Tracking provenance information with the OPUS job controller

Mathieu Servillat

Observatoire de Paris - LUTH
Paris Astronomical Data Centre

IVOA Santiago - Oct. 2017

OPUS: Observatoire de Paris UWS Server

Main features

- IVOA standards
  - Universal Worker System (**UWS**)
  - **Provenance** data model
- REST architecture
  - Python micro-framework: bottle.py
- Collaborative development
  - Git server at PADC (gitolite)
  - GitHub: https://github.com/mservillat/OPUS
- Tests and quality
  - Unit tests with `unittest` and webtest
  - Activity history with `logging`

Prototype available

- https://voparis-uws-test.obspm.fr

Job management at PADC

Available structure

- Work cluster (Tycho)
- Job scheduler (SLURM)

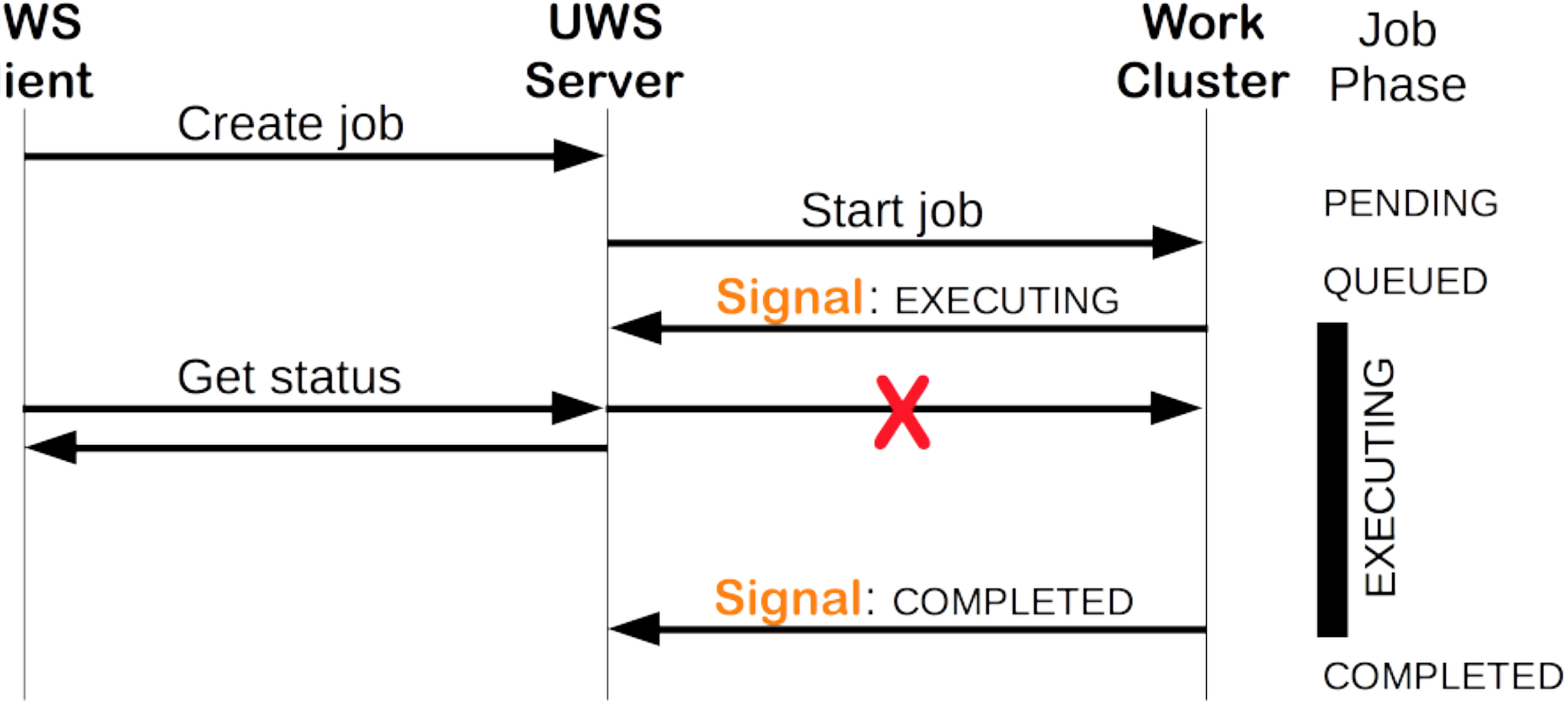
PADC projects

- Web based clients
  - Data access
  - Online data processing
  - Wrap simulation codes

