

UTILISING SEMANTIC WEB ONTOLOGIES TO PUBLISH EXPERIMENTAL WORKFLOWS

SIMPLIFYING EXPERIMENTS THROUGH DECENTRALISATION

Harshvardhan Pandit
Ensar Hadziselimovic
David Lewis



About Us



- The ADAPT research centre focuses on developing next generation **digital technologies** that transform how people communicate. Adapt excels in areas of **Natural Language Processing, Machine Learning, AI** and many more.
- ADAPT researchers are based in four leading universities: Trinity College Dublin, Dublin City University, University College Dublin and Dublin Institute of Technology.
- **115** Expert Researchers **80** Industry Partners **327** Published Papers

Background and Motivation

- Diversity in publishing methods
- Centralised nature of publications
- No clear guidelines when reproducing an experiment
- Semantic web and linked data as decentralised approach
- How to enable anyone to publish their experimental workflows?

Project Plan

- Workflows -> Steps to perform the experiment
- Data -> What goes in, properties, conditions for usage
- License -> Conditions for reproduction / repeatability / reuse
- Publish using Linked Data

Open Publications

The Benefits

- Workflows can be used as tools of documentation
- Published under (author) self-controlled environment
- Decentralise knowledge and connect through using linked data

Considerations

- How do researchers conduct their research?
- How does the research map into an ontology?
- What existing ontologies can be used here?
- How can we license data sets?
- Can we create a user study?

Main Concepts in Experimental Workflows

— — —

Reproducibility and Repeatability

- ‘Sharing of data’ is a key tenet of scientific publications
- Peer reviews are a decentralised form of validating research
- Open Access needs access to data and implementation steps

Variations in experiments

- Experiments are similar and use (almost) similar datasets
- Variations of some common template or of previous research
- Using linked open data, these experiments can be ‘linked’ together

Licensing and Intellectual Property

— — —

Starting Point

- Decision on best licensing practices
- Need for comprehensive method to describe not just an experiment, but all of its components and the relations
- Licensing model to suggest the best license
- Decentralised approach to the above

Challenges

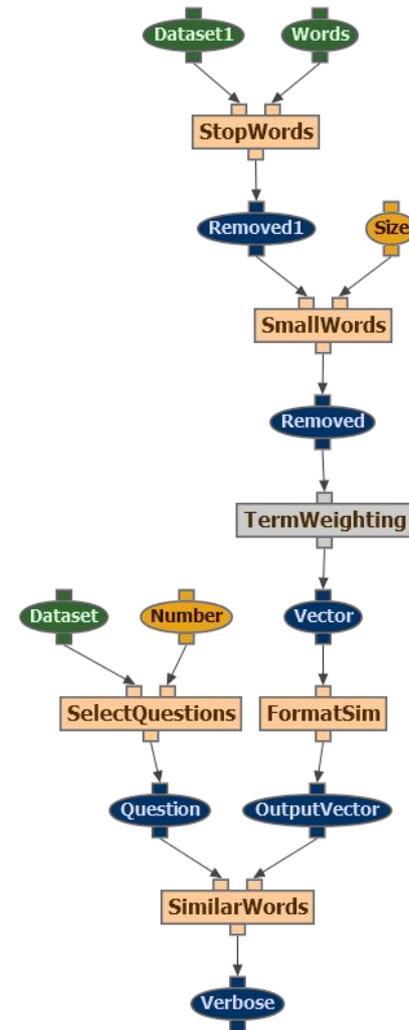
- Explore the best practices
- Avoid over-complicated legal requirements
- Final experiment's license is not sum of its parts
- Inheritance: inclusion vs exclusion (parent-child relationship)

Ontologies

- OPMW (Open Provenance Model for Workflows)
 - Based on PROV-0 and P-Plan
 - Track provenance in scientific workflows
- ODRL (Open Digital Rights Language)
 - Expressing digital rights management
 - Use for access and usage conditions

