everest

Workflows and Research Objects models in Earth Science

Workpackage	**WP 4**	**Research Objects in Earth Science**	
Task	4.2	Research Objects Models and Support Technology in Earth Sciences	
Author (s)	Jose Manuel Gomez-Perez		ESI
	Raul Palma		PSNC
	Nuria García		ESI
Reviewer (s)	Simone Mantovani		MEEO
Approver (s)	Jose Manuel Gomez-Perez		ESI
	Cristiano Silvagni		ESA
Authorizer	Mirko Albani		ESA
Document Identifier	EVER-EST DEL WP4-D4.2		
Dissemination Level	Public		
Status	Approved by the EC		
Version	1.1		
Date of Issue	14 December 2018		

Abstract

Deliverable D4.1. produced an analysis of the main concepts related to research objects under the light of the initial requirements elicited from the EVER-EST VRCs. This analysis enabled the understanding of the necessary adaptations, customizations and extensions of research object methods, tools and models to a new domain such as Earth Science disciplines and showed the way for future work in the project. This document focuses on the latter of such areas, workflow and research object models, and introduces the necessary extensions and customizations that have been implemented in the research object models and vocabularies for Earth Science. In addition to the final ontologies this document provides an overview of the methodology followed, including further analyses and the joint work carried out with the EVER-EST user communities in WP3 and technical partners in WP5 and WP6. Throughout this work, it has been sought to leverage the familiarity of the Earth Science community with existing resources, maximizing the reuse of existing terminology and metadata used in the field and ensuring the compliance of the newly introduced terms with existing standards and potential future extensions. In doing so, it is sought to facilitate adoption, reducing the learning curve. This work also pushes the boundaries of the generic models available for research object representation and produces significant contributions and extensions to their specification. Furthermore, a new branch has been created specifically for the Earth Science disciplines, which provides specialized modeling support for research objects in such scientific areas.

Document Log

Date	Author	Changes	Version	Status
28/03/2016	Raul Palma	Initial draft metadata guidelines	v0.1	Draft
15/04/2016	Nuria García	Analysis of available Earth Science metadata and standards	v0.1	Draft
19/05/2016	Nuria García	Internal metadata specification document released within consortium	v0.2	Draft
20-May-2016	Jose Manuel Gomez-Perez	Content reviewed, new comments included, further discussion	v0.3	Draft
25/05/2016	Raul Palma Jose Manuel Gomez-Perez	Addition of new terms. Initial implementation of ontology with updated metadata	v0.4	Draft
24/06/2016	Jose Manuel Gomez-Perez	Interim review (M9) feedback collected and incorporated	v0.5	Draft
07/07/2016	Raul Palma Jose Manuel Gomez-Perez	2nd RO Hackathon complete. New requirements and metadata gaps identified	v0.5	Draft
28/07/2016	Raul Palma	Ontologies updated and merged back into the research object model repository. Metadata guidelines updated	v0.5	Draft
29/07/2016	Jose Manuel Gomez-Perez	First complete version of the document generated and released for preliminary review and next iteration	v0.6	Draft
10/08/2016	Raul Palma Jose Manuel Gomez-Perez	Review comments incorporated and additional extensions	v0.7	Draft
11/08/2016	Jose Manuel Gomez-Perez	Additional review comments incorporated. Document finalized and delivered for review by technical coordinator	v0.8	Draft

16/08/2016	Jose Manuel Gomez-Perez	Review comments from technical coordinator incorporated	v0.9	Draft
18/08/2016	Jose Manuel Gomez-Perez	Final version released	v1.0	Draft to be approved by the EC
14/12/2018	Cristiano Silvagni	Document status	1.1	Approved by the EC

Table of Contents

List of Figures

List of Tables

Definitions and Acronyms

Acronyms	Description
CNR-ISMAR	Consiglio Nazionale delle Ricerche – Istituto di Scienze Marine
DCAT	Data Catalog Vocabulary
DOI	Digital Object Information
EO	Earth Observation
ESA	European Space Agency
EVER-EST	European Virtual Environment for Research - Earth Science Themes
FOAF	Friend Of A Friend
GERO	Golden Exemplar Research Object
JSON	JavaScript Object Notation
GUI	Graphical User Interface
OA	Open Annotation ontology
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
OWL	Ontology Web Language
PROV	Provenance Data Model
RDF	Resource Description Framework
RDFS	RDF Schema
RO	Research Objects
RODL	Research Objects Digital Library
SATCEN	European Union Satellite Center
SIP	Submission Information Package
SKOS	Simple Knowledge Organization System
SMART	Specific, Measurable, Achievable, Relevant, Time-bound
URI	Uniform Resource Identifier
VRC	Virtual Research Community
W3C	World Wide Web Consortium
WP3	Work Package 3 - VRE Use Cases
WP5	Work Package 5 - VRE Infrastructure, Service Design & Development
WP6	Work Package 6 - VRE Deployment, maintenance and operations

Reference Documents

Document ID	Document Title
EVER-EST DEL WP4-D4.1	Workflows and research objects in Earth Science - Main concepts and definitions
ESA Metadata Draft_1.0	ESA EO Assets Metadata Definition_DRAFT_1.0.doc
ESA SEASAT_v1.3	EO SIP-Scene Specification for SEASAT
Metadata Guidelines	Wf4Ever Metadata Implementation Guidelines
RO ontology stack	Using a suite of ontologies for preserving workflow-centric research objects
User actions	VRC User Actions

1 Introduction

1.1 Overall introduction

Deliverable D4.1-Workflows and Research Objects in Earth Science - main concepts and definitions [2] introduced the notion of Research Object (RO) to the EVER-EST Earth Science Virtual Research Communities (VRCs), providing a survey of the main concepts, technologies and models related to research objects. The analysis of such concepts combined with the initial requirements elicited from the VRCs enabled the understanding of the necessary adaptations, customizations and extensions of research object methods, tools and models to a new domain such as Earth Science disciplines and showed the way for the development of future work within the project. This document focuses on the latter of such areas, workflow and research object models, and introduce the updates that have been implemented in the research object models and vocabularies for Earth Science disciplines. In addition to the final ontologies resulting from this work, this document provides an overview of the process whereby the previous requirements were drilled down, including further analyses and joint work with users in WP3. Throughout this work, two main directives have been followed, seeking to leverage the familiarity of the Earth Science community with existing resources: i) maximize the reuse of existing terminology and metadata used in the field and ii) ensure the compliance of the newly introduced terms when existing standards do not provide suitable alternatives. In doing so, it is sought to facilitate adoption and reduce the learning curve.

1.2 Relation to other work packages and deliverables

This deliverable follows up on the work started in D4.1 and represents a link between the information modeling requirements of users in WP3 and the infrastructure being engineered on top of such models in WP5. The work on research object modeling described herein contributes to and benefits directly from work conducted in WP5 and WP6 related to the analysis of the sequences of user actions throughout different scenarios, with the objective of translating such actions into actual calls to the infrastructure APIs and components execution.

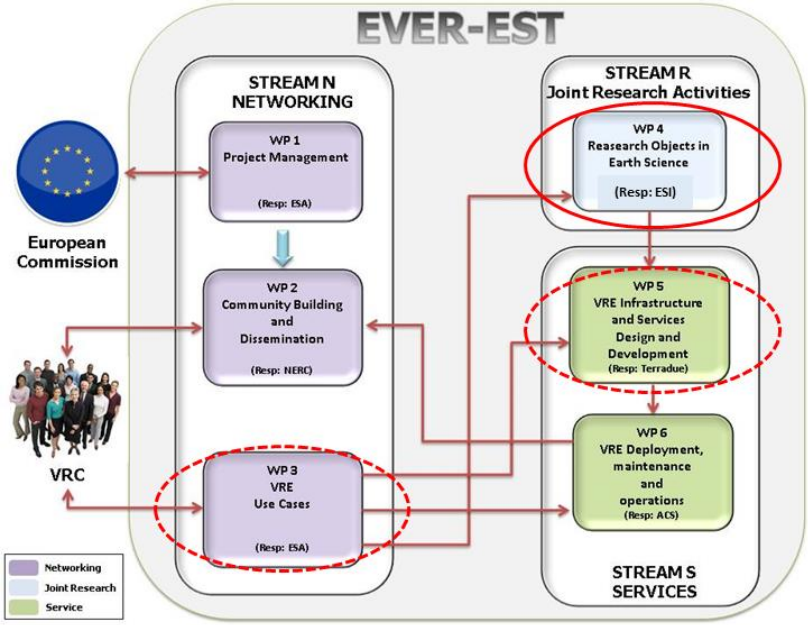

Figure 1: WP Dependencies

1.3 Compliance to the Smart Objectives and Key Performance Indicators

The content of this deliverable is in line with the EVER-EST project Objective 3, i.e. the implementation and validation of the use of research objects in Earth Science and the SMART objective that is mainly addressed is SM_OB#3.1 (Representation of research investigations and scientific experiments in the VRCs and Earth Science communities as research objects):

SM_OB#3.1	Representation of research investigations and scientific experiments in the VRCs and Earth Science communities as research objects.
Measured by	Availability of the model as a specification and as a set of extensible ontologies, implemented in OWL, uploaded into ontology repositories and in use by Earth Science communities. Number of research objects aggregating information, metadata, and computational resources including data and workflows in Earth Sciences.
Achievable	Availability of initial models for Workflow-centric research objects that will be used as the basis and will be adapted and extended to cover the Earth Science experiments needs.
Relevant	Research objects are a key enabler for the representation, sharing and preservation of scientific experiments; they aggregate all the resources associated to the experiment and the metadata necessary for their understanding and interpretation.
Timely	The first version of the research object model adapted for Earth Science observations will be delivered in M3.

This work contributes directly to the following Key Performance Indicators related to the above mentioned SMART objective.

KPI number	Objective / Result	Indicator	KPI achieved	Verification method
5	Establish a VRE for Earth Science	Facilitated sharing of experiences: number of positive feedbacks from the user community	All VRCs engaged in the generation of research objects with ROHUB at the moment	RO Hackathon and user actions
7	Implement and validate the use of research objects in Earth Science	Number of research objects implemented in Earth Science	+24 research objects already available in ROHUB after the 2nd RO hackathon	RO Hackathon Direct collaboration with the VRCs

Deliverable D4.2 releases the first extended version of the research object model and its ontology stack in the form of a new branch for Earth Science which are both publicly available through GitHub. New terms have been added to the research object vocabulary that enable the generation of Earth Science-specific metadata, with special emphasis on geographic and time information, data access policies and intellectual property. The entities corresponding to processing units in the model have been generalized to cover not only workflows but also other computational resources widely used in Earth Science disciplines. Likewise, the types of research objects available have been extended, based on the requirements from the project VRCs, now supporting not only workflow-centric

research objects but also new ones like e.g. data-centric research objects, service-centric research objects, and documentation and bibliographic research objects. It is also worth mentioning that during the analysis and extension of the research object model, one of the first tasks carried out was the update and alignment of the research object core and some of its existing extension with the latest model of the Open Annotation ontology. Furthermore, all the ontologies were cleaned and properly annotated with provenance and metadata information.

The resulting model is being used in the VRE and VRC portals under development at the time of writing this document. The research object model is the "glue" that joins together the components and enables interplay. This will enable the generation of a large number of research objects by the user communities in a way completely transparent for them. The new research object model for Earth Science is also being used by the VRCs to model a number of Golden Exemplars through the current available tools, namely the ROHUB.org platform.

1.4 Document overview

Section 2 provides an overview on the main building blocks used as a basis for the development of the research object model for Earth Science, providing a detailed insight on the starting point of the project work. Section 3 describes the methodology followed to identify the gap between such basis and the needs of the EVER-EST VRCs, including direct inspection of existing resources and the analysis of real-life scenarios and user actions from the VRCs. Section 4 continues with a summary of the outcomes of the gap analysis and the conceptual extensions required. Section 5 introduces the namespaces relevant for the updated models. Section 6 provides a detailed description of the new terms, customizations and extensions implemented in the research object model, extracted from existing metadata specifications in Earth Observation and their translation in RDF vocabularies and ontologies. Section 7 presents the core of this newly created branch of the research object model for Earth Science through updated metadata implementation guidelines associated to it. Throughout the document two main principles are conveyed: reuse of existing vocabularies and compliance with standards, and include a number of examples of the actual metadata produced by the application of the model for the description of specific research objects. Section 8 concludes the document and provides an overview of future work.