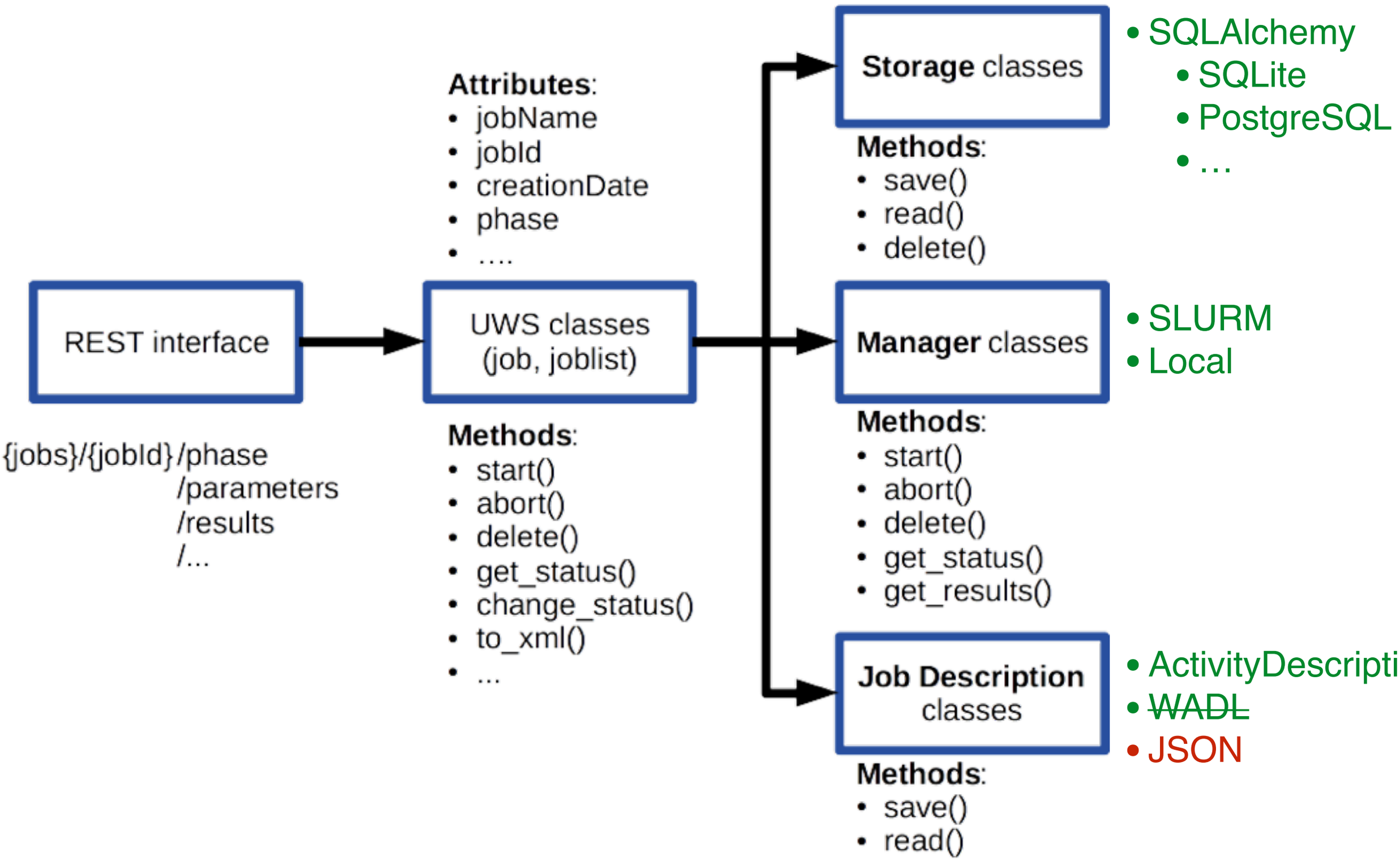
Need a simple interface to computational resources