OPMW <http://www.opmw.org/>

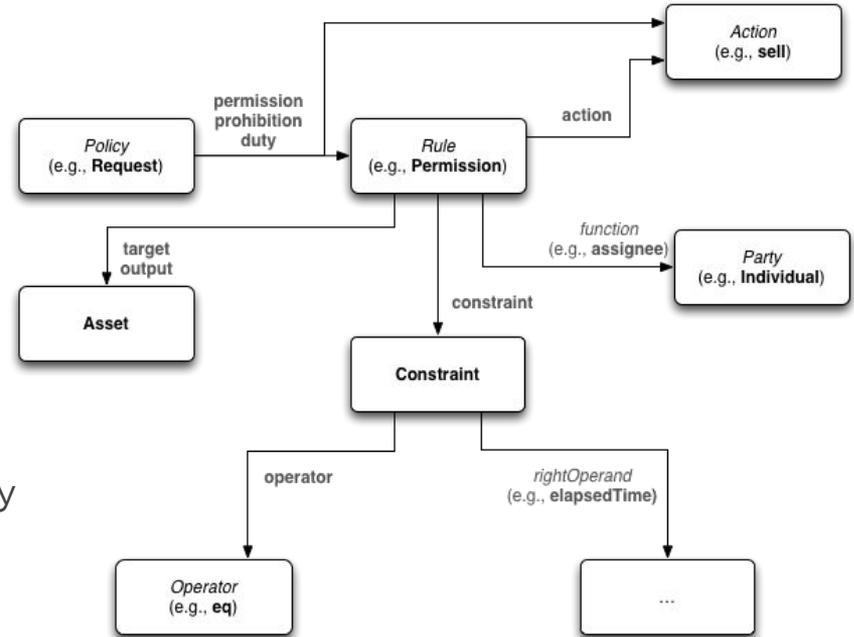
- Template
 - Artifact
 - Data Variable
 - Parameter Variable
 - Process/Step
- Execution
 - Artifact
 - Process
 - Controller / Agent



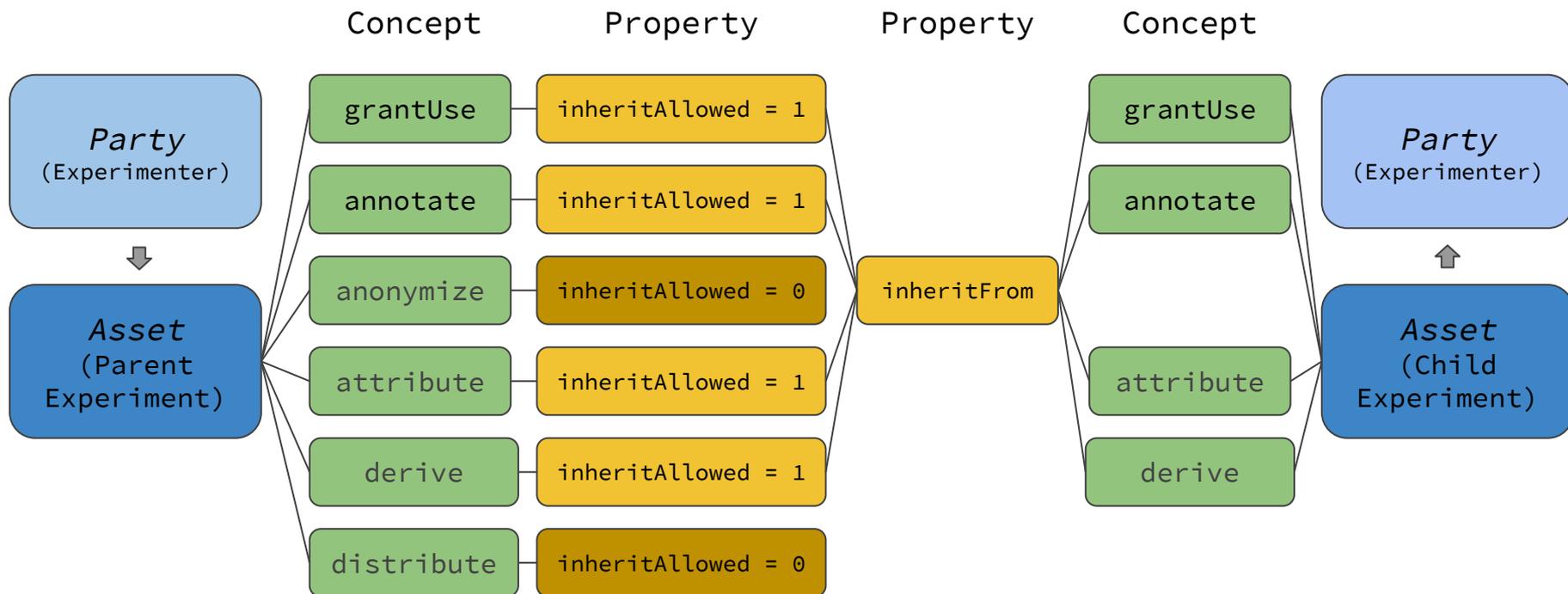
ODRL

<https://www.w3.org/community/odrl/>

- Classes
 - Permission, Prohibition
 - Asset, Party
 - Policy, Privacy
- Properties
 - inheritAllowed, inheritFrom
 - constraint, relation
 - attributedParty, consentingParty
- Concepts
 - grantUse, annotate
 - anonymize, attribute
 - derive, distribute



