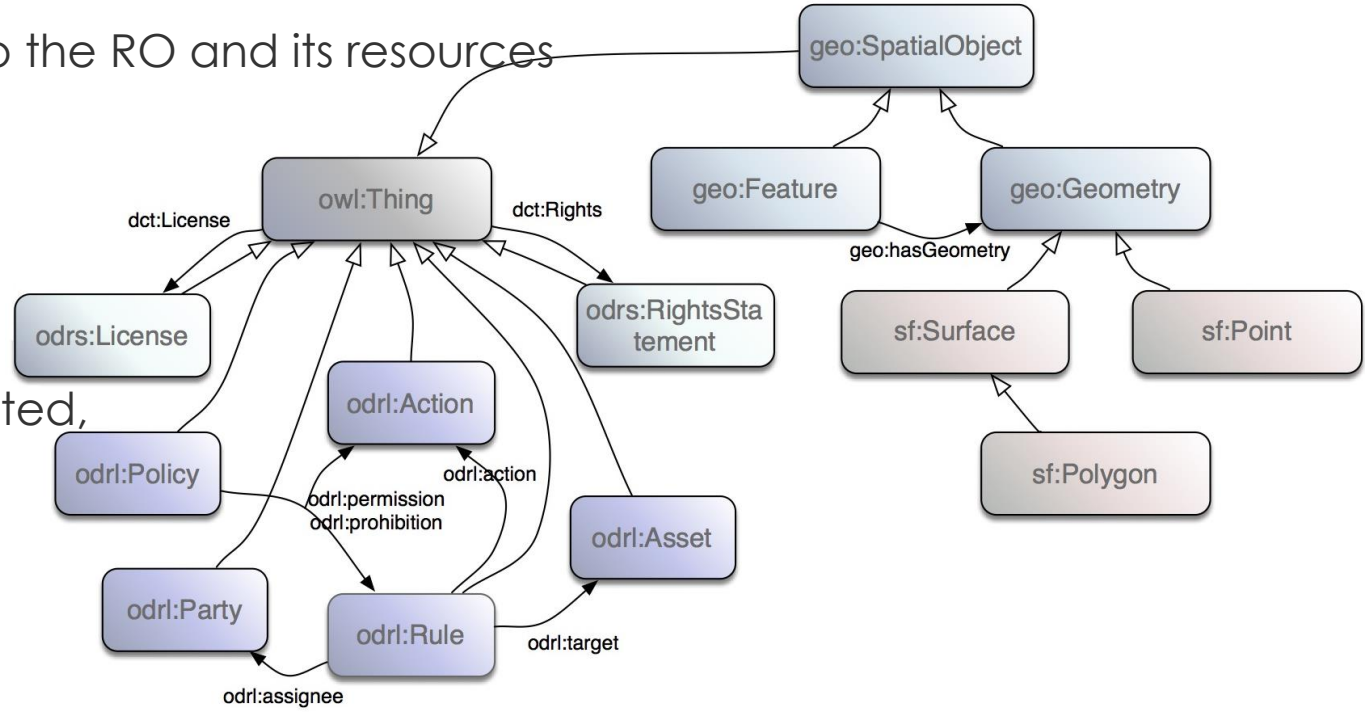

## RO for Earth Science (roes)

Emphasis on **Geo**, **access rights and IP** policies
https://github.com/wf4ever/ro/tree/earth-science

This project is co-funded by the European Union

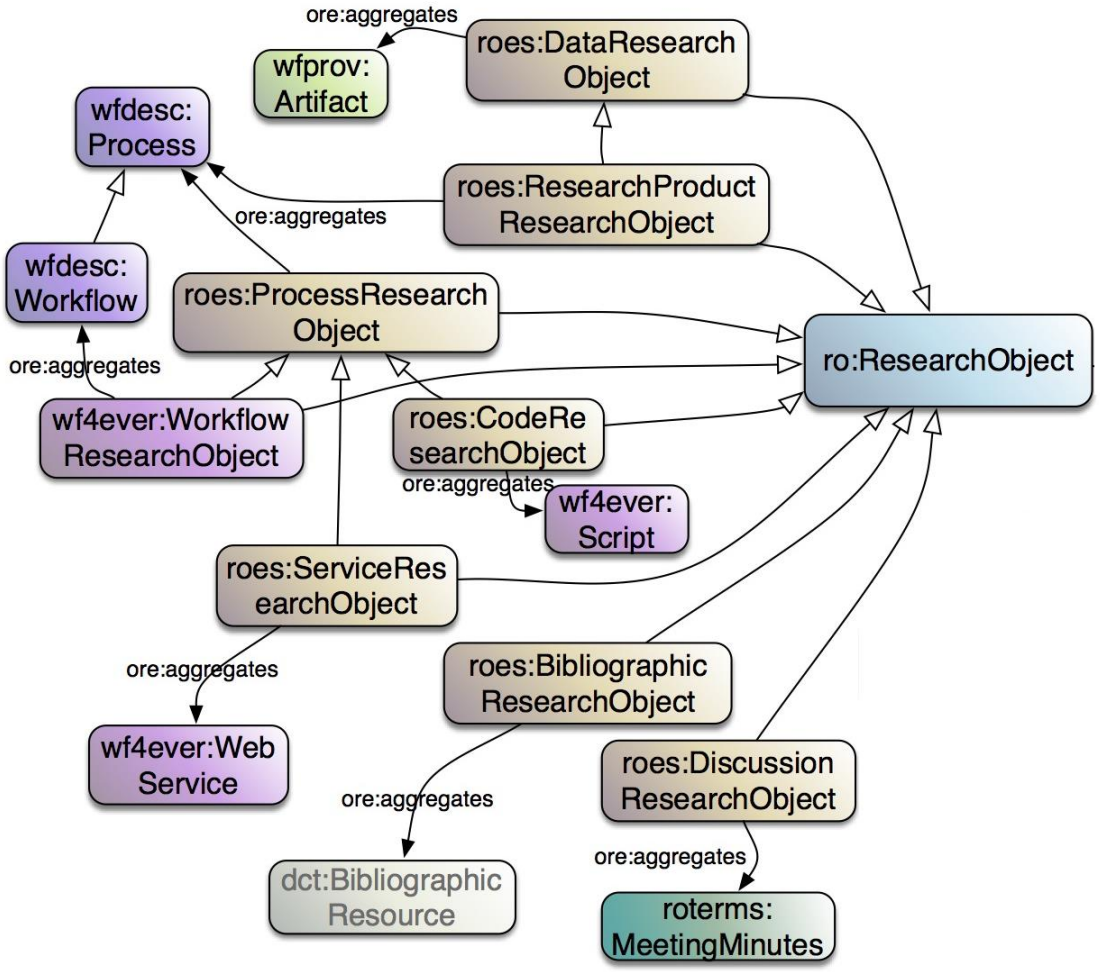
# Earth Science extensions (1/3)

- RO model should cover metadata regarding
  - Geospatial information
    - Coordinates of the region relevant for the RO and the observation it represents
  - Time-period coverage
    - Indication of the time span covered in the observation
  - Data access policies
    - Detailed information for controlling access to the RO and its resources
  - Intellectual property rights
    - Detailed information about for licensing and attribution
  - General metadata relevant to ES
    - E.g. the main scientific discipline of the RO, the size and format of the resource aggregated, the RO submission date, its digital object identifier (DOI), the main target community, etc.



