

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

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



## The Research and Information Lifecycle in Earth Science

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

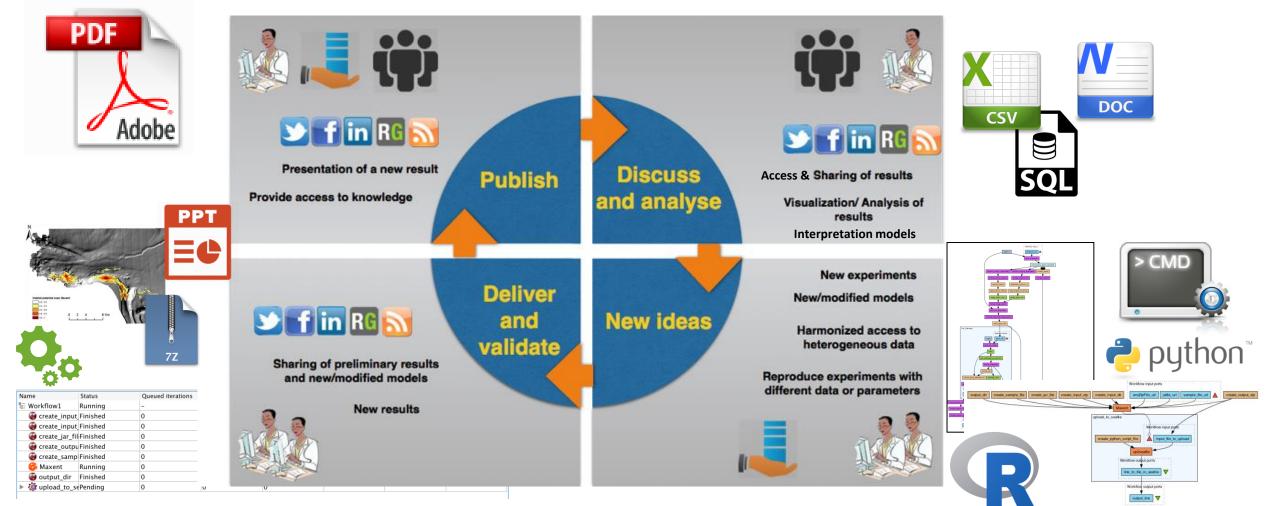




## The Research and Information Lifecycle in Earth Science



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





# Research in data and process intensive disciplines



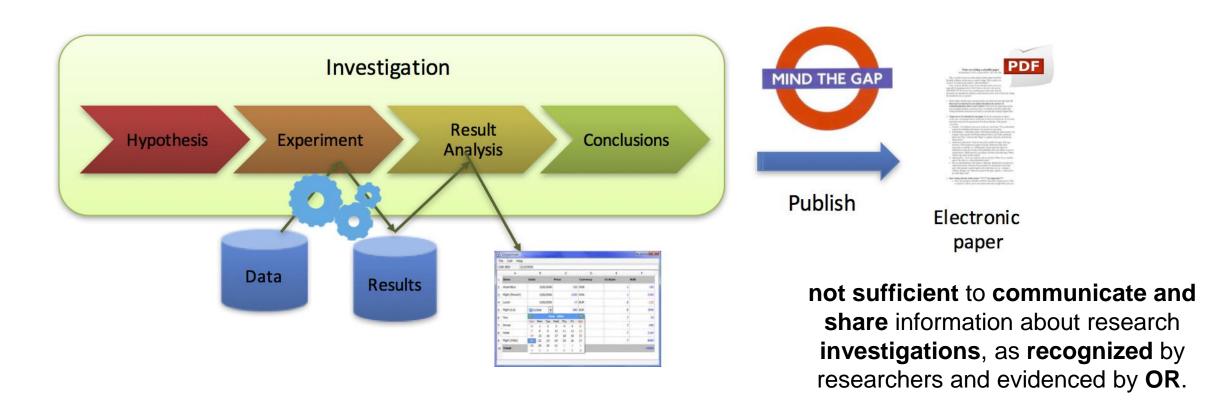
- How to provide a single-entry point to all the related resources ?
- How to capture the lifecycle of scientific investigations ?
- How to validate and disseminate scientific knowledge ?
- How to preserve the scientific knowledge for their reuse?







## Electronic paper is not enough!



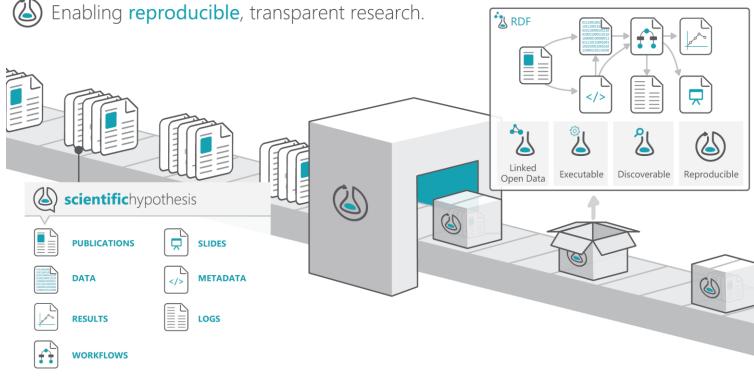
**Open Research** movement: Openly share the data of your experiments