2 Workflow and Research Object Modeling

The work described in this document extends the ontology stack for research object modeling [1] resulting from the Wf4Ever project[1], where the research object concept was put forward for the first time as a mechanism to package, uniquely identify and preserve scientific investigations, for different purposes including communication of scholar findings, sharing of materials and their reuse for incremental science. The basic objective of research objects is therefore to account, describe and share everything about particular scientific investigations, including the way in which those things are related and the context in which they were used. As such, a research object bundles and relates all kinds of digital resources related to data-intensive research, including:

- The data used and the results produced in an investigation in the form of e.g. new data products.
- Scientific methods employed to produce and analyse the data.
- Scientific workflows and other computational resources implementing such methods, like e.g. online web services or virtual machines available in cloud environments or HPC infrastructures.
- Provenance information resulting from the execution of the scientific routines over the infrastructure and the configuration settings used.
- The researchers involved in the investigation, e.g. through their account identifiers (e.g. ORCID[2]), and the roles they play in it.
- Annotations about these resources, that are essential to the understanding and interpretation of the scientific outcomes captured by a research object.

The main benefits of a generalized use of research objects can be summarized as:

- Systematic machine-readable organization and description of the resources, materials, and methods of an investigation.
- Enable the accountability and reproducibility of the findings claimed in scholarly articles through evidence based both on the data and the provenance of the execution of the associated scientific methods.
- Enable tracking the evolution and progress of the scientific investigation, keeping snapshots (like tags) of the state of the research work at different points in time, which can be used for comparison, rollback, or to start (fork) different directions of work.
- Being uniquely identified by a digital identifier like e.g. a DOI, the research object becomes a citable, unique scientific object. This enables sharing research materials with other scientists at discrete milestones of a particular investigation, i.e. releasing the current status of the research versus its actual publication in a conference or journal.
- As a consequence, research objects are recognizable by the community and can be cited, lowering the barriers for scientific collaboration.
- Support reproducibility and reuse, alleviating decay of the scientific methods and their implementations.
- Support the collaboration among scientists, providing a shared unit of work where different scientists can contribute and comment dynamically and incrementally, keeping track of the people involved and the inputs from each of them. People involved can follow and get notifications of changes in the research object.

Such benefits come to a large extent from the ability of research objects to aggregate and provide scientific or technical context to the materials related to an investigation through annotations. These, aggregation and annotations, are the key aspects that the research object model aims to support from a representational point of view. Since the general research object model was introduced in detail in D4.1, this document only provides a short overview of the main ontologies comprising the model, and then introduces (not presented in D4.1) the

[1] http://www.wf4ever-project.org

[2] http://orcid.org

existing extensions and vocabularies, which specify terms borrowed and adapted from existing vocabularies and standards. When needed specific new terms are also defined, to be used within annotations.

The main ontologies comprised by the research object model are:

- **Research Object Core Ontology (ro, http://purl.org/wf4ever/ro)**: provides the basic structure for the description of aggregations, aggregated resources and annotations on those resources.
- **Workflow Definition Ontology (wfdesc, http://purl.org/wf4ever/wfdesc)**: describes computational processes and their implementations, particularly focused on scientific workflows
- **Workflow Execution Provenance Ontology (wfprov, http://purl.org/wf4ever/wfprov)**: describes the provenance information about the execution of a computational process, particularly focused on scientific workflows.
- **Research Object Evolution Ontology (roevo, http://purl.org/wf4ever/roevo)**: extends ro core to capture the research object evolution, including their lifecycle, versioning and related changes.

The research object model core ontology leverages two main ontologies for aggregation and annotation: the Open Archives Initiative Object Reuse and Exchange ontology[3] (OAI-ORE) and the Open Annotation ontology[4] (OA)

- **OAI-ORE** defines standard ways to describe and exchange aggregations of Web resources. These aggregations combine distributed resources with multiple media types including text, images, data, and video. The goal is therefore to expose the rich content of these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. OAI-ORE is unmistakably focused on the changing nature of scholarship and scholarly communication, and the required cyber infrastructure to support such scholarship. However, it is also generic enough to spread across a broad spectrum of web-based information.
- On the other hand, **OA** specifies an interoperable framework for creating associations between related resources, i.e. annotations. An annotation is considered to be a set of connected resources, typically including a body and target, where the body is somehow about the target. The full model supports additional functionalities, enabling semantic annotations, embedding content, selecting segments of resources, choosing the appropriate representation of a resource and providing styling hints for consuming clients. Under the auspices of the W3C, as of July 2016 the Open Annotation ontology was superseded by the Web Annotation ontology and the RDF implementation of the corresponding vocabulary[5]. This announcement required additional work in the extensions to the research object model described in this document in order to keep consistency with the new standard.

Figure 2 illustrates how the research object model organizes the encapsulated information, and provides a concise view of the use of the different ontologies and vocabularies comprised by the model. The core is the concept of the research object itself, in the form of an ORE aggregation. This is described by the research object manifest, an RDF description of the content and the structure of the specific research object. The research object aggregates a series of resources, which are referred to and linked to it by means of their URIs. Additionally, the research object aggregates a set of annotations both about a specific research object and the resources it aggregates.

Annotations are essential to the understanding and interpretation of the scientific outcomes captured by a research object as well as the reuse of the resources within. Annotations can refer among other things to e.g. provenance information about the experiments or observations, the study or any other resource, evolution information about the research object and its resources, descriptions of computational methods or processes, and dependency information or settings about the executions. In the research object model an annotation (expressed in the Web Annotation ontology) describes the link between a target resource (here aggregated in the research

[3] https://www.openarchives.org/ore

[4] http://www.openannotation.org/spec/core

[5] Web Annotation ontology: https://www.w3.org/ns/oa

Web Annotation vocabulary: https://www.w3.org/TR/annotation-vocab

object), and a body resource, which is typically provided as a separate RDF graph in order to enable the use of existing vocabularies to describe and relate each individual resource. Such annotation body comprises a description of the target resource in the form of a set of RDF metadata. The use of the term metadata in the remainder of this document will refer to this.

The Wf4Ever project produced a number of guidelines concerning the recommended properties and relations for the annotation of research objects and their aggregated resources[6]. These guidelines aimed at ensuring interoperability between the different systems using the research object model so that the annotations created in one tool or platform can be used and interpreted in the same way in another, preventing metadata ambiguity and duplicated efforts by the adopters. The metadata identified was specified through extensions of the research object model and new vocabularies, as follows:

- **Research Object Wf4Ever Ontology (wf4ever, http://purl.org/wf4ever/wf4ever)**: Defines more specific subclasses of terms from the main research object ontologies ro, wfdesc and wfprov, e.g. wf4ever:Image, wf4ever:WebServiceProcess in addition to associated properties for describing services and tools, e.g. wf4ever:command and wf4ever:script.
- **Research Object Terms vocabulary (roterms, http://purl.org/wf4ever/roterms)**: Specifies terms borrowed/adapted from existing popular vocabularies and standards, and defines new terms that are used (i) to specify annotations that are, to our knowledge, not catered for by existing ontologies, e.g., roterms:Hypothesis and roterms:exampleValue; (ii) to specify shortcuts that make the ontology easy to use and more accessible.

As mentioned above, the outcomes of the gap analysis conducted on the previously research object model and existing metadata conventions are presented in this document, together with an update and extension of the research object model for the Earth Science domain. This includes a customized version of metadata implementation guidelines, which have been implemented in the latest release of the ro ontologies and vocabularies.

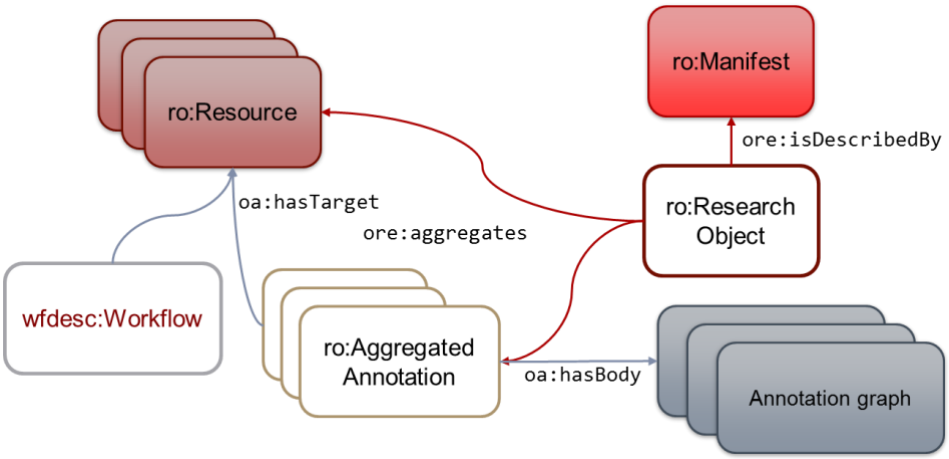

Figure 2: Research object model at a glance, including resource annotation and aggregation

The following example uses the model to describe in RDF a workflow-centric research object containing information about the workflow itself (described as a Taverna workflow specification in the file helloworld.t2flow) and its executions, through the wfdesc and wfprov vocabularies[7]. The annotation *ann1* contains information expressed in the Dublin Core Terms vocabulary about its creation time and author. The namespaces used in the example are described in section 5.

[6] https://confluence.man.poznan.pl/community/display/docs/Annotations+implementation+guidelines+release+1

[7] Detailed information about the general research object model and further examples can be found in the research object primer (http://wf4ever.github.com/ro-primer) and specification (http://wf4ever.github.com/ro).

```
<> a ro:WorkflowResearchObject ;
    ore:aggregates <helloworld.t2flow> ;
    ore:aggregates <artifact/hello> ;
    ore:aggregates :ann1 ;
    dct:created "2011-12-02T15:01:10Z"^^xsd:dateTime ;
    dct:creator [ a foaf:Person; foaf:name "John Doe" ] .

<helloworld.t2flow>    rdf:type    wfdesc:WorkflowDefinition .
<artifact/hello>       rdf:type    wfprov:Artifact .
```

Research objects can also be serialized as so called research object bundles, which allow users to dump the research object content in a compressed format for portability across different systems. A research object bundle is a regular zip file, which can be explored and unpacked with standard tools, enabling offline manipulation and exchange of the research object through conventional means. Figure 3 shows the anatomy of the research object bundle corresponding to the execution of a workflow. The bundle represents as a single .zip file a research object that aggregates workflow outputs, workflow definition, inputs used for execution, a description of the execution environment, external URI references (such as the project homepage) and attribution to scientists who contributed to the investigation. The resulting zip file is a research object where all those resources are tied together by the research object bundle manifest, which is in JSON-LD[8] format, a regular JSON representation that is also valid RDF.