ODRL - Concept of Inheritance Through Inclusivity



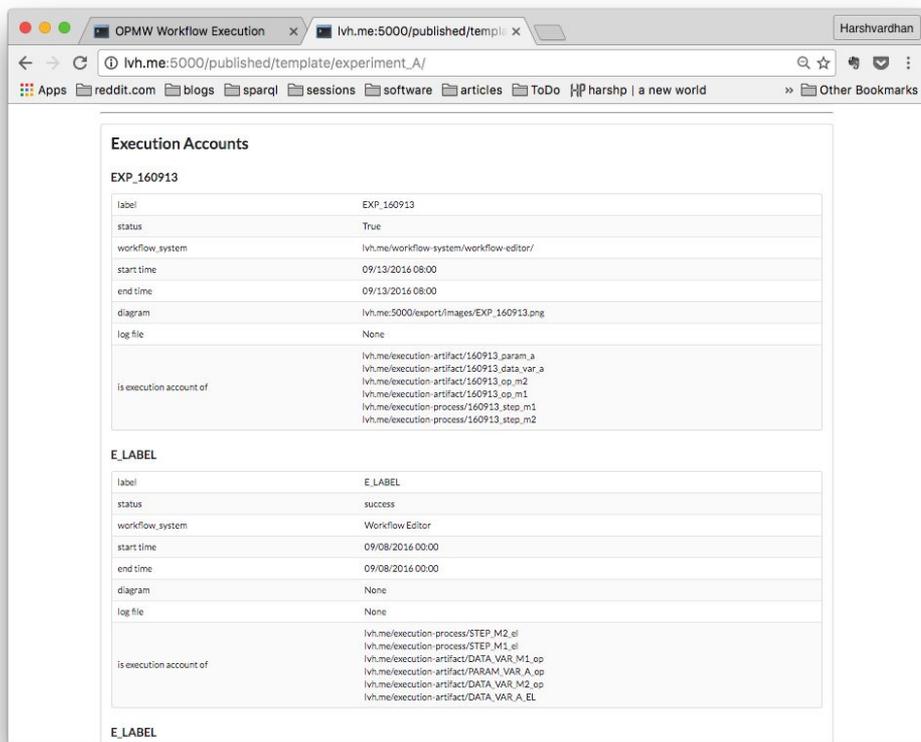
Create template

The screenshot displays the OPMW Workflow Creation interface. The browser address bar shows 'lvh.me:5000/create/'. The interface is divided into three main sections: Execution Objects, Object properties, and Object diagram.

- Execution Objects:** Contains sections for Template (EXPERIMENT_X), Data Variables / Execution Artifacts (DATA_VAR_A), Parameter Variables / Execution Artifacts, and Steps / Execution Processes (STEP_X).
- Object properties:** Features a 'save form' button, a 'cancel form' button, and a green notification box stating 'Step saved successfully'. Below this, there is a 'Step Label' field containing 'step_x', a description of the label's purpose, a 'Data Variables / Parameters used' dropdown menu with 'data_var_a' selected, and an 'Additional properties' section with an 'add' button.
- Object diagram:** Shows a simple flow diagram with a green oval labeled 'DATA_VAR_A' connected by a blue arrow to an orange rectangle labeled 'STEP_X'.

- Experiment contains 3 tasks, based on user's pre-existing knowledge on experimental workflows as well as linked data principles
- All tasks converge on the documentation generated for the workflows which the users are encouraged to explore at the end of their task.

Generate some form of documentation



The screenshot shows a web browser window with two tabs. The active tab is titled 'lvh.me:5000/published/template/experiment_A/'. The browser's address bar shows the URL 'lvh.me:5000/published/template/experiment_A/'. The page content is titled 'Execution Accounts' and contains two tables. The first table is titled 'EXP_160913' and the second is titled 'E_LABEL'. Both tables have columns for 'label', 'status', 'workflow_system', 'start time', 'end time', 'diagram', 'log file', and 'is execution account of'.

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_op lvh.me/execution-artifact/PARAM_VAR_A_op lvh.me/execution-artifact/DATA_VAR_M2_op lvh.me/execution-artifact/DATA_VAR_A_EL

- The documentation generated follows the principles of linked open data
- Each resource has its own corresponding properties and attributes.
- Comprehensive overview of the entire workflow as well as the ability to follow the links to the documentation for a particular resource.

Benefit to the user

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 the users to linked open data and experimental workflows
- Provide motivation to publish experiments using linked open data

Challenges

- What exactly is a variation?
- Different experiments have differing views
- How to track provenance? PROV may not be sufficient.
- Licensing issues are multi-dimensional problem
- Ambiguity in attached licenses, or non-existing license
- Intellectual Property can affect part of the experiment and the experiment as a whole - dissection needed

Future Work: User Study

1. (pre-)Questionnaire to gauge familiarity with linked data concepts and area of research
2. Ask participants to use the tool, generate documentation
3. (post-)Questionnaire to identify alignment with research methods and use of automatically generated documentation
4. More participants for user study
5. Formally declare variation of experiment as a workflow ontology term rather than through provenance
6. Attach licenses to datasets and generate documentation listing terms of usage

that's all for
today

questions?

Slides and comments:

<https://goo.gl/ARd2wJ>

Source code:

https://github.com/___