## Research objects



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



http://www.researchobject.org



ROHUB





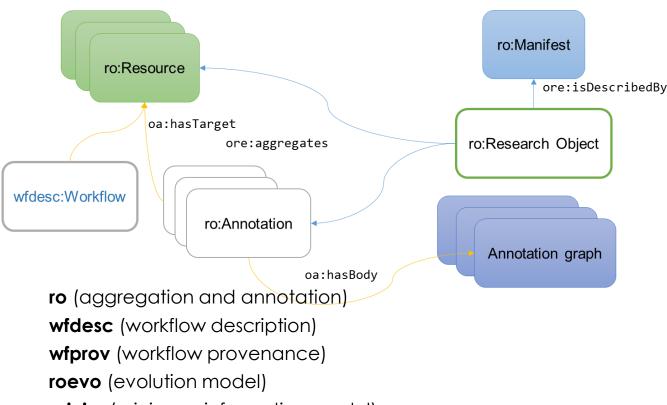


## Why research objects

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

## Representing research objects: The RO Model





minim (minimum information model)

This project is co-funded by the European Union

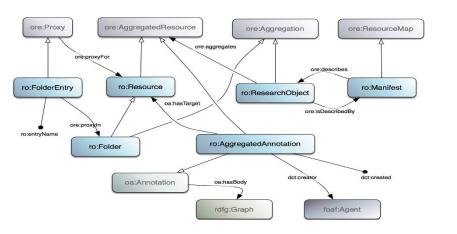
> **RO primer:** <u>http://wf4ever.github.com/ro-primer</u> **RO specification:** <u>http://wf4ever.github.com/ro</u>

- 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

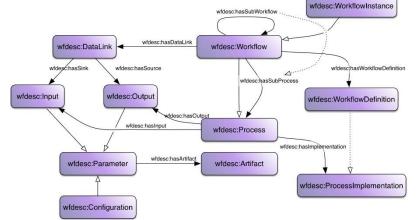
W3C Semantic Web

## This project is co-funded by The RO ontology stack



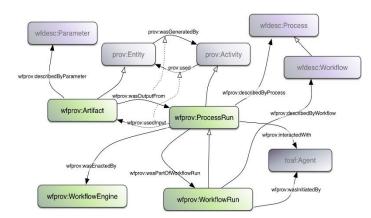


### Research object (ro)

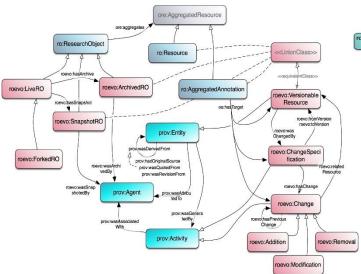


Workflow description (wfdesc)

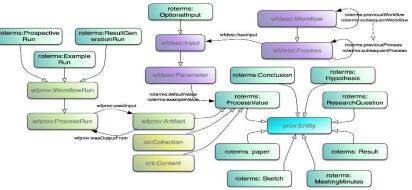
### Workflow provenance (wfprov)



### RO evolution (roevo)



### RO terms (roterms)



### RO for Earth Science (roes)

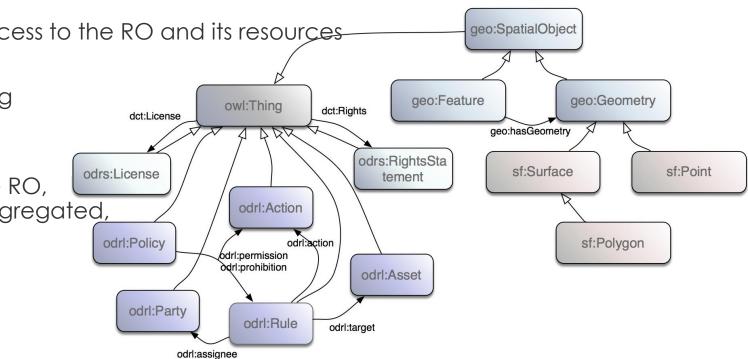


Emphasis on **Geo**, access rights and IP policies <u>https://github.com/wf4ever/ro/tree/earth-science</u>





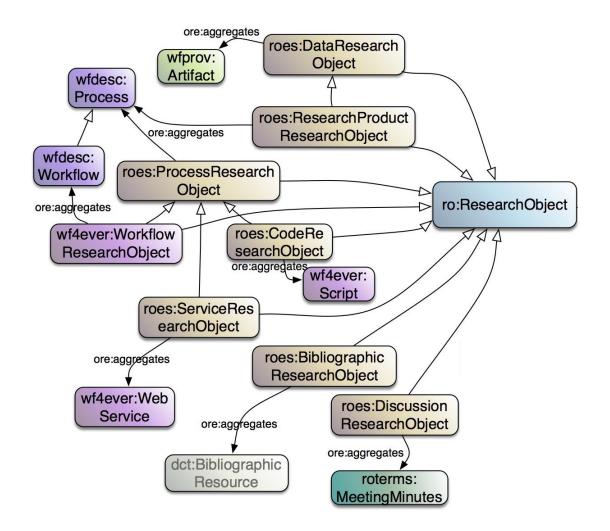
- RO model should cover metadata regarding
  - Geospatial information
    - Coordinates of the region relevant for the RO and the observation it represents
  - Time-period coverage
    - Indication of the time span covered in the observation
  - Data access policies
    - Detailed information for controlling 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.