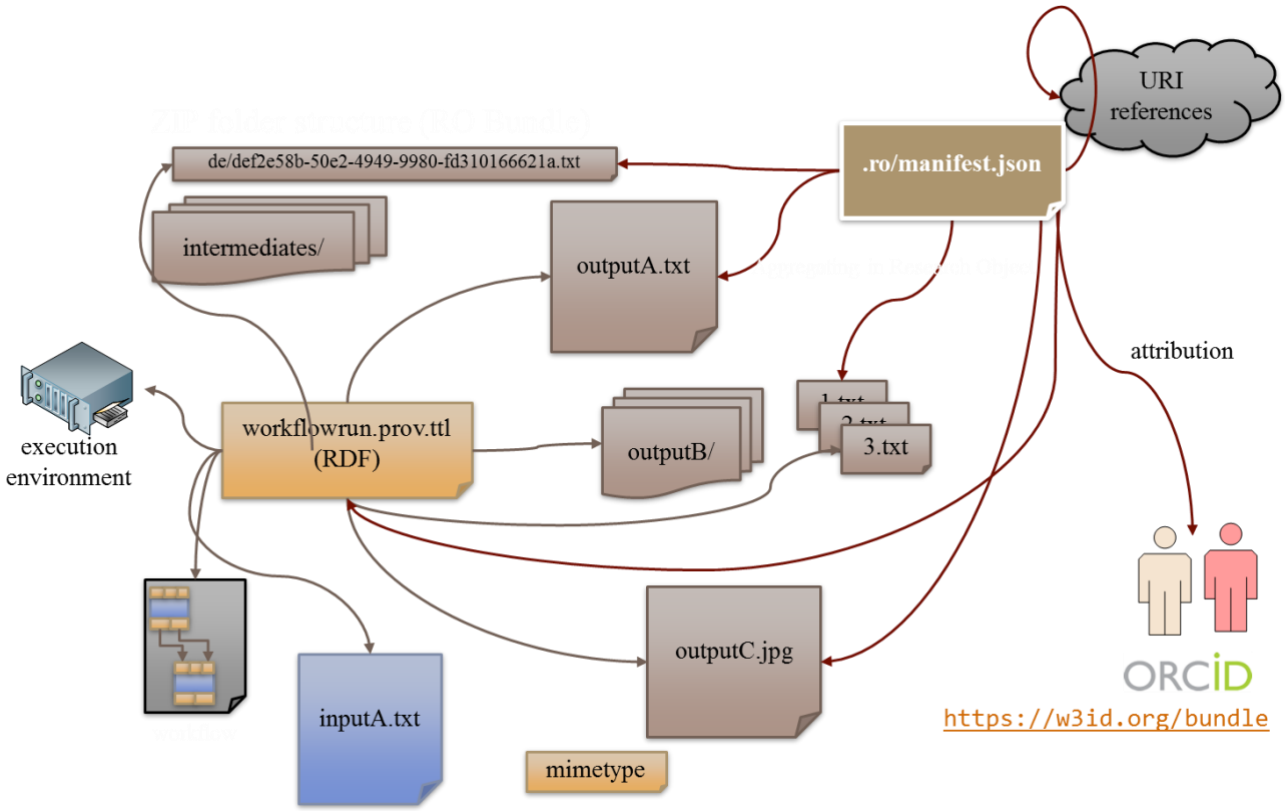

Figure 3: Example research object bundle – workflow execution results

[8] JSON-LD is a lightweight Linked Data format. A complete specification can be found at: http://json-ld.org

3 Research Object Model Customization and Extension for Earth Science - Methodology

3.1 Overall methodology

Figure 4 shows the overall methodology followed in order to generate the necessary extension and customization of the research object model for the Earth Science domain. The starting point is the existing research object ontology stack already mentioned in this document. Then, a gap analysis was performed that allowed surfacing the needs of the Earth Science communities which are not covered by the general model and the necessary extensions and updates were developed. It should be noted that as part of this work, the necessary updates in the main branch of the research object model, available through GitHub[9], were agreed and shared with the wider research object community. The extensions and updated ontologies (derived from the gap analysis) that have been produced were released as a new branch of the research object model repository in Github, called Earth Science.[10]

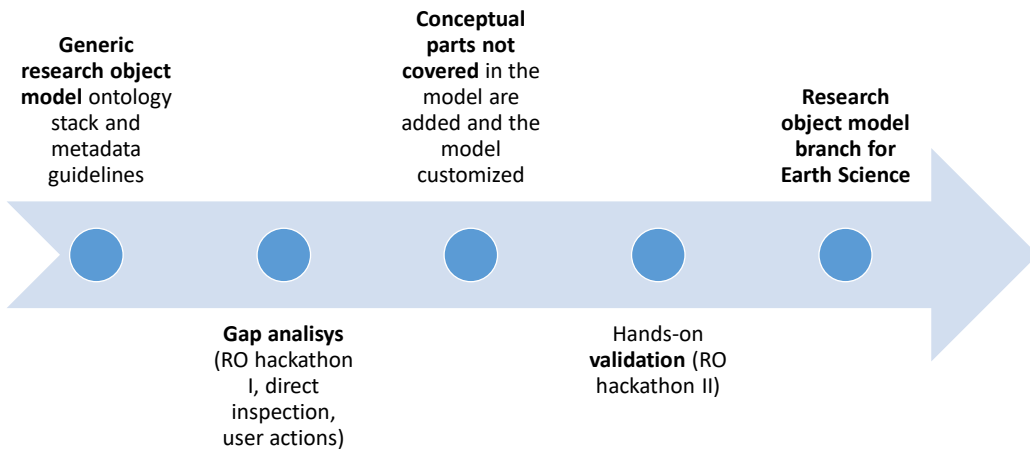

Figure 4: Work methodology for the research object model branch for Earth Science

Following the strategy defined in D4.1, the development of the research object model for workflow and research object representation in Earth Science adopted a learning-by-doing approach, where interaction with both technical partners and user communities has been constant. The objective was on the one hand to ensure that all the annotations required by the VRCs were appropriately expressed as metadata in the research object model and on the other hand that the on-going development of the infrastructure in WP5 was performed by taking into account such model as the basic information representation guidelines to be supported. This approach will ensure interoperability and portability between the VRC portals and the VRE as well as with new or third party tools and systems to be developed in the future.

Figure 5 shows the overall schedule for such process and the

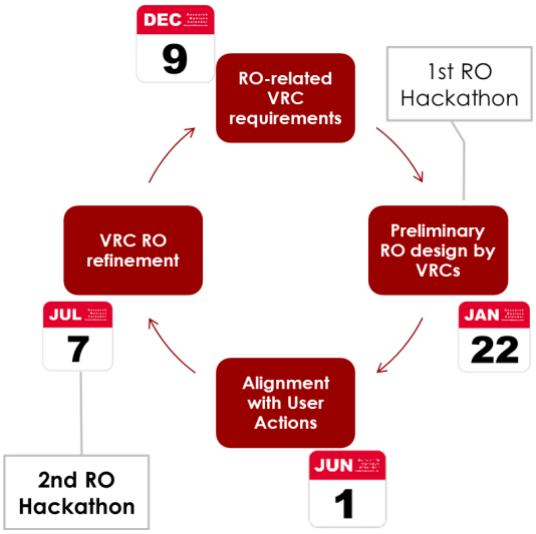

Figure 5: Research object model development and GERO modeling schedule

[9] Original research object model available at: https://github.com/wf4ever/ro.

[10] The new Earth Science extensions and ontologies are available at: https://github.com/wf4ever/ro/tree/earth-science

different interaction points with the user communities, particularly through the two **research object hackathons** held until now. The RO Hackathons proved once again extremely valuable in this regard. By having the VRC delegates model Golden Exemplar Research Objects (GERO), certain gaps in the expressivity of the model became apparent, particularly related to geographic information. Subsequent communication with the VRCs on the basis of their specific GEROs and **direct inspection** helped confirm incipient representational gaps that needed to be addressed. In this setting, the latter (direct inspection) was done as a pair activity between members of WP4 and a (group of) representative(s) of each VRC, typically by Skype. During such sessions various drafts of the metadata guidelines for modelling and describing research objects in Earth Science were successively proposed where the metadata was classified in different thematic areas representing their intended use, e.g. geographic information. Subsequently, the VRCs made observations highlighting possibly wrong interpretations or additional conceptual gaps that had not been appropriately covered.

In addition to hands-on work during the hackathons and direct inspection on the proposed metadata, an additional activity concerning the analysis of the **user actions** [4] typically carried out by the VRCs was performed and represented the third instrument of the gap analysis.

The following two chapters provided an overview of the analysis performed on the user actions and scenarios of two of the EVER-EST VRCs, namely in the domains of Land and Sea Monitoring.

3.2 Analysis of Land Monitoring user actions and the research object model

The Land Monitoring VRC identified the following three scenarios to be analysed:

1. **Land Monitoring:** change detection of selected areas of interest using image pairs from a specific dataset.
2. **Land Monitoring time series:** change detection of selected areas of interest using time series from heterogeneous datasets.
3. **Land Monitoring supported by social media:** change detection and characterization of selected areas of interest using time series from heterogeneous datasets and validation against ancillary information from social media.

3.2.1 User actions and research objects

The user actions associated to such scenarios and the corresponding implications from the research object modeling point of view are shown in **Table 1**.

Table 1: User actions and research objects in Land Monitoring VRC

Action ID	Action Description	ROs Metadata Comments
US_LM_03	User1 wants to define the access rights to the change detection map algorithm.	Very important in these cases seems to be the use of data access policy vocabularies, like e.g. ODRL [11] in order to generate metadata about concrete access to resources by user restrictions.
US_LM_09	User2 wants to define the access rights to the published change detection map.	
US_LM_10	User2 wants to define the access rights to the input data products of the change detection map.	
US_LM_04	User2 wants to define a specific area of interest within a specific time and search for products.	Metadata for geo-localisation and a map interface necessary to select areas. Time information specification also important.
US_LM_12	User2 or User3 want to upload an auxiliary information (e.g. raster, vector, another change detection) layer to visualize together with change detection map.	Updates in the research object infrastructure (upload new resources, edit the existing ones, etc.) supported by the platform.

[11] https://www.w3.org/community/odrl

US_LM_13	User3 wants to annotate the discovered change detection map.	Including annotations and comments related to a resource.
US_LM_14	User3 wants to reproduce the change detection map.	Mechanisms to reproduce observation outcomes encapsulated in a research object.
US_LM_15	User4 wants to subscribe to the change detection service.	Research object notification and subscription.
US_LM_16	User4 is notified when a given change detection is produced.	
US_LM_21	User2 wants to add the information coming from Social Sensing sources related to the specified area of interest.	Include social sensing annotations metadata with subject and geo-localisation data.

3.2.2 User actions GUI requirements and research objects

Table 2 shows the requirements, from the perspective of the user interface, associated to the scenarios and the research object counterparts.

Table 2: GUI requirements and research objects in Land Monitoring VRC

GUI Requirement	Description	ROs Comments
GUI_LM_01	The search user interface allows a keyword, spatial and temporal search.	Research object browsing must allow the introduction of keywords, spatial and temporal data to make searches.
GUI_LM_06	The user interface gives access to logs, metadata and SS events of the change detection map.	Research object platforms must allow downloading and accessing to logs, metadata, resources, the research object itself, etc.
GUI_LM_07	The user interface allows the download of the (enriched) change detection map.	
GUI_LM_08	The user interface allows the user to subscribe to the notification service.	Research object notification service.
GUI_LM_09	The user interface allows the user to upload maps to verify the changes.	Research objects are agnostic from the type of resource to be aggregated.

3.3 Analysis of Sea Monitoring user actions and the research object model

The Sea Monitoring VRC identified the following three scenarios to be analysed:

1. **Habitat suitability model for CWCs:** A user develops the habitat suitability model for the cold water corals (CWC) of the Bari canyon (and creates a research object when saving the model).
2. **The Citizen science and jellyfish distribution:** A crowdsourcing app sponsored by an 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 to generate meaningful indicators in MSF perspective.
3. **Trend in the evolution of invasive jellyfish distribution:** Starting from jellyfish sightings, explicit geographic data is produced concerning trends about the evolution and distribution of alien species according with MSF directive descriptors.
4. **Habitat suitability model for jellyfish species, study cases *Pelagia noctoiluca* and *Velella velella:*** Starting from jellyfish sightings according with environmental variables a suitability model is produced (and added it to the previously created research object) to evaluate habitat extent for *Pelagia noctoiluca* and *Velella velella.*

In such scenarios, the input data and the outcomes generated shall be encapsulated following the research object approach. Particularly, in scenario 3 it can be seen once more the importance of the geolocalisation metadata as in the previous VRC.

3.3.1 User actions and research objects

The user actions associated to such scenarios and the corresponding implications from the research object modeling point of view are shown in Table 3.

