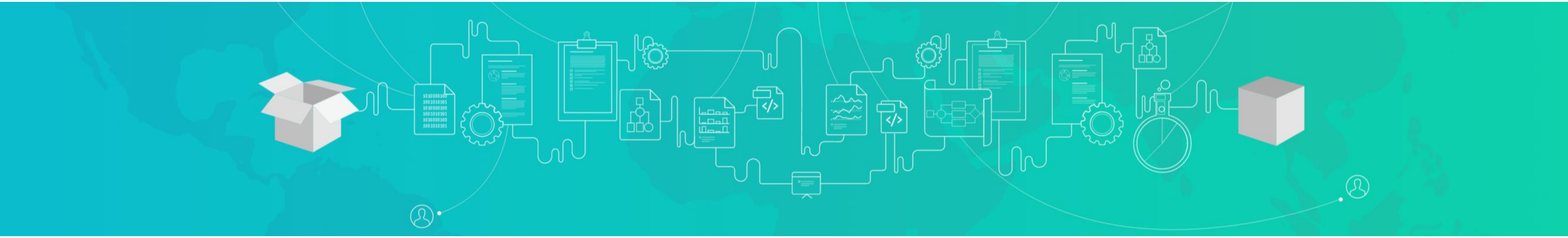


Research Objects and ROHub - A journey from theory to practical infrastructure

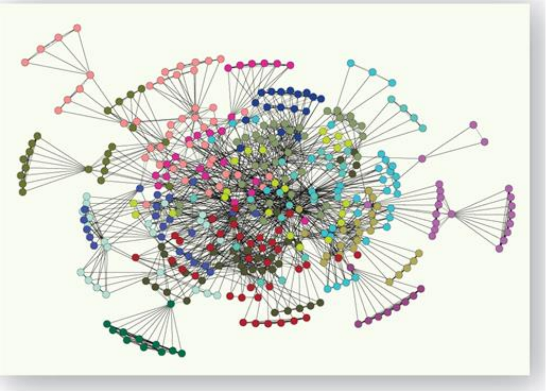
Jose Manuel Gomez-Perez (Expert System)
Raul Palma (PSNC)

EGU – 12th April 2018

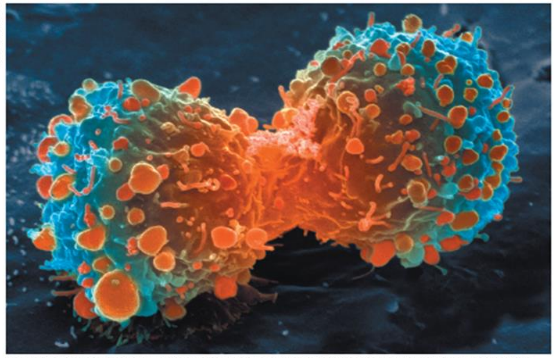
This project is co-funded by
the European Union

The Scientific Enterprise

Collaborate

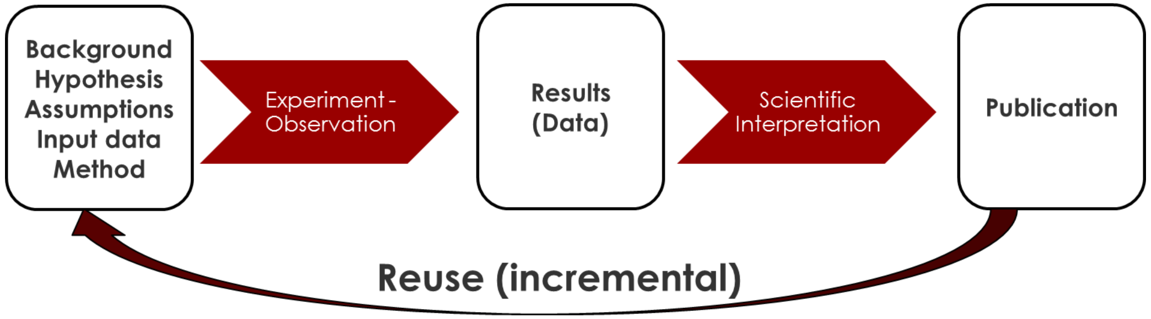


Validate


S. GSCHMEISSNER/SPL

Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for
preclinical cancer research


Background
Hypothesis
Assumptions
Input data
Method

Experiment - Observation

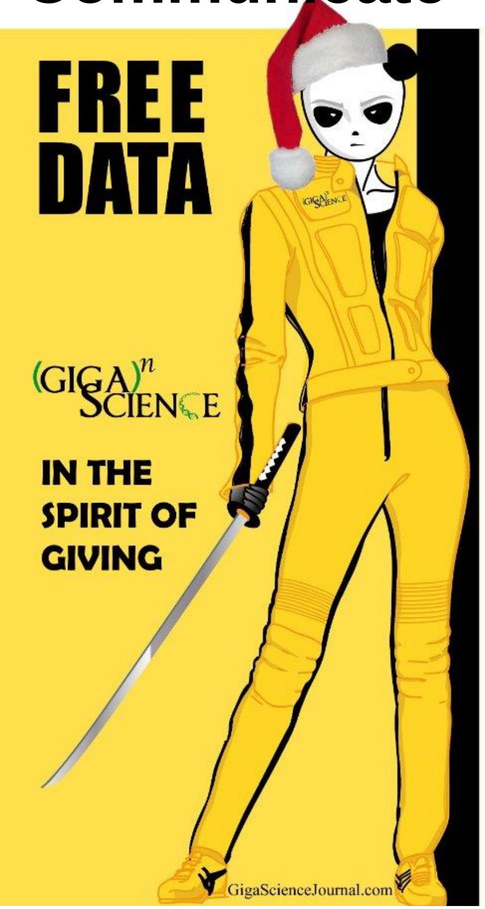
Results
(Data)

Scientific Interpretation

Publication

Reuse (incremental)

