Materials of Towards Human-Machine Scientific Partnership Based on Semantically Rich Research Objects

Jose Manuel Gomez Perez *null Palma* *Raul Palma* *Andres Garcia-Silva*

***Abstract*—Materials that support the developments, experiments, surveys presented in the paper "Towards Human-Machine Scientific Partnership Based on Semantically Rich Research Objects" submitted for review in escience 2017 conference. The materials include: ¡br¿ - Link to the research object vocabularies ¡br¿ - Link to the research object model extension for Earth Science ¡br¿ - Reference to the research object used as example in the paper ¡br¿ - EVER-EST project deliverables referenced in the paper ¡br¿ - Link to the semantic analysis platform used in the paper (Cogito) ¡br¿ - Link to the vocabulary used to describe the semantic annotations (contentdesc) ¡br¿ - Results of the survey to assess the semantic enrichment ¡br¿ - Link to the web page: Research Object Components ¡br¿ - Link to the research object recommender system ¡br¿ - Java code to reproduce the similarity evaluation reported in the paper. ¡br¿ - Link to the solr index containing the dataset that we used to perform the similarity experiments. ¡br¿ - Excel file with the similarity evaluation data collected in our experiments. ¡br¿ - Bib tex file containing all the bibliographic entries used in the paper.**

***Index Terms*—Research Objects; Earth Science; Recommender System; Knowledge Sharing and Resue; Knowledge preservation; Semantic Enrichment**

I. Introduction

This document provides a paper-style view of the Research Object (RO) "*Materials of Towards Human-Machine Scientific Partnership Based on Semantically Rich Research Objects*"[1] generated . The RO has been created, managed and preserved via ROHub platform [1]. Please refer to [2] for a general introduction to the RO concept, to [3] for a detailed description of the RO model, and to [4] for more information about ROHub platform.

The RO is of type *"Basic", which represents a general aggregation of related resources.* [2].

An overview of this RO is depicted in Figure 1. In summary, the hypothesis of this work is described in *hypothesis.txt*[3]; Additionally, this RO has been enriched automatically with the following annotations:

- concepts (most frequently mentioned in the RO): *methodology, metadata, Earth Science, Objects, ONTOLOGY, Research*
- domains (fields of knowledge in which the main concepts are commonly used): *statistics, literature*
- frequent expressions (most frequently mentioned noun phrases): *Supersites VRC Research Object content structure, comparator requirements question, evolution ontology, research object model, VRC Research Object, appendix H Supersites VRC Research Object, requirements questionnaire, exemplar RO schedule, Research Object metadata, content structure, B sea monitoring VRC, appendix E geohazards Supersites VRC*
- named entities (most frequently mentioned):
  - Places: *Everest, Spain*

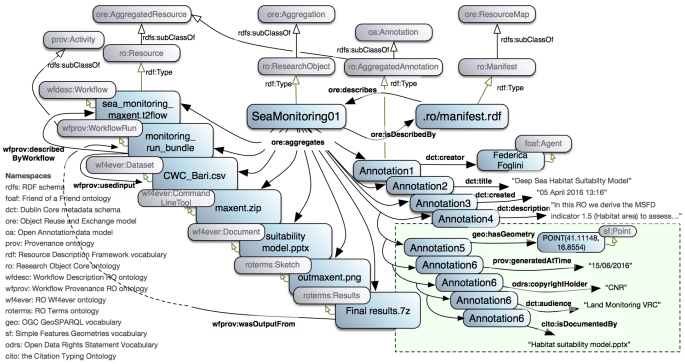


Fig. 1. Research Object Sketch

II. Resources

The resources encapsulated by the RO are summarized in table I

A. Key Resource details

- Name: *illustrationRO1.png*
  Description: *The Figure shows a graphical representation of an existing research object (http://sandbox.rohub.org/rodl/ROs/SeaMonitoring01/) that uses the core vocabulary and earth science specific metadata. ¡br¿ This research object uses several modules of the research object ontology suite. It contains a habitat suitability model to derive the Marine Strategy Framework Directive indicator 1.5 (habitat area), assessing a descriptor of biological diversity. ¡br¿*

