Research Objects

Kevin Page
Daniel Garijo

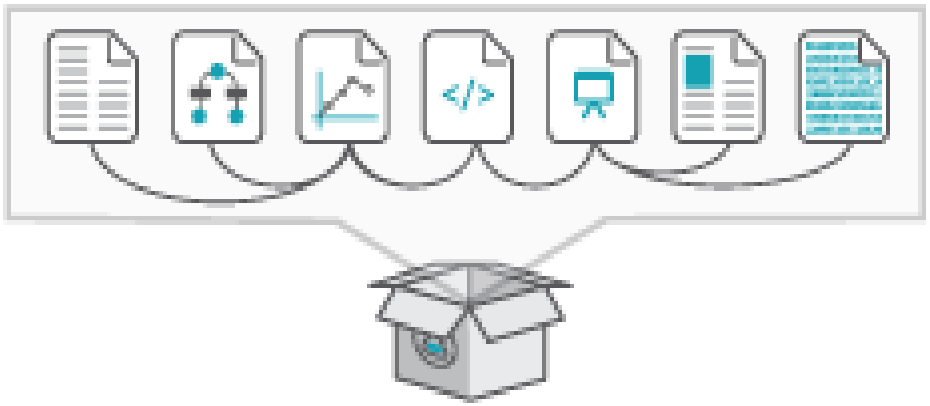
From Preserving Data to Preserving Research: Curation of Process and Context

iPRES tutorial, 2nd September 2013

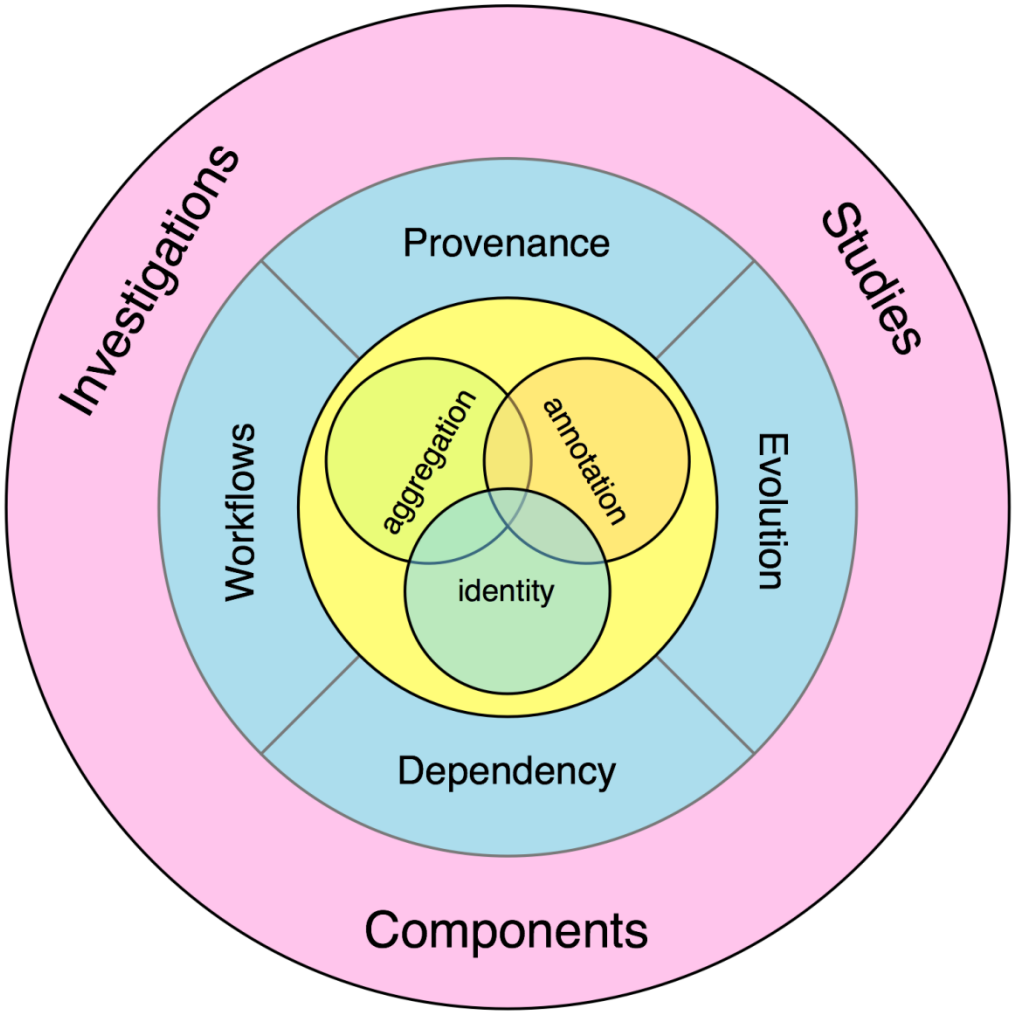
What is a Research Object?

•Aggregation of resources that bundles together the contents of a research work:

- •Data
- •Experiments
- •Examples
- •Bibliography
- •Annotations
- •Provenance
- •ROs
- •etc.



RO at 5000 feet



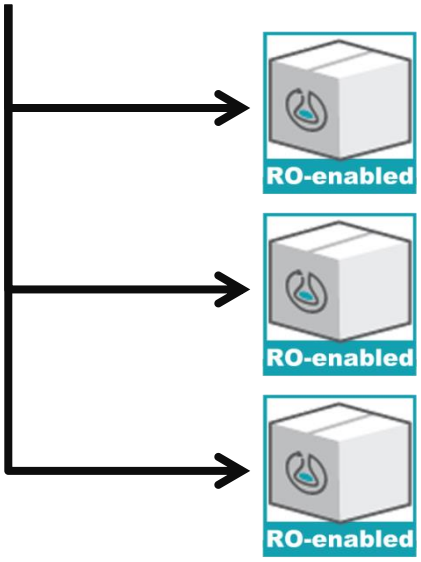
Why Research Objects?

- **Process preservation:** URIs/dois for referencing resources.
- **Reusability** of any part of the RO
- **Repeatability /Reproducibility:** redeployment of the method
- **Traceability** and error detection.

- **Attribution**: able to cite data and publications of the RO
- **Understandability:** Links between data, results and annotations.
- **Curation:** by explicitly exposing the methods of the experiment.