## Earth Science extensions (2/3)



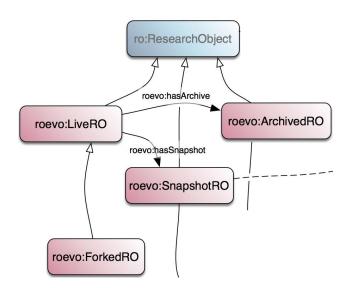
- We also defined different types of research objects according to the purpose and hence the kind of the resources it would aggregate:
  - Workflow-centric
  - Data-centric
  - Research Product centric
  - Process-centric
  - Code-centric
  - Service-centric
  - Bibliography-centric
  - Discussion-centric

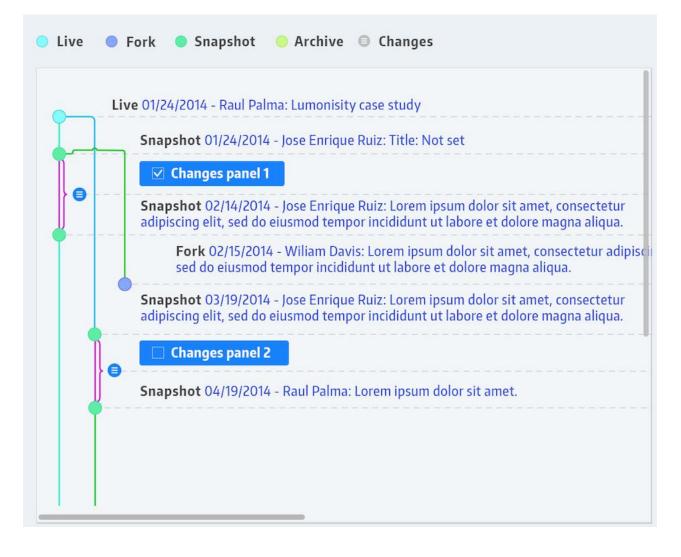




## Earth Science extensions (3/3)

- 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

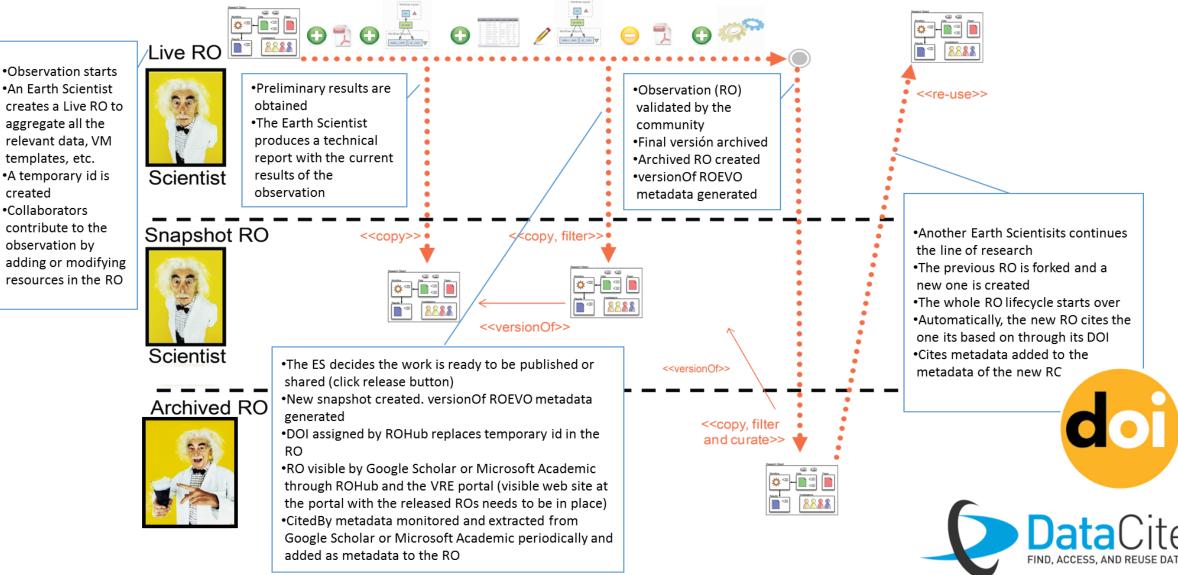






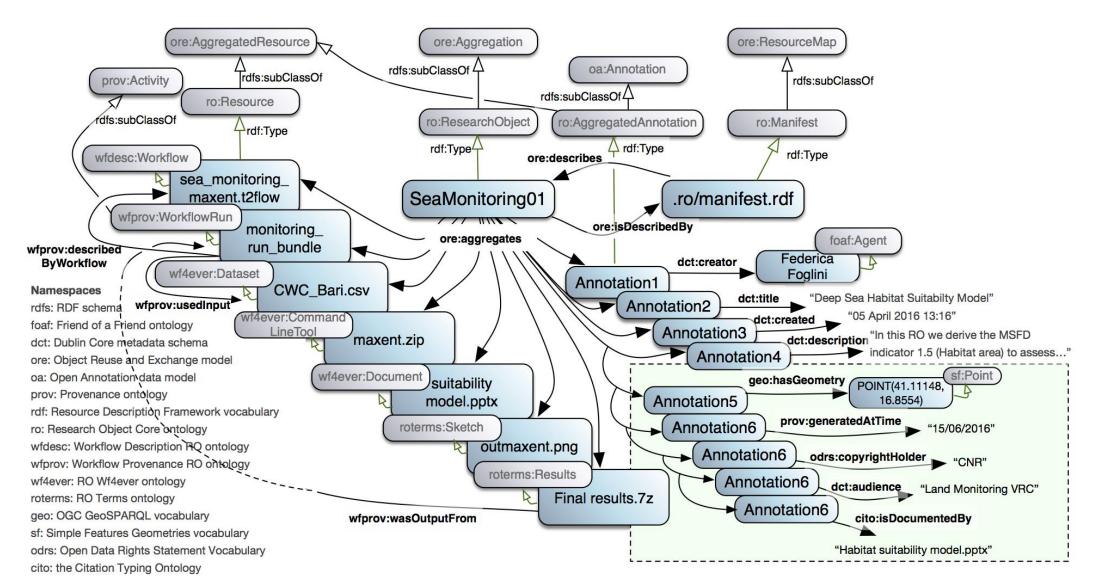


## Exemplary RO lifecycle





# Exemplary research object (sea monitoring use case)







# And what is the existing RO supporting technology ?







## Research object management platform

Controller

Ontology

Store AP

MysqL

### ROHUB comprises both

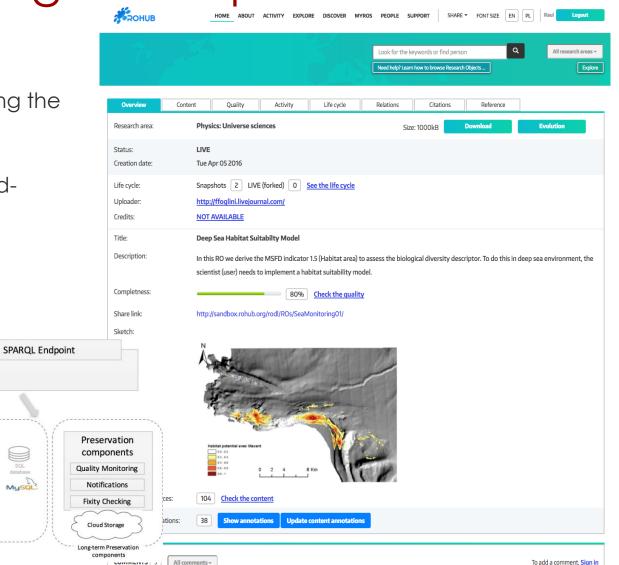
a backend service (RODL), implementing and exposing the set of APIs

File System

- a reference web client application (ROHUB portal), exposing the research object functionalities to the endusers.
- Combination of digital libraries, long term-preservation and semantic technologies RESTAPI RO API, ROEVO API,
- 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.







## 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)





## 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



the European Union



