

# Enhancing Workflow with a Semantic Description of Scientific Intent

Edoardo Pignotti, Peter Edwards, Alun Preece, Nick Gotts and Gary Polhill

Computing Science  
University of Aberdeen

The Macaulay Institute, Aberdeen



# Content

---

- Background
- Workflow in eScience
- Example
- Research Challenges
- Semantic Workflow Architecture
- Case-study
- Evaluation & Conclusions

# Background

---

- Semantic Grid Vision
  - Support for researchers to publish, share and re-use resources, integrate heterogeneous information and collaborate. ([www.semanticgrid.org](http://www.semanticgrid.org))
- Central to this view is the integration of computational Grid technologies with Semantic Web technologies.
- Fearlus-G Project
  - Deployed existing social simulation model of land-use change onto the Grid.
  - Created metadata tools to support annotation/sharing of social simulation resources.
- Fearlus-G was not designed to be a flexible problem-solving environment.

You are logged in as: fearlus

### Options

[My Workspace](#)  
[My Grid Tasks](#)  
[Public Repository](#)

[Questionnaire](#)

[Log Out](#)

### My Projects

[My Entire Workspace](#)

[Imitative vs non imitative strategies](#)  
(info)

[+ New Project](#)

### My Hypotheses

[Innovators do better than Imitators](#)  
[Imitators do better than Innovators](#)

[+ New Hypothesis](#)

### My Experiments



#### Experiment Sets

[Experiment2.2-Polhill-et-al-2001](#)

#### Experiments

[SI-v-HRYI-c-II-env1-set2.2](#)

[SI-v-II-c-HYI-env1-set2.2](#)

[SI-v-II-c-HRYI-env1-set2.2](#)

[II-v-HRYI-c-HYI-env1-set2.2](#)

[MORE...](#)

[+ New Type 2 Experiment Set](#)

### My Experimental Hypotheses



[HYI-better-than-SI-against-HRYI-env1](#)

[SI-better-than-II-against-HYI-env1](#)

[HYI-better-than-II-against-HRYI-env1](#)

[SI-better-than-HRYI-against-HYI-env1](#)

[MORE...](#)

### My Simulation Parameters



#### Models

[modelo-6-6unix](#)

#### Environments

[Environment3](#)

[Environment2](#)

[Environment1](#)

#### Subpopulations

[OD](#)

[OS](#)

[HYI](#)

[II](#)