Examples of Research Objects

ROs propose a flexible structure that can be extended

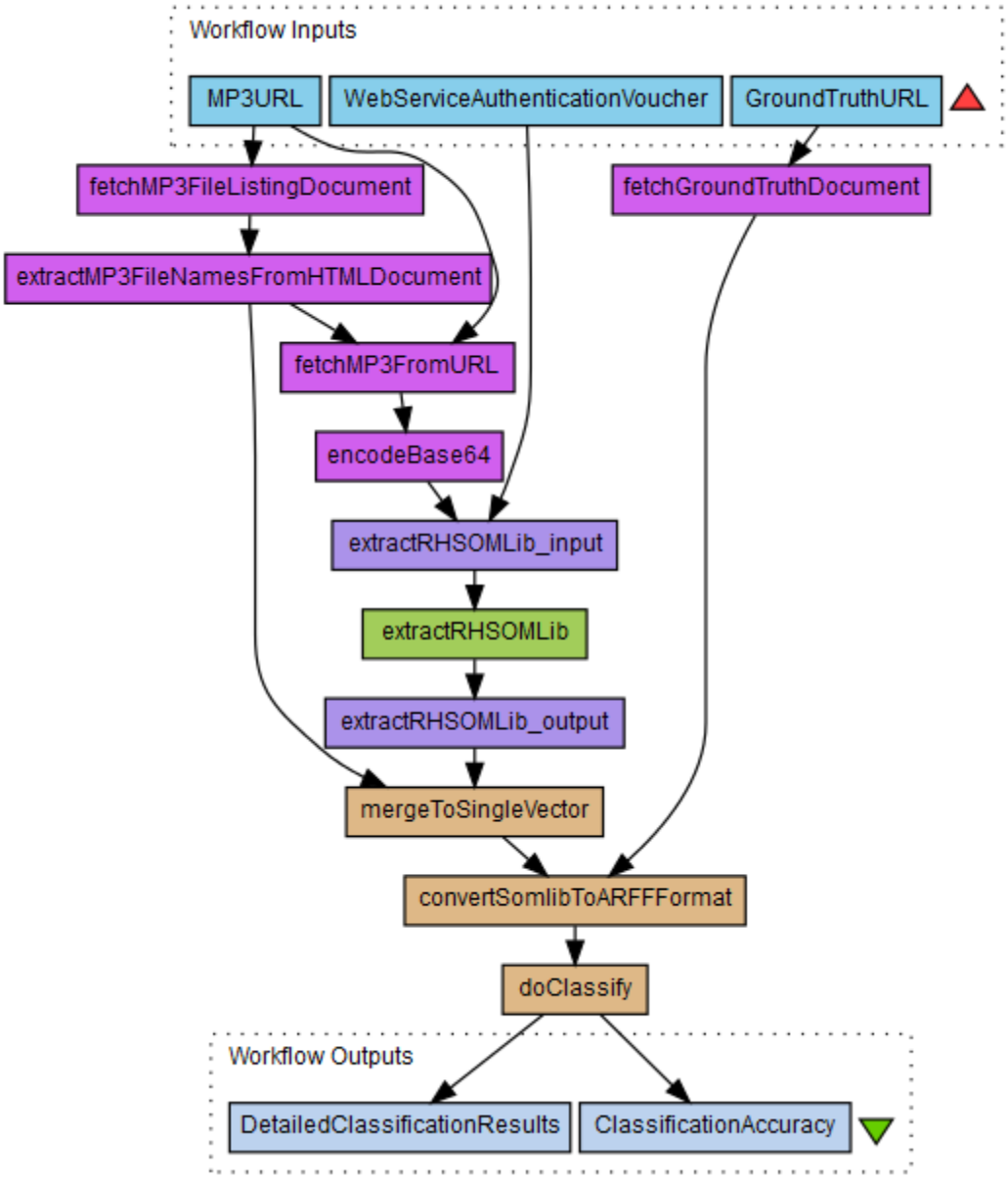


Experiments: inputs, outputs, methods, schemas, hypothesis, etc.

Papers: Bibliography, Sections, Charts, Tables, Images, etc.

Web pages: Data, images, people, etc.

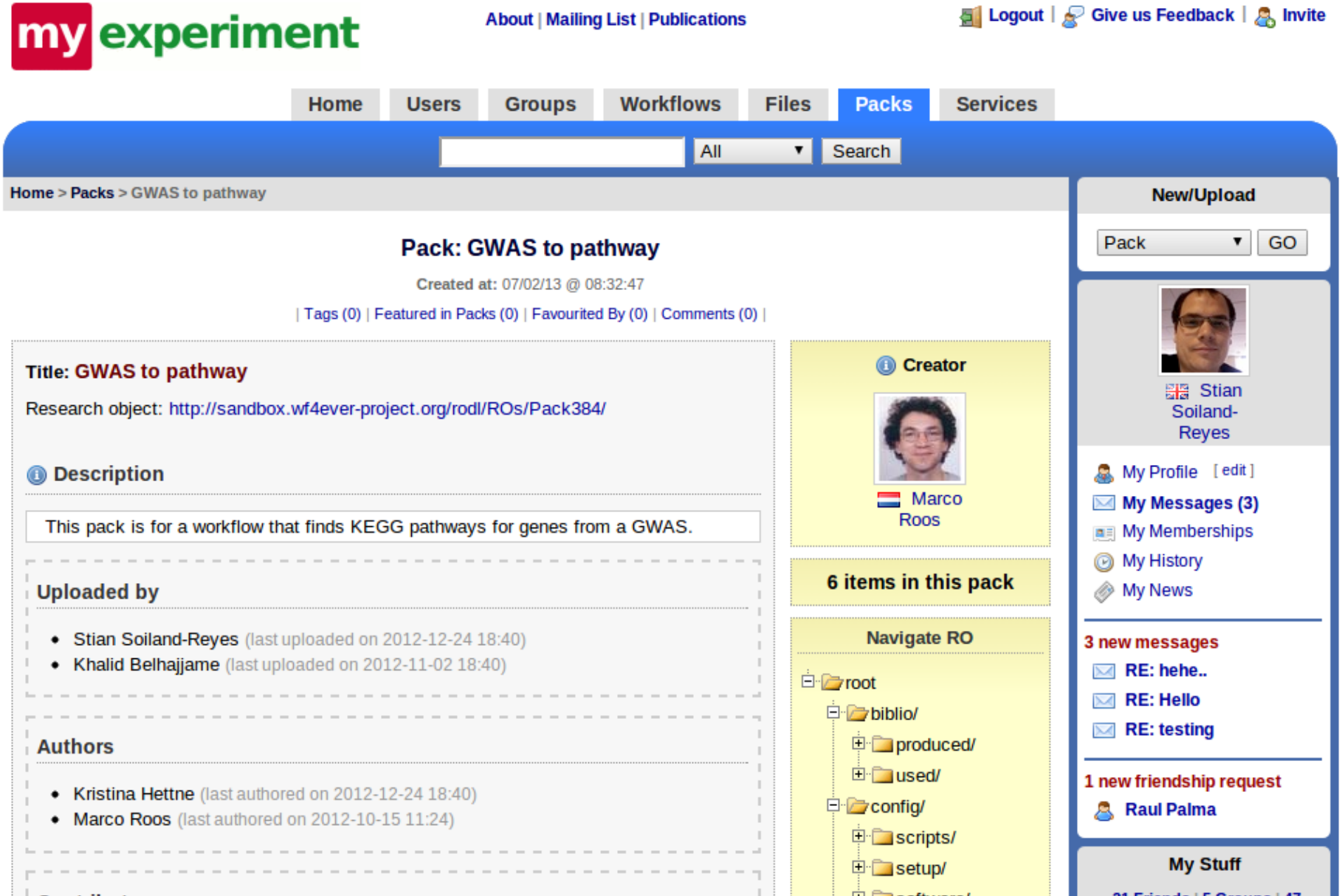
Workflows



- Digital instrument that allows scientists to represent computational and data manipulation steps.

- Coordinates the execution and links the resources together.

ROs: A real example (1)



ROs: A real example (2)

WORKFLOW MOTIF ANALYSIS AND CATALOGUE
ABSTRACT
INPUTS AND EXAMPLES
RESULTS
ABOUT THE AUTHORS

COMMON MOTIFS IN SCIENTIFIC WORKFLOWS: AN EMPIRICAL ANALYSIS

This page represents a bundle for the contents of the analysis currently submited to the Future Generation Computer Systems Journal. The analysis is an expansion of a paper published in eScience 2012 (link to the paper), and it is currently under review. The purpose of this web page is to make accessible, link and describe the inputs and outputs of the analysis, which are stored as a **Research Object (pack) in myExperiment**.