#### A local file Resource type A web resource Resource type Workflow Clear Workflow Process Workbench Dataset Documen Image File to upload Choose file test-bundle.bundle.zip Is this an RO bundle? ROHUB SHARE - FONT SIZE EN PL The Taverna Suite of Tools Web Portals / Gateways Client User Interfaces Workflow Repository User Interfaces Q ook for the keywords or find person ed help? Learn how to browse Research Objects ... Third Party Tools Workbench Taverna Lite SEAMONITORING-CNR-TESTING-SEA\_MONITORING\_MAXENT\_V6.BUNDLE Galaxy Workflow Engine Submit Changes 02 Service Catalogues Google refine Workflow Provenance Workflow Server Relations Citations Reference Overview Ouality Activity Life cycle Player home Resource information BiodiversityCatalogue W3C Virtual Activity and Service Machine Workflow inputs intermediates Plug-in Manager nteraction Components Server Command **UNIC**®RE Line Nome : worldlow wfbundlo Secure Service Access SEAMONITORING-CNR-TESTING-SEA\_MONITORING\_MAXENT\_V6.BUND ReanShel lo: Prog OAuth1 & 2, username/password, OSGi APIs Submit Changes certificates. Overview Quality Activity Life cycle Relations Citations Reference Conten home > outputs > Q Resource information output\_link.txt Server Name: output\_link.txt output\_link.txt ~ https://box.everest.psnc.pl/seafhttp/files/ 303de0b7-44b3-4275-87dd-3c7dc2cf4824/maxentoutput7129193146824760280.zip



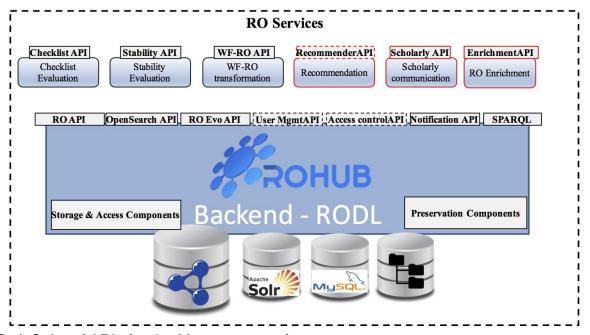


### A full suite of RO APIs provided by ROHUB and other RO services

- 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.