Table 3: User Actions and Research Objects in Sea Monitoring VRC

Action ID	Action Description	ROs Metadata Comments
CNR_CWC_03	User1 downloads the bathymetry, the occurrences layer and the absence data layer.	In these steps, the action of creating a research object is involved, uploading the data, building workflows to model the processes and aggregating any annotation to specify the relations between resources
CNR_CWC_04	User1 produces the EGVs.	
CNR_CWC_05	User1 carries out a statistical analysis to evaluate which EGVs are relevant for the suitability model.	
CNR_CWC_06	User1 choose the EGVs useful for the analysis.	Updates in the current Live research object, editing the existing EGV data or uploading a new resource.
CNR_CWC_07	User1 uses the selected EVGs and the occurrences layer to generate the habitat suitability model.	Workflows are built to specify and enact these processes. Human interaction needed.
CNR_CWC_08	User1 checks the results.	
CNR_CWC_09	User1 updates the live research object and creates a snapshot of the research object and define access rights to external users.	Modifications in the Live research object and definition of access rights metadata.
CNR_JFD1.1_07	User1 updates the map in the Live research object by adding a new resource, creates a snapshot of the research object and defines access rights to external users.	
CNR_CWC_10	CNR-ISMAR and EU access the research object.	
CNR_JFD1.1_02	User2 is notified when new sightings were sent out.	Research object notification and subscription.
CNR_JFD1.1_07	Upon notification, User3 updates the CoCoNet database with the new validated sightings.	
CNR_JFD1.1_03	User2 creates the initial Live research object and checks and validates the new data within ROHub/VRC GUI.	Creation of the research object and data upload.
CNR_JFD1.1_04	User2 processes the data to produce a daily map of distribution for each species sighted.	Workflows are modelled to execute these analyses automatically.
CNR_JFD1.1_05	User2 processes the data to produce map of distribution for each species to inform Institutional users.	
CNR_JFD2.1_02	User2 and User1 elaborate sightings to produce density raster map for each invasive species for each year.	
CNR_JFD2.1_03	User2 and User1 analyse density rasters and made a quantitative analysis to represent change in time distribution of each alien/invasive species.	

CNR_JFD2.1_04	User3 receives results and produces the final map.	
CNR_JFD2.2_04	User2 and User1 perform a statistical analysis using Maxent and R.	
CNR_JFD1.1_06	User2 shares the new data to User3.	Modifications in the Live research object to upload the new data and maybe data access policy updates to include User3's permissions.
HI_CNR_01/ HI_CNR_02/ HI_CNR_03	Selection of the variables, validation of the results and validation of the data.	Need of human interaction to address these actions

4 Gap Analysis Outcome

The gap analysis revealed four main areas where the gap between the coverage provided by the previously existing research object model and the needs of the VRCs were significant: Geographic and time information, intellectual property rights and data access policies, plus general information which was considered to be useful. In some cases, such information was not covered at all by the previous version of the research object model (geographic and time, data access policies), and in other cases it was not covered with sufficient detail as required by the project VRCs (intellectual property rights). Table 4 shows the main conceptual areas where such gaps were identified and a (non exhaustive) list of specific examples. The aim has been to bridge such gaps through standards deriving from the Earth Science disciplines themselves in order to maximize the familiarity of the user communities with the resulting terminology and leverage previous work by consortium members. This is particularly the case of ESA and their contribution to the Earth Observation SIP-Scene Specification for the SEASAT mission[12]. The following section provides further detail on such updates.

Table 4: Main representational gaps in the research object model – metadata types and examples

Metadata Types	Metadata Examples
General	Type - When the user creates a research object it should be possible to specify the type of the scientific discipline. In the case of the EVER-EST VRCs, the Earth Observation type should be added and the corresponding machine-readable annotation generated.
	Size – Calculated automatically from the resources aggregated by the research object.
	Date of submission – Generation of the annotation automatically when the user rereleases the research object.
	Digital Object Identifier (DOI).
	Status – Corresponding to the lifecycle of the research object.
	Distribution Category and Designated Community to whom the research object is aimed.
	Is disseminated, is catalogued – Include binary checks in each resource to mark this information by users (this can be done either at the level of the model or at the system implementation level).
GEO and Time	Time period – Include an annotation to define the time span covered by the observation referred by the research object.
	Geoposition coordinates – Supports the selection of a space region relevant for the research object and the observation it represents, as well as the generation of the respective geographical metadata.
Intellectual Property Rights	Support for annotations on the research object and its resources dealing with terms like e.g. copyright, copyright year, attribution text, license, etc.
Data Access Policy	Data access restrictions. These annotations will involve e.g. privacy, permission, access scope, access groups, etc.

[12] https://directory.eoportal.org/web/eoportal/satellite-missions/s/seasat

5 Namespaces

The following namespaces of the ontologies and vocabularies are used along the remainder of the document, and in the final implementation:

- ore=http://www.openarchives.org/ore/terms
- ao=http://purl.org/ao[13]
- oa=http://www.w3.org/ns/oa#
- dc=http://purl.org/dc/elements/1.1
- dct=http://purl.org/dc/terms
- rdfs=http://www.w3.org/2000/01/rdf-schema#
- rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
- foaf=http://xmlns.com/foaf/0.1
- ro=http://purl.org/wf4ever/ro#
- wfdesc=http://purl.org/wf4ever/wfdesc#
- wfprov=http://purl.org/wf4ever/wfprov#
- roevo=http://purl.org/wf4ever/roevo#
- wf4ever=http://purl.org/wf4ever/wf4ever#
- roterms=http://purl.org/wf4ever/roterms#
- pav=http://purl.org/pav
- swrc=http://swrc.ontoware.org/ontology#
- cito=http://purl.org/spar/cito
- dbo=http://dbpedia.org/ontology
- ov=http://open.vocab.org/terms
- bibo=http://purl.org/ontology/bibo
- prov=http://www.w3.org/ns/prov#
- geo=http://www.opengis.net/ont/geosparql#
- sf=http://www.opengis.net/ont/sf#
- gml=http://www.opengis.net/ont/gml#
- odrs=http://schema.theodi.org/odrs#
- cc=http://creativecommons.org/ns#
- odrl=http://www.w3.org/ns/odrl/2
- geo-wgs84=http://www.w3.org/2003/01/geo/wgs84_pos#
- roes=http://w3id.org/ro/earth-science#

Additionally, the following xml namespaces are used for reference in some sections of the document:

- sar: http://www.opengis.net/sar/2.1 (Synthetic Aperture Radar)[14]
- gml-spec: http://www.opengis.net/gml/3.2 (Geography Markup Language)
- om: http://www.opengis.net/om/2.0 (Observations and Measurements)
- eop: http://www.opengis.net/eop/2.1 (Earth Observation Product)[15]

[13] Currently deprecated, ao is kept and aligned with oa for backward compatibility purposes.

[14] During the analysis this namespace could not be resolved. Nevertheless, this convention has also been adopted herein since it seems to be accepted by the user communities.

[15] Same as above.

6 Research Object Model Updates

Based on the outcome of the gap analysis, the research object model has been updated and extended as follows:

- Included the required vocabulary terms for describing geographic and time information, data access policies and intellectual property.
- Updated some existing concepts corresponding to processing units to cover not only scientific workflows but also other computational resources widely used in Earth Science disciplines.
- Extended the research object types, to characterize not only workflow-centric ones, but also data-centric research objects, service-centric research objects, and documentation and bibliographic research objects.
- Extended the research object lifecycle states to include the fork state, in addition to live, snapshot and release/archive, which will serve to characterize different branches of the research object for new/different lines of work.
- Updated and aligned the research object core ontology and required extensions with the latest model of the Annotation Ontology, called Open Annotation Ontology (and since July 2016 Web Annotation Ontology, W3C Candidate Recommendation).
- Cleaned and properly annotated all the ontologies with provenance and metadata information.

The last two actions were the first ones carried out, and they were agreed and shared with the wider research object community. In particular, before extending the model the required updates were analysed in order to (i) clean and annotate the existing ontologies, (ii) align them with the new annotation model, and (iii) align the different branches of the research object model in GitHub[16]. The outcome was discussed and agreed with other key contributors of the original model. Then, the extensions required by the other actions were conducted and implemented, and the resulting ontologies were agreed as well.

The following sections detail the changes that were applied for each ontology. When appropriate, a diagram is included to illustrate the main classes and properties in the latest version. Each colour used in the diagrams corresponds to a specific namespace.

6.1 Research object core ontology (ro)

- Align terms from Open Annotation Ontology and Annotation Ontology.
    - object property ro: annotatesAggregatedResource subproperty of oa:hasTarget.
    - object property ao:annotatesResource equivalent to oa:hasTarget.
    - object property ao:body equivalent to oa:hasBody.
    - class ao:Annotation equivalent to oa:Annotation.
    - class ro:AggregatedAnnotation subclass of oa:Annotation.
- Deprecated terms.
    - object property ro: annotatesAggregatedResource.
    - object property ao:annotatesResource.
    - object property ao:body.
    - class ao:Annotation.
- Removed terms.
    - class ro:SemanticAnnotation.
- Removed imports external ontologies (dc,ao), import newly created metadata ontology.
- Add definedby statement for each term.
- Cleaned and properly annotated ontology .

[16] https://github.com/ResearchObject/specifications/issues/13

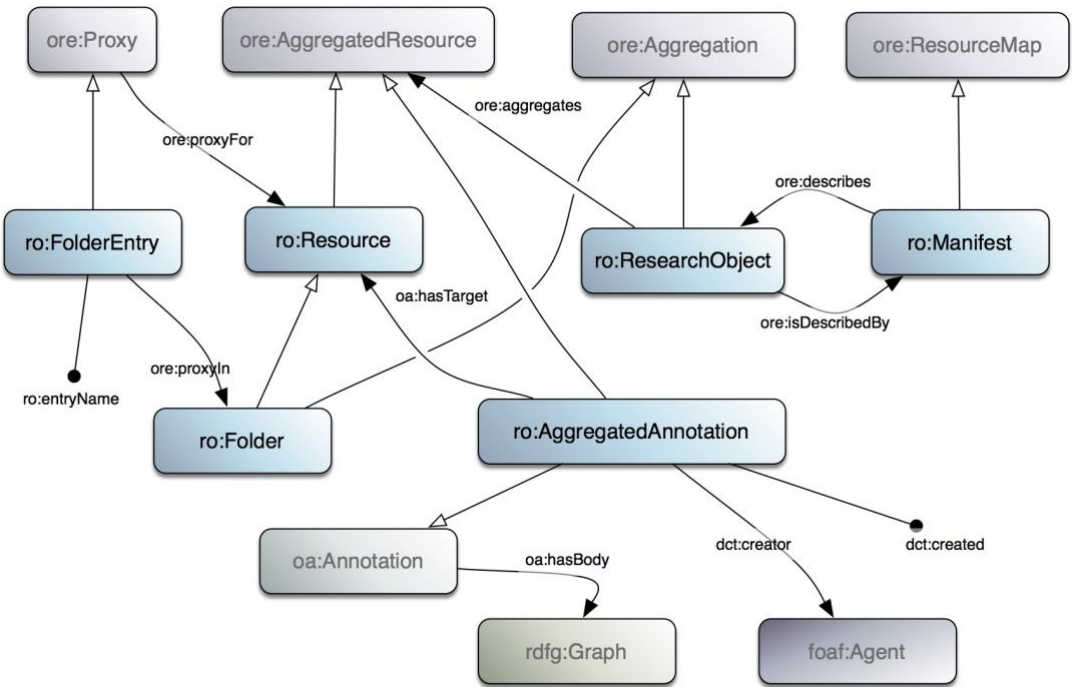

Figure 6. ro ontology overview

6.2 Workflow description ontology (wfdesc)

- Removed terms.
  - wfdesc:Description
- Add definedby statement for each term.
- Cleaned and properly annotated ontology.

