

ROHub— A Digital Library for Sharing and Preserving Research Objects

Raúl Palma¹, Piotr Hohubowicz², Oscar Corcho², José Manuel Gómez-Pérez³, Cezary Mazurek¹ and the Wf4Ever consortium



¹ Poznan Supercomputing and Networking Center, Poznan, Poland

² Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain

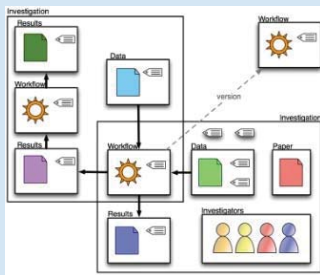
³ ISOCO, Madrid, Spain

URL	http://www.rohub.org/	http://www.rohub.org/rodl/
Demo video	http://youtu.be/TxW2wvreyoQ	http://youtu.be/gSEUswMmr8E
Source code	https://github.com/wf4ever/rodl	https://github.com/wf4ever/portal

ABSTRACT

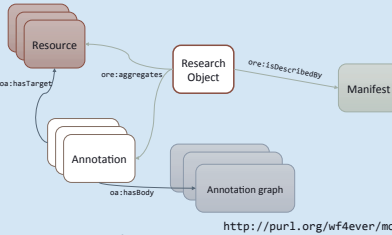
Research Objects (ROs) are semantic aggregations of related scientific resources, their **annotations** and research context. They provide the means to refer a **bundle** of research artifacts supporting an **investigation**, and the mechanisms to associate human and **machine-readable metadata** to these artefacts.

ROHub is a digital library system for ROs that supports their **storage, lifecycle management and preservation**. ROHub enables the **sharing of scientific findings** via ROs and includes features that help scientists throughout the research lifecycle to create and maintain **high-quality ROs** that can be **interpreted and reproduced** in the future.



RESEARCH OBJECT MODEL

A research object (RO) is described in an RDF **manifest** which lists the **aggregated resources** and their **annotations** as separate RDF graphs containing user annotations (*title, description, example value*), typing information (*hypothesis, workflow, input data, etc*) and automatically extracted metadata (*provenance, workflow structure*).



The RO model consists of a core ontology and several extensions

- **ro-core** provides the basic structure for the description of aggregated resources and annotations on those resources
- **ro-ev** captures the RO evolution, including the different stages during their lifecycle, the corresponding version along with their associated changes.
- **wf-desc** provides the vocabulary for the description of workflows
- **wf-prov** provides the vocabulary for the description of workflow execution provenance

The ontologies for the RO Model are based on standards for aggregations (**OAI-ORE**), annotations (Annotation Ontology, W3C Open Annotation Core **OAC**) and provenance (W3C **PROV** ontology).

ROHUB INTERFACES

ROHub implements a set of open REST APIs, being the two primary ones:

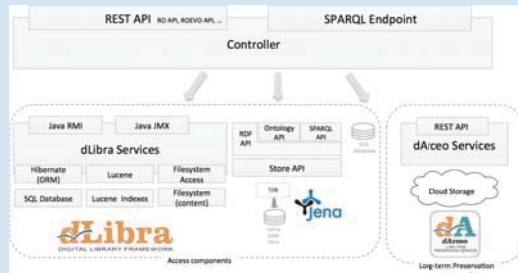
- **RO API** - defines the formats and links used to create and maintain ROs in the digital library, according to RO model
- **RO Evolution API** - defines the formats and links used to change the lifecycle stage of a RO, as well as to retrieve their evolution provenance

ROHub also exposes a Notification API, User Management API, Access Control API, Solr API, and provides a SPARQL endpoint and a Web Interface

IMPLEMENTATION

ROHub has a **modular structure** that comprises **access components**, **long-term preservation components** and the **controller** that manages the flow of data. ROs are stored in the access repository once created, and periodically the new and/or modified ROs are pushed to the long-term **preservation repository**.

The storage backend can be based on **dLibra** (as shown below), or it can use a built-in module for storing ROs directly in the **filesystem**.

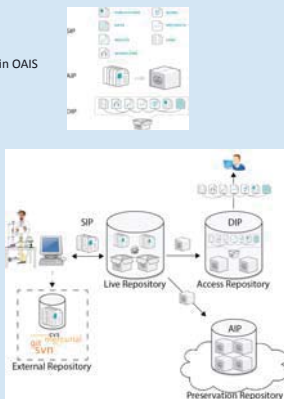


KEY FUNCTIONALITIES

- Create, manage and share ROs:** ROHub provides different methods for creating ROs and different **access modes** to share them: open, public or private. ROHub provides a **faceted search** interface, in addition to a keyword search box.
- Assessing RO quality:** In the Overview panel, ROHub shows a **progress bar** of the RO quality based on set of predefined basic RO requirements. Additional **quality information** can be visualized in the Quality panel.
- Managing RO evolution:** Users can create **RO snapshots** at any point in time, and release and **preserve** the RO when the research has concluded. Users can visualize the **evolution of the RO** from the History panel.
- Navigation of workflow run:** Scientists can **aggregate** any type of resource, including **links to external resources** and **RO bundles**, which are structured ZIP files representing self-contained ROs that facilitate their transfer and integration with 3rd party tools (e.g., Taverna)
- Monitoring ROs:** ROHub includes monitoring features, such as fixity checking and **RO quality**, which generate **notifications** when changes are detected (e.g., workflow decay). Users can visualize this information in the Notification panel and they can subscribe to the **atom feed**.

Preservation according to OAIS model

RO as Information Packages in OAIS



The archival system establishes three different storage repositories (or logical repositories) to support the preservation and access to preserved ROs



This work was enabled by the Wf4Ever project funded by the European Commission's 7th FWP (FP7-ICT-2007-6 270192), and partially by the Polish National Centre for Research and Development (NCBiR) under Grant No. SP/I/1/77065/10