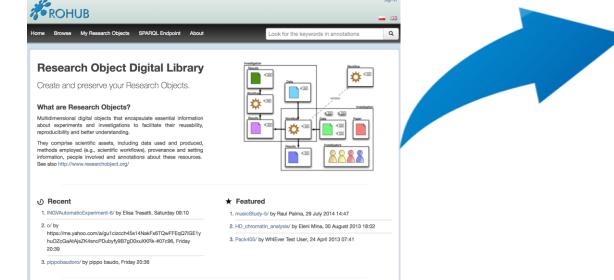


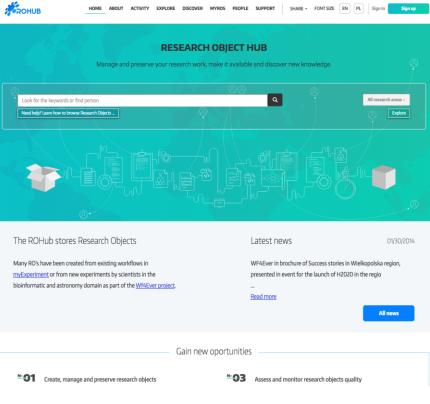
## The RO user interface – ROHUB portal

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

Find other Research Objects or Sign in to create your own

OHUB











## But there are other user interfaces...

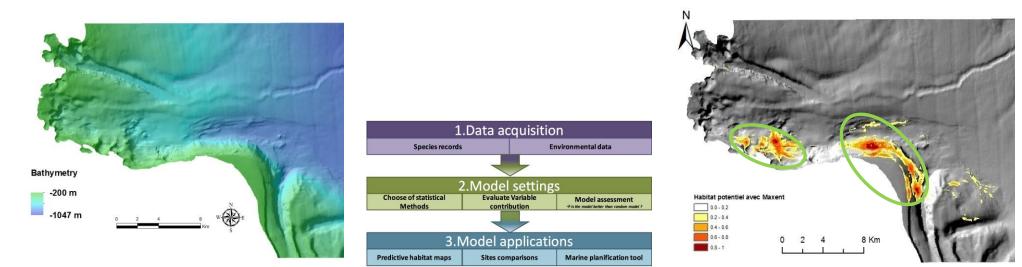
RODL can have other client applications, e.g., VRC portals, collaboration spheres, command-line (ro-manager), myExperiment (alpha),...

← → C ☆ O Not Secure ever-est.services.meeo.it/seamonitoring/		२ 🛧 :				
everest		2	← → C ☆ O Not Secure alpha.myexperiment.org			☆ :
Virtual Globe Advanced Search   Saarch for.		arch x dentifier	my exper		makes it easy to find, use and share scientific s and other Research Objects, and to build communities.	
		Status Live -			All   Search	
P.A.S.		Date To Description Search Experiments	First time visitor? Try Project Introduction Bioinformatics Case Use myExperiment to	a Study	plore Register \$ ./ro_sample.sh ro config -b /usr/workspace/wf4ever-ro-manager/src/rocommand/tes -r http://calatola.man.pornan.pU/robbx/dropbox_account -p d4idscd9f9bb24e9809996cf427e -u Test user -e testuser@example.org	
A second second	or everest.expertsystemilab.com/spheres/index.html     ch Objects     scientists     search a scientist/specialis     search a scientist/specialis     arch Objects     collaboration     Related		te i Share Your Workflows C Share Your Workflows C Create and Find Pao B Find People and Ma A Find People and Ma A Create and Join Gro	ws and Files  ks of Items ke Friends upps About my	ro configuration written to /Users/graham mkdir: rocommand/test/robase/test-create-R0: File exists cp: /ro-test-1/x: No such file or directory ro create "Test Create R0" -d "rocommand/test/robase/test-create ro status -d "rocommand/test/robase/test-create-R0" Research Doject status identifier: R0-id-testCreate, title: Test Create R0 creator: Test user, created: 2011-09-15153:12:17	
Details	Arth Cojects Al Scientists	Bonnell ( Control of the second of the seco	S Tag and Rate things Write Reviews and	ryExperime For D Comments Give us	<pre>path: /usr/workspace/wf4ever-ro-manager/src/rocommand/t uri: file://usr/workspace/wf4ever-ro-manager/src/roco description: Test Create R0 ro list -d "rocommand/test/robase/test-create-R0" test-create-R0,ro/manifest.rdf test-create-R0/RLRMEo-test-1 test-create-R0/subdirl/subdirl-file.txt test-create-R0/subdirl/subdirl-file.txt</pre>	
	an research on freeCapability		myExpe	iment has over 5000 members, 250 groups	<pre>ro annotate rocommand/test/robase/test-create-R0/subdir1/subdir1 ro annotations -d "rocommand/test/robase/test-create-R0/subdir1" file:///usr/workspace/wf4ever-ro-manager/src/rocommand/test/roba suble compared to the test of test of</pre>	' rocommand/test/robase/test-create-R0/subd
Man tapies spe Anad familia Mai tapies spe Mai tapie spe	Allen species in the Mediterranean Sea by 2012. A contribution to the application of European Union's Marine Statesty Printerwork Director (MSDP), Part 2019 (1999)	And	About		<pre>title: subdirl-file.txt title ro annotate rocommand/test/robase/test-create-R0/subdir2/subdir2 ro annotate rocommand/test/robase/test-create-R0/subdir2/subdir2 rocommand/test/robase/test-create-R0/subdir2/subdir2 rocommand/test/robase/test-create-R0/subdir2 rocommand/test/robase/test-create-R0/subdir2 rocommand/test/robase/test-roco</pre>	2-file.txt keywords "subdir2-file.txt foo,bu 2-file.txt description "subdir2-file.txt des 2-file.txt format "subdir2-file.txt format" 2-file.txt file "subdir2-file.txt title" 2-file.txt created "20110914T12:00:00" ' rocommand/test/robase/test-create-R0/subdi