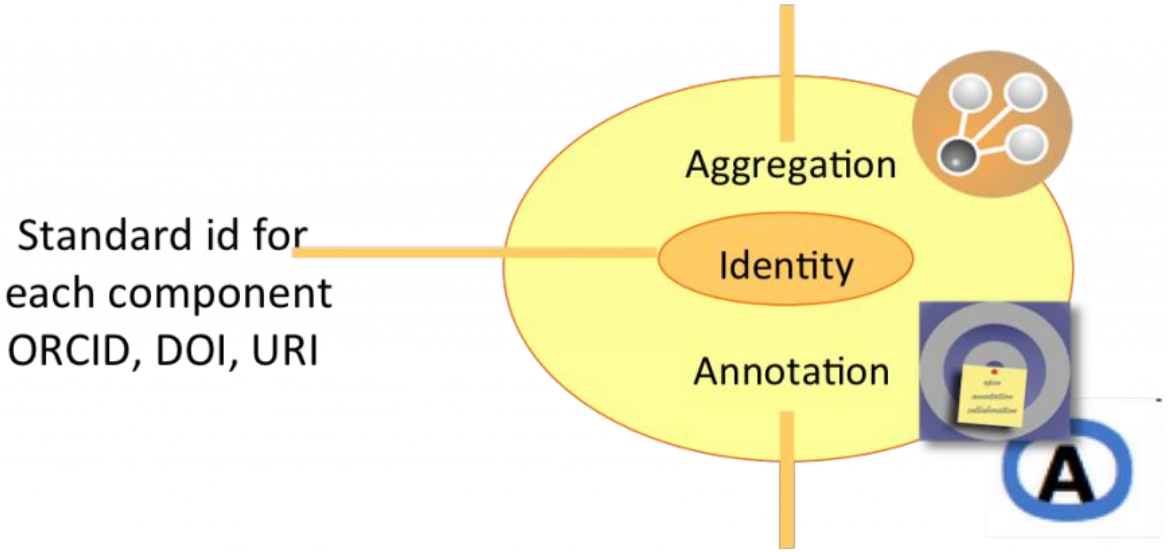
ABSTRACT

Workflow technology continues to play an important role as a means for specifying and enacting computational experiments in modern Science. Reusing and repurposing workflows allow scientists to do new experiments faster, since the workflows capture useful expertise from others. As workflow libraries grow, scientists face the challenge of finding workflows appropriate for their task, understanding what each workflow does, and reusing relevant portions of a given workflow. We believe that workflows would be easier to understand and reuse if high-level views (abstractions) of their activities were available in workflow libraries. As a first step towards obtaining these abstractions, we report in this paper on the results of a manual analysis performed over a set of real-world scientific workflows from Taverna, Wings, Galaxy and Vistrails. Our analysis has resulted in a set of *scientific workflow motifs* that outline i) the kinds of data-intensive activities that are observed in workflows (*data-oriented motifs*), and ii) the different manners in which activities are implemented within workflows (*workflow-oriented motifs*). These motifs are helpful to identify the functionality of the steps in a given workflow, to develop best practices for workflow design, and to develop approaches for automated generation of workflow abstractions.

INPUTS AND EXAMPLES OF THE ANALYSIS

Research Objects: An Overview

OAI-ORE
http://www.openarchives.org/ore/
Structuring and Bundling descriptions and components.

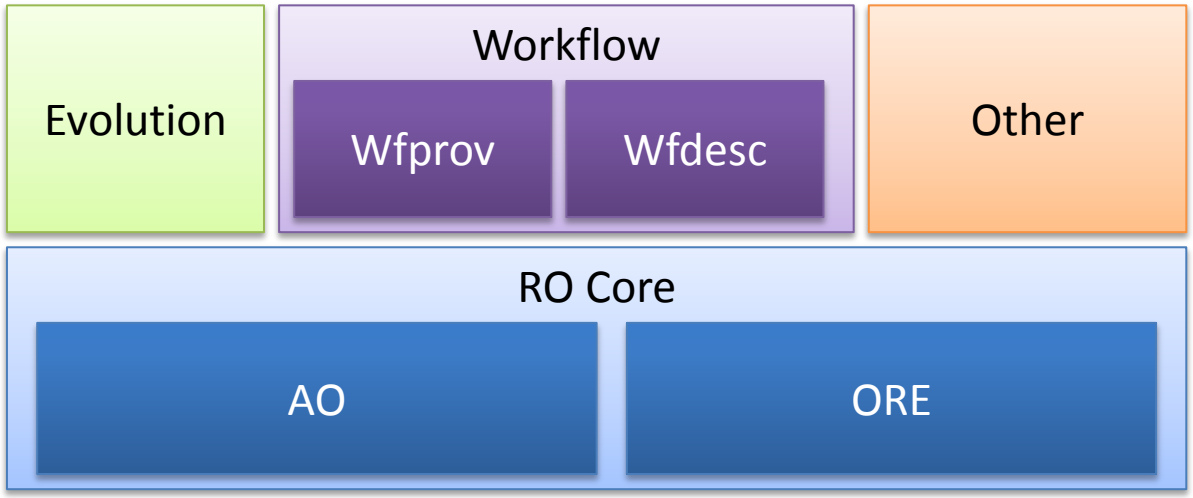


Annotation Data Ontology (AO)
http://code.google.com/p/annotation-ontology/
A generic, domain-neutral annotation framework

