
Evaluation of workflows for repeatability and sharing of experimental templates, execution records, and data

— Harshvardhan J. Pandit —

Supervised by David Lewis & Declan O'Sullivan
Theme E, ADAPT, Trinity College Dublin

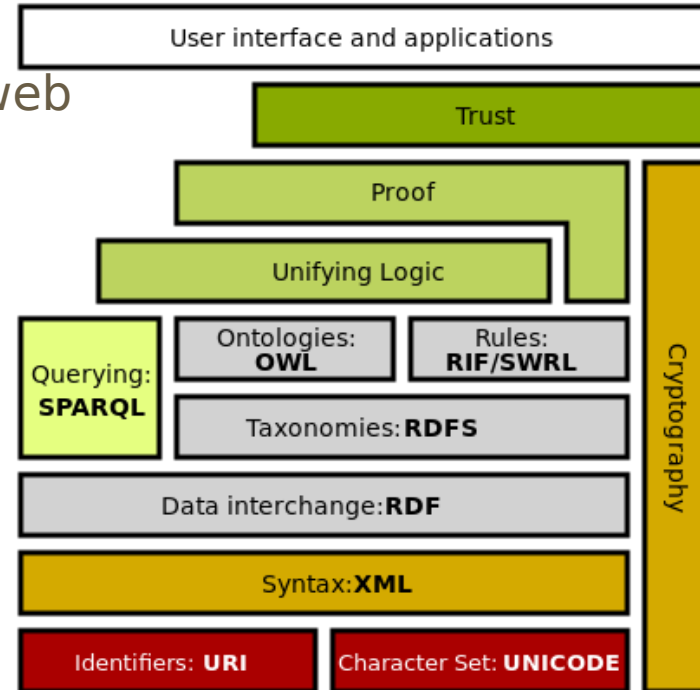
Problem domain

- Researchers in areas related to NLP do a lot of experiments that are similar and use the same or similar data sets.
- Most experiments are a variation of some common template or previous research.
- Using linked open data, experiments and data sets can be 'linked' together.
- There have been some approaches for formalizing NLP related experimental metadata by the community - OpenAIRE, LAPPS Grid / Galaxy