Inner workings

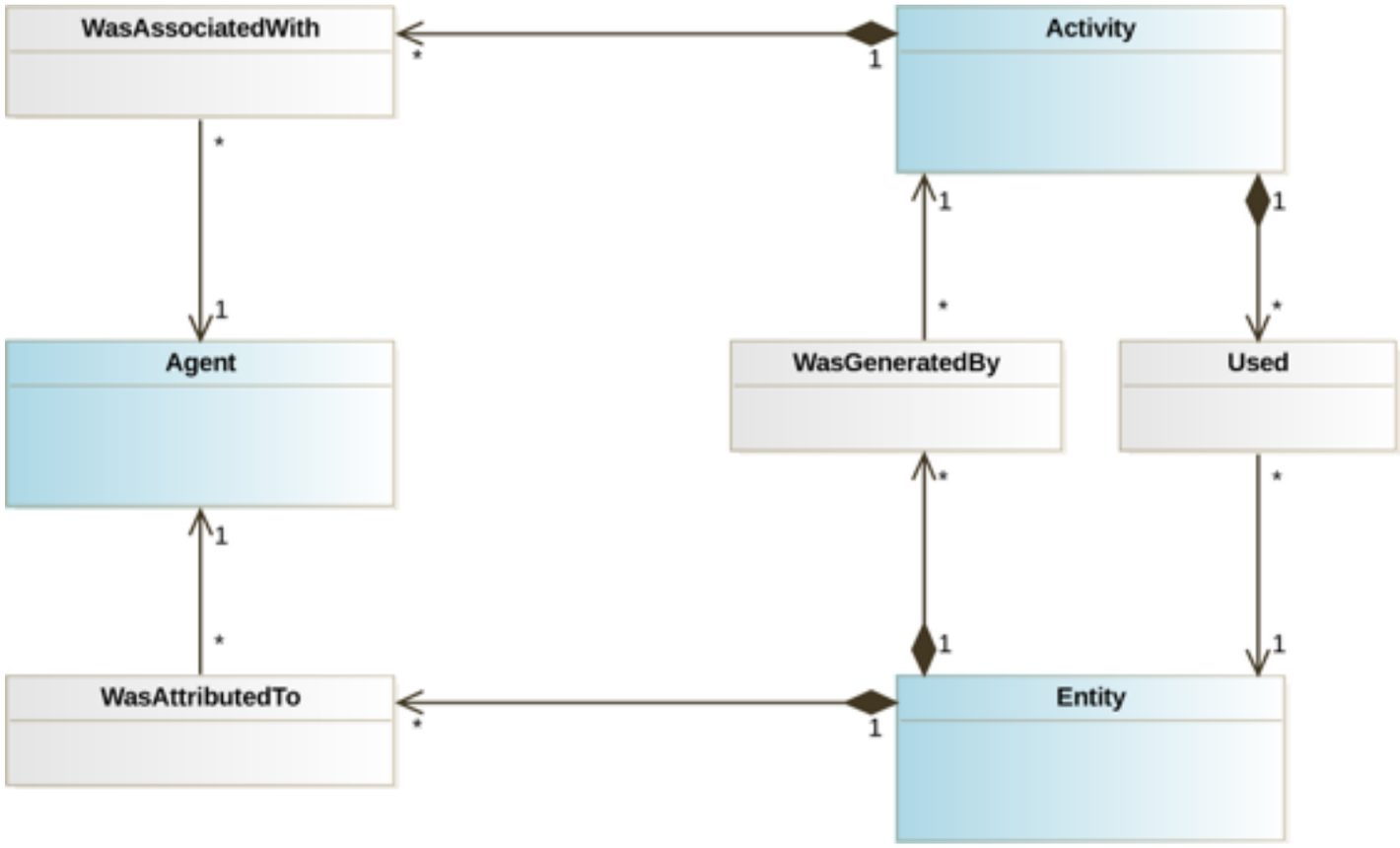
- Separate **job controller** from **work cluster**
  - Wait for work cluster **signals**
  - Avoid (too many) status queries to work cluster