- Tool support
- Interoperability

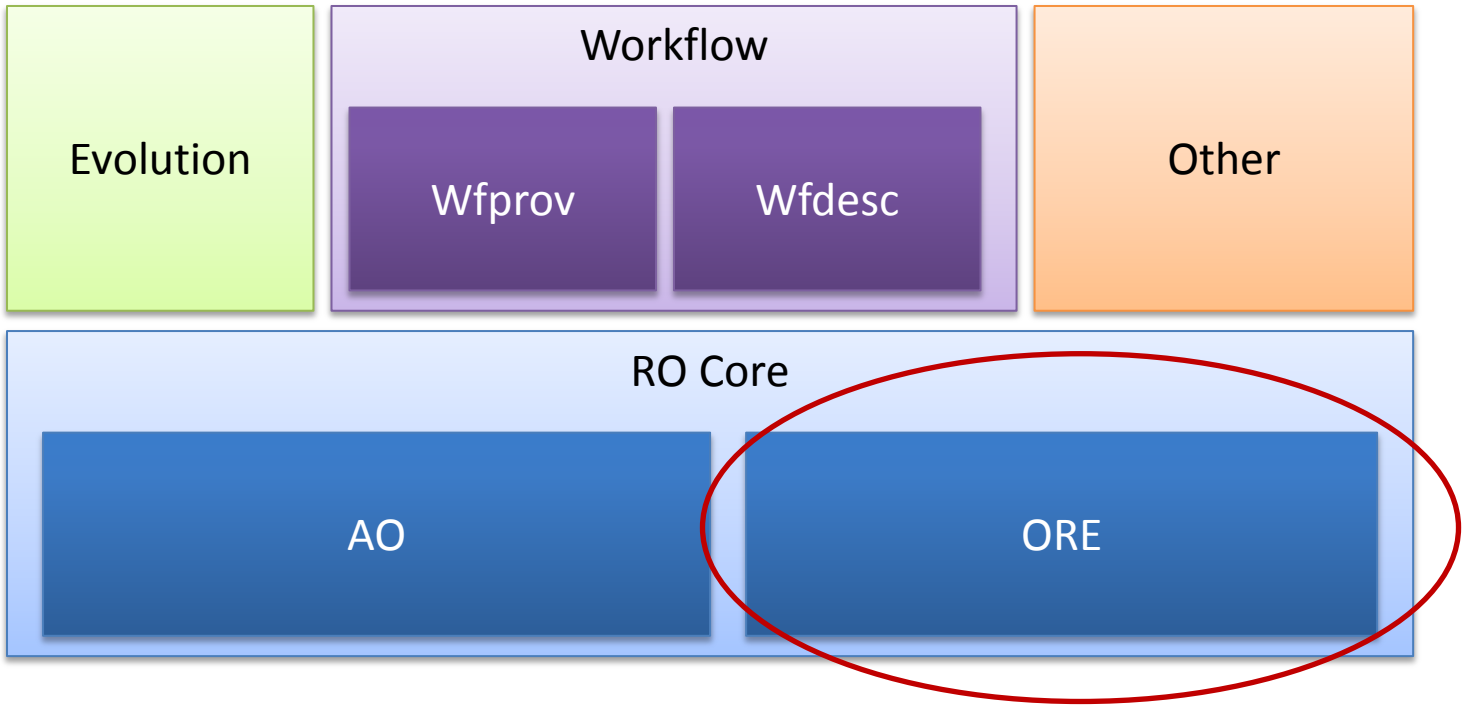
The Research Object Model

- Vocabulary for describing Research Objects
- Generic
- Extensible to multiple domains
- Modular