Communicate



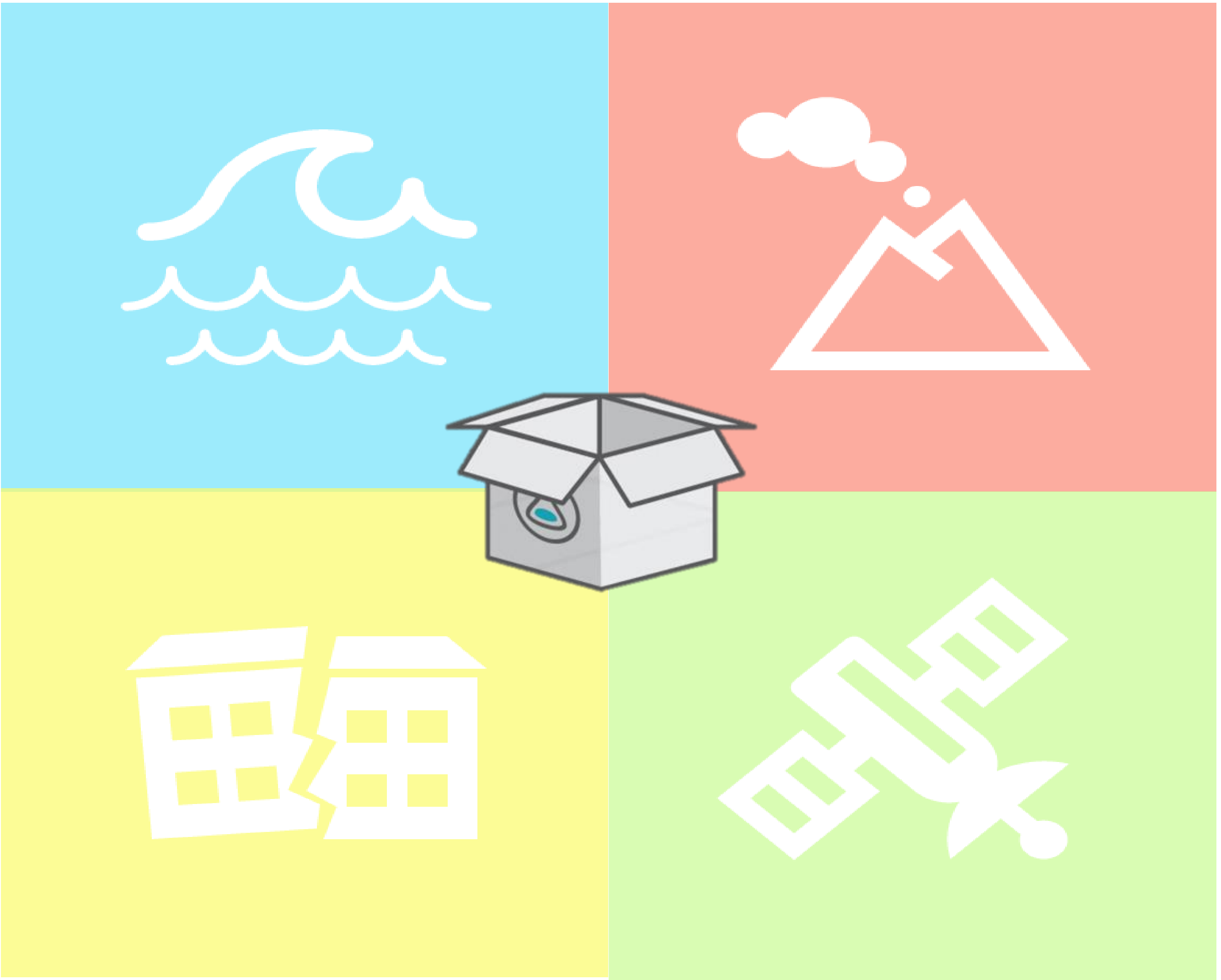
nature.com

PLOS

ELSEVIER

This project is co-funded by the European Union

Research Objects in Earth Sciences



everest

This project is co-funded by the European Union

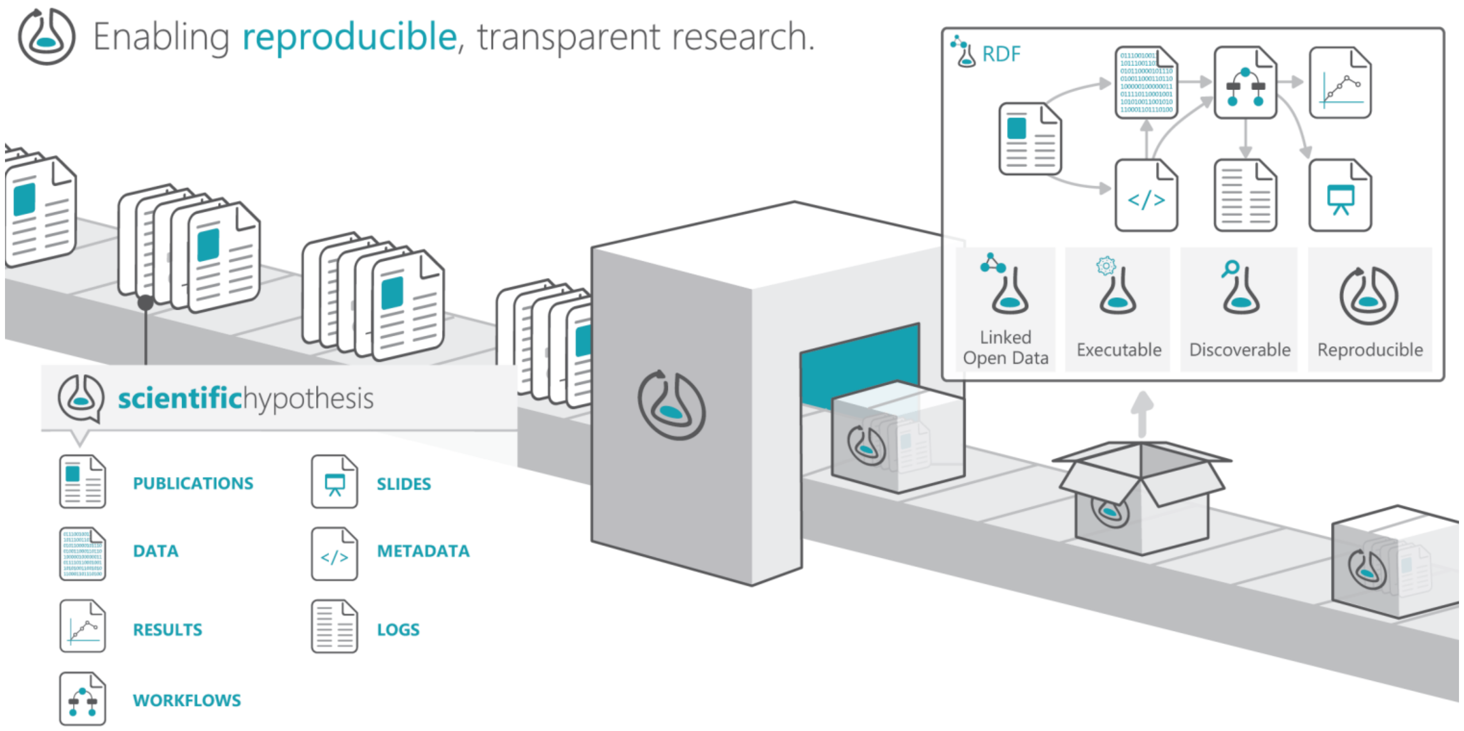
Challenges

- **Long-term preservation**
  - Earth observation missions can cover a long timespan (+30 years)
  - Both data and models need to be preserved (e.g. harvesting of Bathymetry data for Sea Monitoring)
  - Long publication and documentation cycles
- **Sharing & attribution**
  - Reluctance of individual organisations and/or scientists to provide access to their data, methods and tools (IP issues, lack of time or resources, sensitivity of the resources involved, professional rivalry and competitiveness…)
  - Lack of data/methods citation mechanisms to give credit to and incentivize authors to share
- **Automation**
  - Long tail of software and computational resources
  - Limited adoption of scientific workflows for data orchestration and claim validation

This project is co-funded by the European Union

Research Object – The Concept

A **Research Object** is an information artefact that contains and describes *everything* about your research, including how those things are related, in ways that are readable both by humans and machines



http://www.researchobject.org

I. **Logically organize and describe in a single information unit** the resources, materials, methods and outcomes of an investigation

II. **Share** your research materials with other scientists at **discrete milestones of your investigation**. Uniquely identified

III. Enable **reproducibility** and **reuse** of scientific methods

IV. To be **recognized** and **cited**

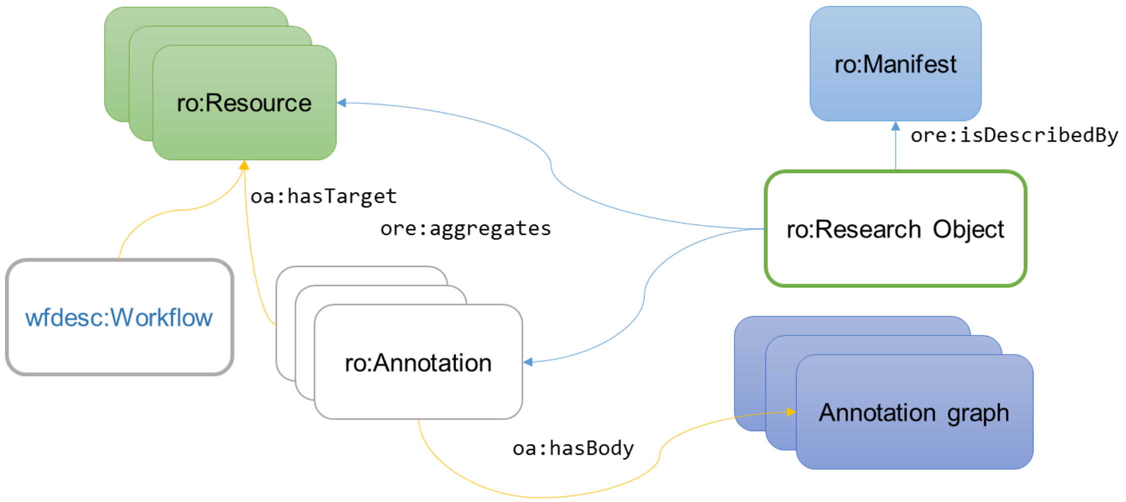
V. **Preserve** results and **prevent decay**

VI. Provide **evidence** to findings claimed in **scholarly articles**

This project is co-funded by the European Union

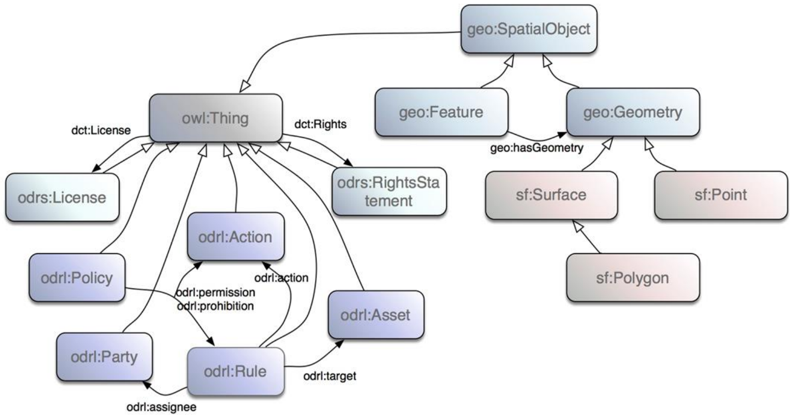
Modeling Research Objects

- **Aggregation** (OAI-ORE) plus **annotation** (Annotation Ontology)
- Other vocabularies used in annotation bodies to provide information about resources, involving types, dependencies, and descriptions

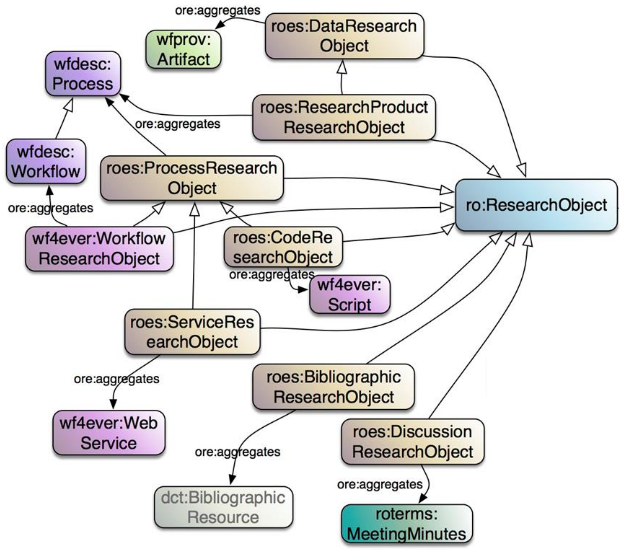


ro (aggregation and annotation)
wfdesc (workflow description)
wfprov (workflow provenance)
roevo (evolution model)
minim (minimum information model)

- **Geospatial information**
- **Time-period coverage**
- **Data access policies**
- **Intellectual Property Rights**
- **General metadata (discipline, size, format, date…)**



- **Eight new types of research object, including:**
  - Workflow-centric, but also process and service-centric
  - Data-centric
  - Research product-centric
  - Documentation
  - Bibliographic