- 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







### 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 1 year 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



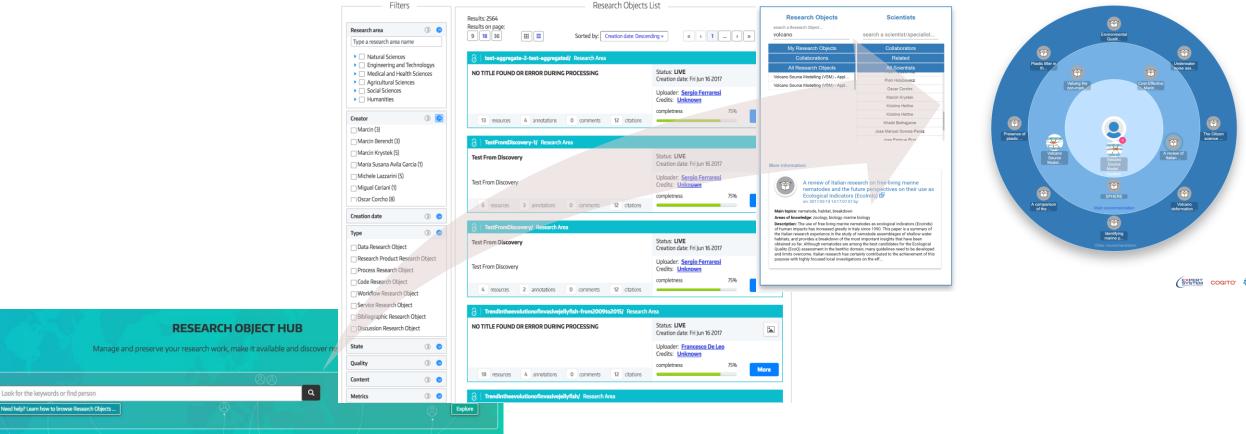
ROHUB enables scientists to encapsulate the data, provenance of workflows executions, results, documentation and other resources related to the particular study through a single information unit

SEAMONITORING01

Create an RO from	m ZIP		×		O rem inve	Content	Quality	A ===== === =	Life surls	Relations	Citations	Reference	
					Overview	Content	Quality	Activity	Life cycle	Relations	Citations	Reference	
Where is the ZIP archive? 💿	A file on my disk				Actions								
$\bigcirc$ $^{\prime}$	A web resource				home > do	ocuments							
Create RO						Juments							
					Name			Created	I	Details	Creato	r	Action
RO identifier	RO identifier				152278-f	doc		1	(	04/14/2016	http://	ffoglini.livejournal.com/	
Title	an optional title												
Description					GeoHAB	2016.docx		1	(	04/14/2016	http://	ffoglini.livejournal.com/	
				3									
RO Type	Basic Research Object -				Habitat	suitability models for I	BARI canyon_2.pp	tx 2	(	04/05/2016	http://	ffoglini.livejournal.com/	
Research area	Natural sciences -				How_to	use desv		2	(	04/05/2016	http://	ffoglini.livejournal.com/	
Template	A workflow-centric RO -				How_to	_use.docx		-		5470572010	Trep.//	nogiminvejournui.com	
					The-cold	l-water-coral-Lophelia	a-pertusa-Scleracti	1	(	04/05/2016	http://	ffoglini.livejournal.com/	
			Cancel										
	Li	pload a resou	Irco	×	"152278-f.doc" d	letails							
	0												
		Resource type			Description		Not out						
			<ul> <li>A web resource</li> </ul>		Description: Type:		Not set Not set						
		Resource type	Workflow		Size:		36352 KB						
			Workflow Process Dataset		Created on:		04/14/2016						
			Document		Uploader:		http://ffoglini.liv	ejournal.com/					
			Image		URI:			ohub.org/rodl/ROs/S	aMonitoring01/1522	78-f.doc			
		File to upload	Choose file No file chosen		Sketch:		No image availal	-					Hide detai
					Number of anno	otations: 1 Show	r the annotations ~						Add adnnotation
				OK Cancel									