This project is co-funded by the European Union

# Earth Science extensions (2/3)

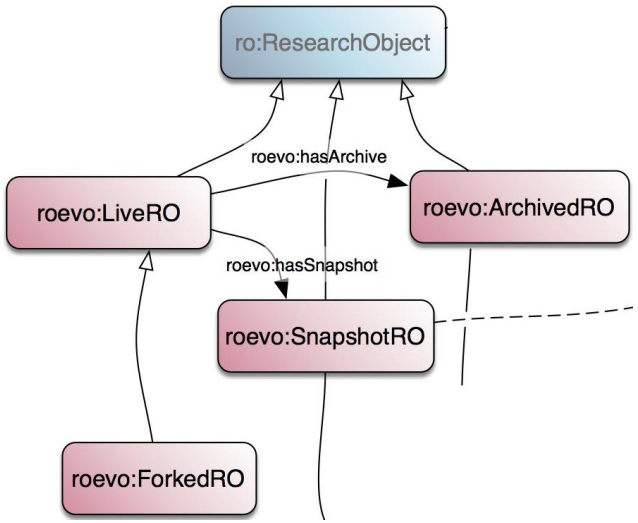


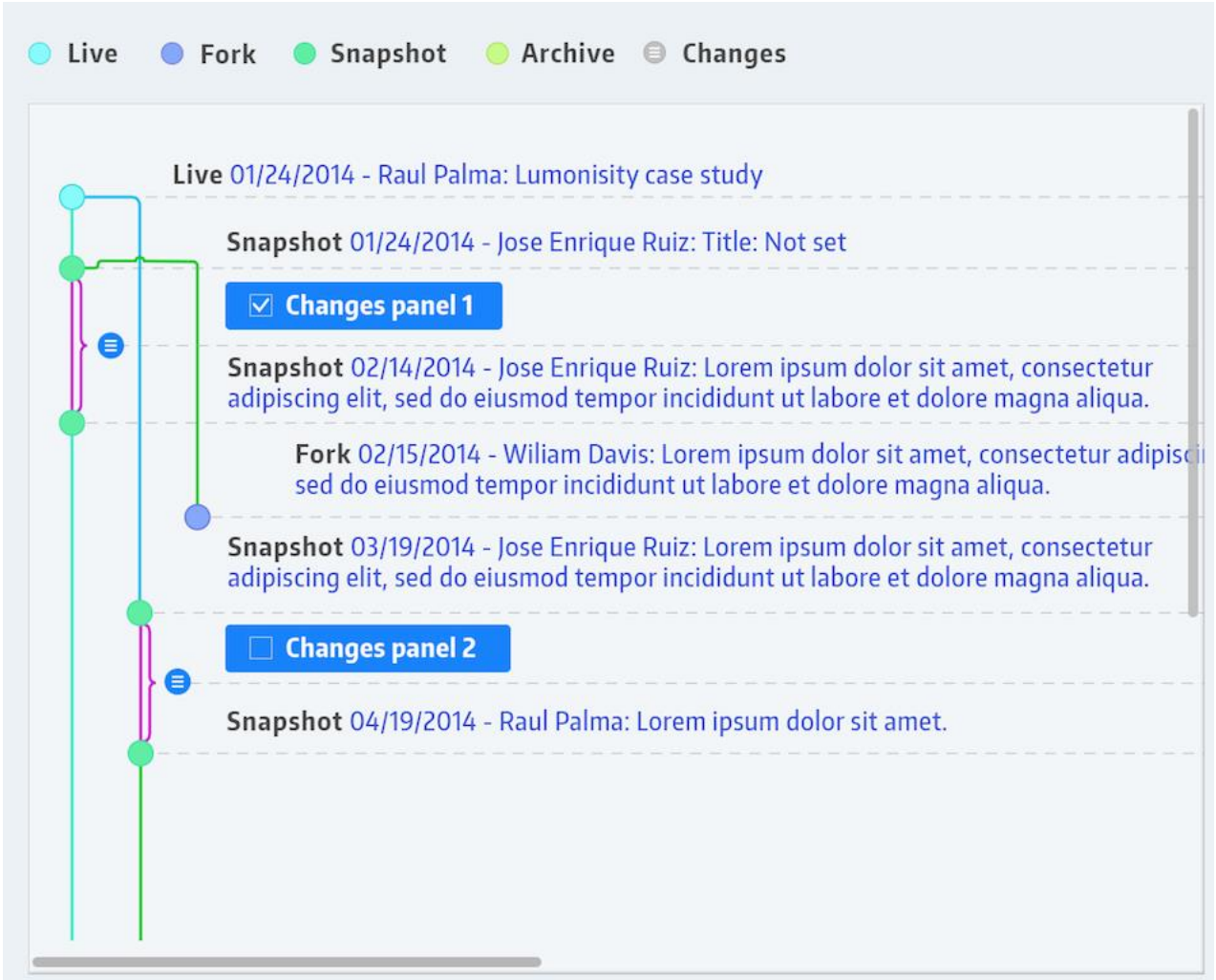

- We also defined different types of research objects according to the purpose and hence the kind of the resources it would aggregate:
  - Workflow-centric
  - Data-centric
  - Research Product centric
  - Process-centric
  - Code-centric
  - Service-centric
  - Bibliography-centric
  - Discussion-centric

This project is co-funded by the European Union

# Earth Science extensions (3/3)

- Also, we made extension to the evolution model:
  - in addition to support snapshot and releases, it was requested support for forking to facilitate creation of new research objects from existing ones, e.g., to extend and build on previous work, or to try alternate lines of work



This project is co-funded by the European Union

# Exemplary RO lifecycle

Live RO

Scientist

- Observation starts
- An Earth Scientist creates a Live RO to aggregate all the relevant data, VM templates, etc.
- A temporary id is created
- Collaborators contribute to the observation by adding or modifying resources in the RO

- Preliminary results are obtained
- The Earth Scientist produces a technical report with the current results of the observation

- Observation (RO) validated by the community
- Final versión archived
- Archived RO created
- versionOf ROEVO metadata generated

<<re-use>>

Snapshot RO

Scientist

<<copy>>

<<copy, filter>>

<<versionOf>>

- Another Earth Scientisits continues the line of research
- The previous RO is forked and a new one is created
- The whole RO lifecycle starts over
- Automatically, the new RO cites the one its based on through its DOI
- Cites metadata added to the metadata of the new RC

Archived RO

- The ES decides the work is ready to be published or shared (click release button)
- New snapshot created. versionOf ROEVO metadata generated
- DOI assigned by ROHub replaces temporary id in the RO
- RO visible by Google Scholar or Microsoft Academic through ROHub and the VRE portal (visible web site at the portal with the released ROs needs to be in place)
- CitedBy metadata monitored and extracted from Google Scholar or Microsoft Academic periodically and added as metadata to the RO

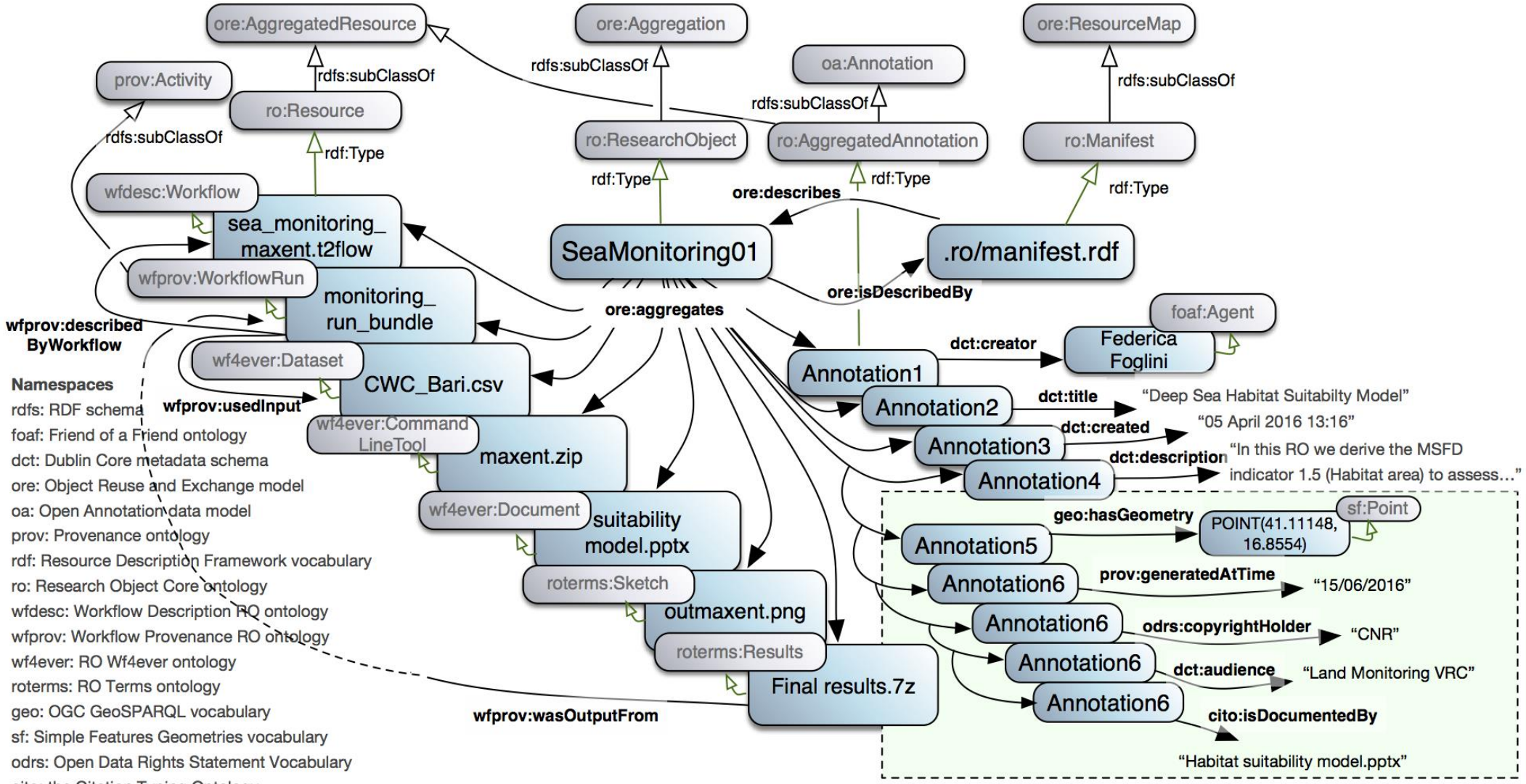
<<versionOf>>

<<copy, filter and curate>>

DataCite
FIND, ACCESS, AND REUSE DATA

This project is co-funded by the European Union

# Exemplary research object (sea monitoring use case)



This project is co-funded by the European Union

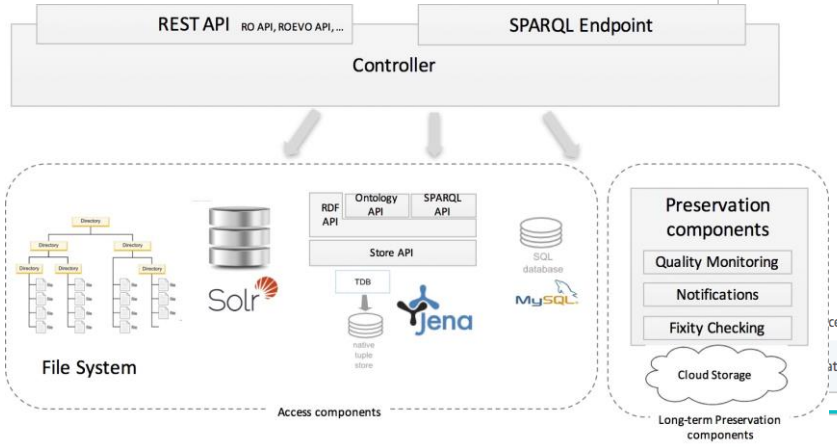
# And what is the existing RO supporting technology ?