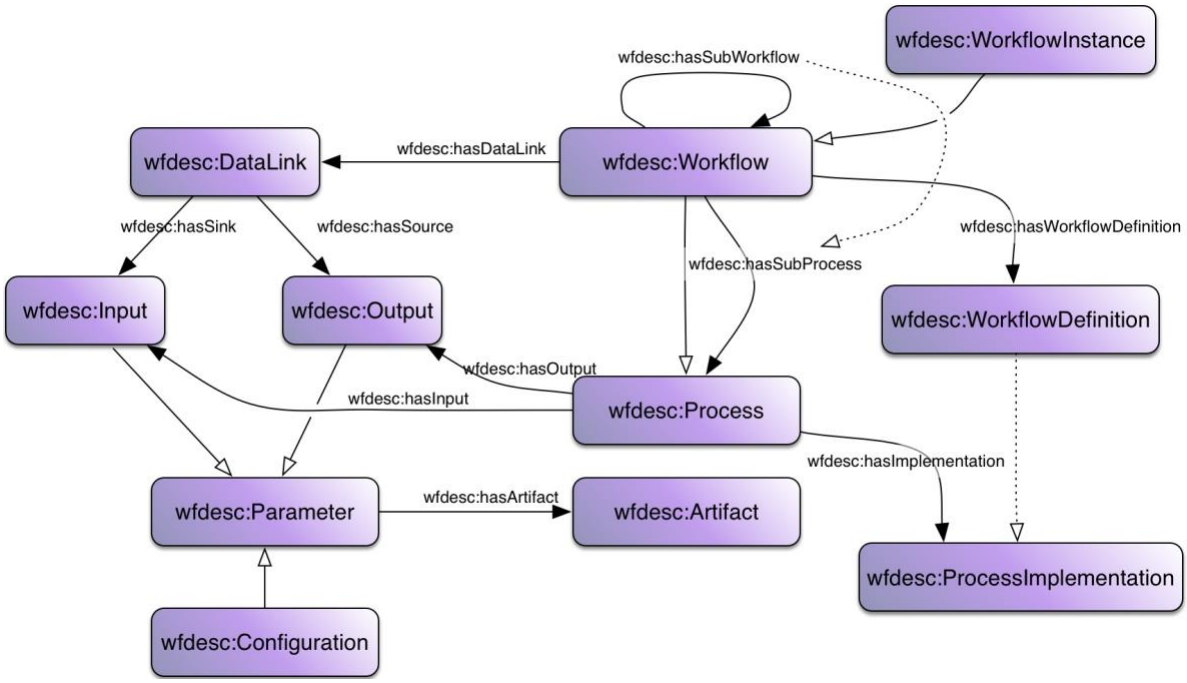

Figure 7. wfdesc ontology overview

6.3 Provenance ontology (wfprov)

- Add definedby statement for each term.
- Cleaned and properly annotated ontology.

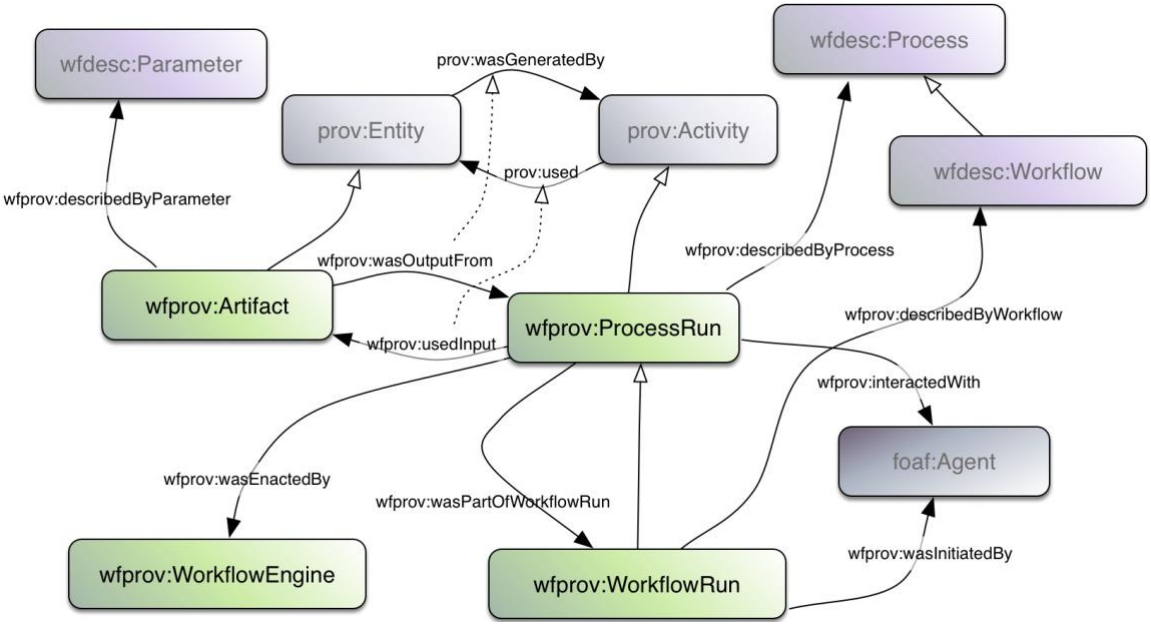

Figure 8. wfprov ontology overview

6.4 Research object evolution and versioning ontology (roevo)

- Align terms from Open Annotation Ontology and Annotation Ontology.
  - Object property ao:annotatesResource equivalent to oa:hasTarget.
  - Object property oa:hasTarget used for specifying annotations on roevo:Change, roevo:ChangeSpecification, roevo:VersionableResource.
- Deprecated terms.
  - Object property ao:annotatesResource.
- New terms.
  - Class roevo:ForkedRO
- Add definedby statement for each term.
- Cleaned and properly annotated ontology.

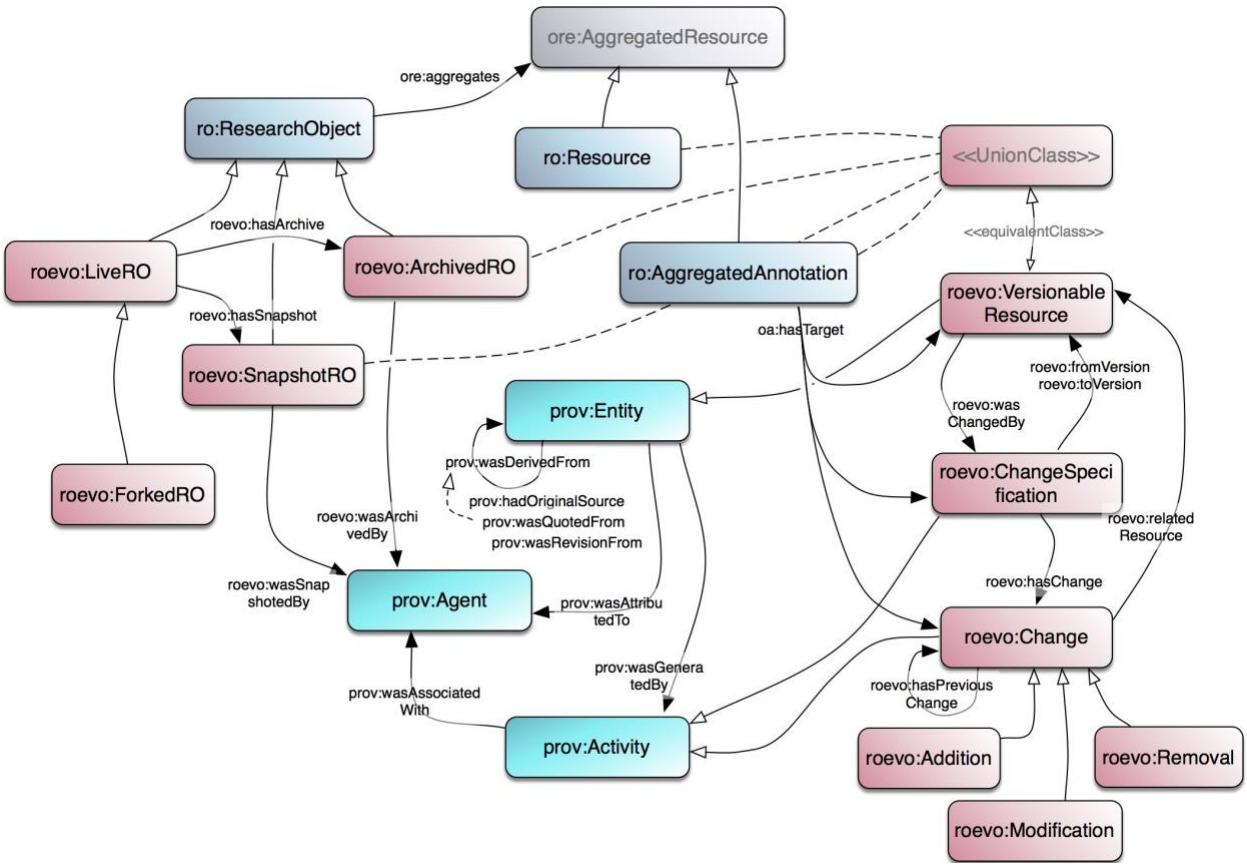

Figure 9. roevo ontology overview

6.5 Wf4Ever ontology (wf4ever)

- Add definedby statement for each term.
- Cleaned and properly annotated ontology.

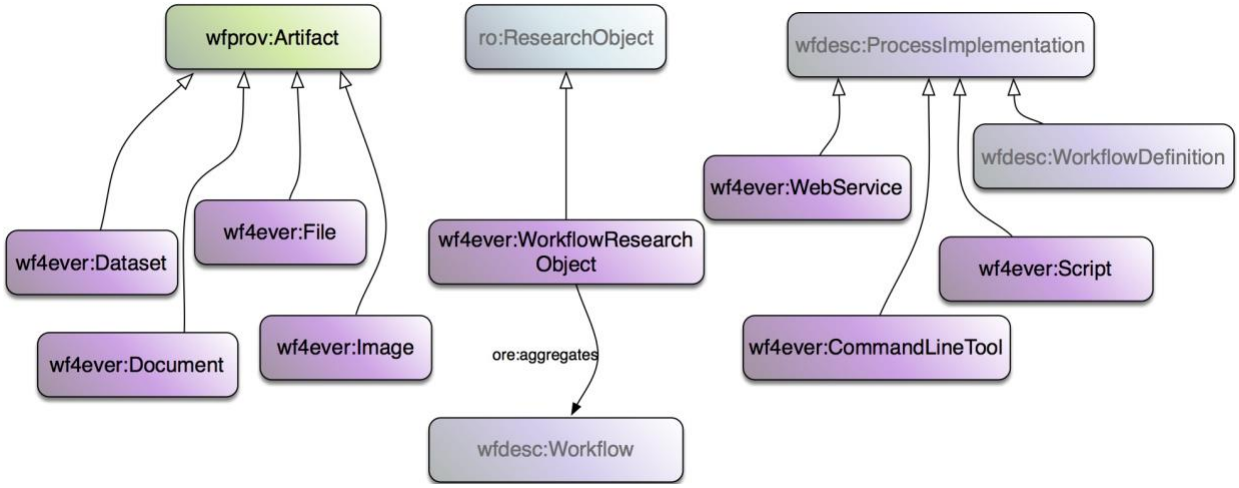

Figure 10. wf4ever ontology overview

6.6 Research object terms ontology (roterms)

- Generalisation of workflow terms
  - Generalise class roterms:WorkflowValue to roterms:ProcessValue
  - Object property roterms:defaultValue to use range roterms:ProcessValue instead of roterms:WorkflowValue.
  - Object property roterms:exampleValue to use range roterms:ProcessValue instead of roterms:WorkflowValue.
  - New class roterms:previousProcess superclass of roterms:previousWorkflow.
  - New class roterms:subsequentProcess superclass of roterms:subsequentWorkflow.
- New class roterms: MeetingMinutes.
- Add definedby statement for each term.
- Cleaned and properly annotated ontology.