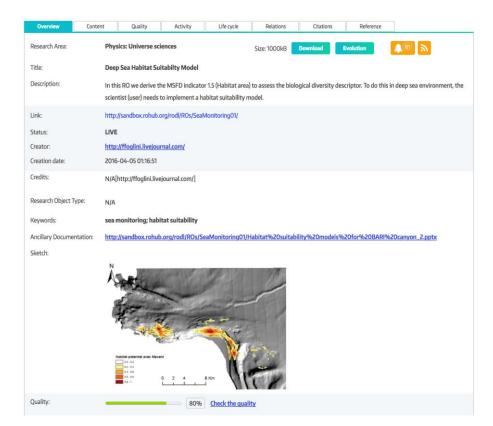


The scientist (& collaborators), then, can search for other research objects aggregating relevant resources that can be re-used, like potential input data, the processes, or other related resources, to reference them from the new research object





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



#### SEAMONITORING01

verview	Content	Quality	Activity	Life cycle	Relations	Citations	Reference	
pletness: 🦲			94	🕑 <u>See qu</u>	ality history with	RO Monitoring Too	<u>ol</u>	
				Quality ch	ecked list —			
Chose the list:	Earth Scie	ence workflow-centri	ic RO key requiremen	nts 🔻	Recomended I	ist: <b>Research Obje</b>	ct basic requirem	ents
✓ Research Ob	ject has title							
✓ Research Ob	ject has descr	iption						
✓ Research Ob	ject has create	or						
✓ Research Ob	ject has copyr	ight holder						
<b>✓</b> Research Ob	ject has keyw	ords						
✓ Research Ob	ject has sketc	h						
<mark>✓</mark> Research Ob	ject has subje	ct						
✓ Research Ob	ject has ancilla	ary documentatio	n					
✓ Workflow re	source is pres	ent						
All workflow kar).	definitions re	ferenced by work	flow reources are	accessible. The w	rf definition is the	resource understo	ood by a workfow	engine (e.g, t2flow, ga



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

SEAMONITORING01

Понив	HOME ABOUT ACTIVITY EXPLORE DISCOVER MYROS PEOPLE SUPPORT SHARE - FONT SIZE EN PL Signin Signup
Evolutio	
snapshot	t snapshot with DOI release release with DOI fork
	Close
EAMONITOR	ING01
Overview Cont	tent Quality Activity Life cycle Relations Citations Reference
Research Area:	Physics: Universe sciences Size: 1000kB Download Evolution
Title:	Deep Sea Habitat Suitabilty Model
Description:	In this RO 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.
Link:	http://sandbox.rohub.org/rodl/ROs/SeaMonitoring01/
Status:	LIVE
Creator:	http://ffoglini.livejournal.com/
Creation date:	2016-04-05 01:16:51
Credits:	N/A[http://ffoglini.livejournal.com/]
Research Object Type:	N/A
Keywords:	sea monitoring; habitat suitability
Ancillary Documentation:	$eq:http://sandbox.rohub.org/rodl/ROs/SeaMonitoring01/Habitat%20suitability%20models%20for%20BARI%20canyon_2.pptx and a set of the sandbox of the sa$
Sketch:	
	N A THE REPORT OF A THE REPORT

#### Quality Activity Life cycle Relations Citations Reference Overview Content ★ Back to life cycle Changes panel 1 -Select changes panel Life cycle chart Changes panel Changes panel 1 Live Fork Snapshot Archive Changes All changes Additions Modyfications Removals http://sandbox.rohub.org/rodl/ROs/SeaMonitoring01/ Date Type of change Title Uploader Raul Palma 2017-05-18 04:58:51 create An annotation for re.. 2017-05-18 04:58:47 create A resource b361f7a2-. Raul Palma 6 Raul Palma 2017-05-18 04:57:35 An annotation for re.. create 2017-05-18 04:57:30 create A resource 2280ac0b-Raul Palma Raul Palma 2017-05-18 04:57:08 create An annotation for re.. 2017-05-18 04:57:05 create A resource 4f4a6363-. Raul Palma 2017-05-18 04:51:02 An annotation for re.. Raul Palma create • 2017-05-18 04:50:59 create A resource 960f5b92-Raul Palma Raul Palma 2017-05-18 04:49:29 create An annotation for re. 2017-05-18 04:49:25 create A resource 0081ee44-. Raul Palma ( 1 2 3 4 5 ... )



- 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
  SEAMONITORINGO1

RO Monitoring										
Compare two evaluation	ons 🗸									
	Stability(0-1)=0.6	l i87 and Cor	Decay Monitorin npleteness(0-1)= paper	<b>g</b> 1.0 -> Reliab	ility(0-1)=0.687	e 1				
Zoom 1m 3m 6m Y	TD 1y All				From Jan 1, 2013	To Feb 23, 2013				
0.75		,	-1		••••					
0.5					$\Lambda \Lambda T$					
0.25	V				VVV	V				
7			4		¥ ¥ ¥					
2013 01-07	2 01 3 01 - 14	2 01 8 01 - 21	2018 01-28	2 01 8 02 - 04	2018 02-11	2018 02-18				
	14. Jan		28. jan		11. feb					