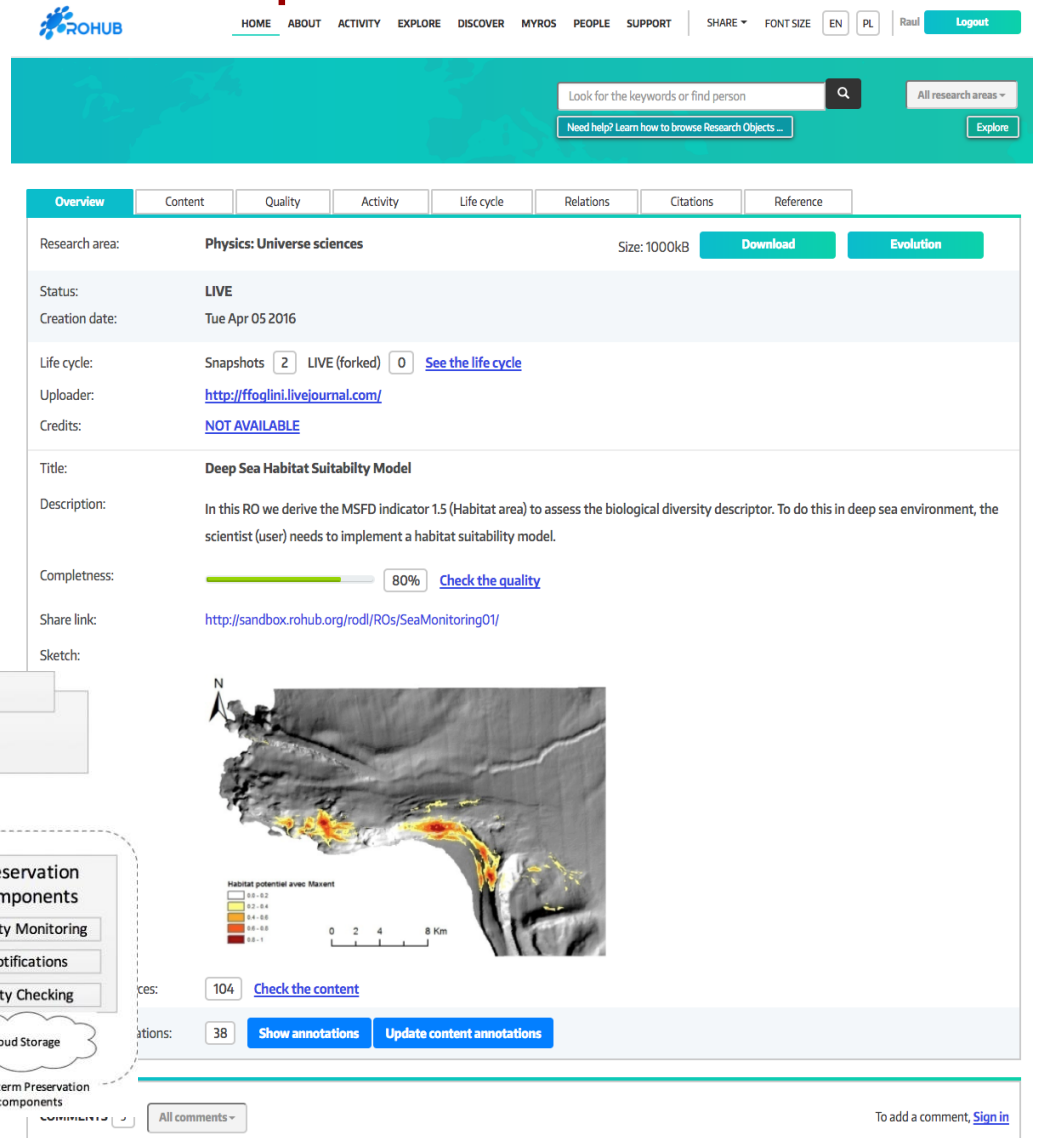
This project is co-funded by the European Union

# Research object management platform

- ROHUB comprises both
  - a **backend service** (RODL), implementing and exposing the set of APIs
  - a reference **web client application** (ROHUB portal), exposing the research object functionalities to the end-users.
- Combination of digital libraries, long term-preservation and semantic technologies
- First released in 2012

Palma R., Corcho O., Gómez-Pérez J.M., Mazurek, C., "ROHub – A Digital Library for Sharing and Preserving Research Objects". Poster Proceedings of iPres 2014.



This project is co-funded by the European Union

# High-level features (1/2)

- **Holistic research object management** platform, enabling
  - scientists throughout the research **lifecycle** to **create** and **manage high-quality ROs** that can be **interpreted** and **reproduced** in the future
  - to **reference**, **share** and **preserve scientific** findings, **campaigns**, and **observations** related resources, including **internal** ones, links to **external** ones as well as other ROs **(nested ROs)**
  - to **collaborate** with colleagues and to **discover new knowledge** through different advanced exploratory search interfaces that exploit **RO metadata** (both explicitly provided and **automatically extracted** from its content)

This project is co-funded by the European Union

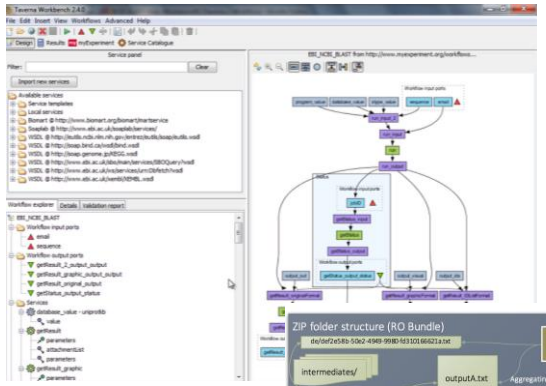
# High-level features (1/2)

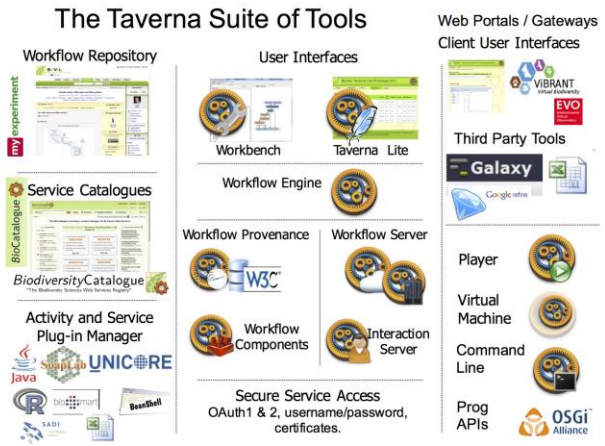
- **Holistic research object management** platform, enabling
  - To **manage** the **RO evolution** including the ability to generate **snapshots**, **releases** and **forks**
  - to **publish** the associated work and assign it a **DOI** for getting **citations**
  - to **monitor and follow** the status of a particular work and to get **notifications** about its progress
  - to **find researchers** relevant in a particular domain, e.g., for possible collaborations or reviews
  - To **supports** different **user roles**
    - Scientists, students and enthusiasts
    - Industry
    - Investors
    - Publishers

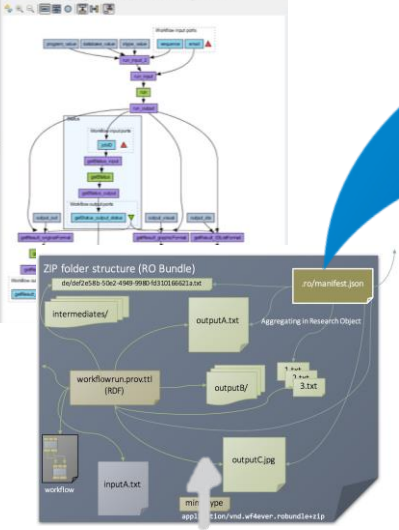
This project is co-funded by the European Union

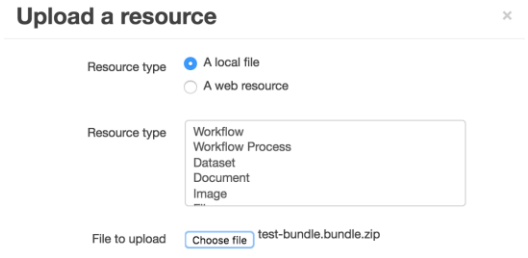
# Interoperability with external systems via RO bundles