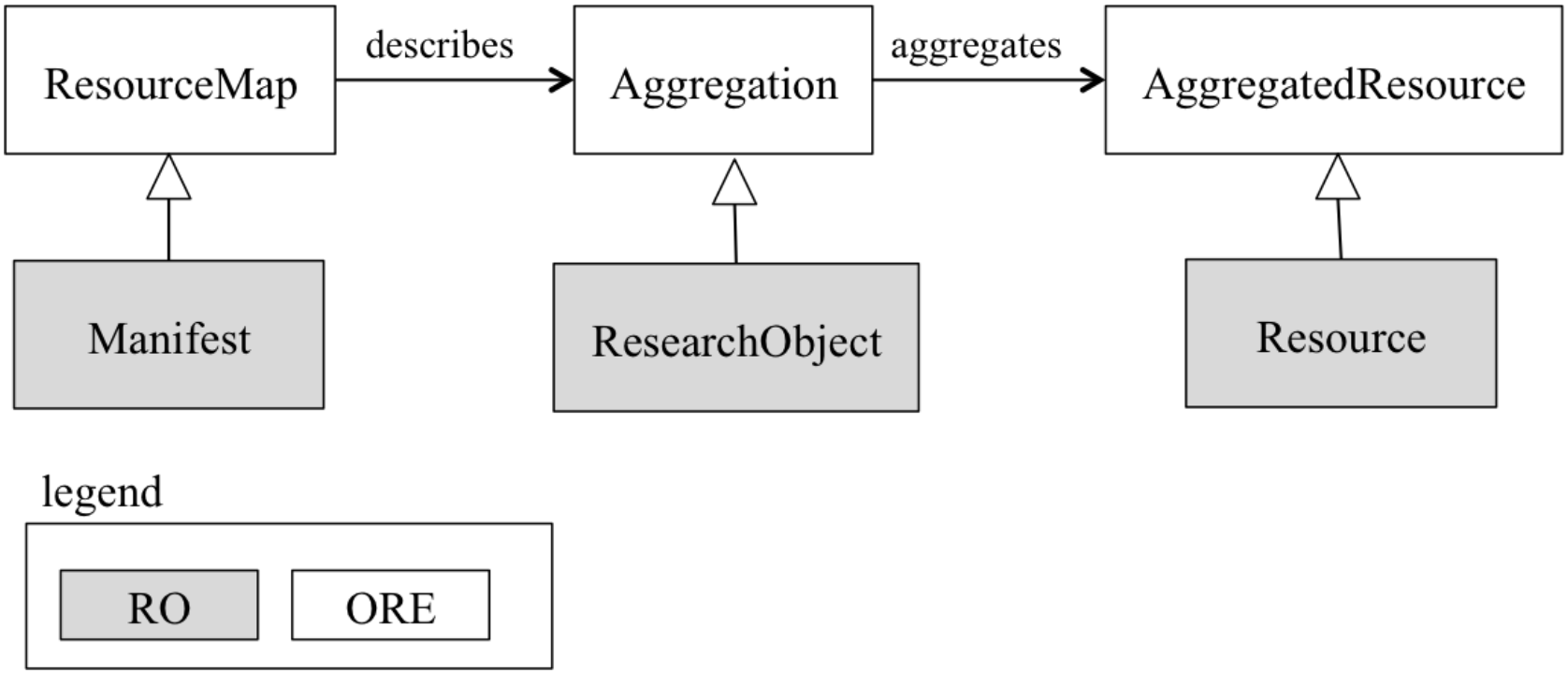


The Research Object Model: ORE

Open Archives Initiative
***O**bject **R**euse and **E**xchange*



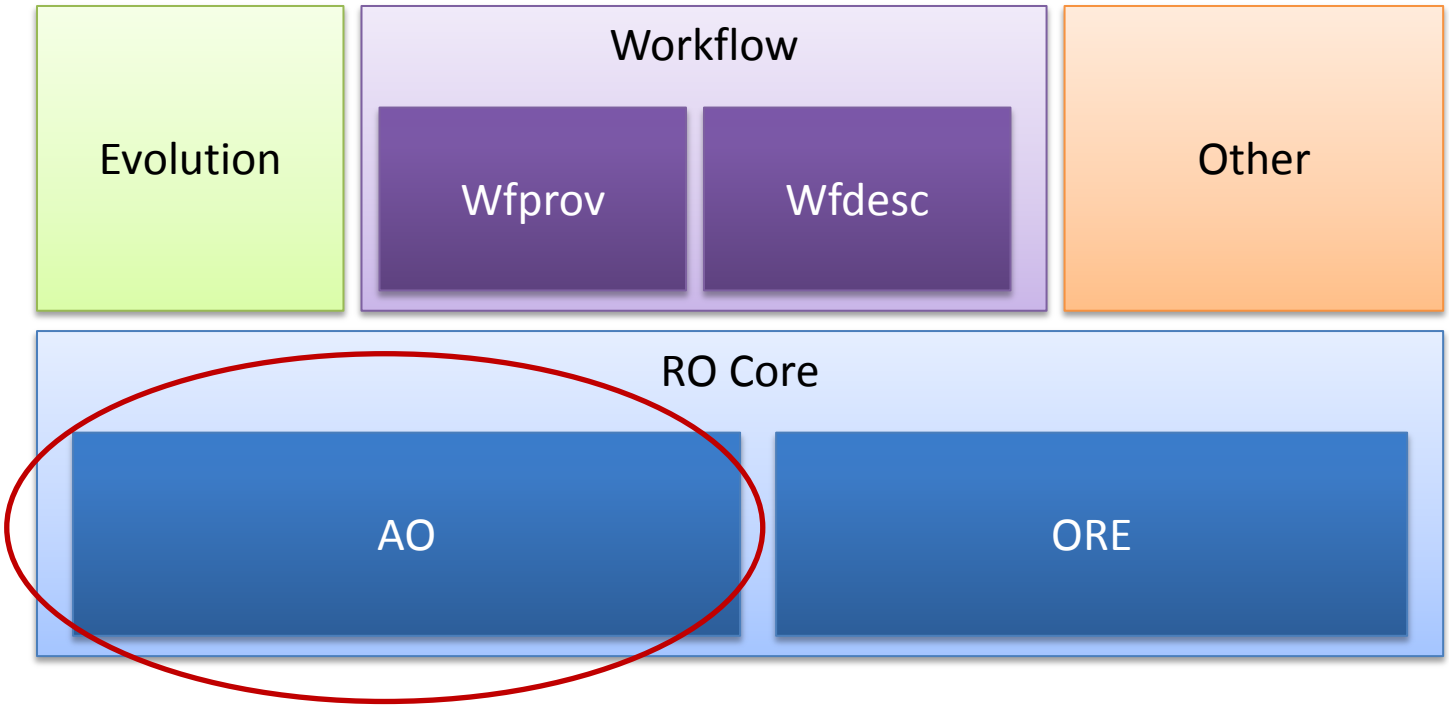
ROs as aggregations of resources



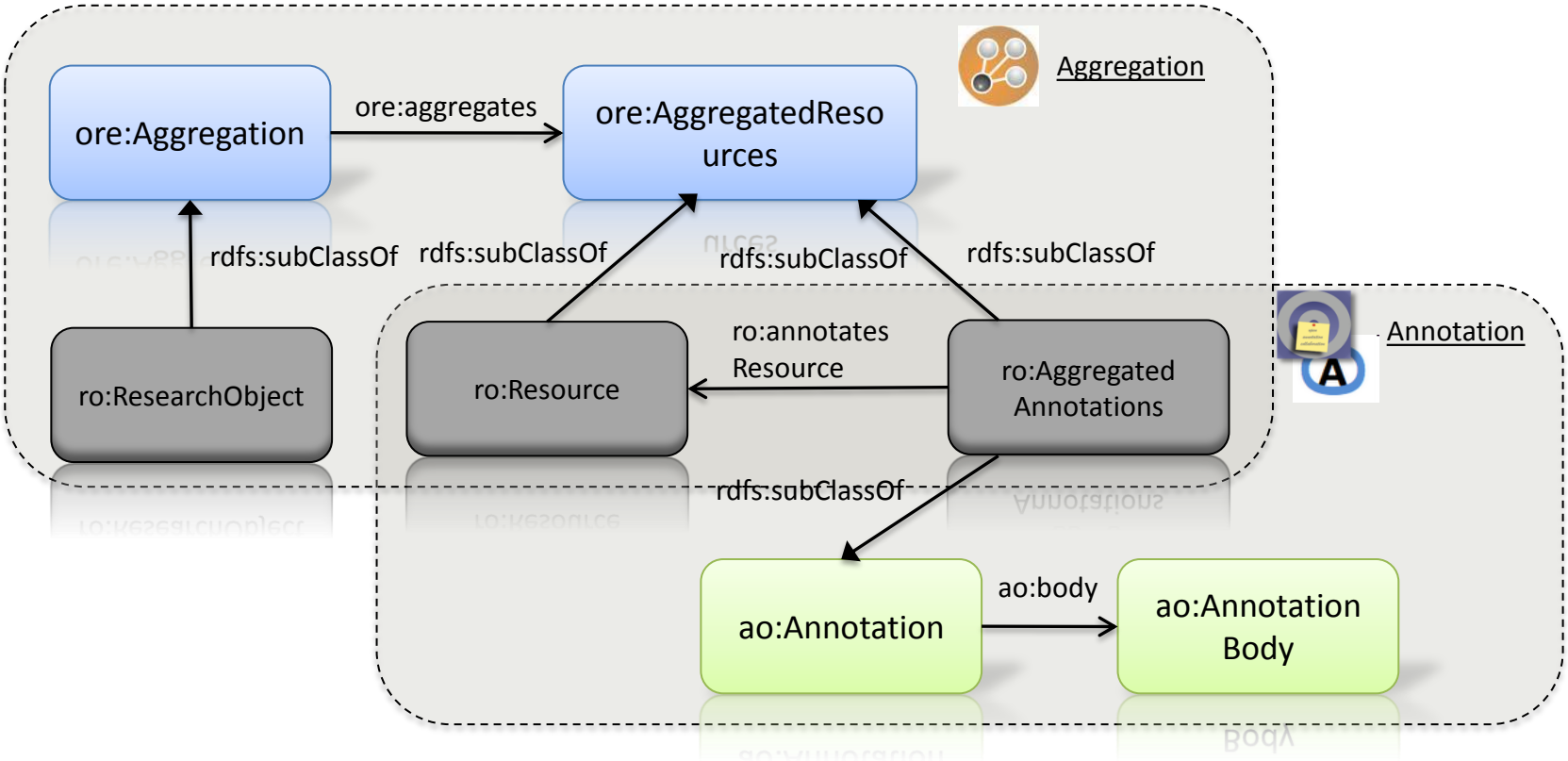
- ORE: Object Reuse and Exchange
- Resources can be further specialized according to the domain