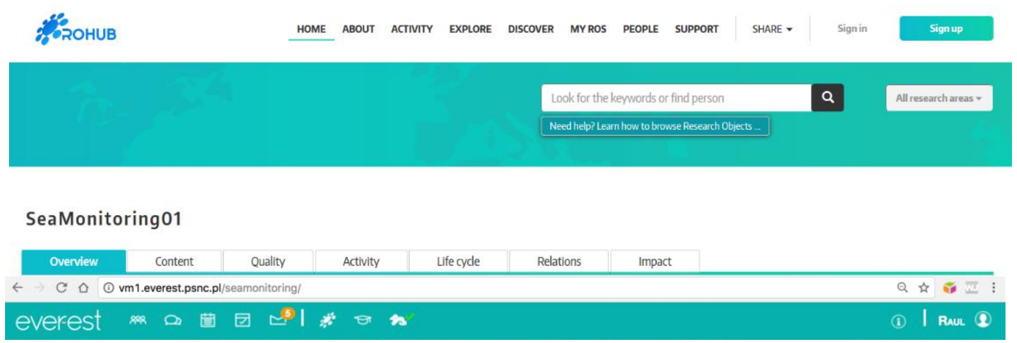


Available at: https://github.com/wf4ever/ro/tree/earth-science

This project is co-funded by the European Union

ROHub.org – The RO Management Platform

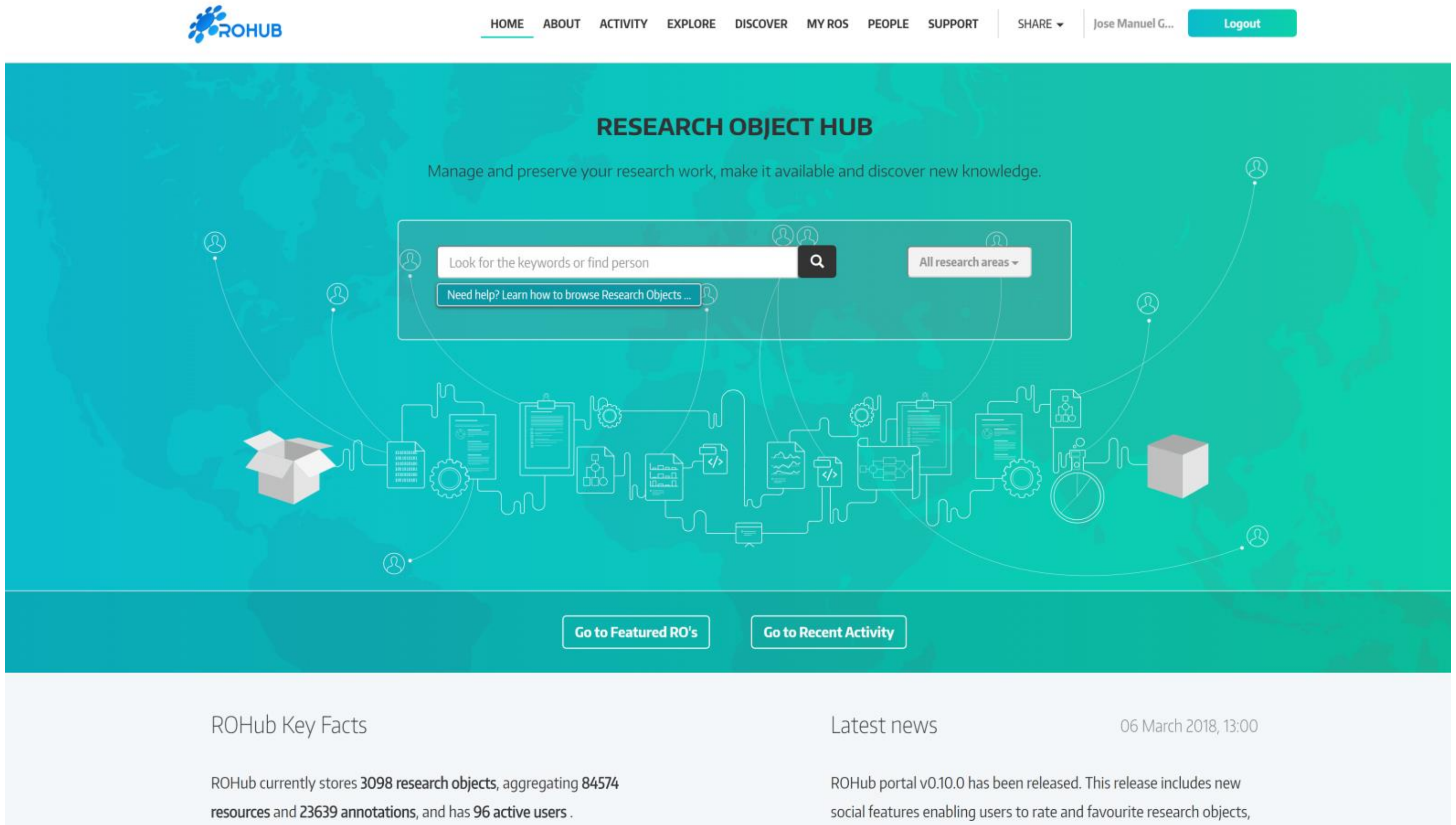
- Comprises both
  - Backend service (RODL) and API
  - A reference web client application (ROHub portal)
- Features include
  - Creation and preservation
  - Lifecycle and version mgmnt.
  - Change tracking
  - Quality and decay monitoring
  - Search, explore, reuse (fork)
  - Follow, subscribe and notifications
  - Likes and ratings



ROHUB
HOME ABOUT ACTIVITY EXPLORE DISCOVER MY ROS PEOPLE SUPPORT SHARE ▾ Sign in Sign up
Look for the keywords or find person
All research areas ▾
Need help? Learn how to browse Research Objects ...
Overview Content Quality Activity Life cycle Relations Impact
everest
RAUL

SeaMonitoring01

vm1.everest.psnc.pl/seamonitoring/



ROHUB
HOME ABOUT ACTIVITY EXPLORE DISCOVER MY ROS PEOPLE SUPPORT SHARE ▾ Jose Manuel G... Logout
Look for the keywords or find person
All research areas ▾
Need help? Learn how to browse Research Objects ...
Go to Featured RO's
Go to Recent Activity

RESEARCH OBJECT HUB

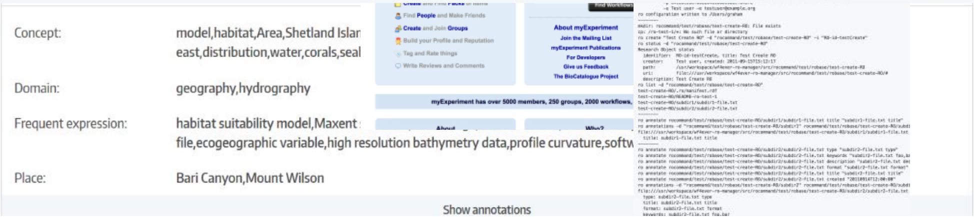
Manage and preserve your research work, make it available and discover new knowledge.

ROHub Key Facts

ROHub currently stores **3098 research objects**, aggregating **84574 resources** and **23639 annotations**, and has **96 active users** .

Latest news 06 March 2018, 13:00

ROHub portal v0.10.0 has been released. This release includes new social features enabling users to rate and favourite research objects,



Concept:	model,habitat,Area,Shetland Islan east,distribution,water,corals,seal
Domain:	geography,hydrography
Frequent expression:	habitat suitability model,Maxent file,ecogeographic variable,high resolution bathymetry data,profile curvature,softw
Place:	Bari Canyon,Mount Wilson

myExperiment has over 5000 members, 250 groups, 2000 workflows,

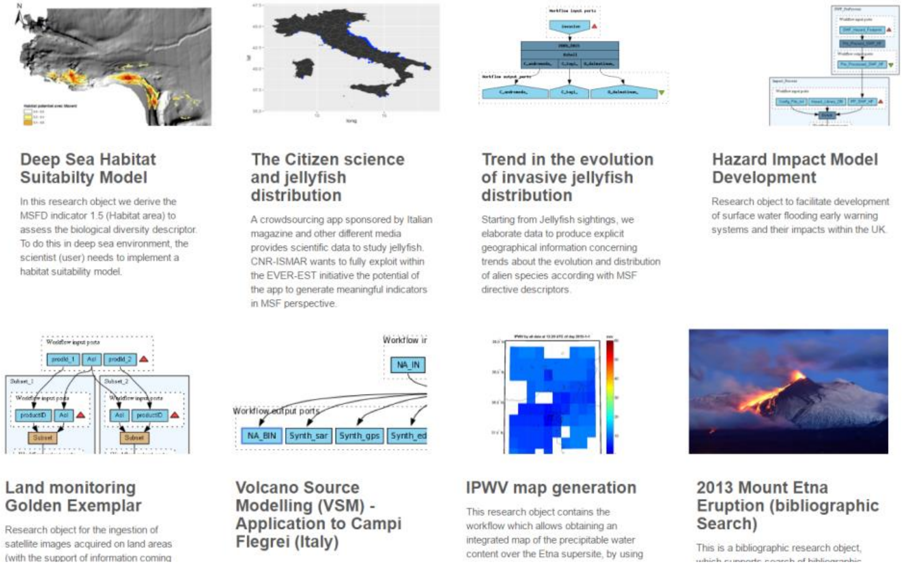
```
ro configuration written to /Users/graham
ro status -d "rocommand/test/robase/test-create-RO"
Research Object status
  identifier:  RO-id-testCreate, title: Test Create RO
  creator:     Test user, created: 2011-09-15T15:12:17
  description: Test Create RO
ro list -d "rocommand/test/robase/test-create-RO"
```

Palma R, Hołubowicz P, Corcho O, Gomez-Perez JM. ROHub - A Digital Library of Research Objects Supporting Scientists Towards Reproducible Science. In Presutti et al. (eds) Semantic Web Evaluation Challenge. SemWebEval 2014, Springer.

This project is co-funded by the European Union

Community Building and Content

Golden Exemplar Research Objects



Deep Sea Habitat Suitabilty Model

In this research object we derive the MSFD indicator 1.5 (Habitat area) to assess the biological diversity descriptor. To do this in deep sea environment, the scientist (user) needs to implement a habitat suitability model.

The Citizen science and jellyfish distribution

A crowdsourcing app sponsored by Italian magazine and other different media provides scientific data to study jellyfish. CNR-ISMAR wants to fully exploit within the EVER-EST initiative the potential of the app to generate meaningful indicators in MSF perspective.