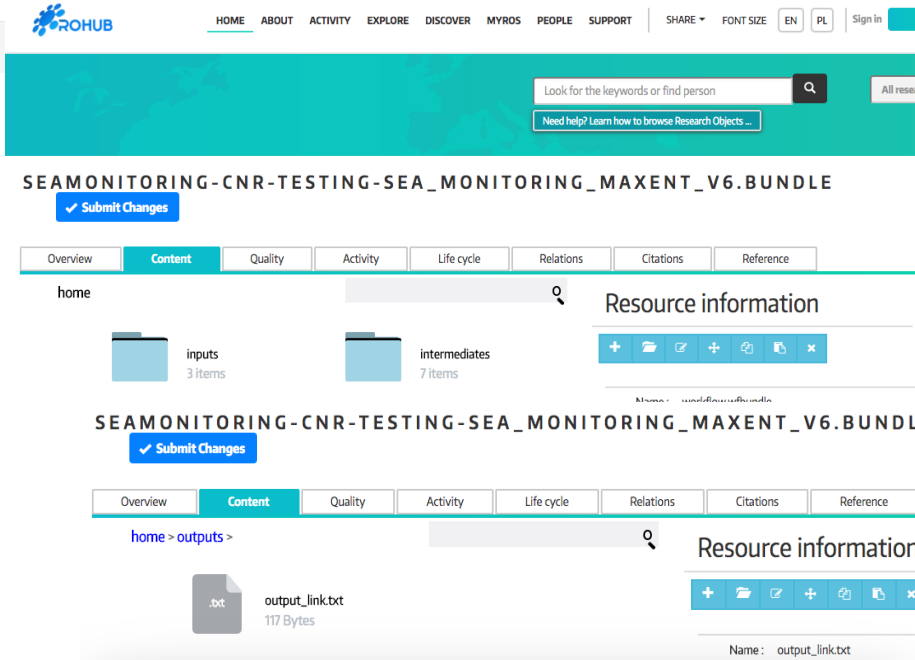
Workbench






Server

This project is co-funded by the European Union

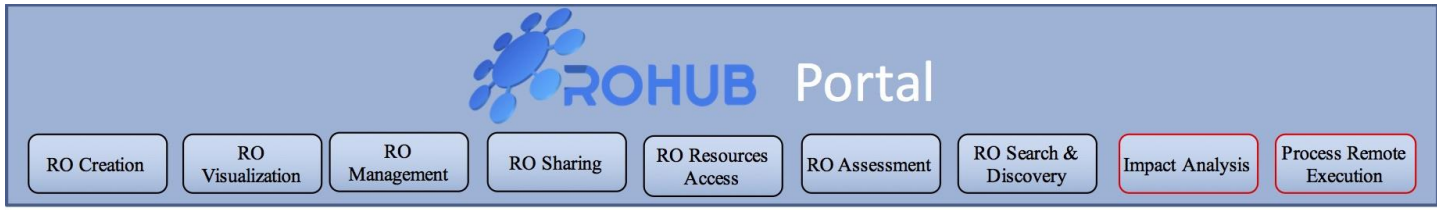
# Behind the scenes

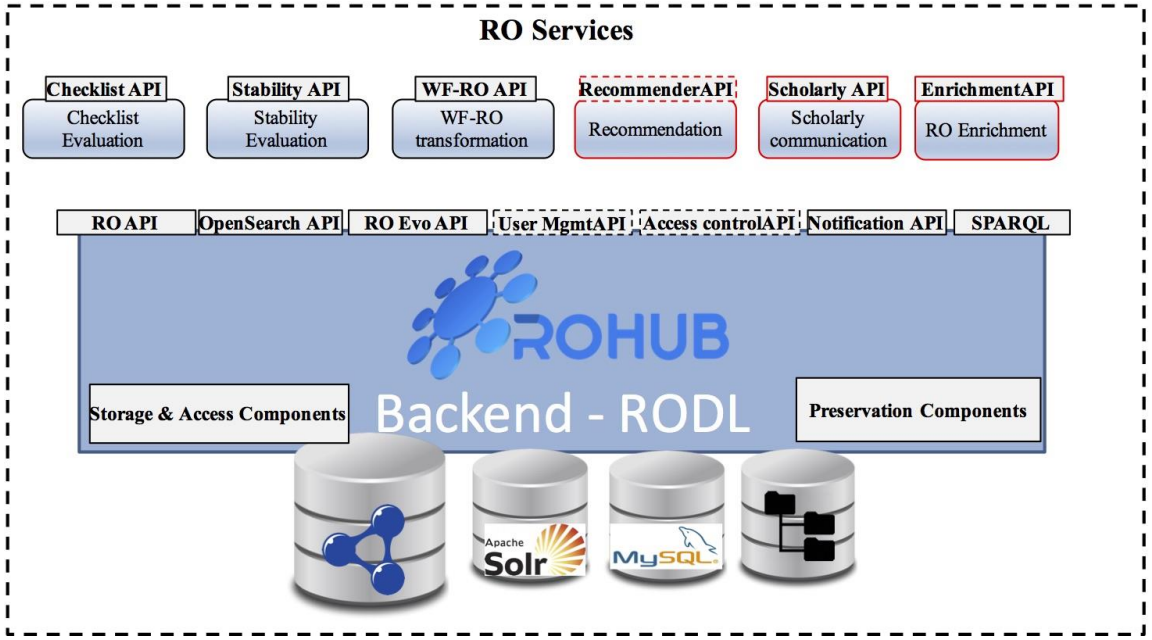
## A full suite of RO APIs provided by ROHUB and other RO services

The RO API suite includes several **specifications for managing different aspects of ROs**. The two main APIs are:

- **RO API** enables the **storage and retrieval** of ROs and their aggregated resources, as well as **annotating** them.
- **RO evolution API** enables to **record the transformation** of ROs based on their **lifecycle**, and to access the **history** of their evolution

Other APIs include: checklist, stability, recommendation, workflow runner, WF-RO, user mgmt, notification, access control, search (OpenSearch)



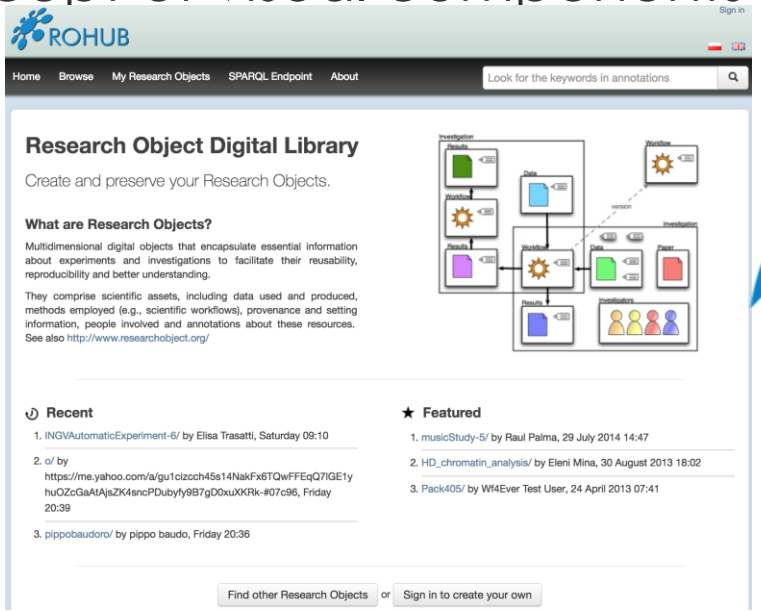



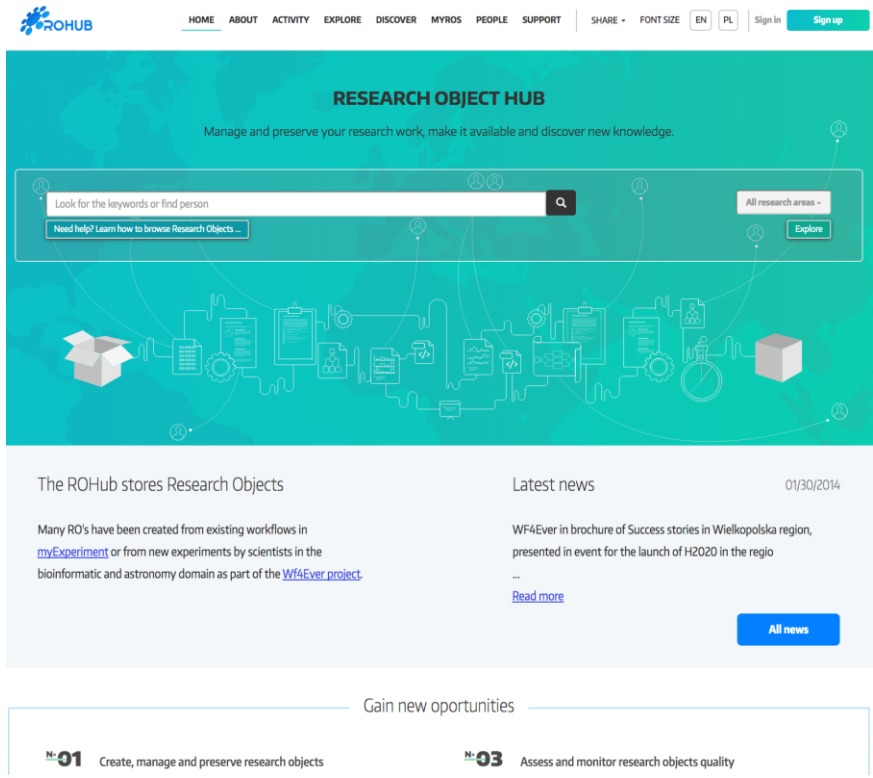

Palma R., Hołubowicz P., Page K., Soiland-Reyes S., Klyne G., Mazurek C. A Suite of APIs for the Management of Research Objects, Proceedings of the Developers Workshop, ISWC. October 2014.

This project is co-funded by the European Union

# The RO user interface – ROHUB portal

- Fully fledged interface for the management of research objects, exposing the services functionalities to the end-users
- A new portal currently under development following the concept of visual components