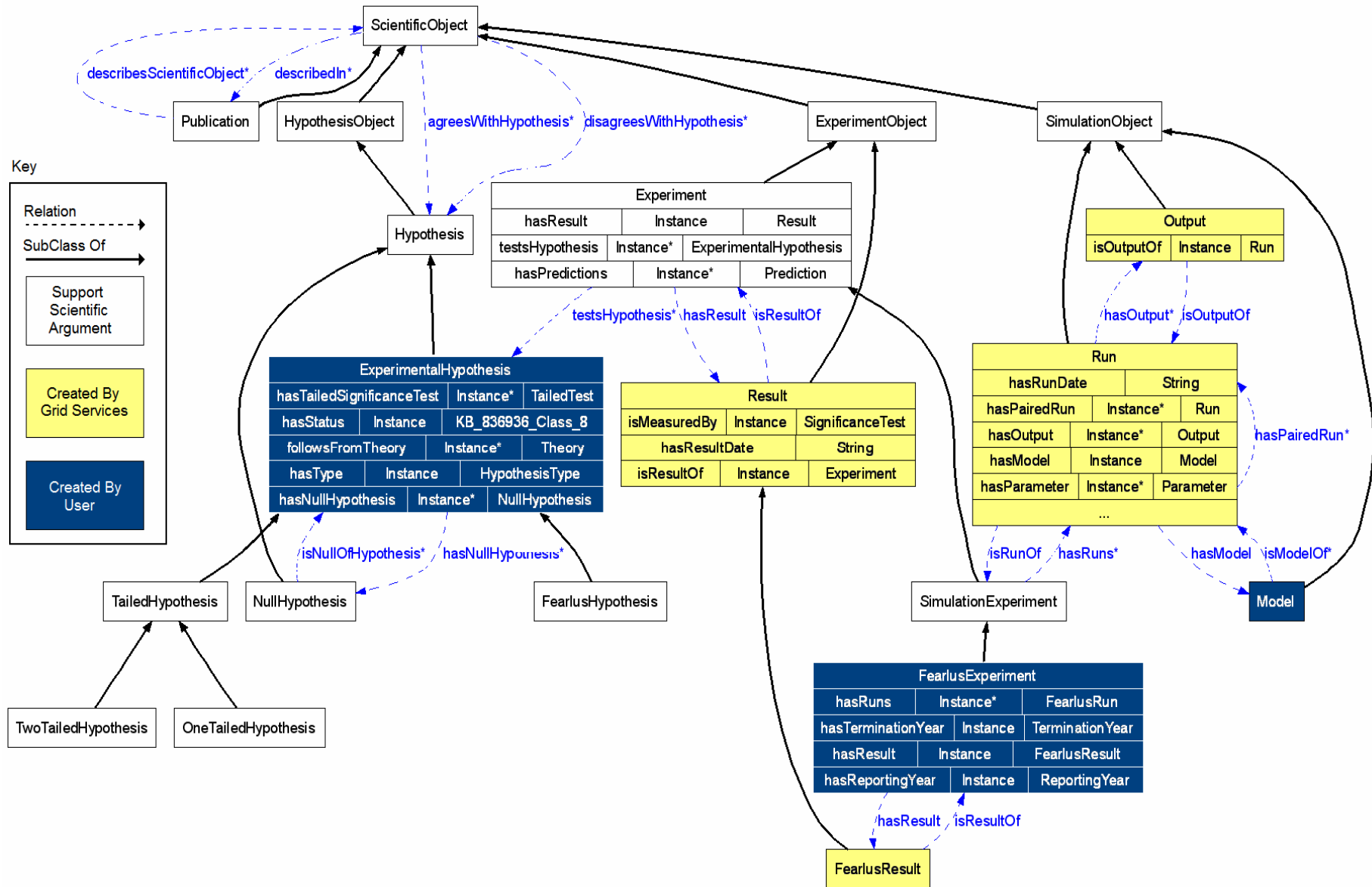
[RS](#)

[LS](#)

[MORE...](#)

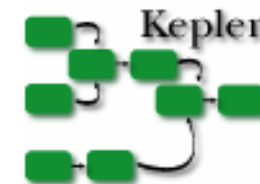
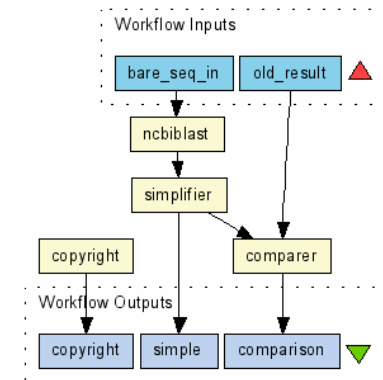
[+ New Model](#)