[1] *http://sandbox.rohub.org/rodl/ROs/experiences-escience-2017/*

[2] See RO types definitions at http://w3id.org/ro/earth-science#

[3] http://sandbox.rohub.org/rodl/ROs/experiences-escience-2017/hypothesis.txt

TABLE I
RESEARCH OBJECT RESOURCES

name	size	type
illustrationRO1.png	1.1 MB	Sketch
EVER-EST_DEL_WP4-D4.1.pdf	4.0 MB	BibliographicResource
EVER-EST DEL WP4-D4.2.pdf	2.0 MB	BibliographicResource
biblio.bib	14.1 KB	BibliographicResource
SemanticEnrichmentSurveyResults.xlsx	30.6 KB	Results
SimilarityEvaluation.xlsx	81.4 KB	Results
everest-experiments-escience.zip	2.9 MB	File
hypothesis.txt	288.0 B	Hypothesis
conclusions.txt	3.0 KB	Conclusions
preprint.pdf	1.4 MB	Paper

The research object encapsulates a scientific workflow, the input dataset, provenance information about the execution of the workflow, the output dataset, ancillary documentation such as images and presentations, and information regarding the author, plus metadata about the research object evolution and quality checks.

- Name: *EVER-EST_DEL_WP4-D4.1.pdf*
  Description: *This document introduces the Research Object (RO) concept into the Earth Science Virtual Research Communities (VRCs) of the EVER-EST project. An overview of the main concepts, tecnologies and models related to research objects is provided along their contextualization and extension under the light of a set of initial requirements elicited from the VRCs about the representation of such objects, their lifecycle management, and exploitation through the EVER-EST VRE. This work has been conducted in close collaboration with WP3 and WP5 partners, and establishes a link between the requirements of the user communities and the technological solutions provided by the VRE infrastructure around the notion of research objects, respectively. The outcomes of this analysis provides the basis for the implementation of the subsequent tasks of WP4 and are expected to evolve over time during the rest of the project, taking into account the continuous software development and integration model that is applied.*
- Name: *EVER-EST DEL WP4-D4.2.pdf*
  Description: *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.*
- Name: *everest-experiments-escience.zip*
  Description: *Maven project that allows to run and produce the evaluation data reported in the paper. ¡br¿ To compile the code first unzip the file and run in side the folder containing the pom.xml file the following instruction: ¡br¿ maven install ¡br¿ then you can run the Classes: ¡br¿ org.everest.experiments.escience.evaluation.OneResInContextEvaluati ¡br¿ org.everest.experiments.escience.evaluation.TwoResInContextEva ¡br¿ the first class runs the evaluation for the similarity using only one research object, while the second class runs the evaluation when similarity is calculated based on two research objects. ¡br¿ Evaluation data is saved in the log file within the /log folder.*
- Name: *preprint.pdf*
  Description: *This the paper version submitted to the escience conference before being reviewed.*

ACKNOWLEDGMENT

The Research Object was uploaded to ROHub by *Raul Palma*. ROHub portal development was supported by EVER-EST EU project (HORIZON 2020 grant 674907).

REFERENCES

[1] The Research Object Management Platform - ROHub http://www.rohub.org/.

[2] K. Belhajjame, O. Corcho, D. Garijo, J. Zhao, P. Missier, D. Newman, R. Palma, S. Bechhofer, E. García Cuesta, J. M. Gómez-Pérez, S. Soiland-Reyes, L. Verdes-Montenegro, D. De Roure, and C. Goble "Workflow-Centric Research Objects: First Class Citizens in Scholarly Discourse", Proceedings of Workshop on the Semantic Publishing, SePublica Crete, Greece 28 May 2012.

[3] Belhajjame K., Zhao J., Garijo D., Gamble M., Hettne K., Palma R., Mina E., Corcho O., Gómez-Pérez J. M., Bechhofer S., Klyne G., Goble C. "Using a suite of ontologies for preserving workflow-centric research objects", Journal of Web Semantics: Science, Services and Agents on the World Wide Web Available online 11 February 2015 ISSN 1570-8268.

[4] Palma R., Corcho O., Gómez-Pérez J. M., Mazurek, C. "ROHub - A Digital Library of Research Objects Supporting Scientists Towards Reproducible Science". In Semantic Publishing Challenge of Proc. Extended Semantic Web Conference (ESWC) Crete, Greece 25-29 May 2014