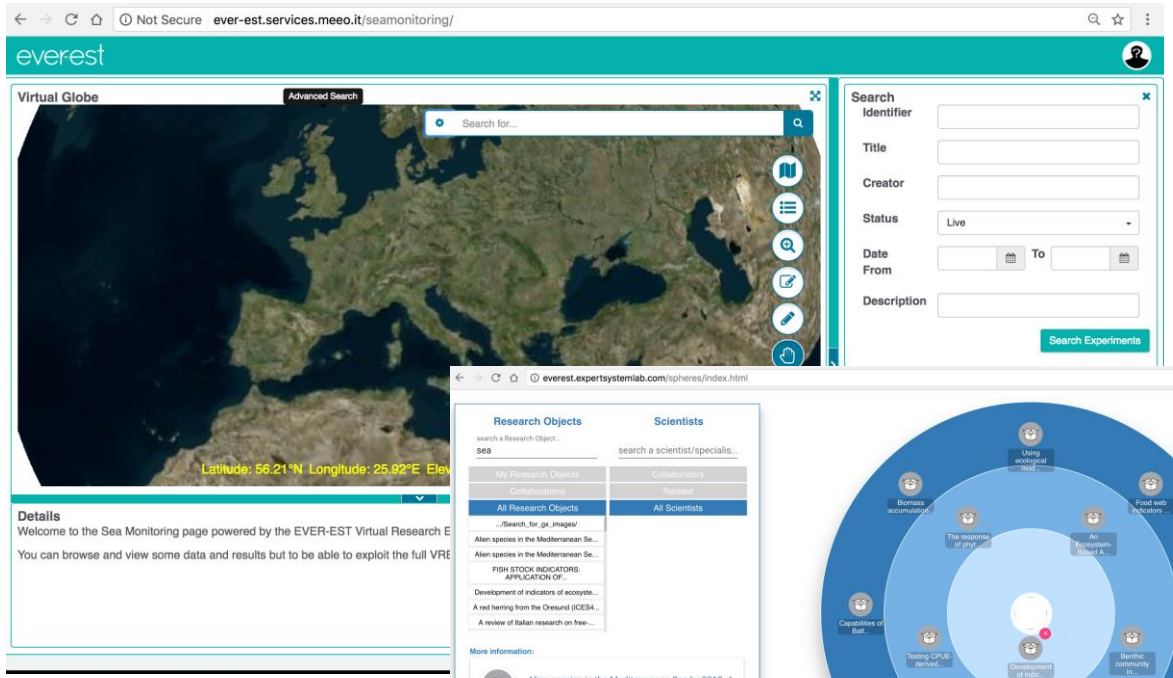


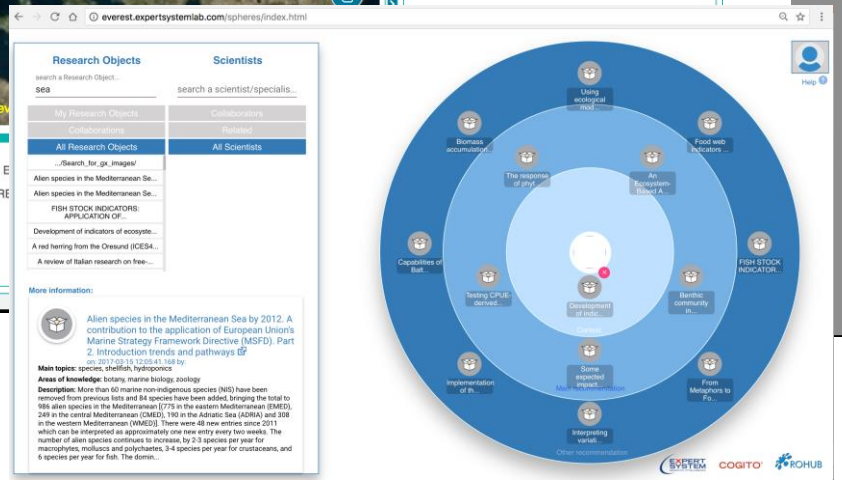



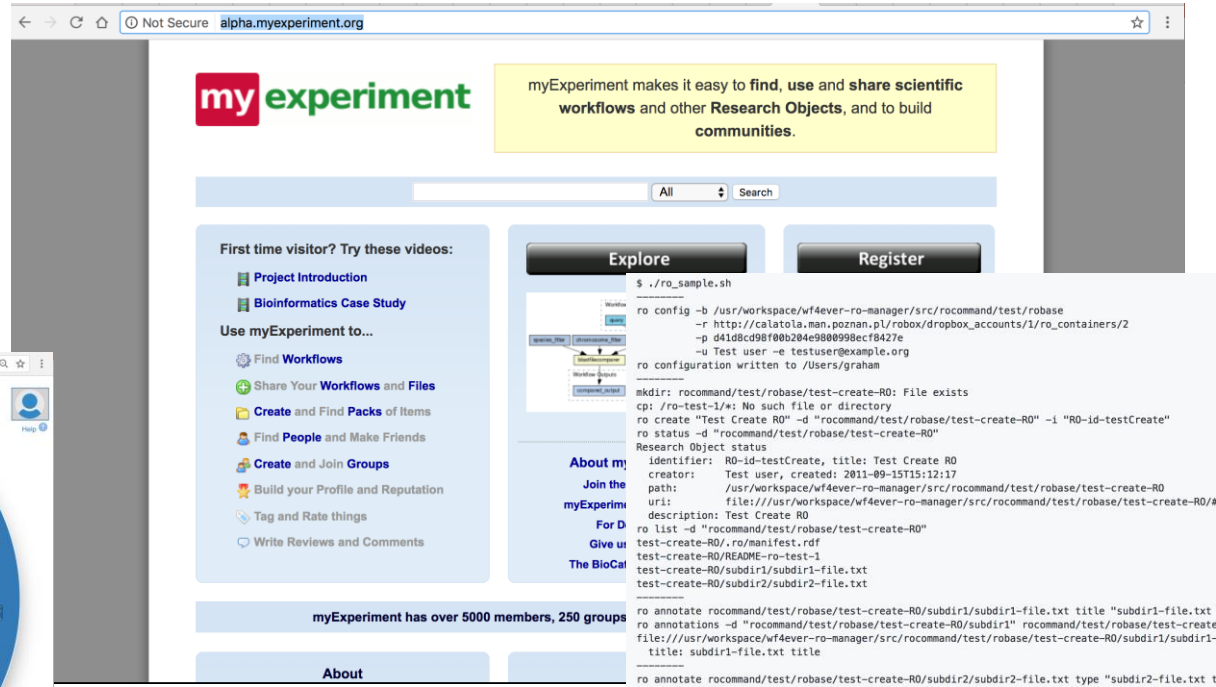
This project is co-funded by the European Union

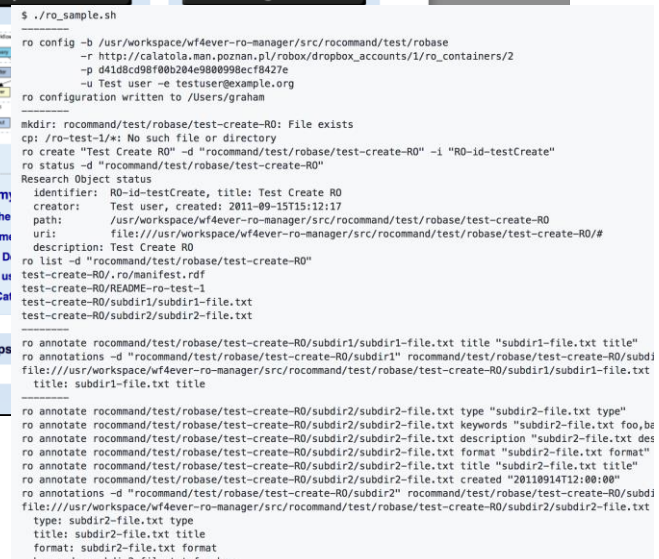
# But there are other user interfaces…

RODL can have other client applications, e.g., VRC portals, collaboration spheres, command-line (ro-manager), myExperiment (alpha),…



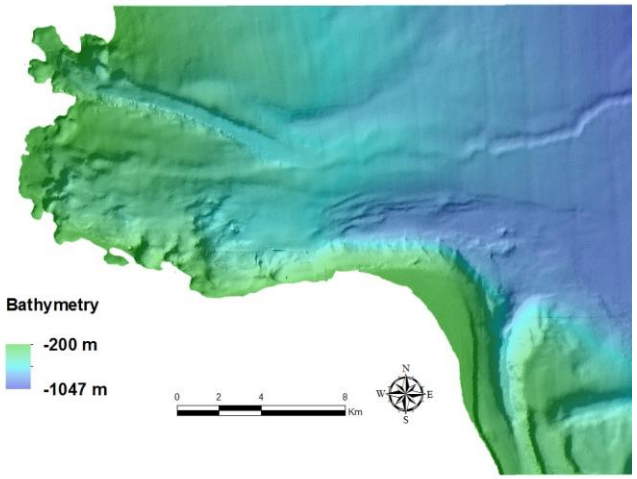



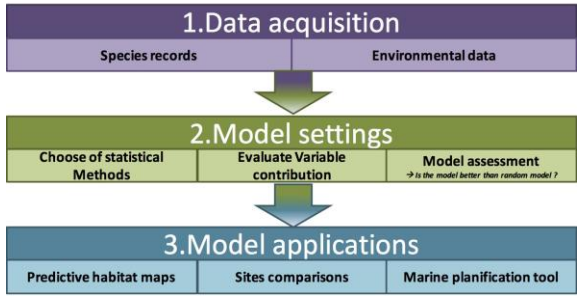



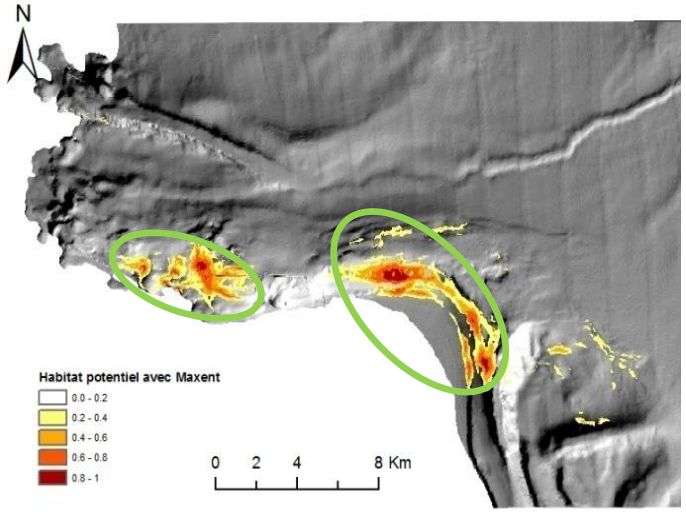
This project is co-funded by the European Union

# Exemplary use case

- Study: predictive habitat mapping of cold-water coral (CWC) in the Bari canyon
  - Relate **species occurrence data** (**CWC distribution**) with **environmental predictor variables** (**EGVs** = Eco-geographic variables)
  - Explain the contribution of each environmental variable to the species distribution
  - Produce **continuous maps** of potential species or habitat
  - **Automatize** this process by formalize it in a scientific **workflow**