# (Fearlus-G) Scientific Object Ontology



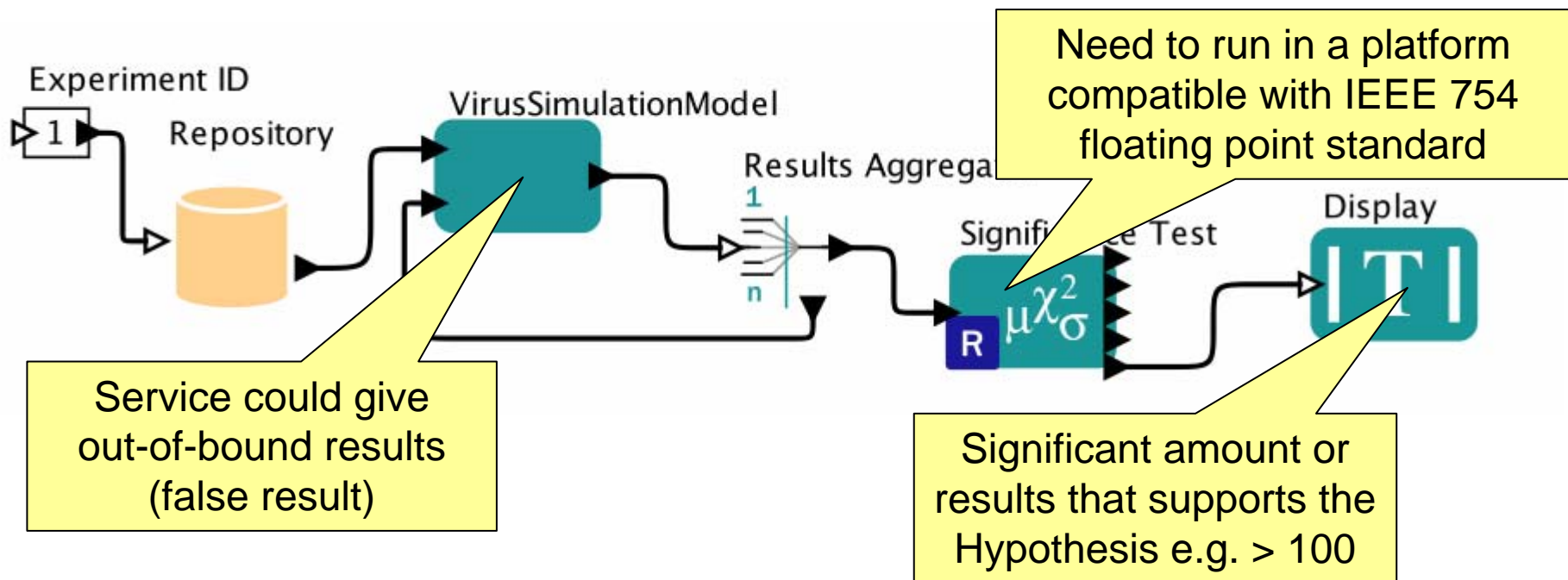
# Workflow in e-Science

- Scientific workflows facilitate the creation and execution of experiments from a pool of available services.
- Important part of *in-silico* experimentation to investigate or verify a hypothesis.
- Taverna** (myGrid, UK eScience)
  - Easy use of workflow and distributed compute technology.
  - Metadata to support the discovery of new services.
- Kepler** (University of California)
  - Provides Director component
    - Controls the execution environment (i.e. the “flow”).
  - Import OWL ontologies.
  - Metadata to describe activities, inputs and outputs.



# Example

- We have identified a number of scenarios through interaction with collaborators from the social simulation community.