The Research Object Model: AO

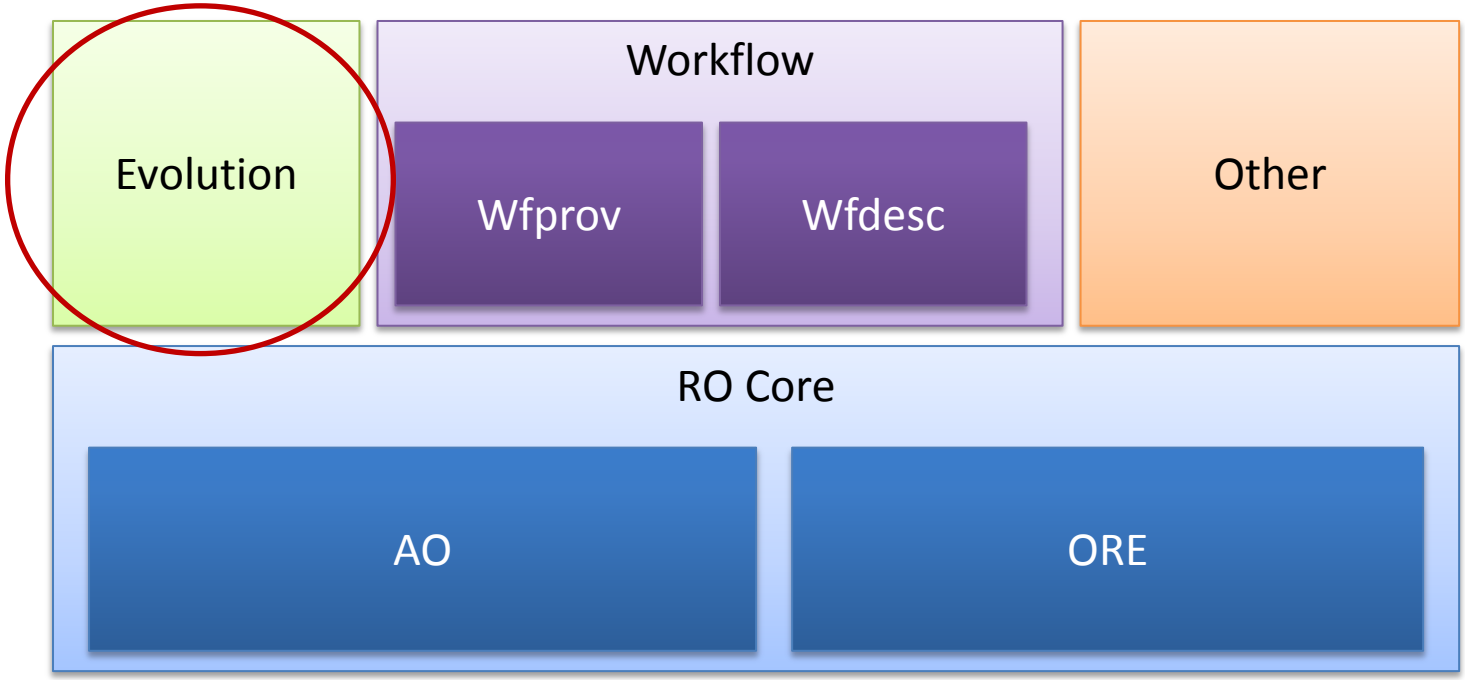
***A**nnotation **O**ntology*



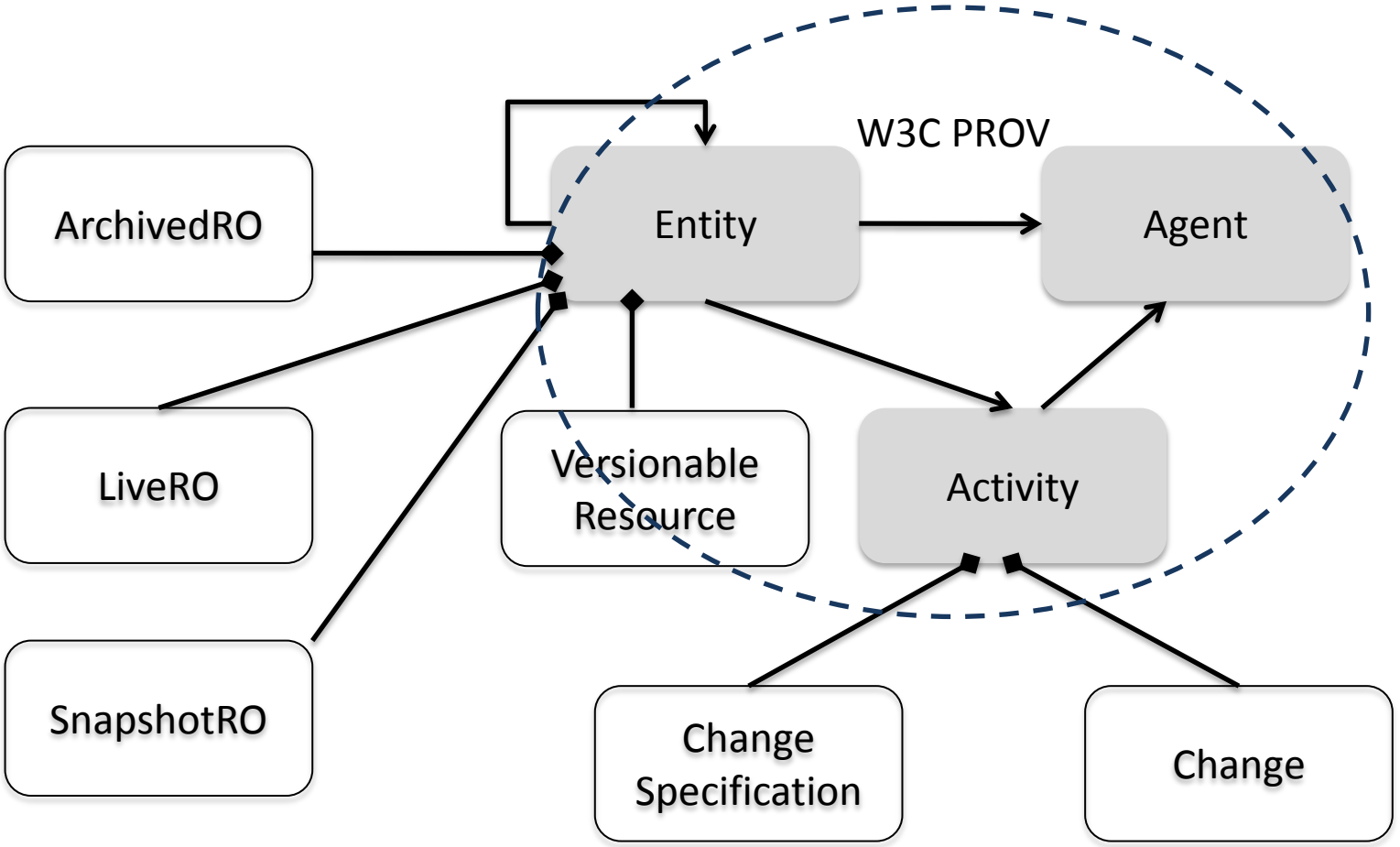
ROs: Annotations



The Research Object Model: Evolution

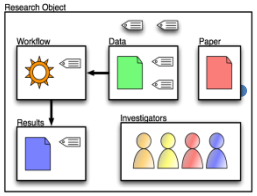


Evolution of ROs

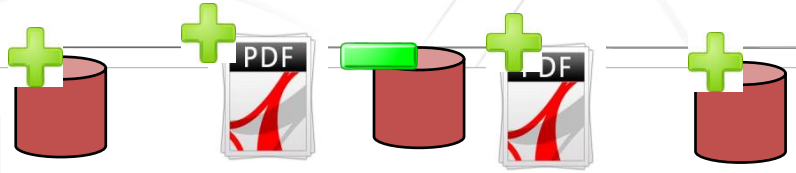


Workflow 4Ever

Scientist



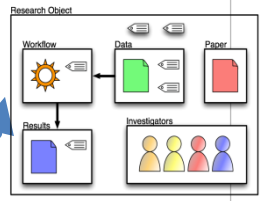
Live RO



My supervisor calls me to report my work

My supervisor calls me again and we decide to publish our RO+paper

Reviews received and final version published



Live RO

A new PhD student continues my work

git

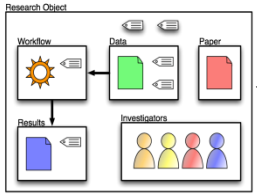
Scientist

<<copy>>

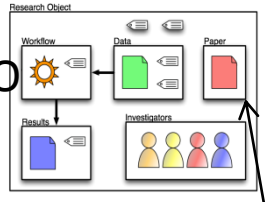
<<copy>>

<<copy, filter and curate>>

<<copy>>



<<versionOf>>



RO snapshot

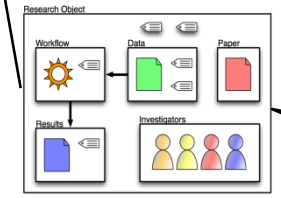
RO snapshot

<<versionOf>>

Librarian/Curator

Identified by a URI
Some metadata
Some curation
Mostly private (for my group)