Server main classes



Core Provenance Data Model



http://www.w3.org/TR/prov-overview/

30 April 2013

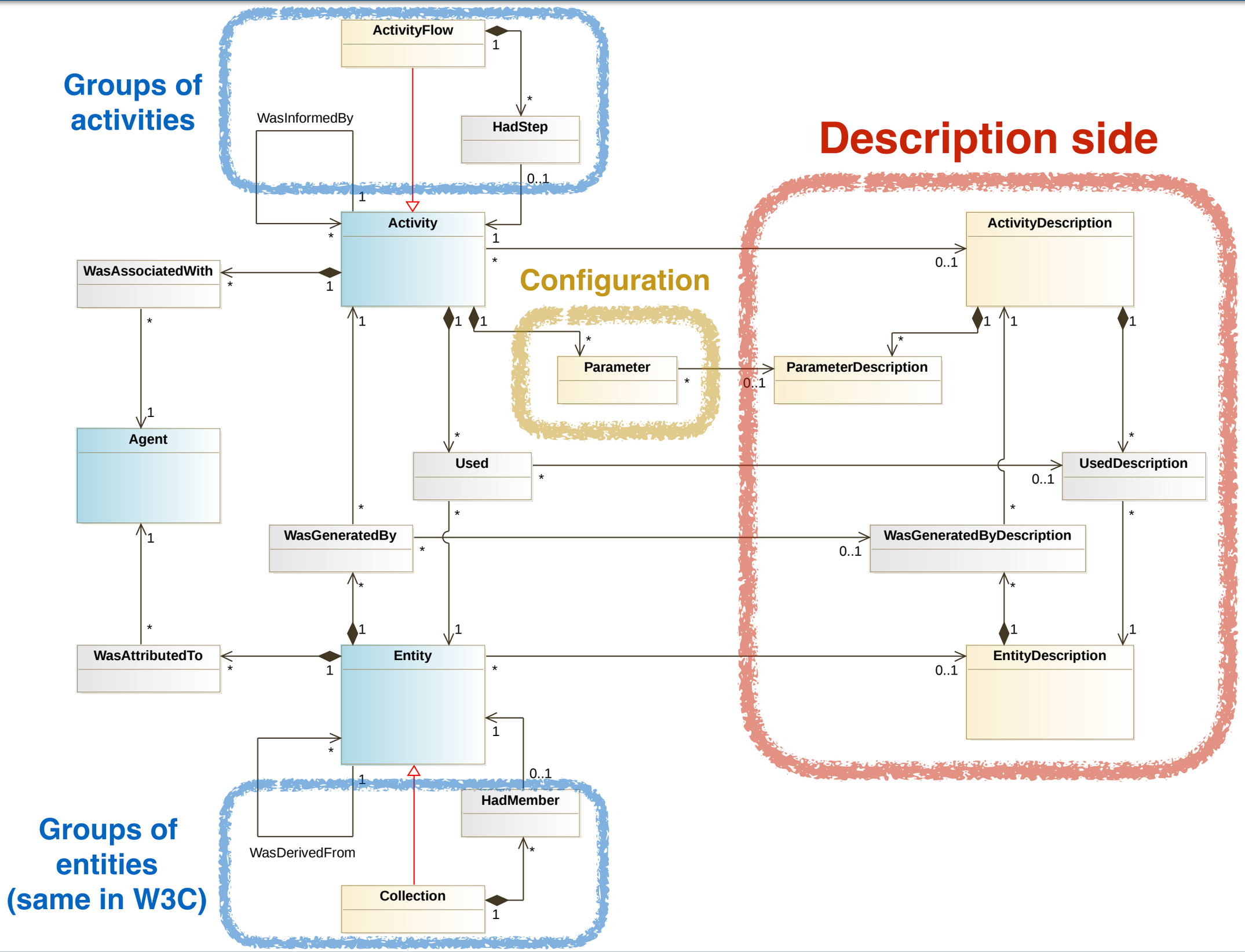
Word Wide Web Consortium

http://www.ivoa.net/documents/ProvenanceDM/

IVOA

- Core concepts from the W3C PROV recommendations
  - **Entity - Activity - Agent**
  - **Relations** and **roles** = provenance information
  - W3C PROV has many more relations
  - IVOA Provenance connected to **VO concepts** and **astronomy needs**

IVOA Provenance Data Model diagram



Serializations - ActivityDescription

```
<?xml version="1.0" encoding="UTF-8"?>
<VOTABLE xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.ivoa.net/xml/VOTable/v1.3" version="1.3"
    xsi:schemaLocation="http://www.ivoa.net/xml/VOTable/v1.3 http://www.ivoa.net/xml/VOTable/v1.3">
<RESOURCE ID="make_RGB_image" name="make_RGB_image" type="meta" utype="voprov:ActivityDescription">
    <DESCRIPTION>Create an RGB image from 3 images</DESCRIPTION>
    <PARAM name="name" datatype="char" arraysize="*" value="make_RGB_image" utype="voprov:ActivityDescription.label" />
    <PARAM name="type" datatype="char" arraysize="*" value="..." utype="voprov:ActivityDescription.type"/>
    <PARAM name="subtype" datatype="char" arraysize="*" value="..." utype="voprov:ActivityDescription.subtype" />
    <PARAM name="version" datatype="float" value="..." utype="voprov:ActivityDescription.version" />
    <PARAM name="doculink" datatype="char" arraysize="*" value="..." utype="voprov:ActivityDescription.doculink" />
    <PARAM name="contact_name" datatype="char" arraysize="*" value="..." utype="voprov:Agent.name" />
    <PARAM name="contact_email" datatype="char" arraysize="*" value="...@..." utype="voprov:Agent.email" />

    <GROUP name="InputParams" utype="voprov:ParameterDescription">
        <PARAM ID="RGB" arraysize="*" datatype="char" name="RGB" type="no_query" value="RGB.jpg">
            <DESCRIPTION>RGB image name</DESCRIPTION>
        </PARAM>
    </GROUP>
```