This project is co-funded by the European Union

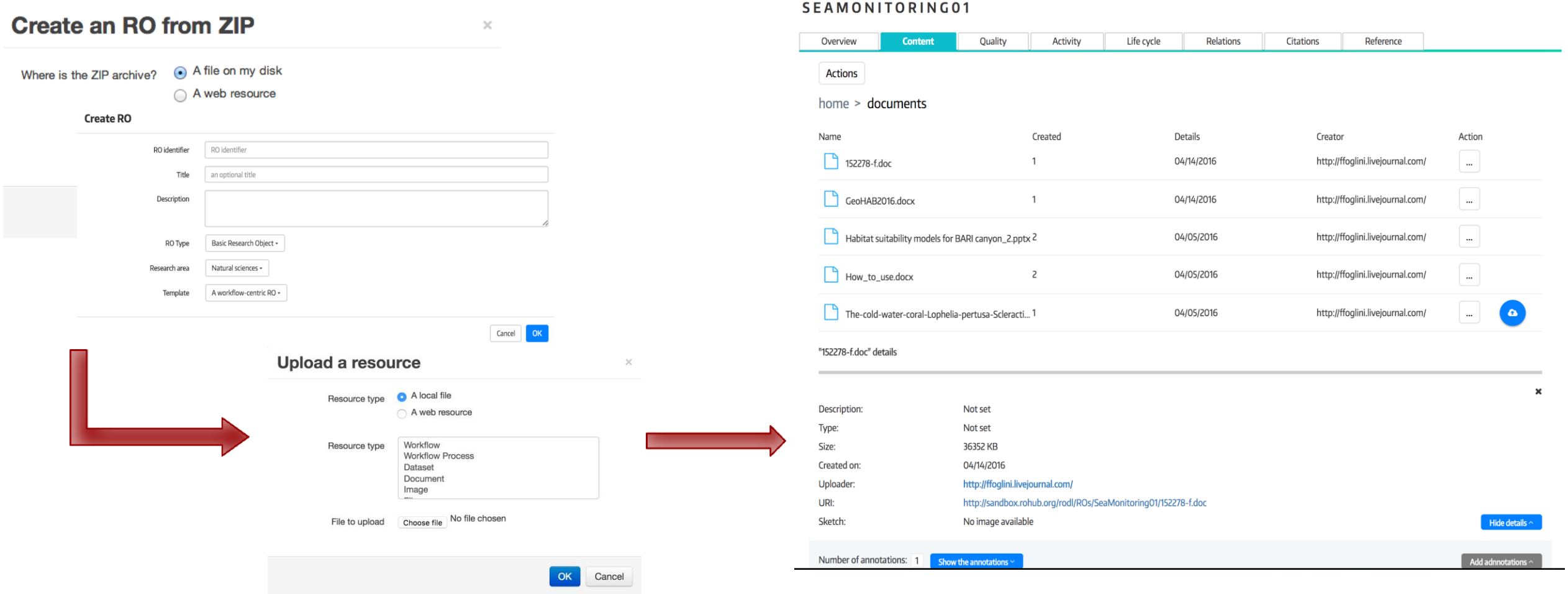
# Exemplary use case

- **Requirements**:
  - To **search** high resolution bathymetric data, and other data, e.g., CWC occurrences data
  - To run a **high quality** model to obtain a reliable map of habitat suitability for CWC.
  - To **share** the results with colleagues from different institutions at the Marine Strategy Framework Directive (**at different points in time**)
  - To **reuse** the model in different locations, and to **re-run the model** after 1year using new data (same location)
  - To **preserve** the results and to **publish** methodologies and final maps
  - To **collaborate** (colleagues should be notified), follow and **keep track of the study lifecycle**
- **Limitations:**
  - **No reference site** where a scientist can **find** publications on this topic, workflows executing the models, links to the data to be used and results (to mention a few).
  - **No specific repositories** that are used to **preserve** and **reuse** all this information
  - **No information** about the **quality** of the models and the methodologies described in the paper.
  - Big **lack** of **communication** and synchronization during the study lifecycle

This project is co-funded by the European Union

# Use case: overcoming existing limitations…

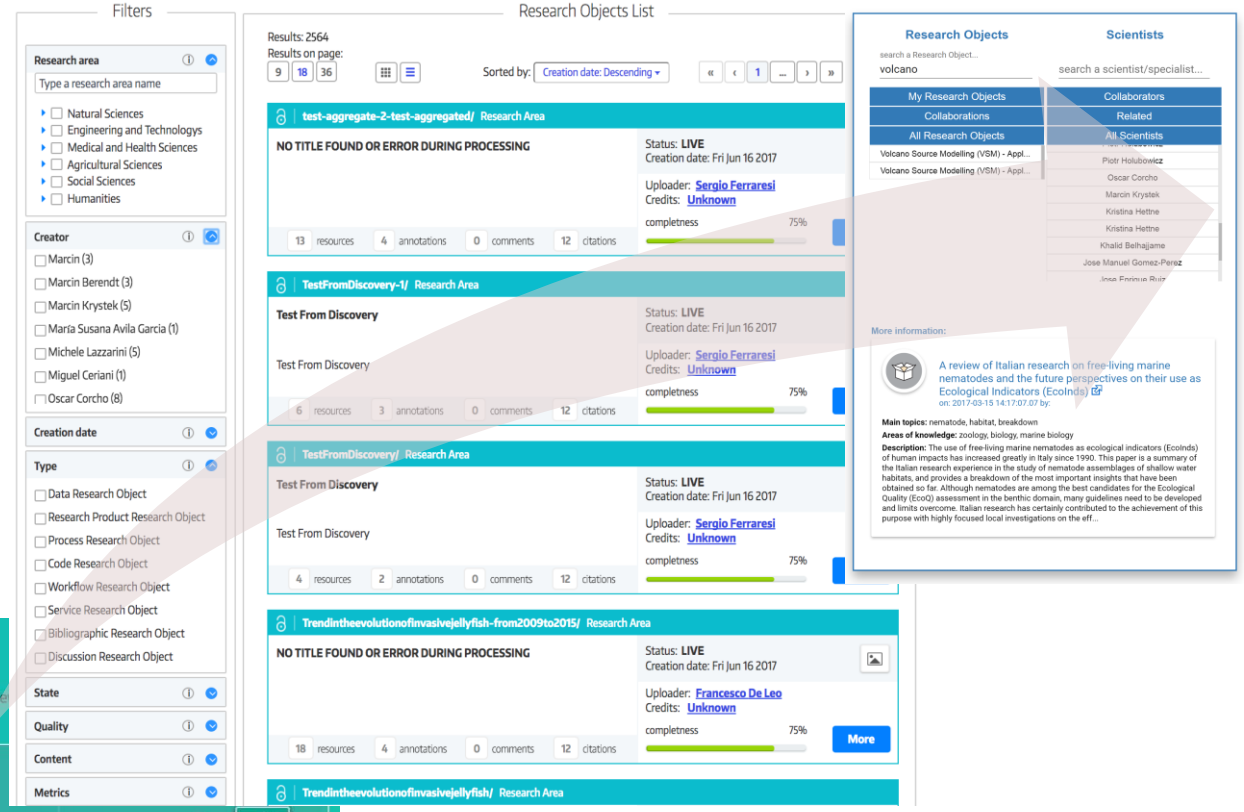
- ROHUB enables scientists to **encapsulate** the data, provenance of workflows executions, results, documentation and other resources related to the particular study through a **single information unit**

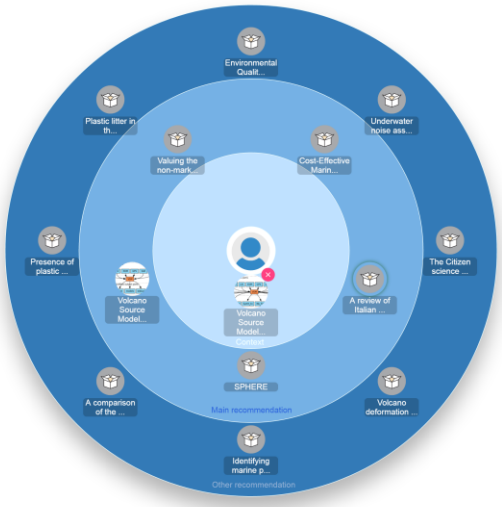


This project is co-funded by the European Union

# Use case: overcoming existing limitations…

- The scientist (& collaborators), then, can **search** for other research objects aggregating relevant resources that can be **re-used**, like potential input data, the processes, or other related resources, to reference them from the new research object