Trend in the evolution of invasive jellyfish distribution

Starting from Jellyfish sightings, we elaborate data to produce explicit geographical information concerning trends about the evolution and distribution of alien species according with MSF directive descriptors.

Hazard Impact Model Development

Research object to facilitate development of surface water flooding early warning systems and their impacts within the UK.

Land monitoring Golden Exemplar

Research object for the ingestion of satellite images acquired on land areas (with the support of information coming

Volcano Source Modelling (VSM) - Application to Campi Flegrei (Italy)

IPWV map generation

This research object contains the workflow which allows obtaining an integrated map of the precipitable water content over the Etna supersite, by using

2013 Mount Etna Eruption (bibliographic Search)

This is a bibliographic research object, which supports search of bibliographic

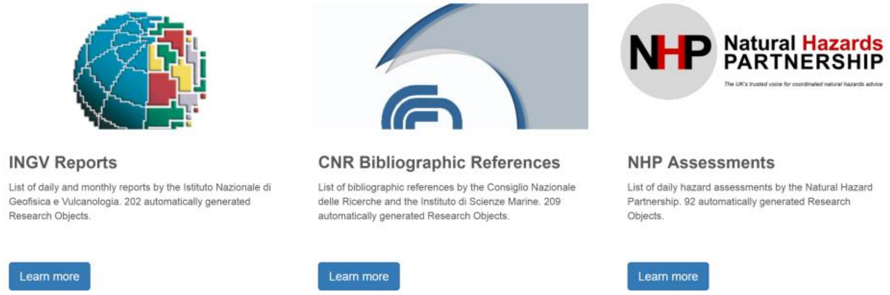
http://everest.expertsystemlab.com/home/#Golden%20Exemplars

- 20+ GEROs produced by the EVER-EST communities and collaboration with the USA National Ecological Observatory Network (NEON) and UNAVCO

ROHub currently stores **3.099 research objects**, aggregating **84.593 resources** and **23.644 annotations**

Automatically generated Bibliographic Research Objects

- ~700 ABROs currently produced
- Semantically annotated
- Include grey literature, field reports, heterogeneous operational information…



INGV Reports

List of daily and monthly reports by the Istituto Nazionale di Geofisica e Vulcanologia. 202 automatically generated Research Objects.

Learn more

CNR Bibliographic References

List of bibliographic references by the Consiglio Nazionale delle Ricerche and the Instituto di Scienze Marine. 209 automatically generated Research Objects.

Learn more

NHP Assessments

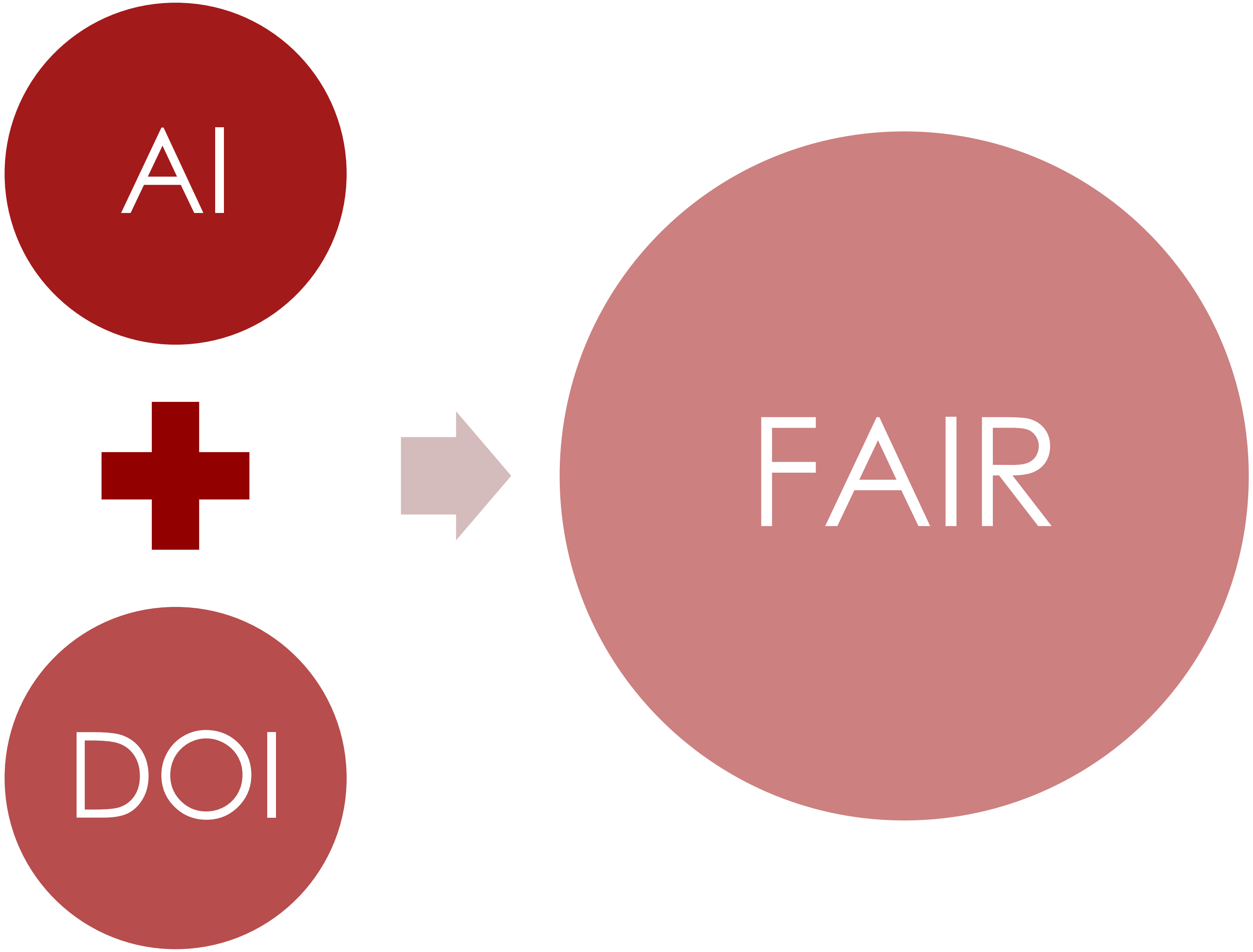
List of daily hazard assessments by the Natural Hazard Partnership. 92 automatically generated Research Objects.

Learn more

NHP Natural Hazards PARTNERSHIP
The UK's trusted voice for coordinated natural hazards advice

This project is co-funded by
the European Union

Two recent and relevant developments



AI

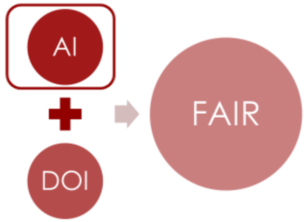
DOI

FAIR

Findable ✓
Accessible ✓
Interoperable ✓
Reusable ✓

This project is co-funded by the European Union

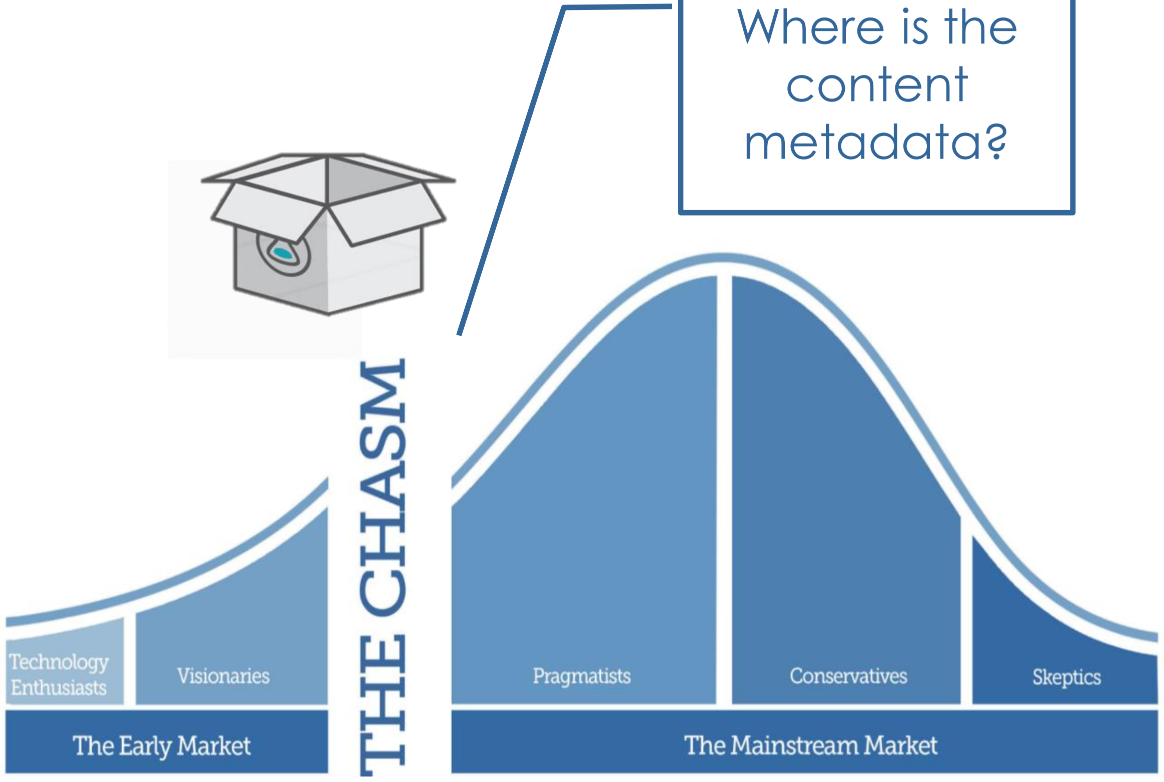
The Metadata Chasm



- **Towards FAIR-ness, metadata is key**
  - **For scientists** *("would this research object fit my investigation, as a whole or partially?")*
  - **For machines,** through machine-readable annotations by search engines or recommendation systems
  - **For both:** To answer scientific questions
- **Research object metadata usually generated manually** (labor-intensive and scarce)
- **Metadata focused on lifecycle, structure and resource types rather than actual payload -** valuable knowledge sources like scientific papers, field notes or technical reports ignored
- **Related information hidden** and non-actionable for machine discovery, search or reasoning
- **Limited diffusion and reuse** of scientific outcomes