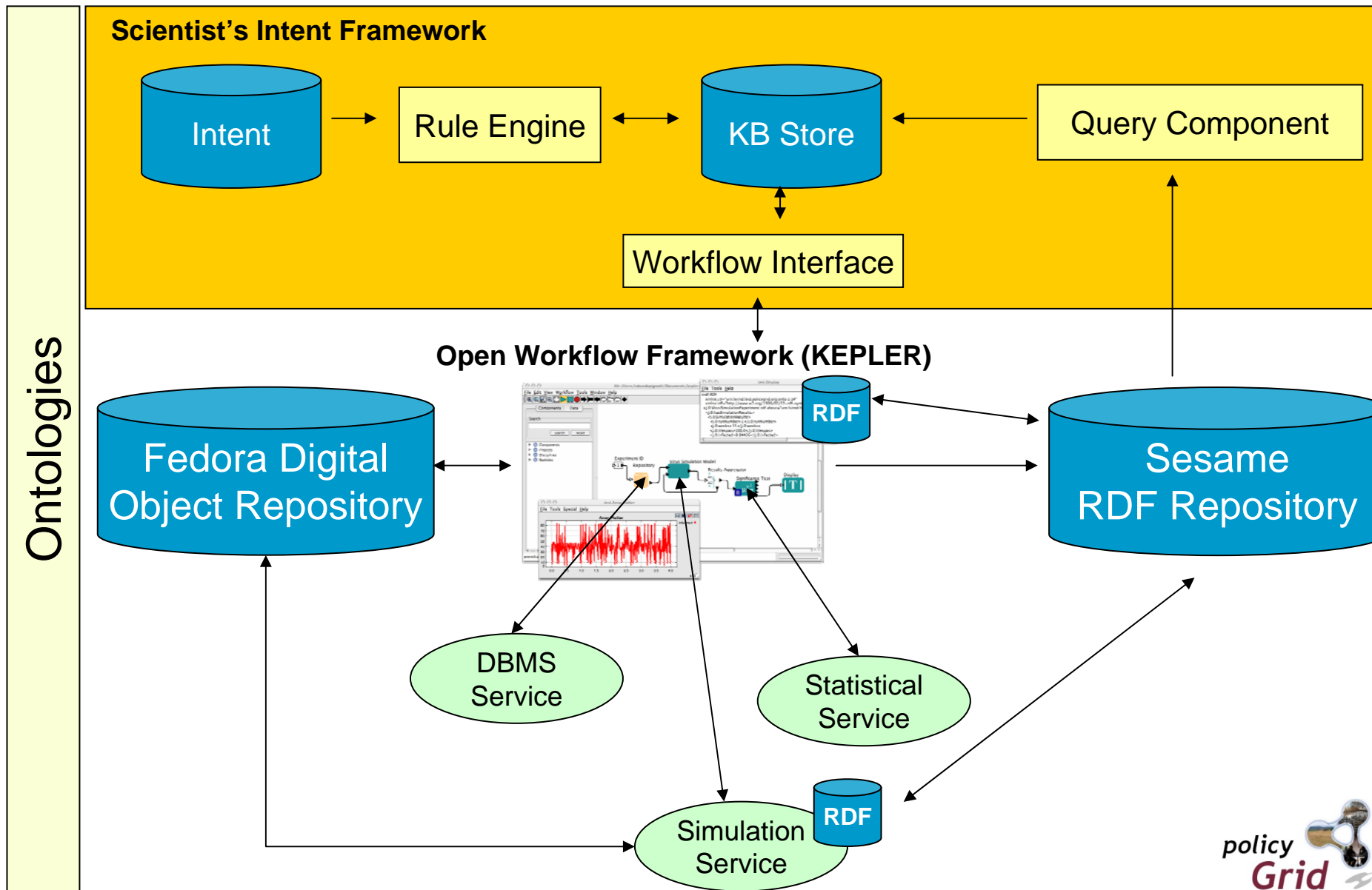
# Research Challenges

---

- Need to go beyond low-level service composition by capturing higher-level descriptions of the scientific process.
- Make the **experimental conditions** and **goals** of the experiment transparent.
- Represent “**scientist's intent**” in such a way that:
  - it is meaningful to the researcher;
  - it can be reasoned about by a software application;
  - it can be re-used across different workflows;
  - it can be used as provenance (documenting the process that led to some result).



# Semantic Workflow Architecture



# Scientist's Intent Framework

---

- Capture the logic behind scientist's intent
- Model of scientist's intent based upon rules
  - **Goal:** Run enough simulations to provide a significant number of results to support the hypothesis.

**(significant-results > 100)**

- **Constraints:** Statistical test has to run on a platform compatible with IEEE 754.