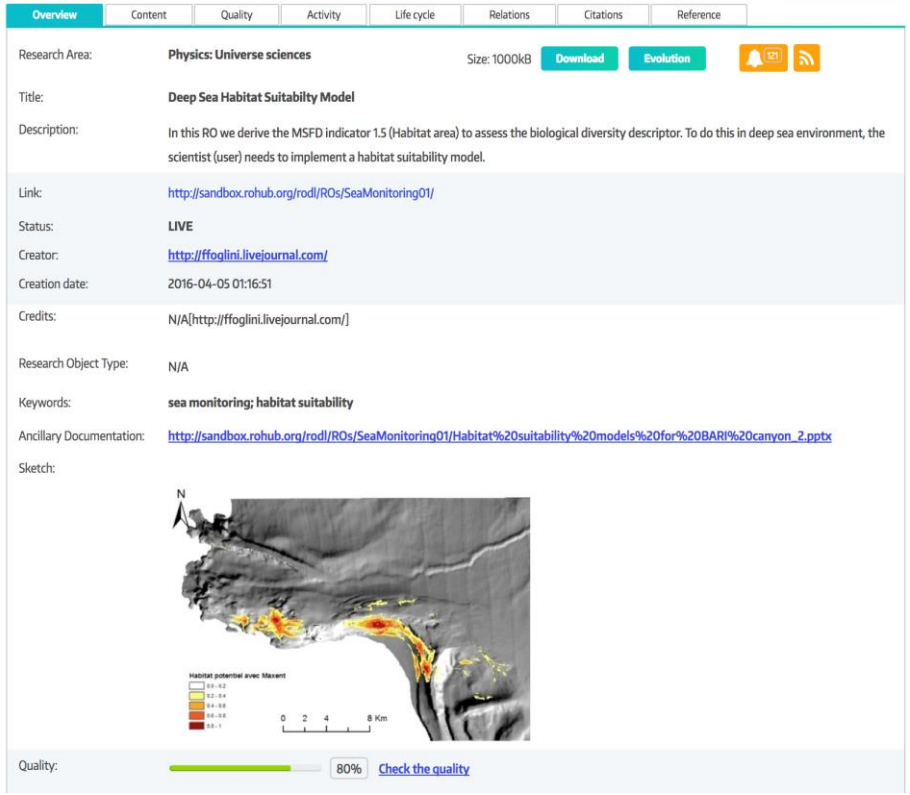



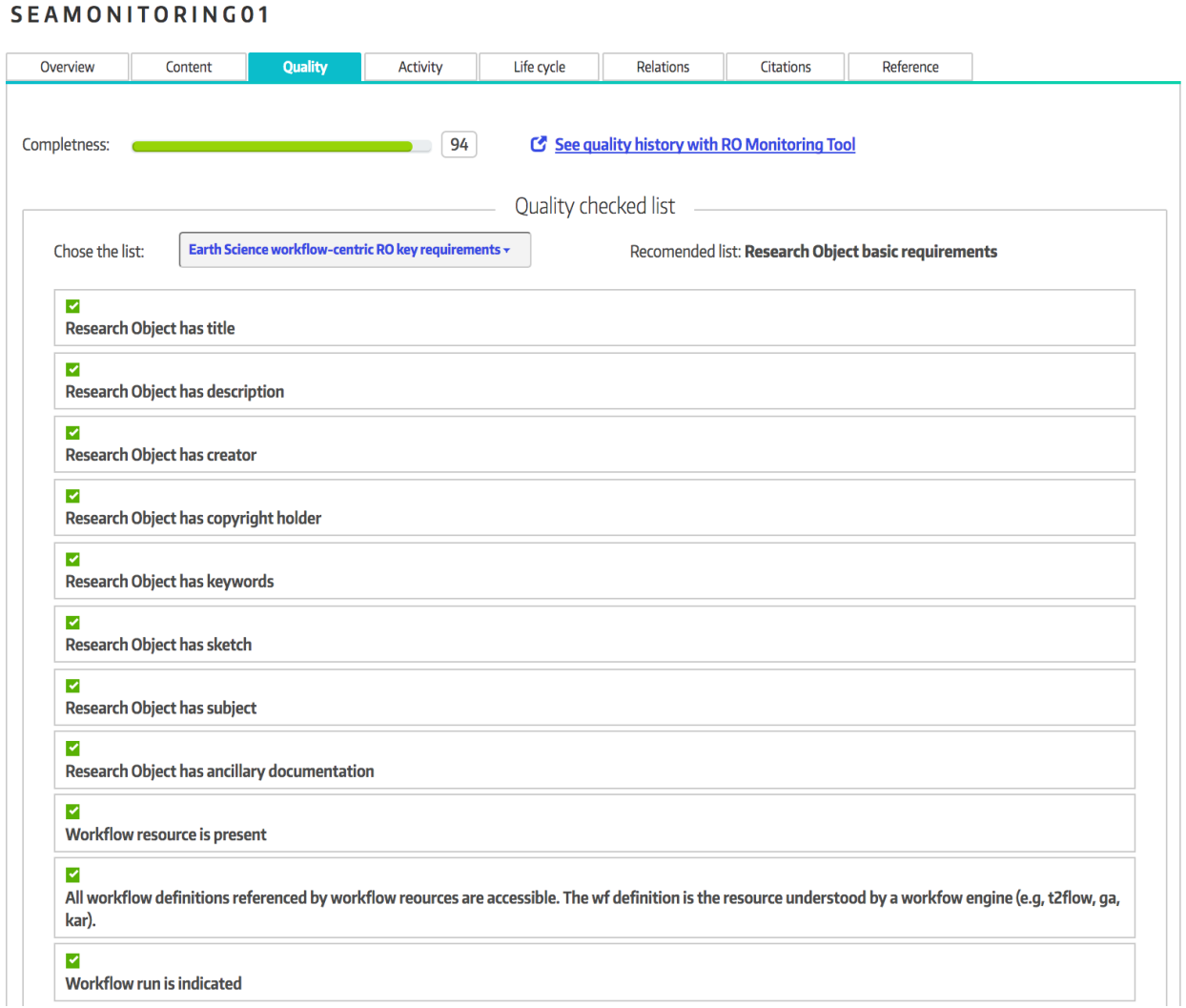



This project is co-funded by the European Union

# Use case: overcoming existing limitations…

- As the research work progresses, the scientist can **verify and get assistance** to build a **high-quality research object**, e.g., that has all the required resources and metadata

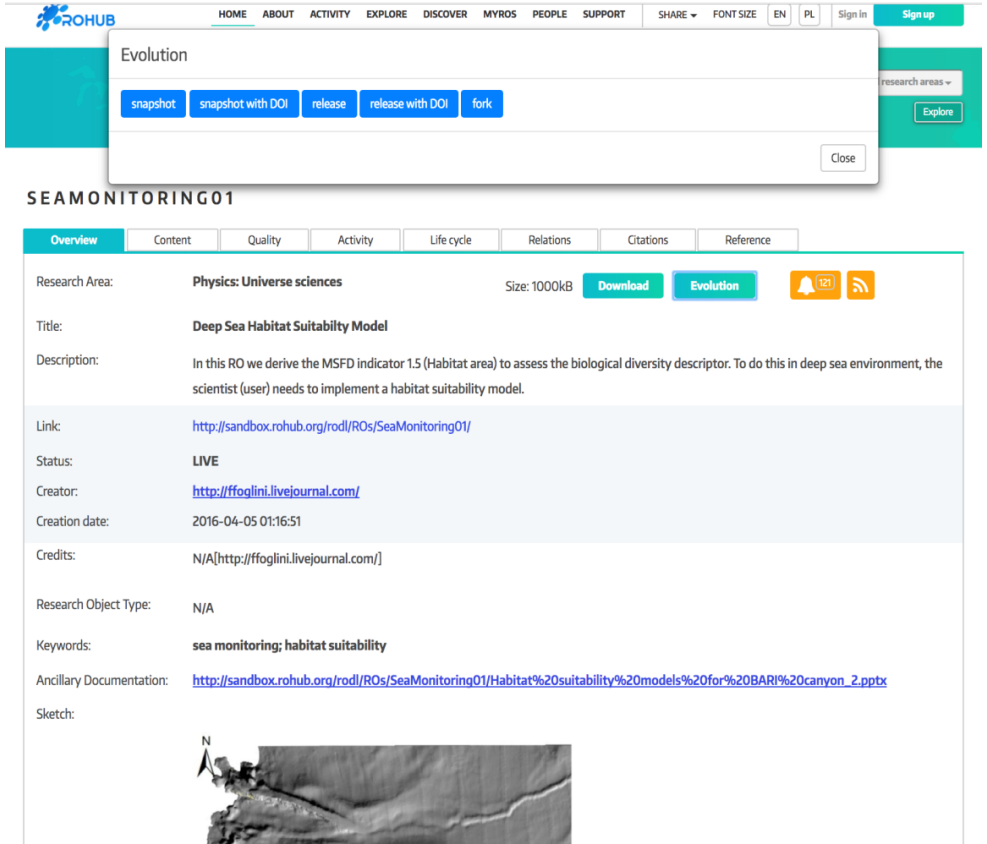


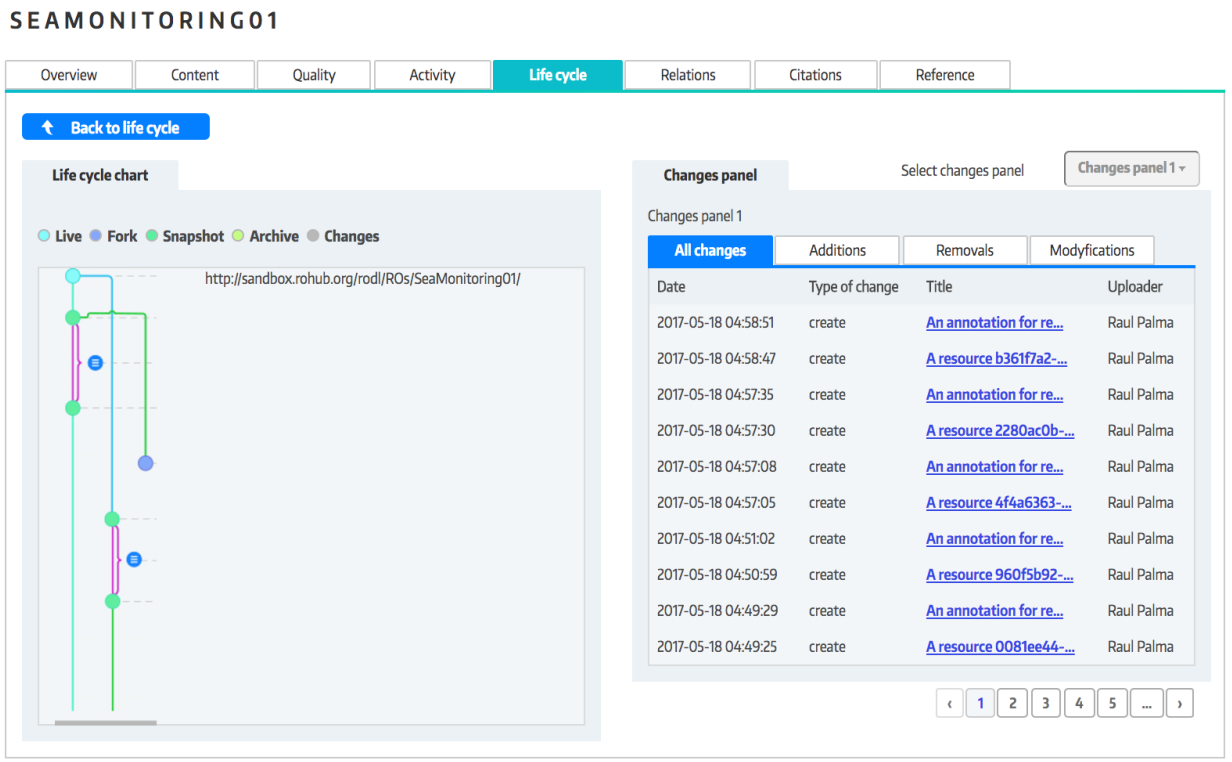



This project is co-funded by the European Union

# Use case: overcoming existing limitations…

- The scientists can i) **release and share** the RO throughout the research lifecycle, so that those releases can then be **review**, **reused** or **re-executed by others;** ii) **visualise** the RO evolution