Linked Data & Semantic Web

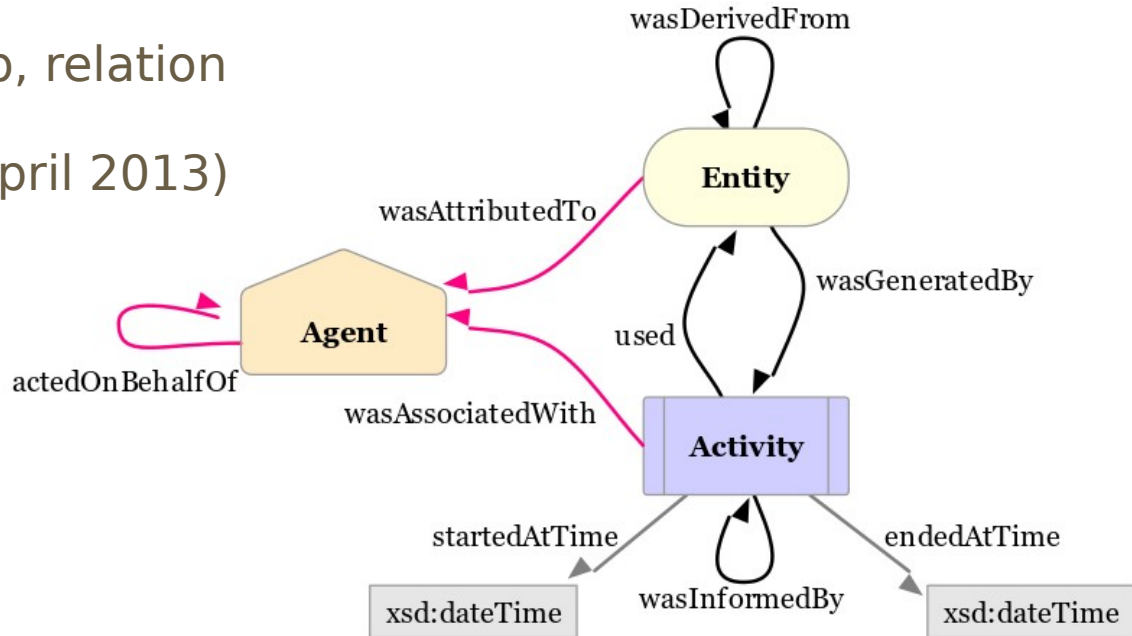
Ontologies

- Linked Data is interlinking of information
- Semantic Web is the extension of current web where information is linked by context
- Ontologies are used to describe the nature of information and its relation



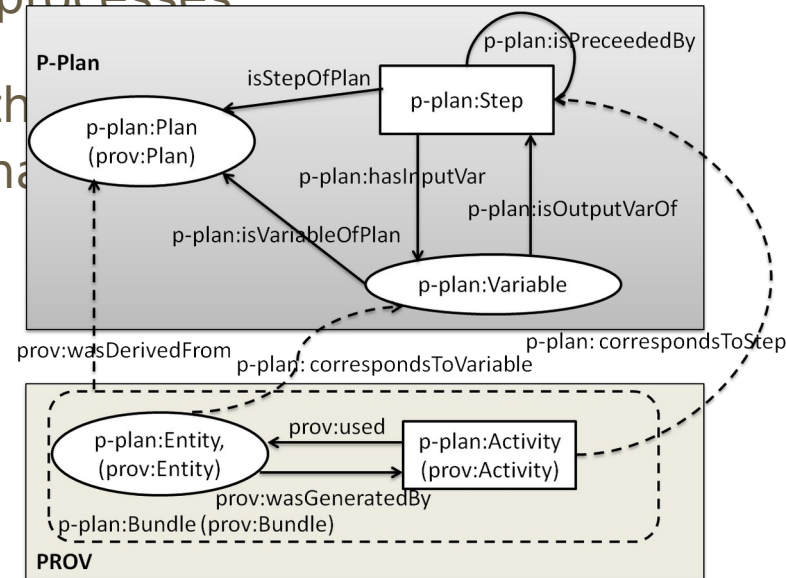
PROV

- Ontology to record and share Provenance Information
- Origin, history, ownership, relation
- W3C recommendation (April 2013)



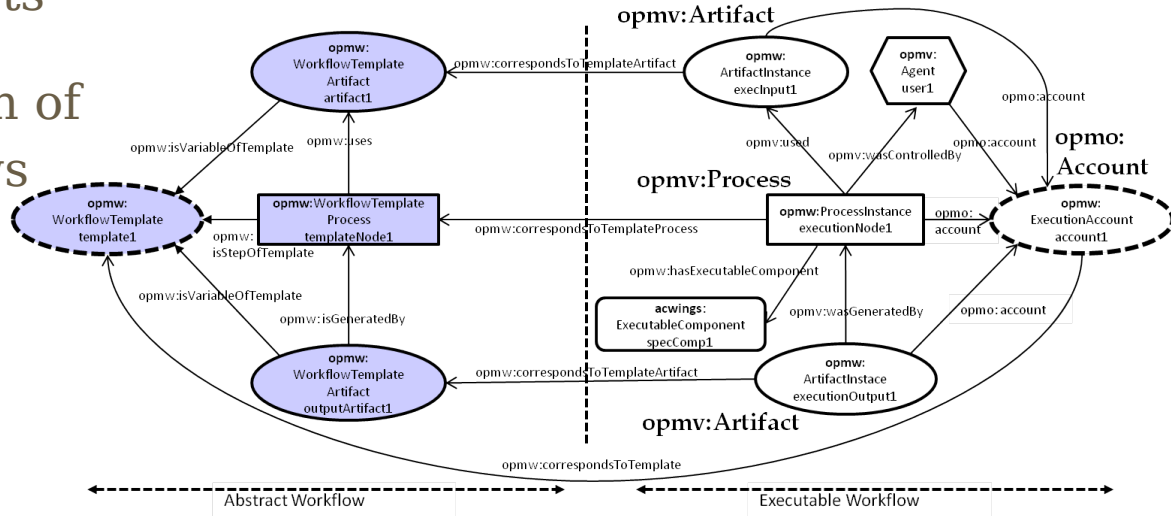
P-Plan

- Extension to PROV
- Represents plans that guide execution of processes
- Describes how plans are composed and the correspondence to provenance records that describe the execution of the plan