**(platform = IEEE 754)**

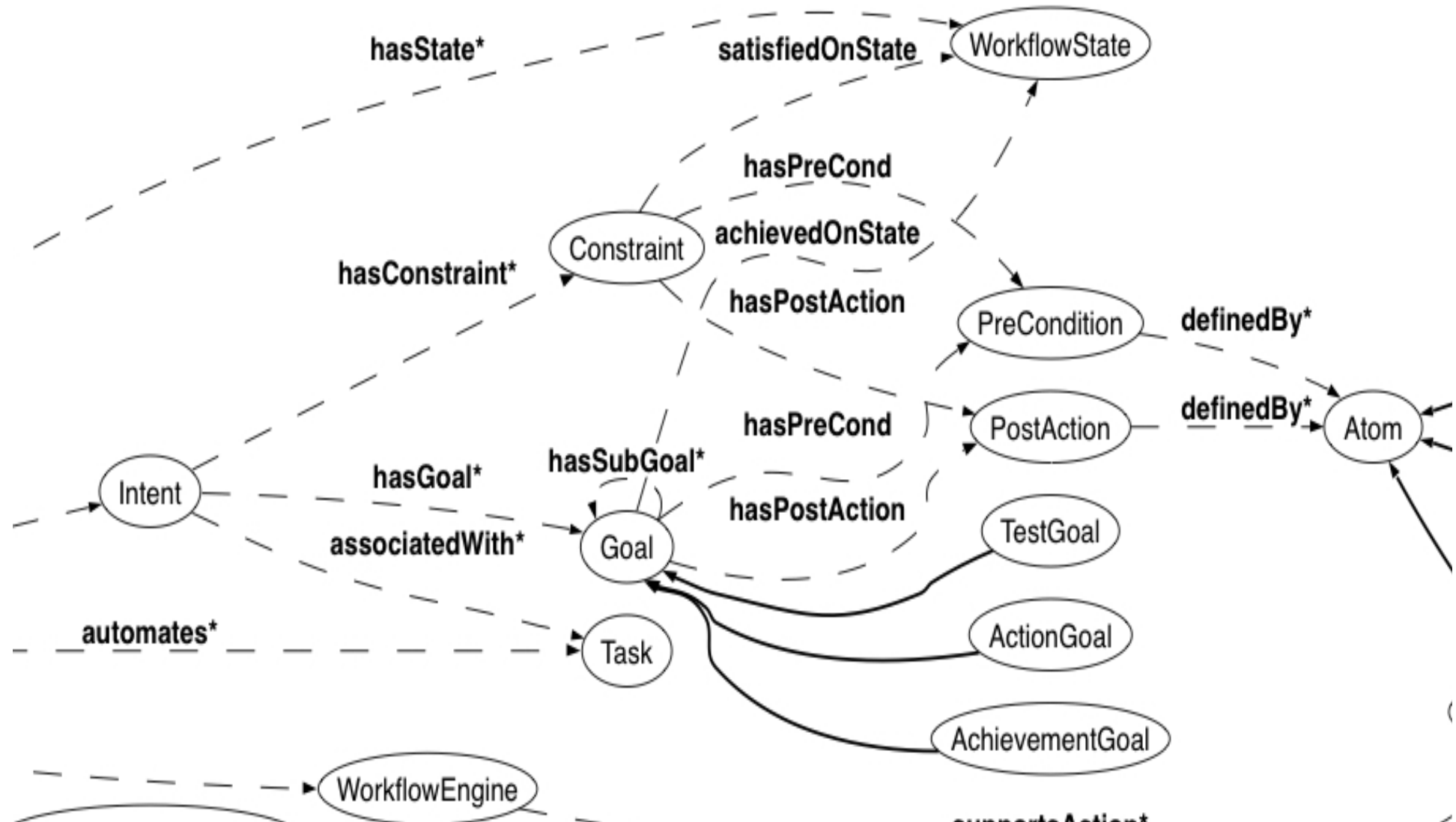
- Based on workflow metadata support and SWRL (Semantic Web Rule Language).

# Workflow Metadata Support

---

- A workflow must produce metadata to be used against scientist's intent rules.
- Possible sources of metadata:
  - metadata about the result(s) generated upon completion of the workflow;
  - metadata about the data generated at the end of an activity within the workflow or sub-workflow;
  - metadata about the status of an activity over time, e.g. while the workflow is running;
  - metadata describing the activity itself (e.g. type of service, platform etc.)

# Scientist's Intent Ontology



# Rules Example

---

*Check if the significance test activity is running on a platform compatible with IEEE 754*

```
platform(?x1,"IEEE754")    ^  
hasResult(?x1,?x2)         =>  
hasValidresult(?x1,?x2)
```

*If the number of infected people in a simulation run is more than 90%, the virus tested is epidemic.*