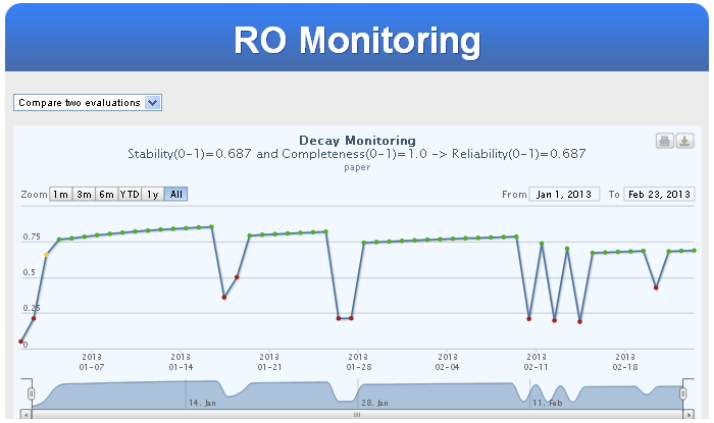


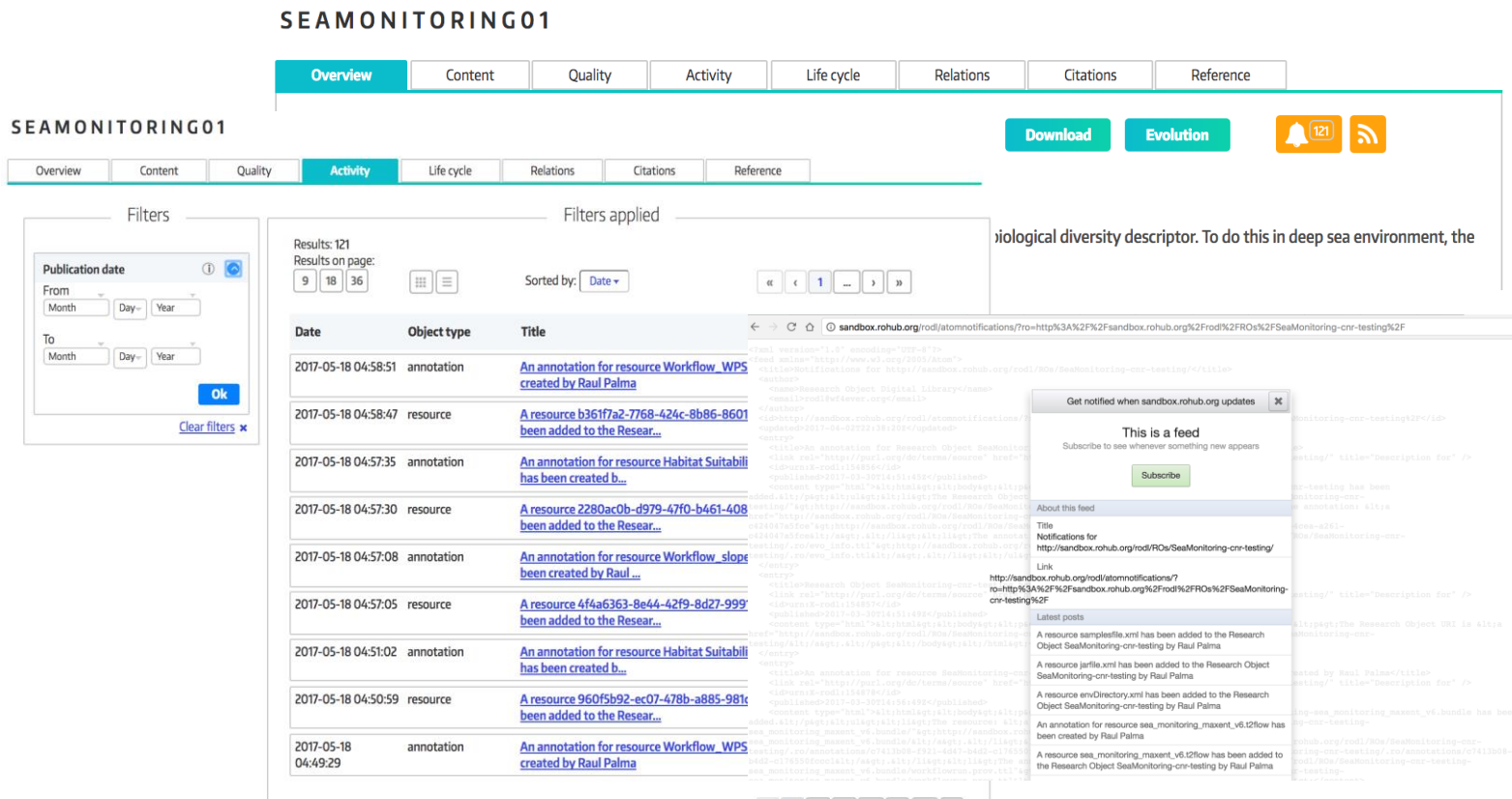



This project is co-funded by the European Union

# Use case: overcoming existing limitations…

- Scientists can **preserve** ROs, so that they can be effectively reused, including the monitoring and notification of changes in the RO quality through time.
- Colleagues can also **subscribe to changes** (quality and content) in the RO **to follow and keep track** of the work progress

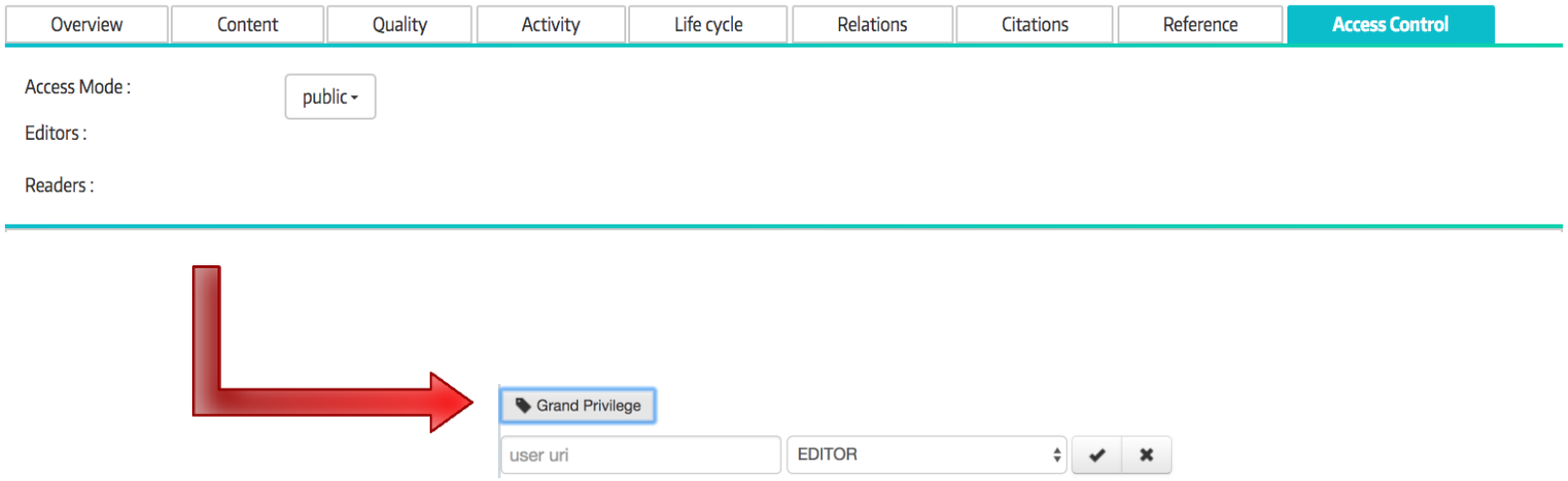


This project is co-funded by the European Union

# Use case: overcoming existing limitations…

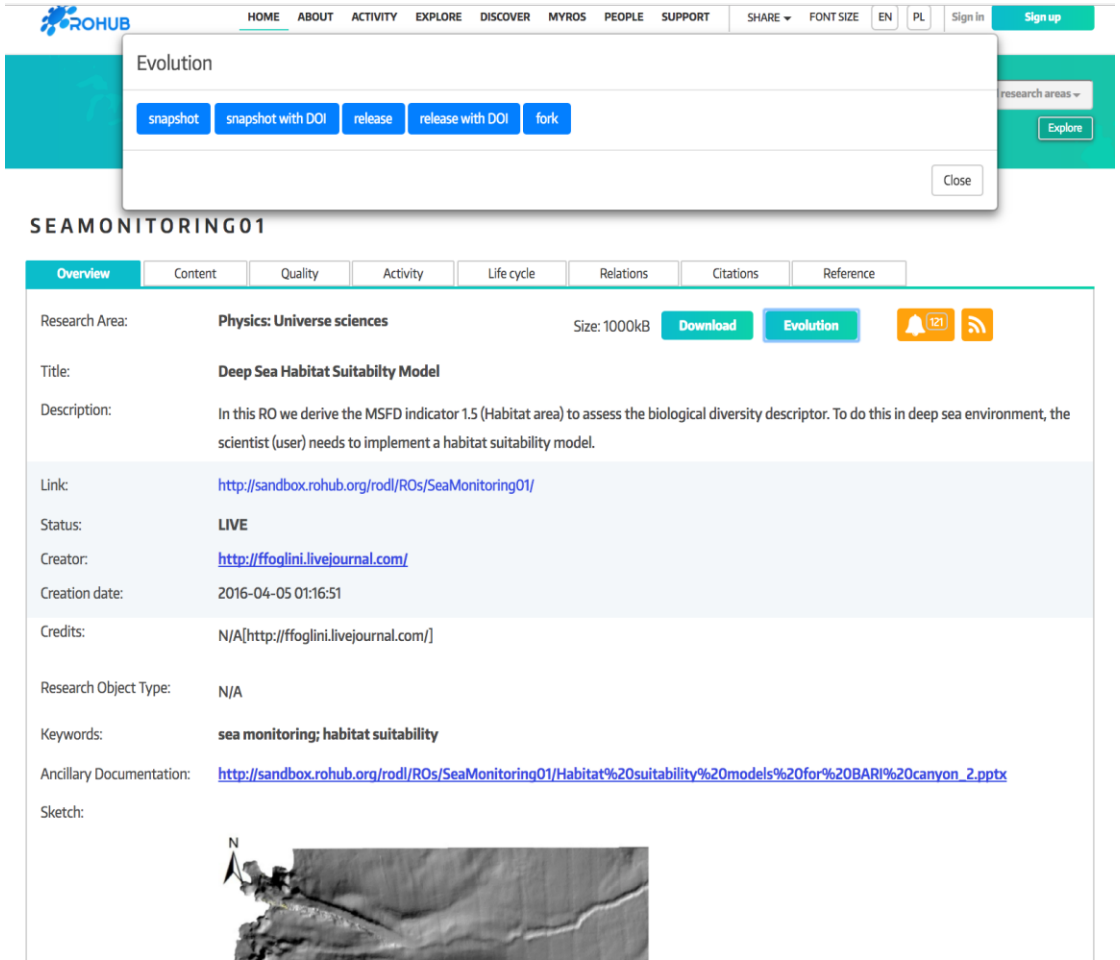
- Scientists can **collaborate** with other colleagues, setting different **access modes** and **permissions** to the RO

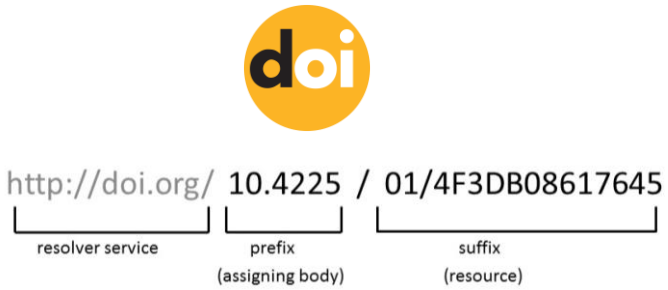


This project is co-funded by the European Union

# Use case: overcoming existing limitations…

- Finally, the results can be **published with DOI** for proper citation and attribution




DataCite
FIND, ACCESS, AND REUSE DATA

This project is co-funded by the European Union


Dr. Raul Palma

rpalma@man.poznan.pl