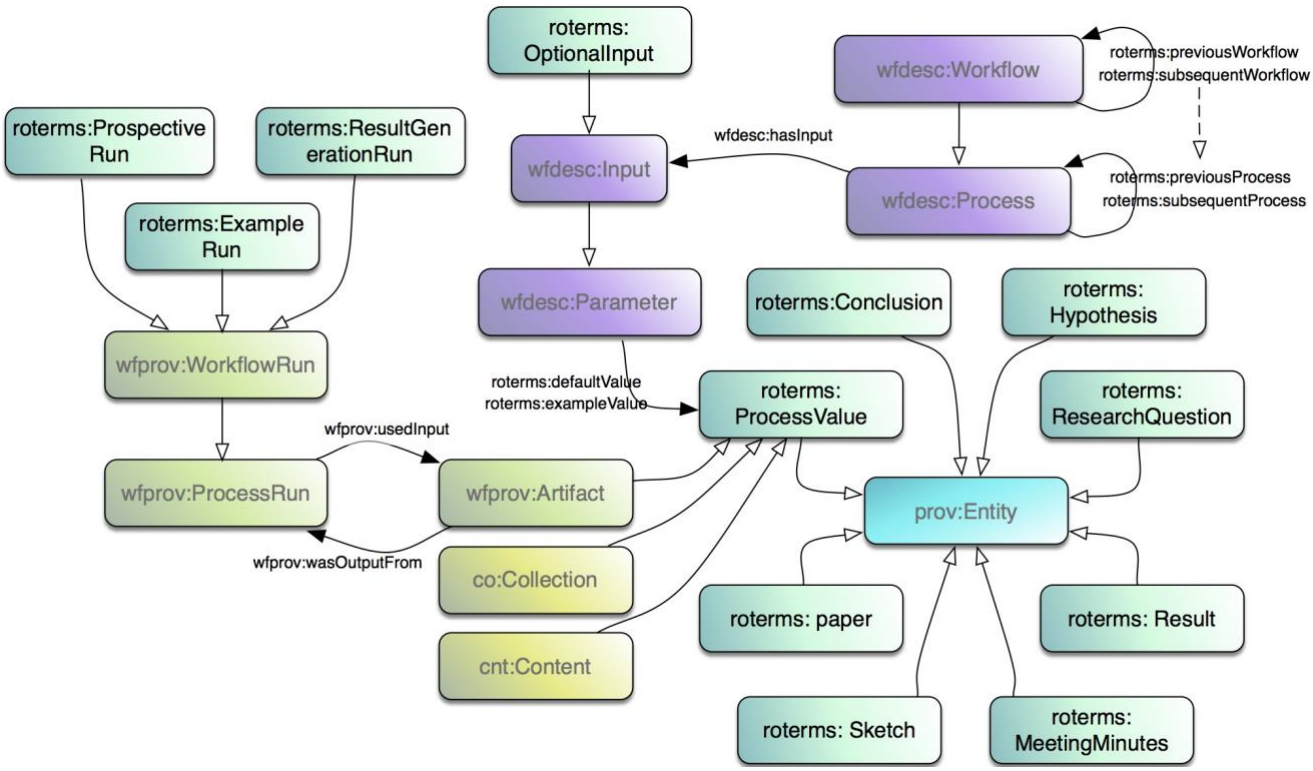

Figure 11. roterms vocabulary overview

6.7 Research object ontology for Earth Science (roes)

- New ontology extending the research object ontology suite with Earth Science-specific terms. This ontology includes both definition of terms, and reference to terms from well-known vocabularies, which are useful for the annotation of resources in earth-science domain.
- This ontology has been assigned the namespace http://w3id.org/ro/earth-science# .

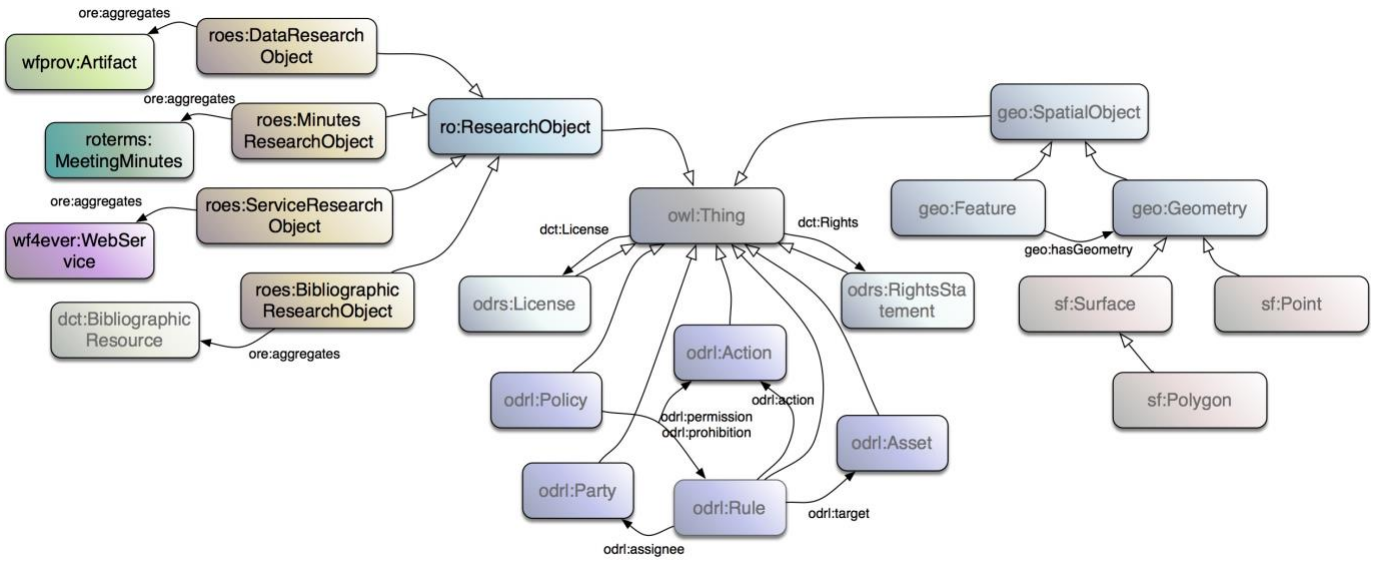

Figure 12. earth-science ontology overview

6.8 Additional metadata

- New vocabulary specifying terms borrowed/adapted from existing popular vocabularies and standards, including: dc, dct, foaf, cito, bibo and others, for specifying metadata information.
- Note: no diagram provided as this vocabulary does not define new terms.

7 Metadata Implementation Guidelines

7.1 Adaptation and integration of existing Earth Observation metadata specifications

This section analyses the metadata that is used in the field of Earth Observation in order to align its specification within the research object metadata ontologies. The metadata included here is information, which was missing from the existing research object model, as shown in the previous section. To this purpose, the metadata specification recommended by the SEASAT Earth Observation mission [3] has proved to be particularly useful. The relevant fragments of such specification have been customized and transformed into ontologies and vocabularies in RDF form, addressing the previously identified gaps. In doing so the Linked Data paradigm[17] for semantic data modeling has been followed in order to enable interoperability with the rest of the research object model ontology stack and potential future extensions.

In the tables in the subsequent sections it shall be noted that **orange relates to new entities in the research object model, including types, properties and relations**, and **^ relates to properties which have been defined in the model but still need to be adopted by the EVER-EST infrastructure**.

7.1.1 General information

Table 5: Additional Metadata for General Information

Name	Term	Description
Earth Observation (type)	dc:subject	To mark the type of a research object which encapsulates an Earth Observation process (value: sar:EarthObservation)
Size	dbo:filesize[18]	To include a whole size reference of the research object (eop:size)
Release Date	roevo: archivedAtTime	Release Date of the research object
DOI Reference^	bibo:doi^[19,20] & dc:identifier^	Include DOI reference for ROs (e.g. "10.5748/369/e4y2k9")
RO Status	roevo ro subclasses: • liveRO ○ forkedRO^ • snapshotRO • archiveRO	Status of the research object (snapshot, release, live, fork, etc.) – (eop:status)
Distribution Category^	roes:distributionCategory^	Intended use of the research object (Official use, Internal use, Releasable to the Public) - filled in by the Data Curator (bibframe:usageAndAccessPolicy)
Designated Community	dct:audience	Intended designated community / target audience (Internal, External) – (bibframe:intendedAudience)

[17] Linked Data W3C specification: https://www.w3.org/standards/semanticweb/data

[18] To evaluate if it would be better to create roterms:rosize

[19] http://inkdroid.org/2011/04/25/dois-as-linked-data/

[20] DOI Linked Data Documentation

7.1.2 GEO and Time metadata

Table 6: Additional Metadata for Geographic and Time Information

Name	Term	Description	Example
Time Period[21] • **Begin Position** • **End Position**	**UML:** gml:TimePeriod (class) •gml:beginPosition •gml:endPosition **ONTO:** • prov:generatedAtTime^ • prov:invalidatedAtTime^	Time coverage of the research object as a period • Start date and time coverage of the research object content • End date and time coverage of the research object content	<gml:beginPosition>1978-09-02T07:59:37Z</gml:beginPosition> <gml:endPosition>1978-09-02T07:59:51Z</gml:endPosition>[22]
Feature of Interest • **Polygon**[23] **List Positions** • **Center Position**	**UML:** om:featureOfInterest (class) • gml:Polygon (class) gml:posList • eop:centerOf (class) gml:Pos	Observed area of interest in a map • **gml:Polygon**, To mark geo-references as polygon shapes in a map, delimiting a space ○ **gml:posList**: Footprint coordinates, closed polygon (last point = first point), Counter-clockwise 5 points[24]; using WGS84 Latitude/Longitude pairs • **eop:centerOf**, Scene center representation structure ○ **gml:Pos**: Center coordinates of a point of Observation (lat/lon)	
Polygon	**UML:** gml:Polygon **ONTO:** sf:Polygon^[25] (class)	To mark geo-references as polygon shapes in a map, delimiting a space	
List Positions	**UML:** gml:posList **ONTO:** geo:hasGeometry^ (some sf:Polygon^) geo:asWKT^[26]	Footprint coordinates, closed polygon (last point = first point), Counter-clockwise 5 points; using WGS84 Latitude/Longitude pairs	<gml:posList>50.536655 0.184164 51.295147 -0.438592 50.869156 -1.717447 50.116280 -1.076581 50.536655 0.184164</gml:posList>
Center	**UML:** eop:centerOf	Scene centre representation structure	

[21] http://www.waterml2.org/

[22] Ontology precursor: https://www.seegrid.csiro.au/subversion/CGI_CDTGVocabulary/trunk/OwlWork/ogc-gml.owl (contains gml:begin, gml:end)

[23] http://www.georss.org/gml.html

[24] At least four pairs with the last being identical to the first

[25] http://www.opengis.net/ont/sf . This class shall provide research objects with a notion of geometry.

[26] http://www.opengis.net/ont/geosparql

	ONTO: roes:centerOf^[27]		
Position	**UML:** gml:Pos **ONTO:** geo:hasGeometry^ (some sf:Point^) geo:asWKT^	Center coordinates of a point of Observation (lat/lon)	<gml:pos>50.7 -0.7</gml:pos> using WGS84 Lat/long

7.1.2.1 Metadata description example for geographic information[28]

```
ex:WashingtonMonument a ex:Monument;
 rdfs:label "Washington Monument";
 geo:hasGeometry ex:WMPoint .
ex:WMPoint a sf:Point;
 geo:asWKT "POINT(-77.03524 38.889468)"^^geo:wktLiteral.

ex:NationalMall a ex:Park;
 rdfs:label "National Mall";
 geo:hasGeometry ex:NMPoly .
ex:NMPoly a sf:Polygon;
 geo:asWKT  "POLYGON((-77.050125  38.892086,  -77.039482  38.892036,  -77.039482
38.895393,
-77.033669  38.895508,  -77.033585  38.892052,  -77.031906  38.892086,  -77.031883
38.887474, -
77.050232 38.887142, -77.050125 38.892086 ))"^^geo:wktLiteral.
```

7.1.3 Intellectual Property Rights

Table 7: Additional Metadata for Intellectual Property Rights

Name	Term	Description
Property rights (0)	dc:rights	Intellectual and Copyright information
License (0)	dct:license (refines rights)	License information
Open Data Rights Statement Vocabulary (ODRS)[29]	odrs:RightsStatement^ (class)	A description of the rights and terms of re-use for a dataset/resource

[27] Term redefined and adapted from ontology gml v1 (gml-1:centerOf) http://schemas.opengis.net/gml/1.0.0/gml.rdfs. Currently it is not resolvable (no namespace) and incompatible with gml v3 used in geosparql ontology.