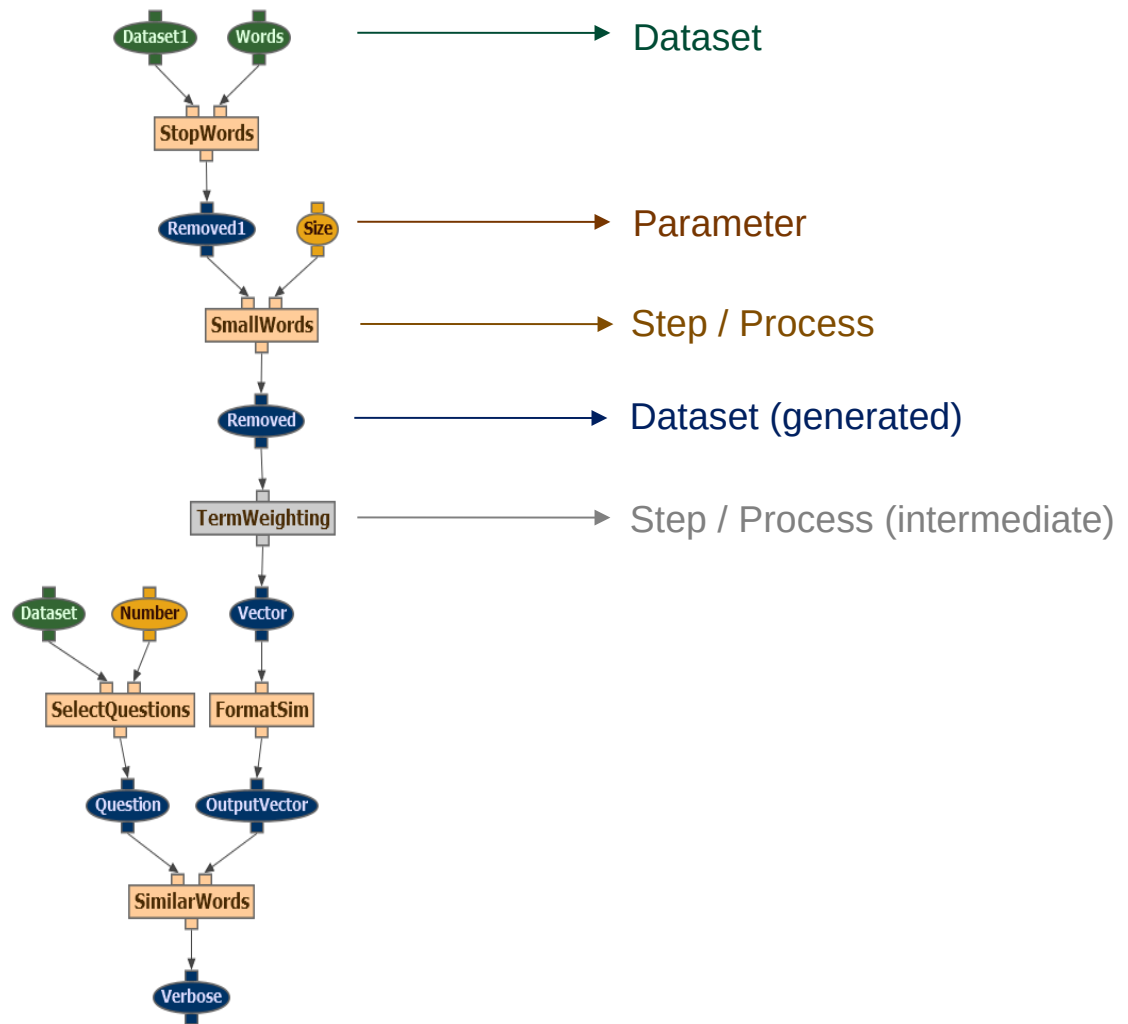
Open Provenance Model for Workflows (OPMW)