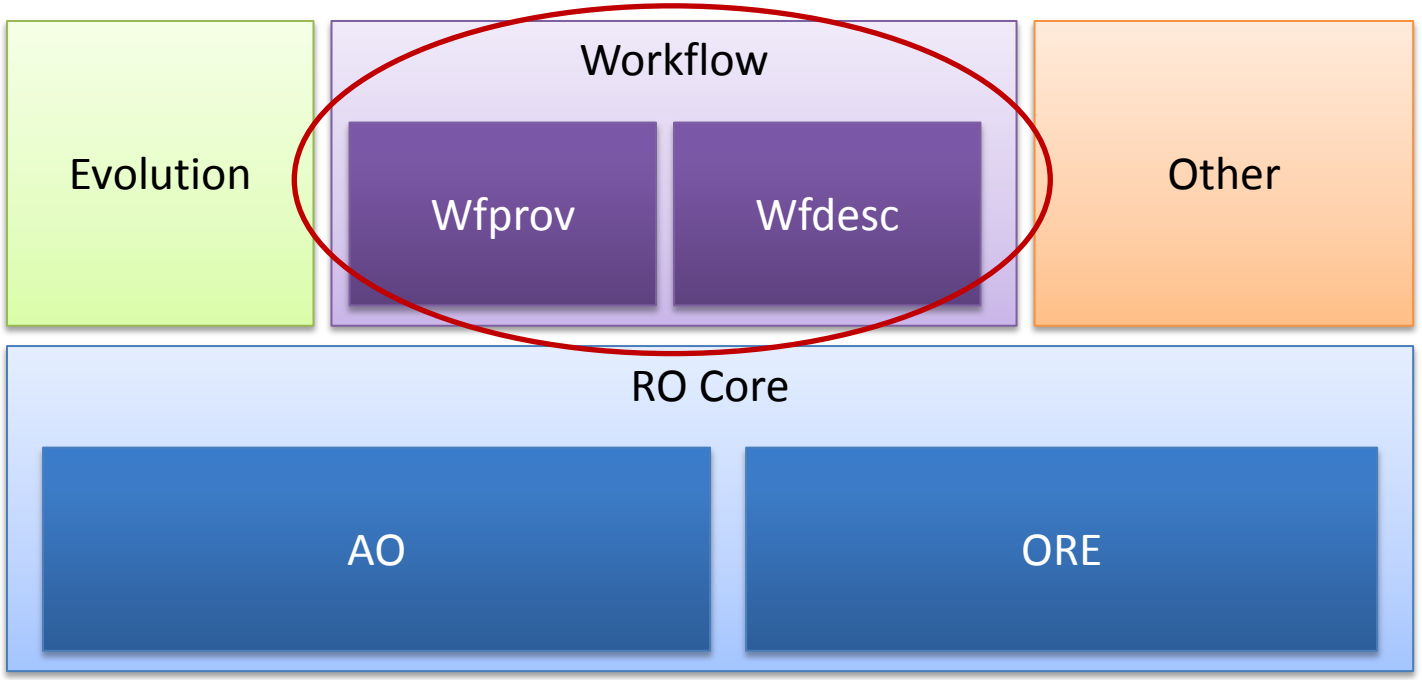
Identified by a URI
Some metadata
Some curation
Mostly private (for my group and for paper reviewers)