		Overview	Content	Quality	Activity	Life cycle	Relations	Citations	Reference	
E A M O N I	TORINGO1 Content Quality	Activity	Life cycle	Relations Cit	tations Refere	ence		Download	Evolution	
Publication da From Month	Filters	Results: 121 Results on page: 9 18 36		Filters applie		« ( 1) ) )	violo	gical diversity des	criptor. To do this in	deep sea environment, the
To		Date	Object type	Title		→ C ☆ ③ sandbox.rohub	b.org/rodl/atomnotifications/?r	o=http%3A%2F%2Fsandbox.	rohub.org%2Frodl%2FROs%2FSe	aMonitoring-cnr-testing%2F
Month	Day- Year Ok	2017-05-18 04:58:51		An annotation for resou created by Raul Palma	rce Workflow_WPS					
	<u>Clear filters</u> ×	2017-05-18 04:58:47		A resource b361f7a2-776 been added to the Resea				This	is a feed ever something new appears	Monitoring-enr-testing%2F
		2017-05-18 04:57:35		An annotation for resou has been created b	rce Habitat Suitabili			1	ubscribe	o> esting/" title="Description for" /> nr-testing has been
		2017-05-18 04:57:30		A resource 2280ac0b-d been added to the Resea				About this feed Title Notifications for		
		2017-05-18 04:57:08		An annotation for resou been created by Raul	rce Workflow_slope			Link dbox.rohub.org/rodl/atomnotific		
		2017-05-18 04:57:05		A resource 4f4a6363-8e been added to the Resea			y/dc/terms/source cm-testin cm-testin %1:49z html> <body><p< td=""><td>g%2F Latest posts</td><td>%2FrodI%2FROs%2FSeaMonitoring</td><td>* esting/" title="Description for" /&gt; slt;p&gt;The Research Object URI is <a< td=""></a<></td></p<></body>	g%2F Latest posts	%2FrodI%2FROs%2FSeaMonitoring	* esting/" title="Description for" /> slt;p>The Research Object URI is <a< td=""></a<>
		2017-05-18 04:51:02		An annotation for resource Habitat Suitab has been created b				Object SeaMonitoring-cnr-te	an added to the Research Object	
		2017-05-18 04:50:59		A resource 960f5b92-ec been added to the Resea					has been added to the Research	_esting/" title="Description for" /> ing-sea_monitoring_maxent_v6.bundle has bee
		2017-05-18 04:49:29	annotation	An annotation for resou created by Raul Palma	sea			been created by Raul Palma A resource sea_monitoring_r	aa_monitoring_maxent_v6.t2flow has naxent_v6.t2flow has been added to hitoring-on-testing by Raul Palma	ng-car-testing- robub.org/rodl/Nos/SeaMonitoring-car- oring-on-testing/.ro/annotations/c7413b08- rodl/ROs/SeaMonitoring-car-testing-

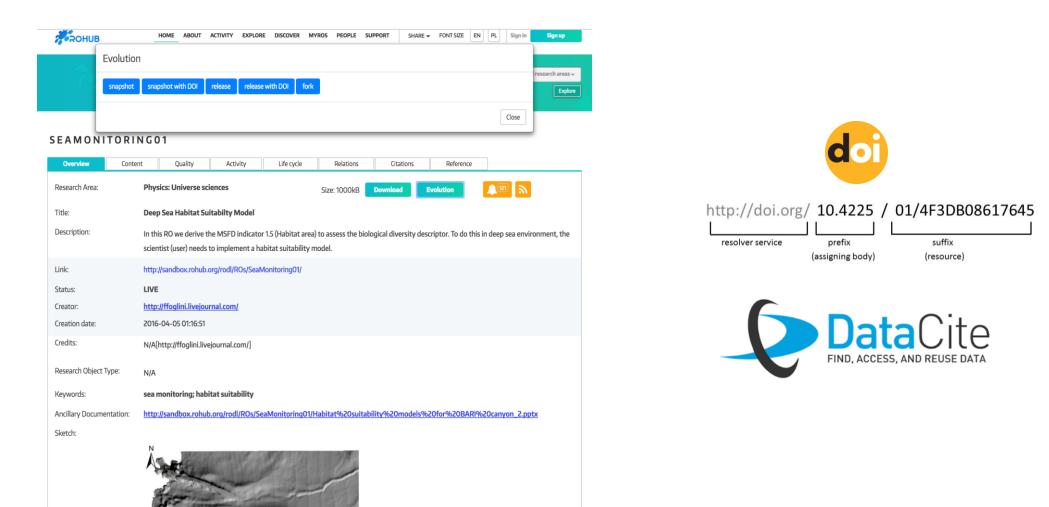


Scientists can collaborate with other colleagues, setting different access modes and permissions to the RO

Overview	Content	Quality	Activity	Life cycle	Relations	Citations		Re	eference	Access Contro	I	
Access Mode : Editors : Readers :	ри	blic -										
			Grand Privileg	je	EDITOR	\$	•	×				



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









Dr. Raul Palma

rpalma@man.poznan.pl