Manual inspection of 2,500 research objects showed only a third have a proper title, with average length of 38 chars. Also, **usually short (138 chars) and non-descriptive descriptions**

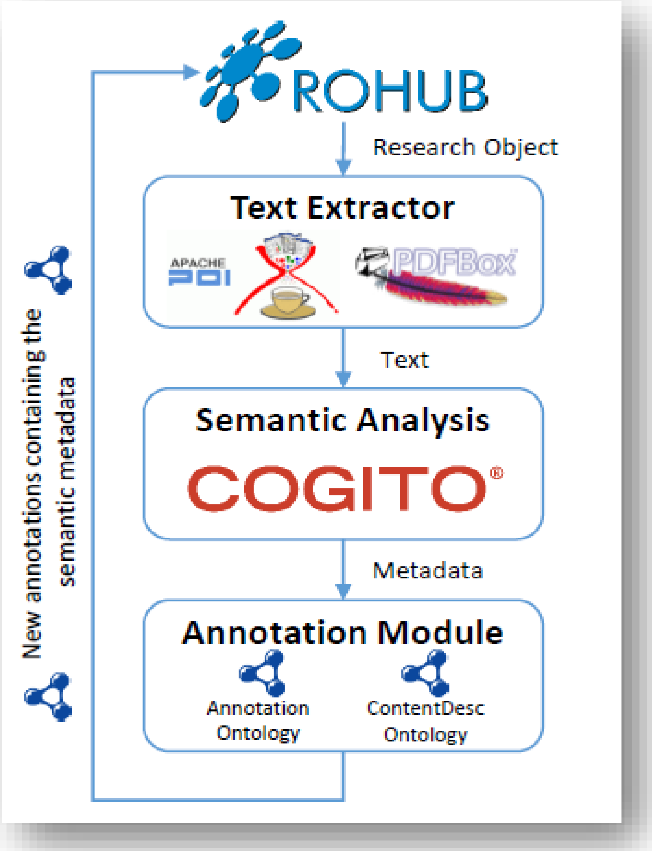
Where is the content metadata?



This project is co-funded by the European Union

Crossing the Chasm – Semantically Enriching Research Objects

AI + DOI → FAIR



- **Automated Semantic Annotation Natural Language Processing**
- **COGITO**, standard version w/o earth science extensions
  - Main entities include concepts, lemmas (canonical representation of a word) and relations (properties, hypernymy, polysemy, synonymy…)
  - ~300K syncons, ~400K lemmas, 80+ relation types (~2.8 million links)
  - Supports word-sense disambiguation based on word context
- Semantic enrichment annotates **the most significant concepts, domains, lemmas, noun phrases and named entities** in research object resources *(titles, descriptions, papers, bibliography…)*

Gomez-Perez, J. M., Palma, R., & Garcia-Silva, A. (2017). **Towards a Human-Machine Scientific Partnership Based on Semantically Rich Research Objects**. In *2017 IEEE 13th International Conference on e-Science (eScience)* (pp. 266–275). IEEE. https://doi.org/10.1109/eScience.2017.40

13th eScience
IEEE INTERNATIONAL CONFERENCE
24 - 27 October 2017 | Auckland, New Zealand

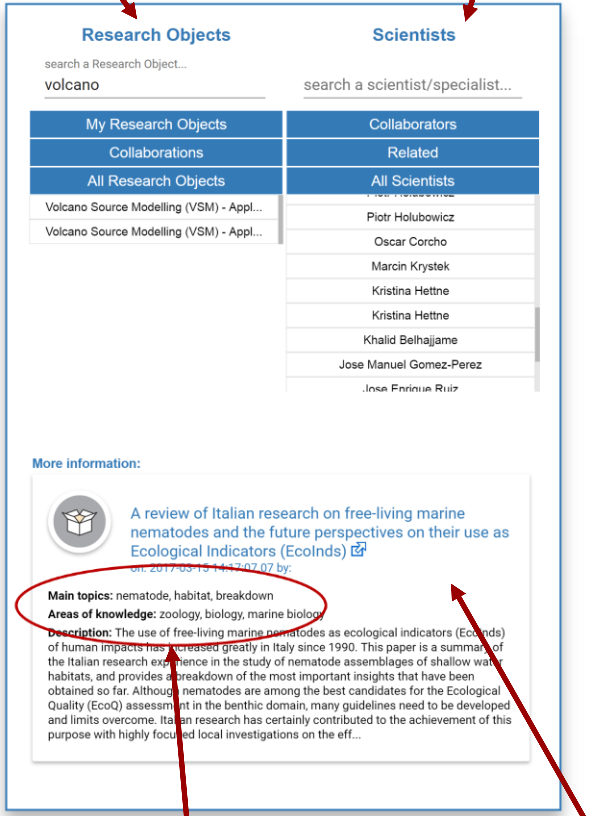
This project is co-funded by the European Union

Richer, Machine-Readable Metadata Enables Cool Apps…

AI ✚ DOI → FAIR

Content

Other users



Research Objects

search a Research Object...
volcano

Scientists

search a scientist/specialist...

My Research Objects | Collaborations | All Research Objects

Collaborators | Related | All Scientists

Volcano Source Modelling (VSM) - Appl...
Volcano Source Modelling (VSM) - Appl...

Piotr Holubowicz
Oscar Corcho
Marcin Krystek
Kristina Hettne
Kristina Hettne
Khalid Belhajjame
Jose Manuel Gomez-Perez
Jose Enrique Ruiz

More information:

A review of Italian research on free-living marine nematodes and the future perspectives on their use as Ecological Indicators (EcoInds)
on: 2017-03-15 14:17:07.07 by:

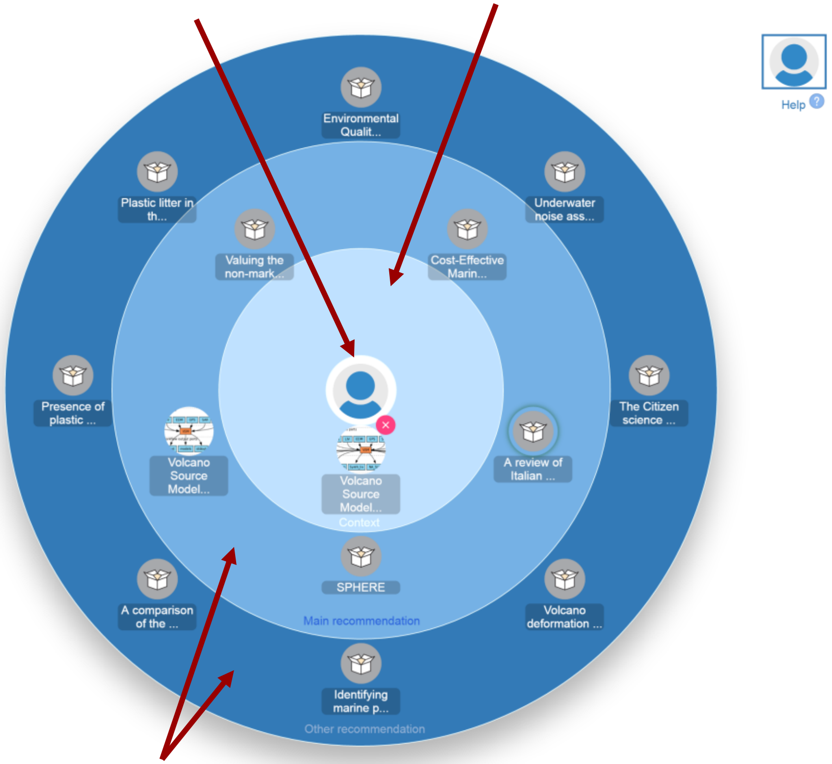
Main topics: nematode, habitat, breakdown
Areas of knowledge: zoology, biology, marine biology
Description: The use of free-living marine nematodes as ecological indicators (EcoInds) of human impacts has increased greatly in Italy since 1990. This paper is a summary of the Italian research experience in the study of nematode assemblages of shallow water habitats, and provides a breakdown of the most important insights that have been obtained so far. Although nematodes are among the best candidates for the Ecological Quality (EcoQ) assessment in the benthic domain, many guidelines need to be developed and limits overcome. Italian research has certainly contributed to the achievement of this purpose with highly focused local investigations on the eff...

