



## **WF4Ever: PDL services integration in Taverna**

Julian Garrido

Instituto de Astrofísica de Andalucía – CSIC

[jgarrido@iaa.es](mailto:jgarrido@iaa.es)

on behalf of the Wf4Ever Team

**INTEROP, October 2012**

**Wf4Ever** - Preservation of scientific workflows in data-intensive science

EU funded FP7 STREP Project  
 December 2010 – December 2013



1. **Intelligent Software Components (ISOCO, Spain)**
2. University of Manchester (UNIMAN, UK)
3. Universidad Politécnica de Madrid (UPM, Spain)
4. Poznan Supercomputing and Networking Centre (PSNC, Poland)
5. University of Oxford (OXF, UK)
6. Instituto de Astrofísica de Andalucía (IAA, Spain)
7. Leiden University Medical Centre (LUMC, NL)



UPM



- » Taverna is an open source domain independent [Workflow Management System](#) – a suite of tools used to design and execute [scientific workflows](#).
- » Available as a command line tool and a server.
- » Allows for the **automation of experimental methods** through the use of a number of different (local or remote) services from a very diverse set of domains.
- » Allows a scientist with limited computing background and limited technical resources and support to **construct highly complex analyses**.
- » Handles parallelism, threading, monitoring, service discovery, provenance tracking.

Available services

The screenshot displays the Taverna Workbench 2.2 interface. On the left is the 'Service panel' with a filter and a list of 'Available services' including Biomart, Biomoby, Soaplab, and various WSDL files. Below this is the 'Workflow explorer' showing a tree view of the 'EBI\_InterProScan' workflow structure, including input and output ports and services. On the right is the 'Workflow diagram' showing a flowchart of the workflow. It starts with 'Workflow input ports' (params\_async\_defaultValue, Job\_params\_crc\_defaultValue, Job\_params\_seqtype\_defaultValue, Email\_address, Sequen) feeding into 'Job\_params'. This leads to 'runInterProScan', which then branches into 'checkStatus' (with 'Job\_ID' as an input) and 'Get\_XML\_result'. 'checkStatus' outputs 'Is\_done' and 'Job\_status'. 'Get\_XML\_result' outputs 'InterProScan\_XML\_result'. 'Job\_status' leads to 'Unpack\_XML\_result'. 'Get\_XML\_result' also leads to 'Get\_text\_result', which then leads to 'Unpack\_text\_result'. 'Unpack\_text\_result' outputs 'InterProScan\_GFF' and 'InterProScan\_text\_result'. 'InterProScan\_GFF' leads to 'Format\_as\_GFF', which outputs 'InterProScan\_GFF'. The final 'Workflow output ports' are 'InterProScan\_XML\_result', 'InterProScan\_GFF', 'InterProScan\_text\_result', and 'Job\_ID'.

Tree view of workflow structure

Workflow

- » Parameter Description Language: A generic grammar for parameters (and relates constraints) description. Presented in the INTEROP as a working draft. (Carlo Maria Zwölf, Franck le Petit, Paul Harrison)
- » PDL aim is to answer to two major issues in scientific services, (description and interoperability needs.

### Description needs

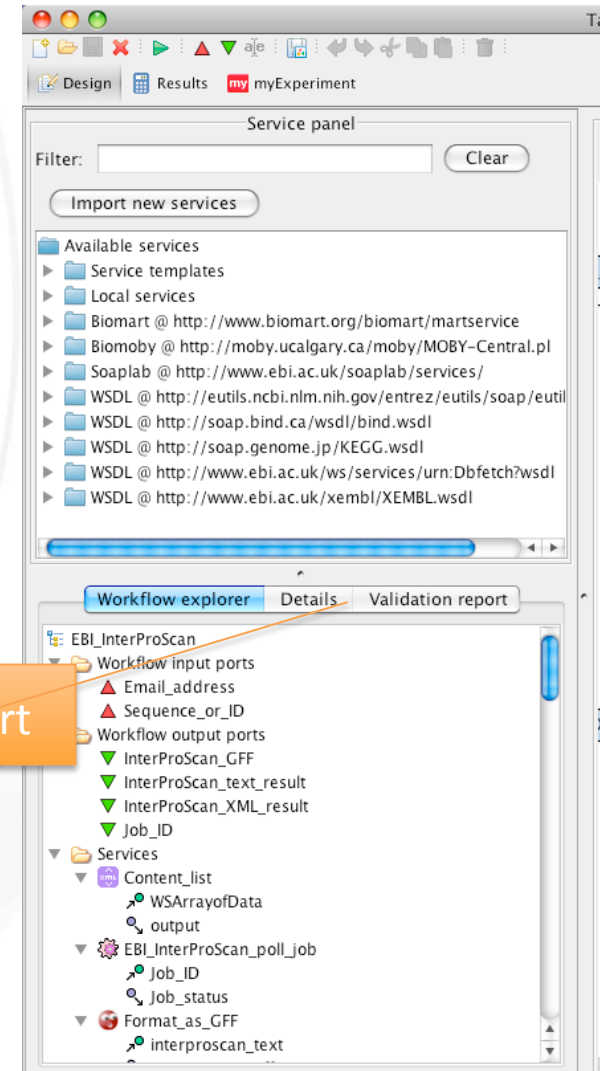
Describe physical properties of parameters				
Nature	Meaning	Unit	Precision	Range
Describe complex relations involving parameters				
Physical constraints		Arbitrary Conditions	Mathematical Conditions	

### Interoperability needs

Interaction of two services has sense if the parameter sent by the first and expected by the second have same		
Computer type	Physical concept	Unit
Interaction of two services has sense if all preconditions of second service are satisfied by output of first one		

- » Running time:
  - › PDL Validation client. Do input/output data fit the restrictions?
- » Design time:
  - › The output of a service is the input of another.
  - › Computing interoperability (precision, SKOS concept, Unit, UCD, Utype)

Validation report



» Missing points in Taverna:

- › Include validation capabilities. Running without errors doesn't mean the result is correct.
- › Improve service discovery and recommendations in Taverna:  
Which services can I connect to the output of another?

???