- Based on P-Plan, which extends PROV-O
- Describes workflows for scientific articles and their results
- Suitable for publication of experimental workflows

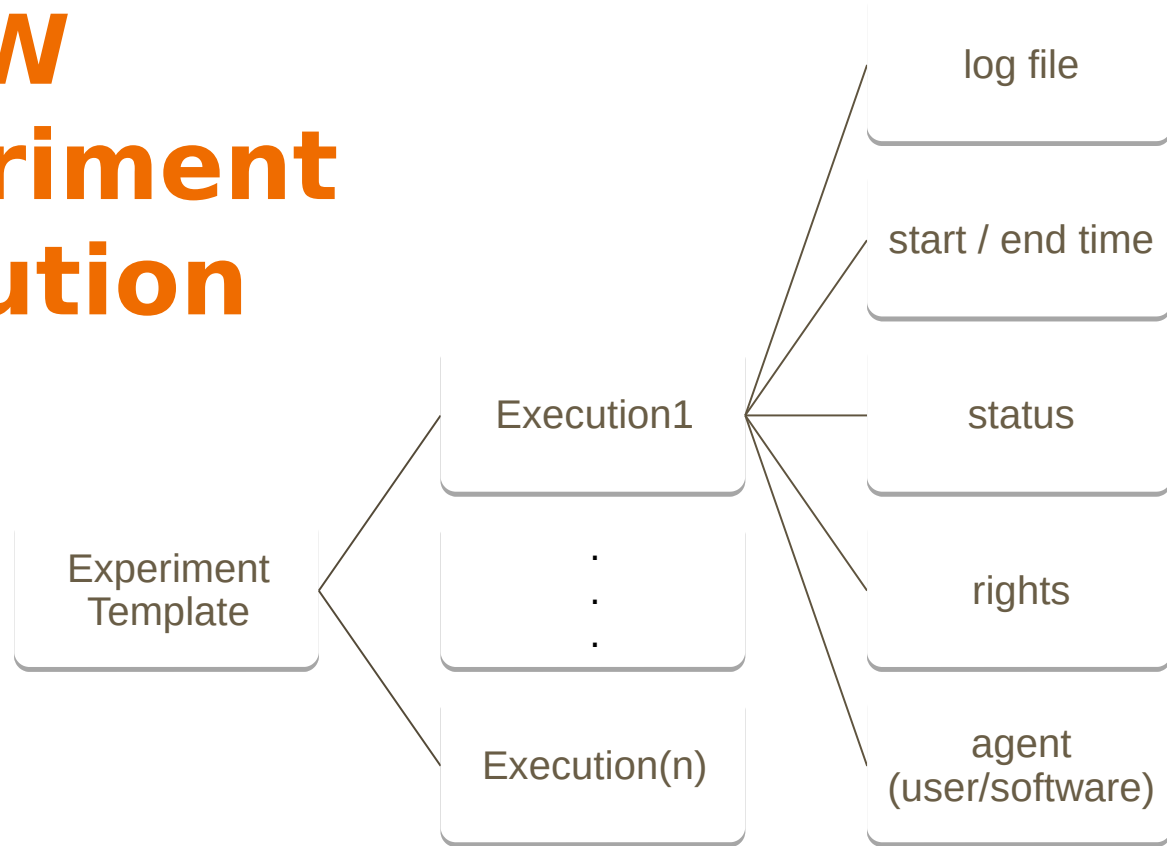


OPMW Experiment Template

Experiment Template



OPMW Experiment Execution



Project Objectives

- Using workflows & tools for documentation of experiments using OPMW

Project Deliverables

- Browser based tool for creating / interacting with experimental workflows
- Declaration of workflows using the OPMW ontology
- User study on evaluation of tasks – search for related components, create execution, fork an experiment