Automatically generated metadata

Information card

Me

Context of interest



Ranked recommendation

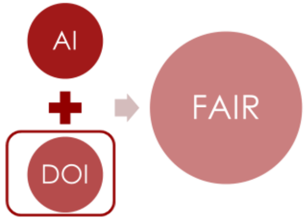
EXPERT SYSTEM COGITO ROHUB

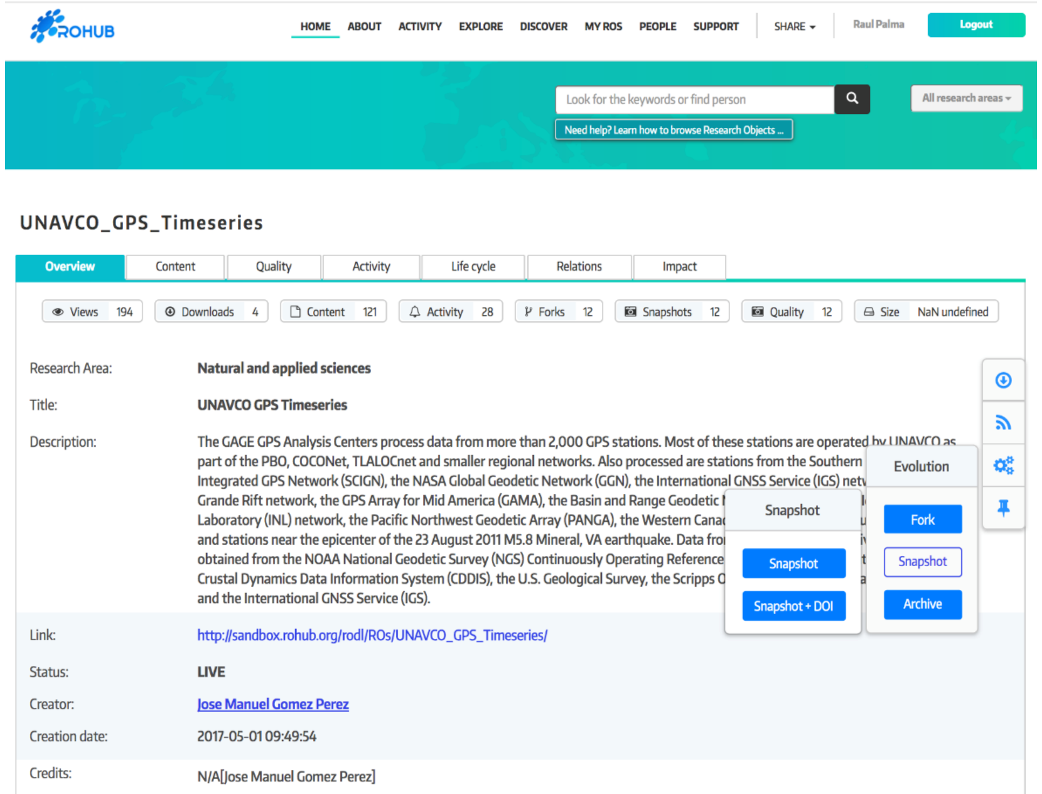
- Exploratory search and recommendation of research objects in scientific social networks
- Recommend by example
- Focused on the aggregated similarity of the context of interest with other items in the repository
- Based on metadata automatically generated from research object content and structure
- Extends findability and reusability
- Reduces cognitive load exploring scientific repositories

Rico M, Gomez-Perez JM, González R, Garrido A, Corcho O. (2017). Collaboration Spheres A Visual Metaphor to Share and Reuse Research Objects. *arXiv preprints https://arxiv.org/abs/1710.05604v1*

This project is co-funded by the European Union

DOI for Citation and Reuse





UNAVCO_GPS_Timeseries

Research Area: Natural and applied sciences

Title: UNAVCO GPS Timeseries

Description: The GAGE GPS Analysis Centers process data from more than 2,000 GPS stations. Most of these stations are operated by UNAVCO as part of the PBO, COCONet, TLALOCnet and smaller regional networks. Also processed are stations from the Southern Integrated GPS Network (SCIGN), the NASA Global Geodetic Network (GGN), the International GNSS Service (IGS) netw Grande Rift network, the GPS Array for Mid America (GAMA), the Basin and Range Geodetic I Laboratory (INL) network, the Pacific Northwest Geodetic Array (PANGA), the Western Canad and stations near the epicenter of the 23 August 2011 M5.8 Mineral, VA earthquake. Data fro obtained from the NOAA National Geodetic Survey (NGS) Continuously Operating Reference Crustal Dynamics Data Information System (CDDIS), the U.S. Geological Survey, the Scripps O and the International GNSS Service (IGS).

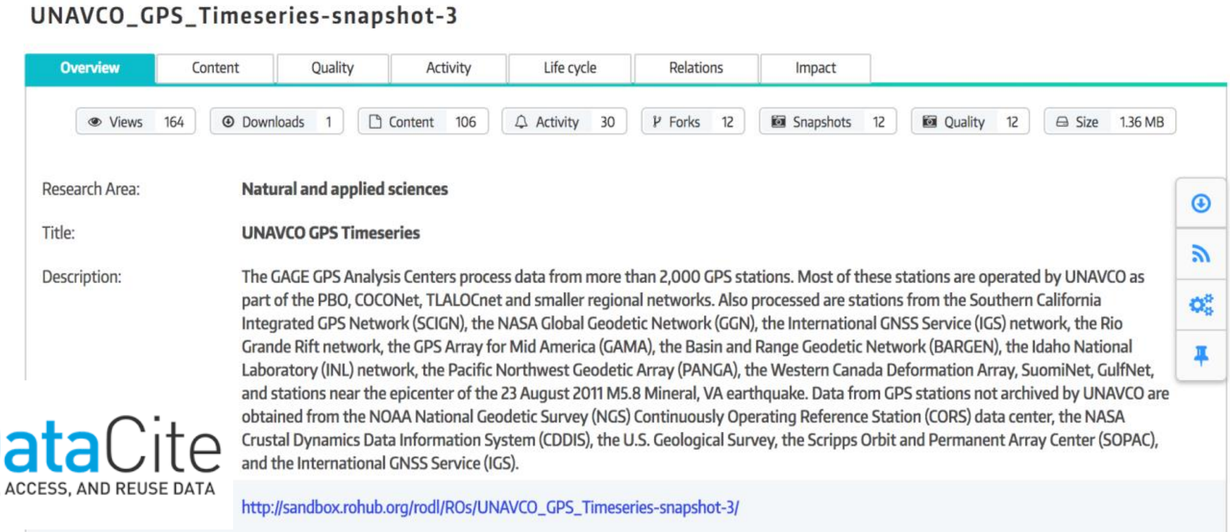
Link: http://sandbox.rohub.org/rodl/ROs/UNAVCO_GPS_Timeseries/

Status: LIVE

Creator: Jose Manuel Gomez Perez

Creation date: 2017-05-01 09:49:54

Credits: N/A[Jose Manuel Gomez Perez]



UNAVCO_GPS_Timeseries-snapshot-3

Research Area: Natural and applied sciences

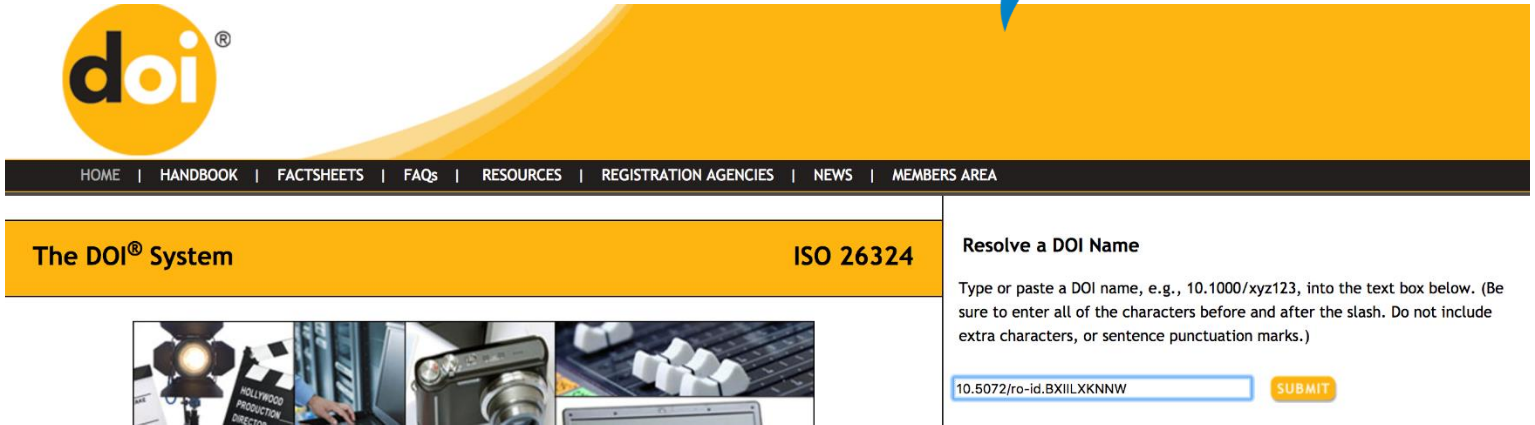
Title: UNAVCO GPS Timeseries

Description: The GAGE GPS Analysis Centers process data from more than 2,000 GPS stations. Most of these stations are operated by UNAVCO as part of the PBO, COCONet, TLALOCnet and smaller regional networks. Also processed are stations from the Southern California Integrated GPS Network (SCIGN), the NASA Global Geodetic Network (GGN), the International GNSS Service (IGS) network, the Rio Grande Rift network, the GPS Array for Mid America (GAMA), the Basin and Range Geodetic Network (BARGEN), the Idaho National Laboratory (INL) network, the Pacific Northwest Geodetic Array (PANGA), the Western Canada Deformation Array, SuomiNet, GulfNet, and stations near the epicenter of the 23 August 2011 M5.8 Mineral, VA earthquake. Data from GPS stations not archived by UNAVCO are obtained from the NOAA National Geodetic Survey (NGS) Continuously Operating Reference Station (CORS) data center, the NASA Crustal Dynamics Data Information System (CDDIS), the U.S. Geological Survey, the Scripps Orbit and Permanent Array Center (SOPAC), and the International GNSS Service (IGS).

http://sandbox.rohub.org/rodl/ROs/UNAVCO_GPS_Timeseries-snapshot-3/

DataCite
FIND, ACCESS, AND REUSE DATA

DOI: 10.5072/ro-id.BXIILXKNNW



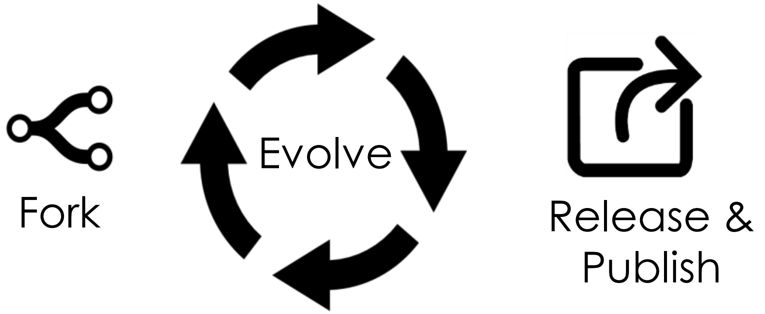
doi®
HOME | HANDBOOK | FACTSHEETS | FAQs | RESOURCES | REGISTRATION AGENCIES | NEWS | MEMBERS AREA
SUBMIT