[28] http://ontolog.cim3.net/file/work/EarthScienceOntolog/2012-12-12_EarthScienceOntolog_session-5/GeoSPARQL_Getting_Started--DaveKolas_20121212.pdf

[29] https://theodi.org/guides/publishers-guide-to-the-open-data-rights-statement-vocabulary

– Copyright Statement		
ODRS -- License	odrs:License^ (class)	A legal document that describes the legal terms for re-use of some information.
ODRS – Attribution Text	odrs:attributionText^	The text to use in an attribution link. This may be the name of the publisher or a reference to a community or group of contributors (e.g. "*Open Data, Ltd*")
ODRS – Attribution URL	odrs:attributionURL^	The link which should be used when attributing a data source. The URL could be a reference to the dataset or publisher homepage, but may also be a dedicated attribution page. This is useful when providing onward attribution to upstream sources
ODRS - Copyright	odrs:copyrightNotice^	The copyright notice associated with a rights statement (e.g. "*Copyright © 2013. Example, Ltd*"). Where a data publisher needs to reference a larger copyright statement and/or related guidance then the copyrightStatement property should be used instead.
ODRS – Copyright year	odrs:copyrightYear^	The year from which copyright over the content of the dataset is asserted
Creative Commons Rights Expression Language (CC REL)[30] - Custom License	cc:permits^, cc:requires^, cc:prohibits^, cc:DerivativeWorks^, cc:Distribution^, cc:Reproduction^, , etc.	Description of copyright licenses in RDF.

7.1.3.1 Metadata description example for a standard license (ODRS)

```
example1
a dcat:Dataset ;
dct:title "Example Dataset" ;
dct:rights :example1-rights-statement;
dct:license <http://reference.data.gov.uk/id/open-government-licence>.

:example1-rights-statement
a odrs:RightsStatement;
rdfs:label "Rights relating to re-use of the Example Dataset" ;
odrs:copyrightNotice "Crown copyright 2016";
odrs:reuserGuidelines <http://gov.example.org/reuser-guide>;
odrs:dataLicense <http://reference.data.gov.uk/id/open-government-licence> ;
odrs:attributionText "Example Department" ;
```

[30] http://creativecommons.org/ns

```
odrs:attributionURL      <https://www.gov.uk/government/organisations/ministry-of-justice.
```

7.1.3.2 Metadata description example for a custom license (ODRS and CC REL)

```
:example2
   a dcat:Dataset ;
   dct:title "Dataset with Custom License" ;
   dct:rights :example2-rights-statement ;
   dct:license :example2-license .

:example2-rights-statement
   a odrs:RightsStatement ;
   rdfs:label "Rights Statement referencing our custom license" ;
   odrs:dataLicense :example2-license .

:example2-license
   a odrs:License, cc:License;
   rdfs:label "Custom Data License";
   cc:legalcode <http://example.org/legal/custom-license> ;
   cc:permits cc:DerivativeWorks, cc:Distribution, cc:Reproduction ;
   cc:requires cc:Notice , cc:ShareAlike.
```

The license of this example allows the creation of derivative works, distribution and reproduction of the data and the re-user must include all copyright notices and license derivatives using similar terms; but there is no *legal* requirement to include attribution.

7.1.4 Data access policy

Table 8: Additional Metadata for Data Access Policy

Name	Term	Description
Open Data Rights Language (ODRL)[31]	odrl:Policy^ (class)	Vocabulary that describes the use of digital content in publishing, distribution, and consumption of digital media across all sectors and communities https://www.w3.org/community/odrl/model/2.0/
Permissions • **Target** • **Action** • **Constraint**	odrl:Permission^ (class) • odrl:target^ • odrl:action^ • odrl:constraint^	A Permission is a rule which indicates the actions the assignee is permitted to perform on the associated asset • odrl:target - The target property specifies the Asset upon which the Action is performed (e.g. a research object) • odrl:action (e.g. odrl:reproduce, odrl:modify) • odrl:constraint a odrl:Constraint

[31] https://www.w3.org/ns/odrl/2/ODRL21

Prohibitions • **Target** • **Action**	odrl:Prohibition^ (class) • odrl:target^ • odrl:action^	A Prohibition is a rule which indicates the Actions that the assignee is prohibited to perform on the related Asset
Agreement	odrl:Agreement^ (class)	Policy expressions that are formal contracts (or licenses) stipulating all the terms of usage and all the parties involved
Offer	odrl:Offer^ (class)	An instance of Offer is a Policy expression that proposes terms of usage from an Asset owner
Privacy	odrl:Privacy^ (class)	An instance of Privacy is a Policy expression that stipulates the terms of usage over personal information
Set	odrl:Set^ (class)	An instance of Set is a Policy expression that consists of entities from the complete model
Party	odrl:Party^ (class)	An entity which can participate in policy transactions (e.g. a User)
Group	odrl:Group^ (class)	Specifies that the scope of the relationship is the defined group with multiple individual members (e.g. Group of users with different scopes; administrators, publishers, common users, etc.)
Action	odrl:Action^ (class)	Instances of Action are things one might be permitted to do or prohibited from doing to a work (e.g., read, write, delete, copy, execute, etc.) – to implement only few actions like read, write, delete, copy.

7.1.4.1 Metadata description Example for data access policy (ODRL)

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .

<http://example.com/policy:5109>
    a odrl:Agreement ;
    odrl:permission [
        a odrl:Permission ;
        odrl:action odrl:display ;
        odrl:target <http://example.com/myPlace:billie:photos:football2010> ;
        odrl:assigner <http://example.com/myPlace:billie> ;
        odrl:assignee <http://example.com/myPlace:network:football>
    ] .

<http://example.com/myPlace:billie> a odrl:Party .
<http://example.com/myPlace:network:football> a odrl:Group.
```

This example shows the instance of an agreement policy for a social network scenario. The target asset are photos posted to a social network site and the assigner is the owner of the photos. The assignee is a group and represents the football network members on the social network, who are each allowed to display the photos.

7.1.4.2 Simplified metadata description example for data access policy (ODRL) – to be implemented

A useful example for implementation: one may abbreviate the expression of the policy and use the Dublin Core rights or license predicates to associate the asset with that policy, when all of the rules associated with the policy have the same target, and do not have any constraints or duties (requirements).

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix dct: <http://purl.org/dc/terms/> .

<http://example.com/asset:9898> dct:license <http://example.com/policy:0099> .
<http://example.com/policy:0099>
    a odrl:Set;
    odrl:permission odrl:reproduce ;
    odrl:prohibition odrl:modify.
```

7.2 General metadata for process/workflow centric research objects in Earth Science

The metadata implementation guidelines propose a set of annotations, properties and relations in the research object paradigm in order to enable user communities from Earth Science disciplines to specify and model research objects and the corresponding resources aggregated within. In the following sections several tables are provided with metadata descriptions of the basic information necessary to define research objects and workflows for Earth Science. This documentation extends the original guidelines produced in the Wf4Ever project, which was focused on experimental sciences.

7.2.1 General properties

Table 9: Metadata Guidelines for General Properties

Description	Property	Domain	Range
Title	dc:title	Any resource	rdfs:Literal
Description	dc:description	Any resource	---
RO creation date	dct:created	Any resource	rdfs:Literal
RO last modification date	dct:modified	Any resource	rdfs:Literal
Comments/notes	rdfs:comment	Any resource	rdfs:Literal
Purpose	roterms:performsTask	Any resource	skos:Concept
RO creator[32]	• dc:creator ○ pav:createdBy^	Any resource	foaf:Agent
RO contributor[33]	• dc:contributor	Any resource	foaf:Agent
Author of RO content^	pav:authoredBy^	Any resource	foaf:Agent

[32] Creator of the research resource. This is automatically added by rohub. Consider also to differentiate rohub creator or uploader (pav:importedBy or pav:providedBy) from creator of digital resource and content author.

[33] Anybody contributing to the research object, e.g., add/update/remove resources. This is automatically added.

Contributor to RO content^	pav:contributedBy^	Any resource	foaf:Agent
Publisher	dc:publisher	Any resource	foaf:Agent
Field/domain	dc:subject	Any resource	selected vocabulary, e.g., LCSH
Property rights[34] - see sec2	dc:rights	Any resource	----Rights Statement
License – see sec2	dct:license (refines rights)	Any resource	---License Doc (URL)
Cited By[35]	dct:isReferencedBy[36]	Any resource	----
Cites[37]	dct:references	Any resource	----
Tags/keywords^	swrc:keywords	Any resource	---
Technical contact	roterms:technicalContact	Any resource	foaf:Agent
General documentation^	cito:isDocumentedBy	Any resource	URI
Format	dc:format	Any resource	Internet Media Types [MIME]
Depiction^	foaf:depiction	Any resource	foaf:Image (equiv wf4ever:Image)
RO release/archive creator	roevo:wasArchivedBy	RO	prov:Agent
Aggregated Resource type	rdf:type	Agg. Resource	rdf:Class (see predefined values below)
Resource name (human-readable)	rdfs:label^	Any resource	rdfs:Literal
Resource size	dbo:filesize^	Any resource	xsd:double - size (B)
Resource checksum	ov:hasChecksum^	Any resource	rdfs:Literal - eg ov:hasChecksum"sha1:a5..."

7.2.2 Predefined resource types

Resource types should be recommended automatically whenever is possible. The main types are described below.

Table 10: Metadata Guidelines for Predefined Resource Types

rdf:Class	Label	Comment
***wfdesc:Process** ****wfdesc:Workflow** *****wfdesc:WorkflowInstance^**	• Process • Workflow • Workflow Instance	• Process is used to describe a class of actions that when enacted give rise to processes. A process can have 0 or more wfdesc:Parameter instances associated using wfdesc:hasInput and wfdesc:hasOutput • Workflow is a directed graph in which the nodes are wfdesc:Process instances and the edges

[34] https://theodi.org/guides/publishers-guide-to-the-open-data-rights-statement-vocabulary

[35] Automatically added after fork, or from API google scholar.

[36] Considered also cito properties but deemed not necessary (cito:isCitedBy^, cito:cites^)

[37] Automatically added after fork

		(wfdesc:DataLink instances) represent data dependencies between the constituent wfdesc:Process descriptions. • WorkflowInstance is a specialisation of a wfdesc:Workflow description which defines all data/parameters/settings that are required to form a wfprov:WorkflowRun. I.e. it is a workflow description ready to run
***wfdesc:ProcessImplementation^** ****wfdesc:WorkflowDefinition^** ****wf4ever:CommandLineTool^** ****wf4ever:Script^** *****wf4ever:BeanShellScript^** *****wf4ever:PythonScript^** *****wf4ever:RScript^** ****wf4ever:WebService^** *****wf4ever:RESTService^** *****wf4ever:SOAPService^**	• Process Implementation • Workflow Definition Command Line Tool • Script • Web Service	• ProcessImplementation is the implementation that is invoked when running the wfdesc:Process, like a wf4ever:Script or wf4ever:WebService. The process specifies the implementation via wfdesc:hasImplementation. • WorkflowDefinition is the resource that contains the workflow definition (say the XML file saved from a workflow designer tool), and can be specified with wfdesc:hasWorkflowDefinition • A command line tool, typically executed locally, through SSH or a job submission system. • A script which can be evaluated • A (typically remote) Web Service, like REST or SOAP. For the purpose of being a wfdesc:ProcessImplementation of a wfdesc:Process, a 'service' is considered more like a specific method then an endpoint.
***wfprov:Artifact** ****wf4ever:Dataset** ****wf4ever:Document** ****wf4ever:File** ****wf4ever:Image**	• Artifact • Dataset • Document • File • Image	• An artifact is a data value or item which wfprov:wasOutputFrom of a wfprov:ProcessRun or that the process run used as input (wfprov:usedInput). Such an artifact might also be a ro:Resource if it has been aggregated in the ro:ResearchObject (typically if the artifact was used or generated by a wfprov:WorkflowRun) - but this might always not be the case for intermediate values from wfprov:ProcessRun. • A dataset or reference database. • A human readable document, like a HTML page, a PDF, plain text file or a LibreOffice document. • A file wfprov:Artifact that was used or generated by a wfprov:WorkflowRun. A wf4ever:File can specify a (possibly relative) wf4ever:filePath. • A visual image or photograph, formats like JPEG, PNG, TIFF
***wfprov:ProcessRun^** ****wfprov:WorkflowRun** *****roterms:ExampleRun** *****roterms:PropectiveRun** *****roterms:ResultsGenerationRun**	• Process Run • Workflow Run • Example Workflow Run • Prospective Workflow Run	• A process run is a particular execution of a wfdesc:Process description (wfprov:describedByProcess), which can wfprov:usedInput some wfprov:Artifact instances, and produce new artifacts. • A workflow run is a process run which has been enacted by a workflow engine, according to a