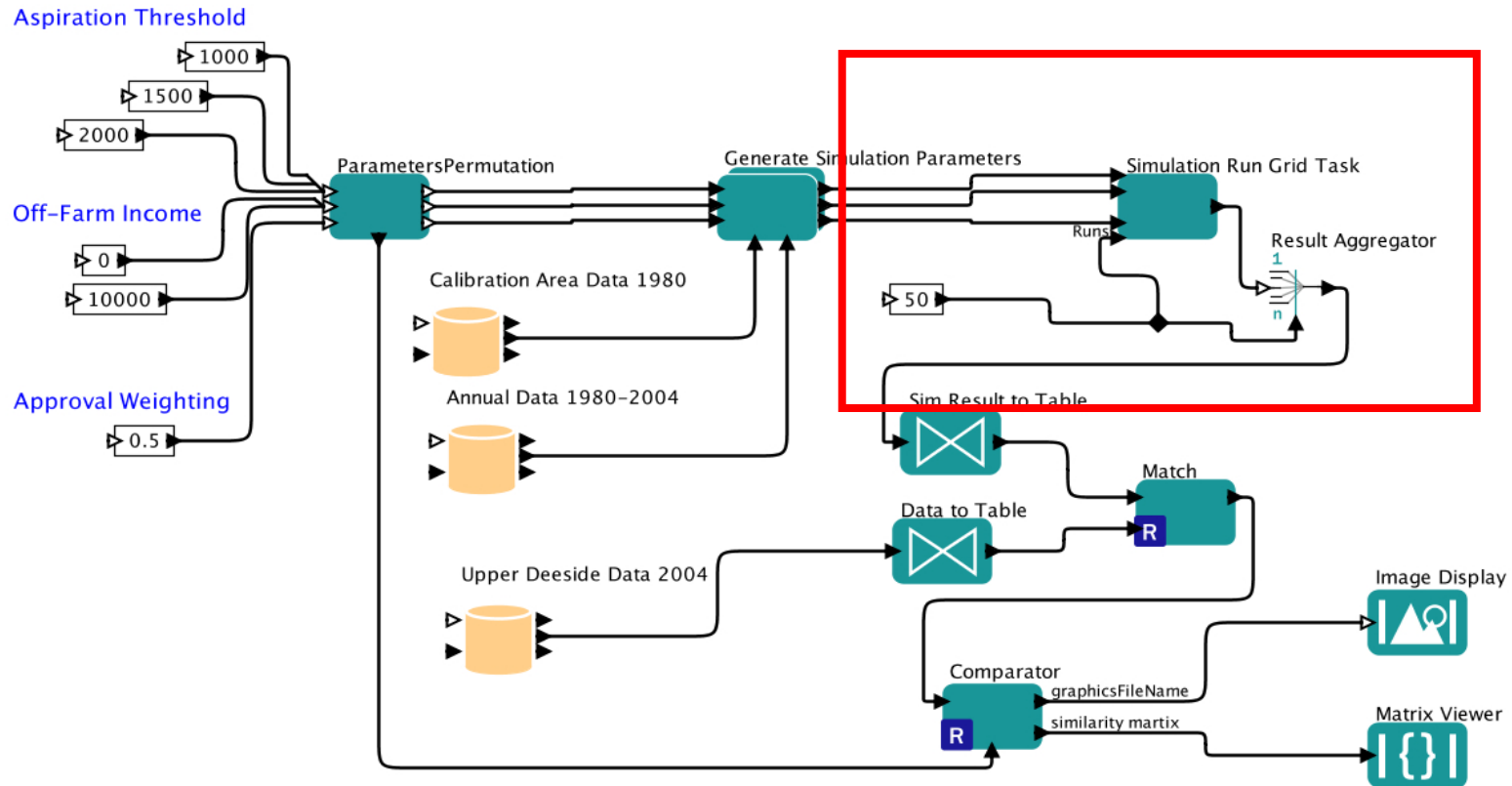
```
virus(?x1)                  ^  
virusModel(?x2)            ^  
testVirus(?x2,?x1)         ^  
hasModelRun(?x2,?x3)       ^  
infectedPeople(?x3,?x4)    ^  
more-than(?x4,90%)         =>  
isEpidemicVirus(?x1)
```

# Case-Study Upper Deeside

---

- Calibration and validation of a land-use model of upland farming in the Grampian region of Scotland.
- Quantitative data representing land-use factors (e.g. farm size, climate, yields, etc.).
- Qualitative data on farmers' decision-making drawn from semi-structured interviews are used in devising the model agents' decision algorithms.

# Case-Study Upper Deeside



**Constraint:** If in any of the 50 runs, one land manager owns more than half of the land, ignore this parameter-set.

**Goal:** Obtain at least one match where the real data falls within 95% confidence interval of the model value.

# Evaluation & Conclusions

---

- Assessing the enhanced workflow representation
  - Expressiveness of the intent formalism
  - Reusability
  - Workflow execution
- Creating and utilizing metadata is a non-trivial task
- Scale issues with simulation metadata (~250.000.000 triples per experiment).
- We aim to provide a closer connection between experimental workflows and the goals and constraints of the researcher, thus making experiments more transparent.



# Questions?