The DOI® System ISO 26324

Resolve a DOI Name

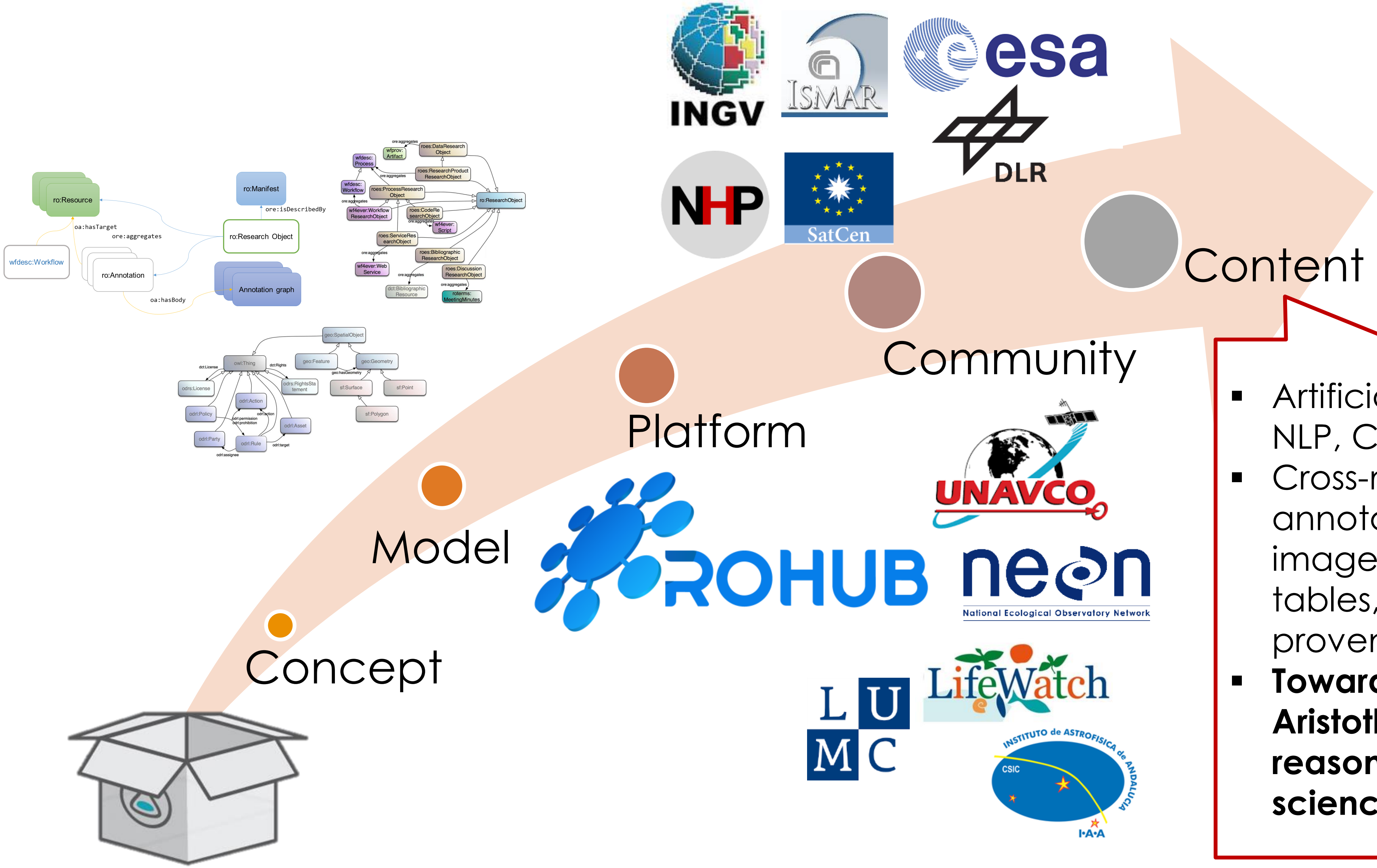
Type or paste a DOI name, e.g., 10.1000/xyz123, into the text box below. (Be sure to enter all of the characters before and after the slash. Do not include extra characters, or sentence punctuation marks.)

10.5072/ro-id.BXIILXKNNW



This project is co-funded by the European Union

The Research Object Journey



Concept

Model

Platform

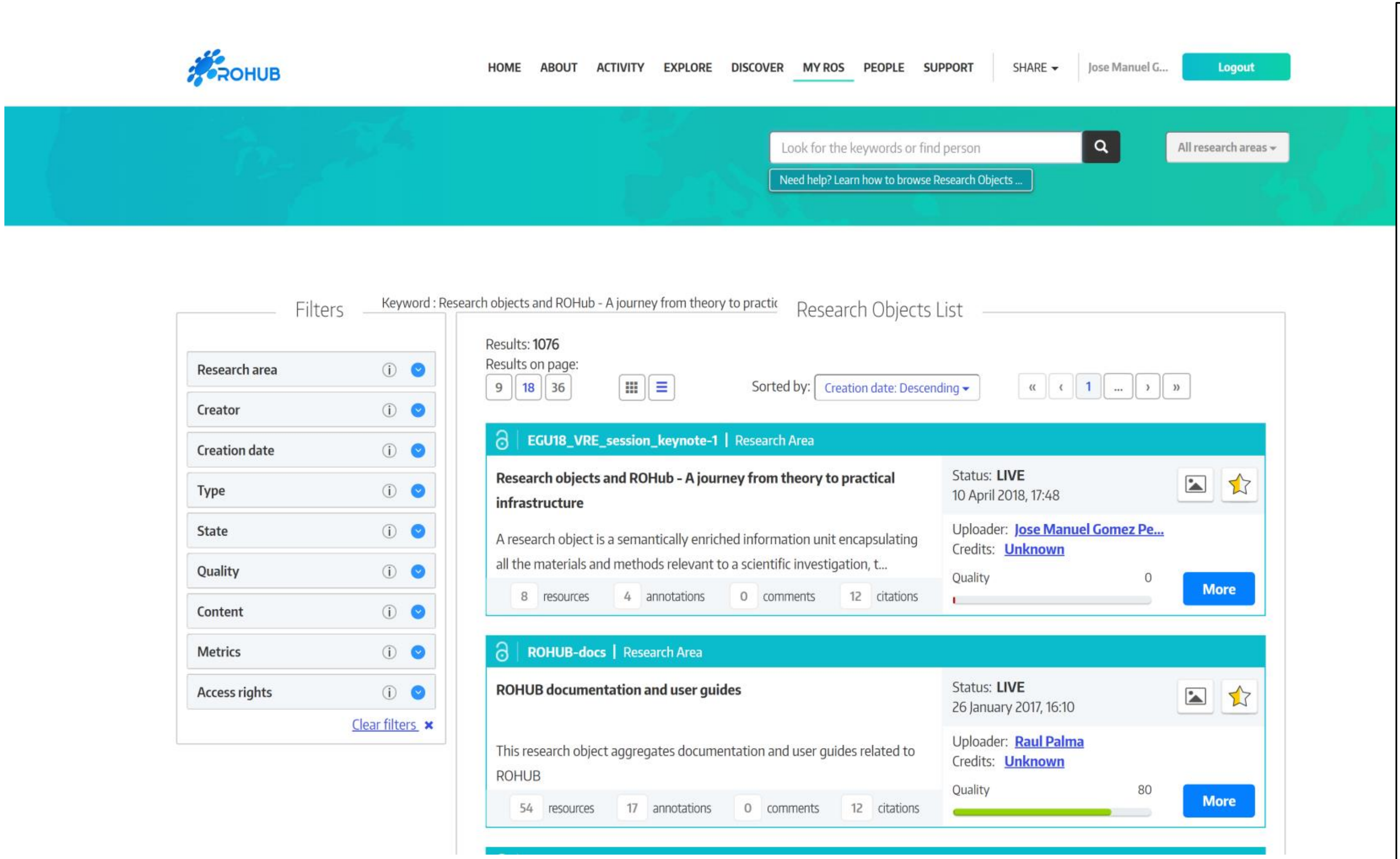
Community

Content

- Artificial Intelligence, NLP, Computer Vision
- Cross-modal content annotation (text, images, diagrams, tables, provenance…)
- **Towards a Digital Aristotle that can reason and answer science questions**

This project is co-funded by the European Union

The research object of this talk



Keyword : Research objects and ROHub - A journey from theory to practic

Results: 1076

Results on page:

Sorted by: Creation date: Descending

EGU18_VRE_session_keynote-1 | Research Area

Research objects and ROHub - A journey from theory to practical infrastructure

A research object is a semantically enriched information unit encapsulating all the materials and methods relevant to a scientific investigation, t...