	• Results Generating Workflow Run	workflow definition. Such a process typically contains several subprocesses corresponding to process descriptions. • A workflow run that serves as an example of how to use this workflow. Example runs typically take only a small subset of inputs and have short execution time. • A workflow run that is ready to start executing, e.g. all workflow inputs and configuration options have been provided, but no outputs are available yet. • A workflow run that generated scientific results. Such workflow runs typically take complete data sets as inputs and may take longer to execute.
prov:Entity** ****roterms:Sketch** * roterms:Hypothesis** **** roterms:ResearchQuestion** ****roterms:Conclusion** **** roterms:Result** **** roterms:Paper** ****roterms:MeetingMinutes^** ****roterms:ProcessValue^**	• Sketch • Hypothesis • Research Question • Conclusion • Result • Paper • Meeting Minutes • Process Value	• A graphical representation of an experiment/investigation • A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation • The methodological point of departure of scholarly research in both the natural and social sciences. It determines (i) where and what kind of research the writer will be looking for and (ii) the specific objectives the study or paper will address • A proposition reached from given premises • An item of information obtained by experiment or some other scientific method; a quantity or formula obtained by calculation • A bibliographic resource about an experiment/investigation (to be replaced by dct:BibliographicResource) • Electronic version of the minutes of a meeting. • A (potential) process (e.g., workflow) value. The value is a piece of data that was used or generated by a wfprov:ProcessRun (e.g., wfprov:WorkflowRun) (as a wfprov:Artifact). It can also be used to specify a value that is suggested for a wfdesc:Parameter given by roterms:exampleValue or roterms:defaultValue.
dct:BibliographicResource	• Bibliographic Resource	• A book, article, or other documentary resource.
wf4ever:WorkflowResearchObject^ **roes:DataResearchObject^** **roes:ServiceResearchObject^** **roes:BibliographicResearchObject^** **roes:MinutesResearchObject^**		• A research object that aggregates at least one workflow description.

7.2.3 Predefined relations

Table 11: Metadata Guidelines for Predefined Relations

Relation	Label	Domain	Range	Description
***prov:wasDerivedFrom** ****prov:wasRevisionOf** ****prov:wasQuotedFrom** ****prov:hadPrimarySource**	• wasDerived From • wasRevisionOf • wasQuoted From • hadPrimary Source	Any resource (prov:Entity)	Any resource (prov:Entity)	Evolution relations in roevo (see PROV). A derivation is a transformation of an entity into another, an update of an entity resulting in a new one, or the construction of a new entity based on a pre-existing entity.
wfprov:usedInput	Input used	wfprov:ProcessRun	wfprov:Artifact	This property specifies that a wfprov:ProcessRun used an wfprov:Artifact as an input.
wfprov:wasOutputFrom	Output produced	wfprov:Artifact	wfprov:ProcessRun	This property specifies that a wfprov:Artifact was generated as an output from a wfprov:ProcessRun.
roterms:defaultValue^[38] **(some Process, e.g., wf)** **wfdesc:hasInput** **(Parameter1)** **(Parameter1)** **roterms:defaultValue (some ProcessValue)**	Default value	wfdesc:Process (e.g., wf)	roterms:ProcessValue e.g., Resource URI (wfprov:Artifact), text (cnt:chars), Base64 encoded bytes (cnt:bytes) or a collection of these (co:Collection)	The default value for a wfdesc:Input or wfdesc:Configuration. It is the value that is inherently assumed if no value is provided by the process (e.g., workflow) run.
roterms:exampleValue^[39] **(some Process, e.g., wf)** **wfdesc:hasInput** **(Parameter1)** **(Parameter1)**	Example Input value	wfdesc:Process (e.g., wf)	roterms:WorkflowValue e.g., Resoure URI (wfprov:Artifact), text	An example value for a given input or output parameter.

[38] This property has been generalised in roterms. In the generic specification, it was for workflow only.

[39] This property has been generalised in roterms. In the generic specification, it was for workflow only.

roterms:exampleValue (some ProcesssValue)			(cnt:chars), Base64 encoded bytes (cnt:bytes) or a collection of these (co:Collection)	
wfdesc:hasSubProcess^	has sub-process	wfdesc:Workflow	wfdesc:Process	specify that the given workflow contains the given process as part of its definition.
wfdesc:hasSubWorkflow	has sub-workflow	wfdesc:Workflow	wfdesc:Workflow	associates a workflow desc to another, specifying that the first one has the given sub-workflow as a contained process
***roterms:previousProcess^[40]** ****roterms:previousWorkflow**	• Previous Process • Previous workflow	• wfdesc:Process • wfdesc:Workflow	• wfdesc:Process • wfdesc:workflow	• Previous process that is or should be executed • Previous workflow that is or should be executed
***roterms:subsequentProcess^[41]** ****roterms: subsequentWorkflow**	• Subsequent Process • Subsequent workflow	• wfdesc:Process • wfdesc:Workflow	• wfdesc:Process • wfdesc:Workflow	• Subsequent proces that is or should be executed • Subsequent workflow is or should be executed
***wfprov:describeByProcess^** ****wfprov:describedByWorkflow**	• described by process • described by workflow	• wfprov:ProcessRun • wfprov:WorkflowRun	• wfdesc:Process • wfdesc:Workflow	• associates a wfprov:ProcessRun to its wfdesc:Process description • associates a wfprov:WorkflowRun to its corresponding wfdesc:Workflow description
***wf4ever:serviceURI^** ****wf4ever:rootURI** ****wf4ever: wsdlURI^**	• Service URI • Root URI • WSDL URI	• wf4ever:WebService • -- • wf4ever: SOAPService	• xsd:anyURI • -- • --	• The URI of a web service • The root URI of a web service • The URI of the WSDL (note:used by taverna)

[40] New super property added for generalisation.

[41] New super property added for generalisation.

wf4ever:command^	Command	wf4ever:CommandLineTool	xsd:string	The executed command, for instance "cp /tmp/file1 /tmp/file2", "/usr/local/bin/analyzestuff" or "C:\Program.exe".
wf4ever:script^	script	wf4ever:Script	xsd:String	scripts of small sizes can be embedded in Script description using wf4ever:script.
wf4ever:filepath^	File path	wf4ever:File	xsd:string	The (possibly) relative filepath of a wf4ever:File artifact that was used or generated by a wfprov:WorkflowRun.
wf4ever:parameterFilePath^	Parameter file path	wf4ever:FileParameter	xsd:string	A (possibly relative) filepath used as a fixed input or output parameter for a workflow.

7.2.4 Workflow (process) specific classes

Table 12 Workflow Specific Concepts

rdf:Class	Label	Comment
***wfdesc:Parameter** ****wfdesc:Input** ****wfdesc:Output** ****wfdesc:Configuration** ****wf4ever:FileParameter**	• Parameter • Input • Output • Configuration • File Parameter	• A parameter of a wfdesc:Process. • Input parameter to a wfdesc:Process. This can be compared to a function parameter, command line argument, files read, or parameter set by a user interface • Output parameter from a wfdesc:Process. This can be compared to functional return values, stdout/stdin, files written, or results shown in a user interface. • parameter to a wfdesc:Process (linked using wfdesc:hasConfiguration) to indicate a setting, flag or customization which somewhat modifies or controls the behaviour of the wfdesc:Process • A parameter (typically an Input or Output) which is read/written from a fixed (possibly relative) file path, specified with wf4ever:parameterFilePath

7.2.5 Workflow (process) specific properties

Table 13: Metadata Guidelines for Workflow Specific Properties

Property	Label	Domain	Range	Description
wfdesc:hasWorkflowDefinition	has workflow definition	wfdesc:Workflow	wfdesc:WorkflowDefinition	The definition of this workflow, typically a file natively understood by the wfdesc:WorkflowEngine.

				Note: used by taverna and checklist
wfdesc:hasImplementation	has implementation	wfdesc:process	wfdesc:ProcessImplementation	The implementation that is invoked when running the wfdesc:Process
wfdesc:hasInput	Has Input	wfdesc:Process	wfdesc:Input	the wfdesc:Input parameter of a given wfdesc:Process.
wfdesc:hasOutput	Has Output	wfdesc:Process	wfdesc:Ouput	The wfdesc:Output parameter of a given wfdesc:Process.
wfdesc:hasConfiguration	Has configuration	wfdesc:Process	Wfdesc:Configuration	wfdesc:Configuration parameter of a given wfdesc:Process.
wfprov:wasInitiatedBy	Was initiated by	wfprov:WorkflowRun	foaf:Agent (prov:Person or prov:SoftwareAgent)	The agent who decided to run the workflow
wfprov: wasEnactedBy	Was enacted by	wfprov:ProcessRun	wfprov: WorkflowEngine	specifies that the execution of the process was enacted by the engine
wfprov:durationInSeconds	Duration in seconds	wfprov:WorkflowRun	xsd:double	How long it took to run the workflow (in seconds)

8 Conclusion

The main objective of deliverable D4.2-Workflows and Research Objects models in Earth Science is to provide Earth Science disciplines with a dedicated set of models and vocabularies for the representation of research objects in such specific disciplines. This document presents the extensions, customizations and newly developed models that have been carried out in the context of the EVER-EST project. The work presented herein pushes the boundaries of the generic models available for research object representation, significantly contributing to the extension and improvement of the current standard specifications. Furthermore, a new branch has been created specifically for the Earth Science disciplines, which provides specialized modeling support for research objects in the scientific areas of interest.

In addition to the models resulting from this work, this document provides an overview of the methodology that was followed, including intensive analysis and joint work with the EVER-EST user communities and technical partners. The work described herein aims at i) maximizing the reuse of existing terminology and metadata used in Earth Science and related disciplines and ii) ensuring the compliance of the newly introduced terms with existing standards and potential future extensions both by EVER-EST consortium partners and by the broader communities. In doing so, it is sought to facilitate wide-spread adoption and the creation of synergies for the evolution of the models. To this respect, the present specification is meant to be a live entity, which will evolve beyond this document as the various stakeholders use it and contribute to it. As part of this strategy there is continuous collaboration with other members of the Semantic technologies and Linked Data community working with research objects as well as with the EVER-EST VRCs and related user communities.

Special emphasis has been placed in leveraging the familiarity of the Earth Science community with relevant existing resources and specifications in their field as well as standards for representing information in a machine readable way in order to enhance representational capabilities. Following such spirit another outcome of this deliverable is a number of guidelines for metadata implementation related to the different annotations in Earth Science research objects and associated resources.

9 References

[1] Belhajjame K, Zhao J, Garijo D, Hettne K, Palma R, Mina E, Corcho O, Gómez-Pérez JM, Bechhofer S, Klyne G, Goble G. Using a suite of ontologies for preserving workflow-centric research objects. Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 32, May 2015, pp 16-42, ISSN 1570-8268, http://dx.doi.org/10.1016/j.websem.2015.01.003

[2] Gomez-Perez JM, Alexopoulos P, García N, Palma R. Workflows and research objects in Earth Science - Main concepts and definitions. February 2016. Deliverable D4.1, project EVER-EST. http://www.ever-est.eu/deliverables/D4-1WorkflowsAndResearchObjects.html

[3] LTDP Team (EOP-GT). EO SIP-Scene Specification for SEASAT. LDPE-GSEG-EOPG-TN-14-0005. October 2014. European Space Agency (ESA).

[4] Marelli F, Foglini F, Albani S, Napier H, Trassatti E, and the EVER-EST consortium. VRC User Actions. May 2016. Technical Report, project EVER-EST.