Browser based tool

Premise

- May not know about linked open data
- May not know about semantic web and ontologies
- Search for related research
- Search for related data sets

Application

- Browser based tool, GUI
- Abstract ontology into simpler terms and forms
- Introduce them to linked open data and experimental workflows
- Provide motivation to publish experiments using linked open data

Browser based tool

The screenshot shows the OPMW Workflow Execution browser interface. The browser address bar is `lvh.me:5000/execute/`. The page has a navigation bar with links: `open template`, `import experiment`, `export experiment`, and `publish experiment`. The main content area is divided into three sections:

- Execution Objects:** Contains sections for Execution Account (EXPERIMENT_TEST_AB), Data Variables / Execution Artifacts (DATA_VAR_A, LABEL NOT SET, LABEL NOT SET), Parameter Variables / Execution Artifacts (PARAM_A), and Steps / Execution Processes (LABEL NOT SET, LABEL NOT SET).
- Object properties:** Features a `save form` and `cancel form` button. A yellow error box states: "There were errors in the form" with a list: label is not set, filename is not set, value is not set, location is not set, and size is not set. Below this is an "Execution Artifact Label" input field with the placeholder "enter label of experiment artif" and a description: "The label of the experiment is used to assign a unique name to the experiment that will distinguish it from other similar experiments." There is also a "Workflow Template Artifact" section with a URL and an "Artifact File Name" label.
- Object diagram:** Shows a simple diagram with a blue box labeled "exp" connected to an orange box, with a yellow oval and a blue oval below.

The screenshot shows the OPMW Workflow Execution browser interface. The browser address bar is `lvh.me:5000/published/template/experiment_A/`. The page displays the "Execution Accounts" section with a table of execution details:

EXP_160913	
label	EXP_160913
status	True
workflow_system	lvh.me/workflow_system/workflow_editor/
start time	09/13/2016 08:00
end time	09/13/2016 08:00
diagram	lvh.me:5000/export/images/EXP_160913.png
log file	None
is execution account of	lvh.me/execution-artifact/160913.param_a lvh.me/execution-artifact/160913.data_var_a lvh.me/execution-artifact/160913.op_m2 lvh.me/execution-artifact/160913.op_m1 lvh.me/execution-process/160913.step_m1 lvh.me/execution-process/160913.step_m2

E_LABEL	
label	E_LABEL
status	success
workflow_system	Workflow Editor
start time	09/08/2016 00:00
end time	09/08/2016 00:00
diagram	None
log file	None
is execution account of	lvh.me/execution-process/STEP_M2_el lvh.me/execution-process/STEP_M1_el lvh.me/execution-artifact/DATA_VAR_M1_co lvh.me/execution-artifact/PARAM_VAR_A_co lvh.me/execution-artifact/DATA_VAR_M2_co lvh.me/execution-artifact/DATA_VAR_A_EL

Below the table is the label "E_LABEL".

How did I implement this?

- Front end - HTML/CSS/JS (the usual suspects)
- Drawing graphs - JointJS (<http://jointjs.com>)
- Backend - flask/python (<http://flask.pocoo.org/>)
- Triple store - SQLite

What's Next?



Directed Study Objectives

- Design, execution and analysis of web application usability study
- Linked open data publish-consume lifecycle
- Documentation requirements for repeatability of NLP experiments and language resource reuse
- SPARQL to query linked data stores

Things I learnt

- How semantic web ontologies and linked open data are used
- How to write (or adapt) an ontology
- The idea of working with triple stores
- Where semantic web ontologies lack and how to deal with it
- The (**IMPORTANT**) difference between “engineering” and “research” projects

Be wise, learn from my shortcomings

- Don't look for a 'software' solution
 - Never lose sight of the aim - better yet, write it and stick it on your desk
 - Always ask someone else rather than be confused or stuck
 - Read, read, read, then revise
 - Make the most minimal working example you can, then work on specific questions you want answered
 - 'Learning' never stops - if you think you're free, you're not learning enough

that's all for today

**Questions?
Suggestions?
Tips & Tricks?**