8 resources 4 annotations 0 comments 12 citations

Status: LIVE
10 April 2018, 17:48

Uploader: Jose Manuel Gomez Pe...
Credits: Unknown

Quality 0

ROHUB-docs | Research Area

ROHUB documentation and user guides

This research object aggregates documentation and user guides related to ROHUB

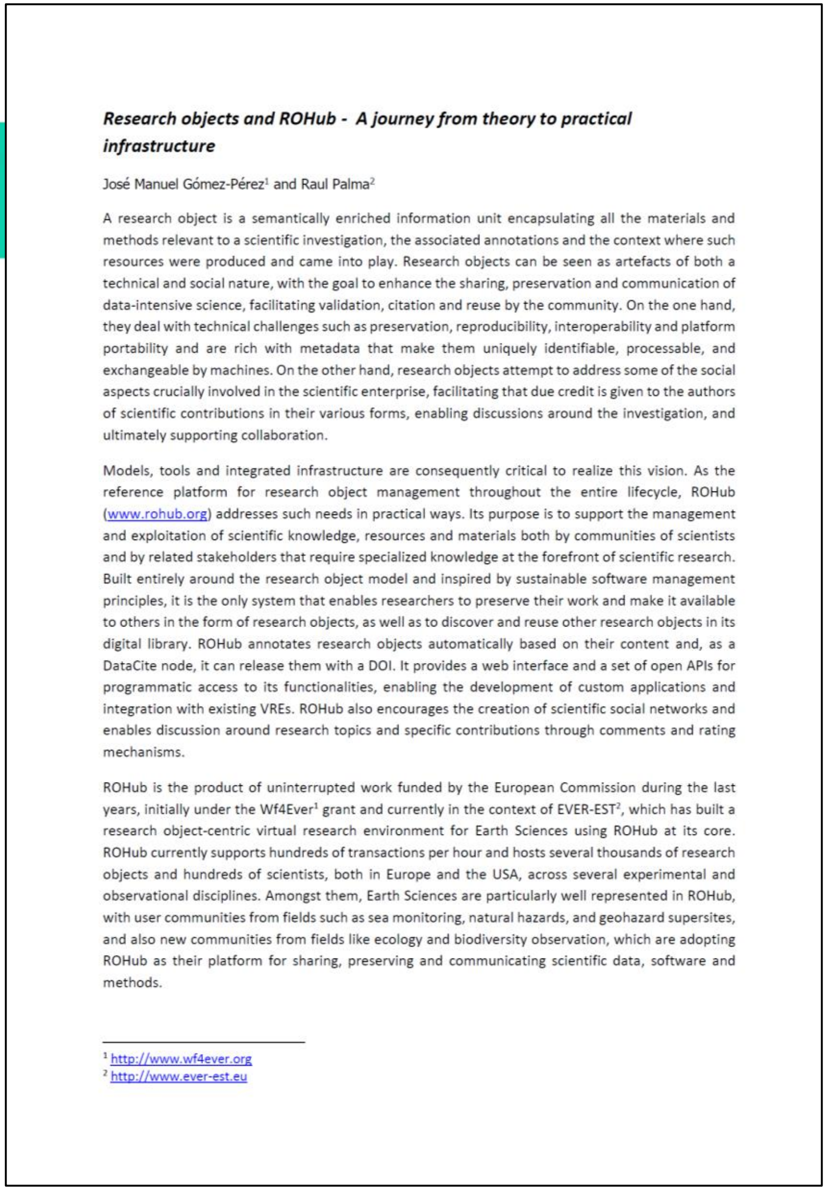
54 resources 17 annotations 0 comments 12 citations

Status: LIVE
26 January 2017, 16:10

Uploader: Raul Palma
Credits: Unknown

Quality 80

http://www.rohub.org/rodetails/EGU18_VRE_session_keynote-1/



Research objects and ROHub - A journey from theory to practical infrastructure

José Manuel Gómez-Pérez[1] and Raul Palma[2]

A research object is a semantically enriched information unit encapsulating all the materials and methods relevant to a scientific investigation, the associated annotations and the context where such resources were produced and came into play. Research objects can be seen as artefacts of both a technical and social nature, with the goal to enhance the sharing, preservation and communication of data-intensive science, facilitating validation, citation and reuse by the community. On the one hand, they deal with technical challenges such as preservation, reproducibility, interoperability and platform portability and are rich with metadata that make them uniquely identifiable, processable, and exchangeable by machines. On the other hand, research objects attempt to address some of the social aspects crucially involved in the scientific enterprise, facilitating that due credit is given to the authors of scientific contributions in their various forms, enabling discussions around the investigation, and ultimately supporting collaboration.

Models, tools and integrated infrastructure are consequently critical to realize this vision. As the reference platform for research object management throughout the entire lifecycle, ROHub (www.rohub.org) addresses such needs in practical ways. Its purpose is to support the management and exploitation of scientific knowledge, resources and materials both by communities of scientists and by related stakeholders that require specialized knowledge at the forefront of scientific research. Built entirely around the research object model and inspired by sustainable software management principles, it is the only system that enables researchers to preserve their work and make it available to others in the form of research objects, as well as to discover and reuse other research objects in its digital library. ROHub annotates research objects automatically based on their content and, as a DataCite node, it can release them with a DOI. It provides a web interface and a set of open APIs for programmatic access to its functionalities, enabling the development of custom applications and integration with existing VREs. ROHub also encourages the creation of scientific social networks and enables discussion around research topics and specific contributions through comments and rating mechanisms.

ROHub is the product of uninterrupted work funded by the European Commission during the last years, initially under the Wf4Ever[1] grant and currently in the context of EVER-EST[2], which has built a research object-centric virtual research environment for Earth Sciences using ROHub at its core. ROHub currently supports hundreds of transactions per hour and hosts several thousands of research objects and hundreds of scientists, both in Europe and the USA, across several experimental and observational disciplines. Amongst them, Earth Sciences are particularly well represented in ROHub, with user communities from fields such as sea monitoring, natural hazards, and geohazard supersites, and also new communities from fields like ecology and biodiversity observation, which are adopting ROHub as their platform for sharing, preserving and communicating scientific data, software and methods.

[1] http://www.wf4ever.org
[2] http://www.ever-est.eu

Jose Manuel Gomez-Perez
R&D Director
jmgomez@expertsystem.com

EXPERT SYSTEM
SEMANTIC INTELLIGENCE

We gratefully acknowledge the EU Horizon 2020 program for research and innovation under grant EVER-EST-674907

everest

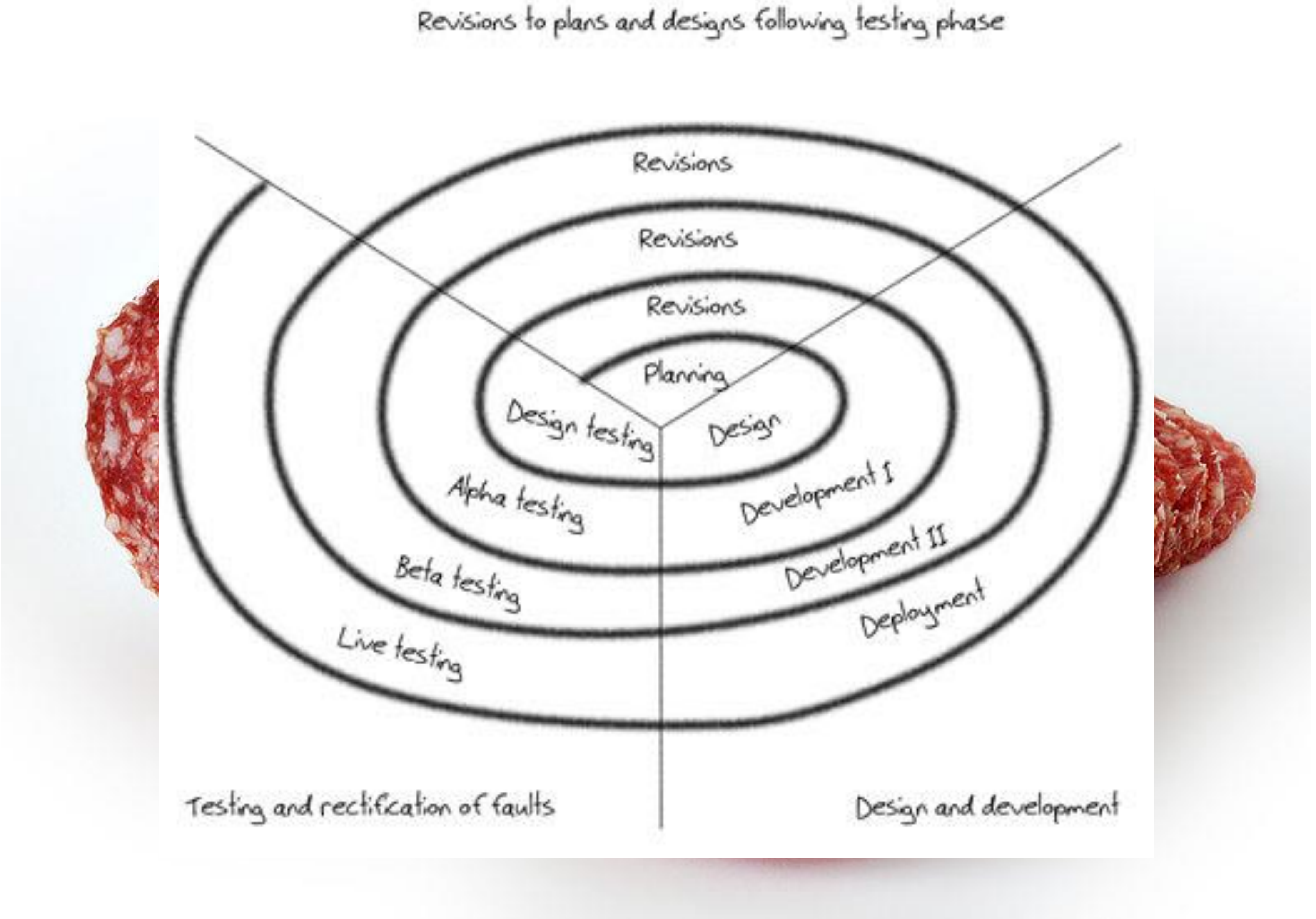
linkedin.com/company/expert-system

twitter.com/Expert_System

info@expertsystem.com

This project is co-funded by the European Union

Don't publish. Release… often!



This project is co-funded by the European Union

Some key facts

- Number of Research Objects: 2,500+ (starting point 1,200)
  - Golden Exemplars: 2 or 3 per VRC
  - Generated automatically: 511
- Number of users: 162
- Activity events: 275K+
  - Last week > 47K (approx. 6K per day)

- ROHUB storage size
  - solr:
    - named_objects: 868 / 912KB
    - notifications: 275620 / 193MB
    - ros: 2648 / 15MB
    - ros-private: 146 / 1.1MB
  - File system: 2.4GB
  - Tripple store: 1.1 GB
  - Database: 300MB

- ROHub access (previous week):

	all	success	ROs	resources
sum	469081	466506	101935	367146
avg	58635	58313	12741	45893

*As of September 2017