Archived RO

Identified by a URI
Good metadata and **curation**
Mostly public

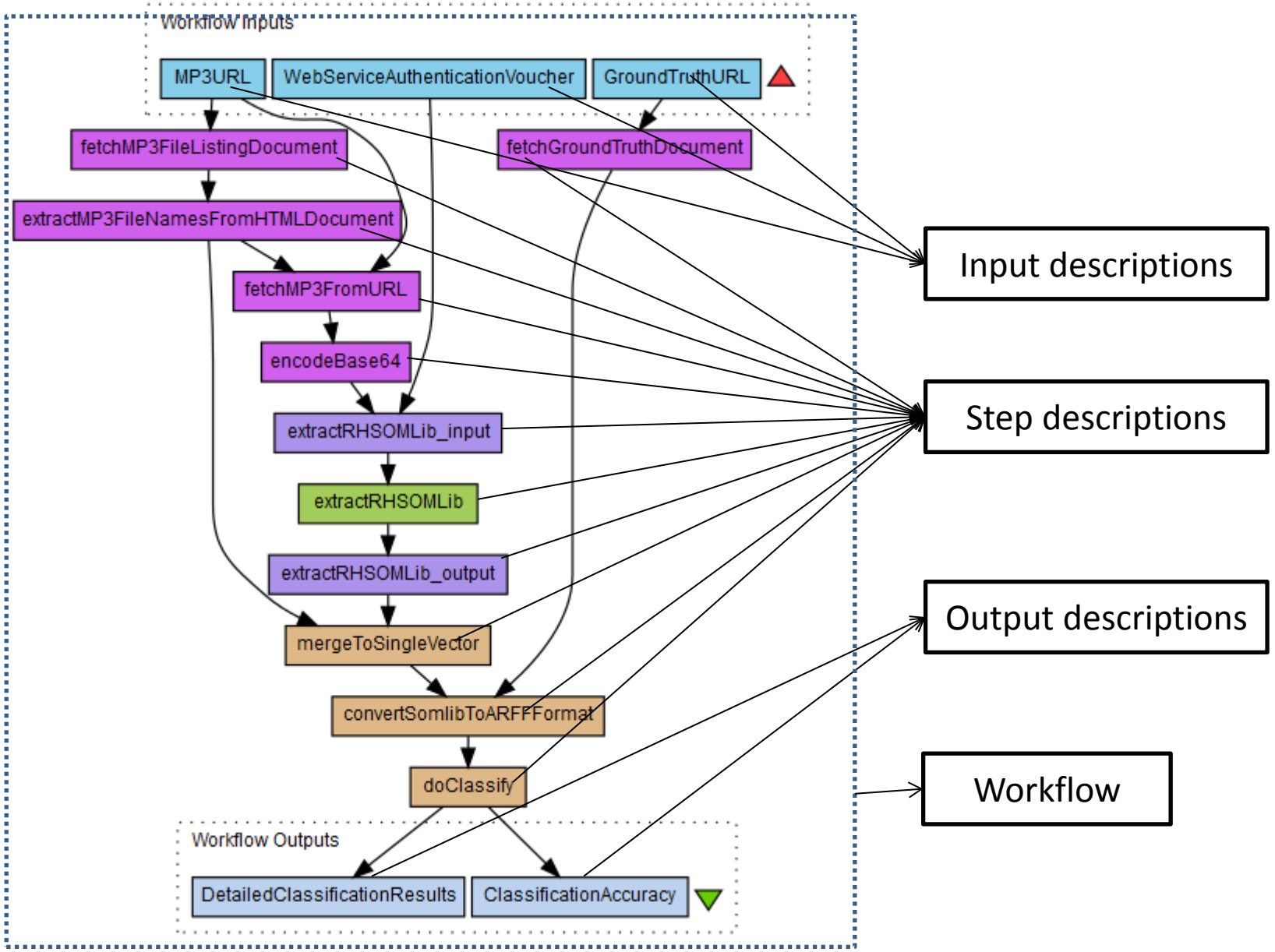
The Research Object Model: Workflow



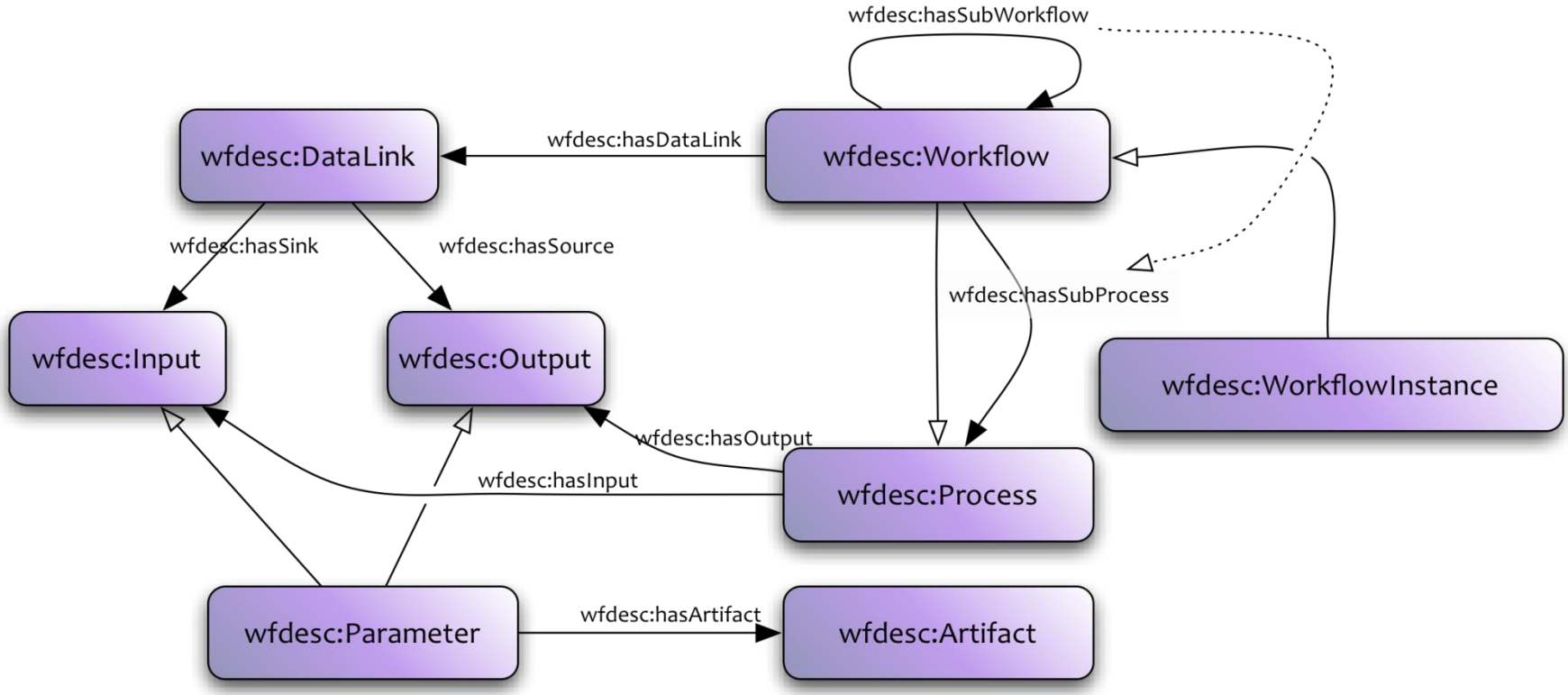
Workflow centric ROs

- Vocabulary for describing scientific workflows as ROs
- Reuse of standards (PROV)
- Compatible with other vocabularies for workflow representation
- Focused on:
  - Workflow description (wfdesc): how the specification of the workflow was planned
  - Workflow provenance (wfprov): how the results have been obtained

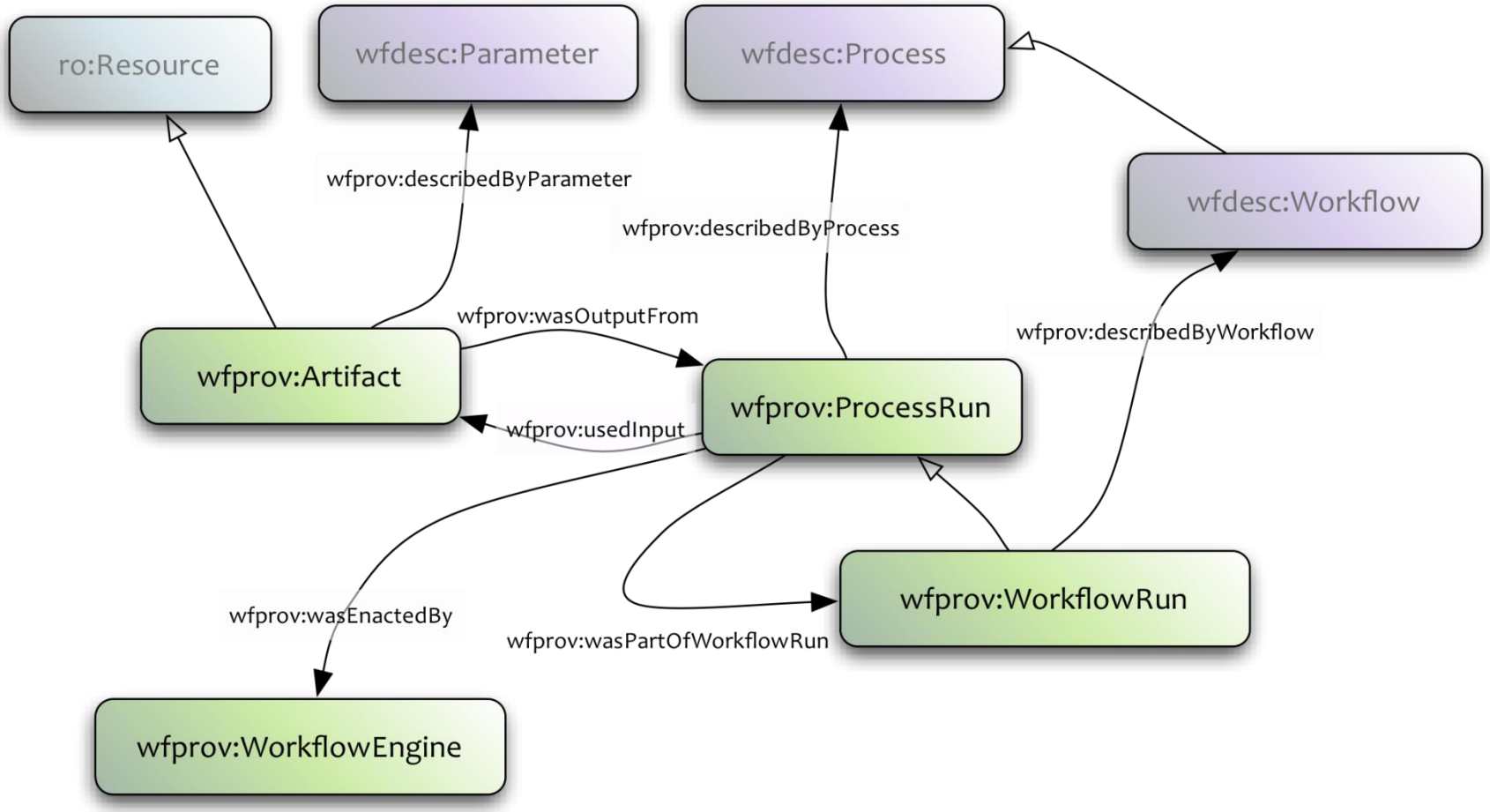
Workflow centric ROs: An example



Workflow centric ROs: Descriptions



Workflow Centric ROs: Provenance



RO-ify your work!

http://researchobject.org/

Join the discussion…

W3C Research Object for Scholarly Communication Community Group
http://www.w3.org/community/rosc/

RO Tools presentation coming after the break…

Acknowledgements: Wf4Ever project

Special thanks to Daniel Garijo for his work on the slides – catch him in the iPRES lobby track!