—> VOTable PARAM + attributes

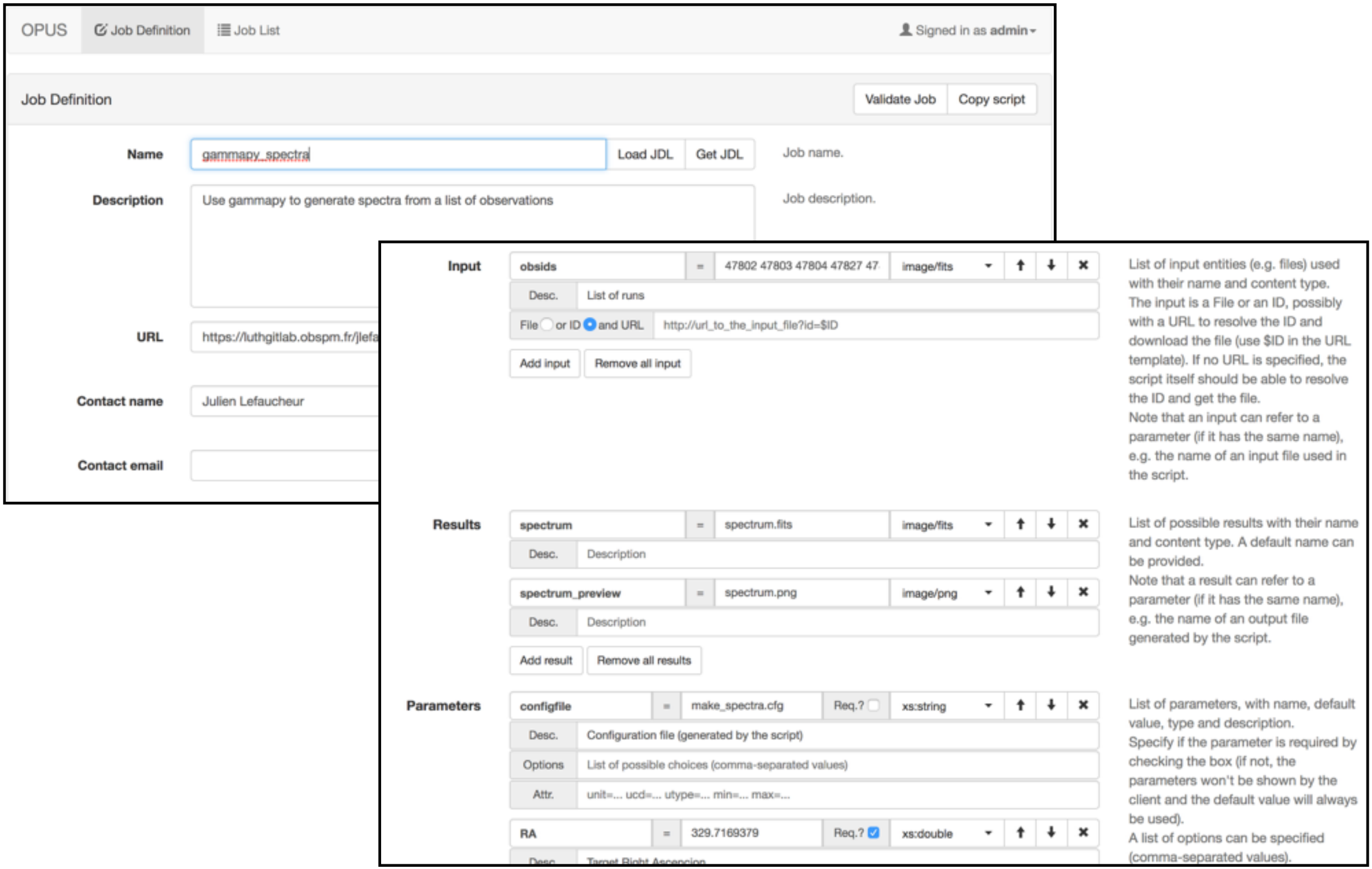
```
    <GROUP name="Used" utype="voprov:UsedDescription">
        <GROUP name="R" utype="voprov:EntityDescription">
            <DESCRIPTION>Image for red channel</DESCRIPTION>
            <PARAM name="default" value="R.jpg" arraysize="*" datatype="char" utype="voprov:Entity.id" />
            <PARAM name="role" value="red" arraysize="*" datatype="char" utype="voprov:UsedDescription.role" />
            <PARAM name="content_type" value="image/jpeg" arraysize="*" datatype="char" utype="voprov:EntityDescription.content_type" />
        </GROUP>
<!--[...]-->
    </GROUP>

    <GROUP name="Generated" utype="voprov:WasGeneratedByDescription">
        <GROUP name="RGB" utype="voprov:EntityDescription">
            <DESCRIPTION>RGB image generated</DESCRIPTION>
            <PARAM name="role" value="RGB" arraysize="*" datatype="char" utype="voprov:WasGenereratedByDescription.role" />
            <PARAM name="content_type" value="image/jpeg" arraysize="*" datatype="char" utype="voprov:EntityDescription.content_type" />
        </GROUP>
    </GROUP>

</RESOURCE>
</VOTABLE>
```

—> DataLink Service Descriptor + Used/Generated groups

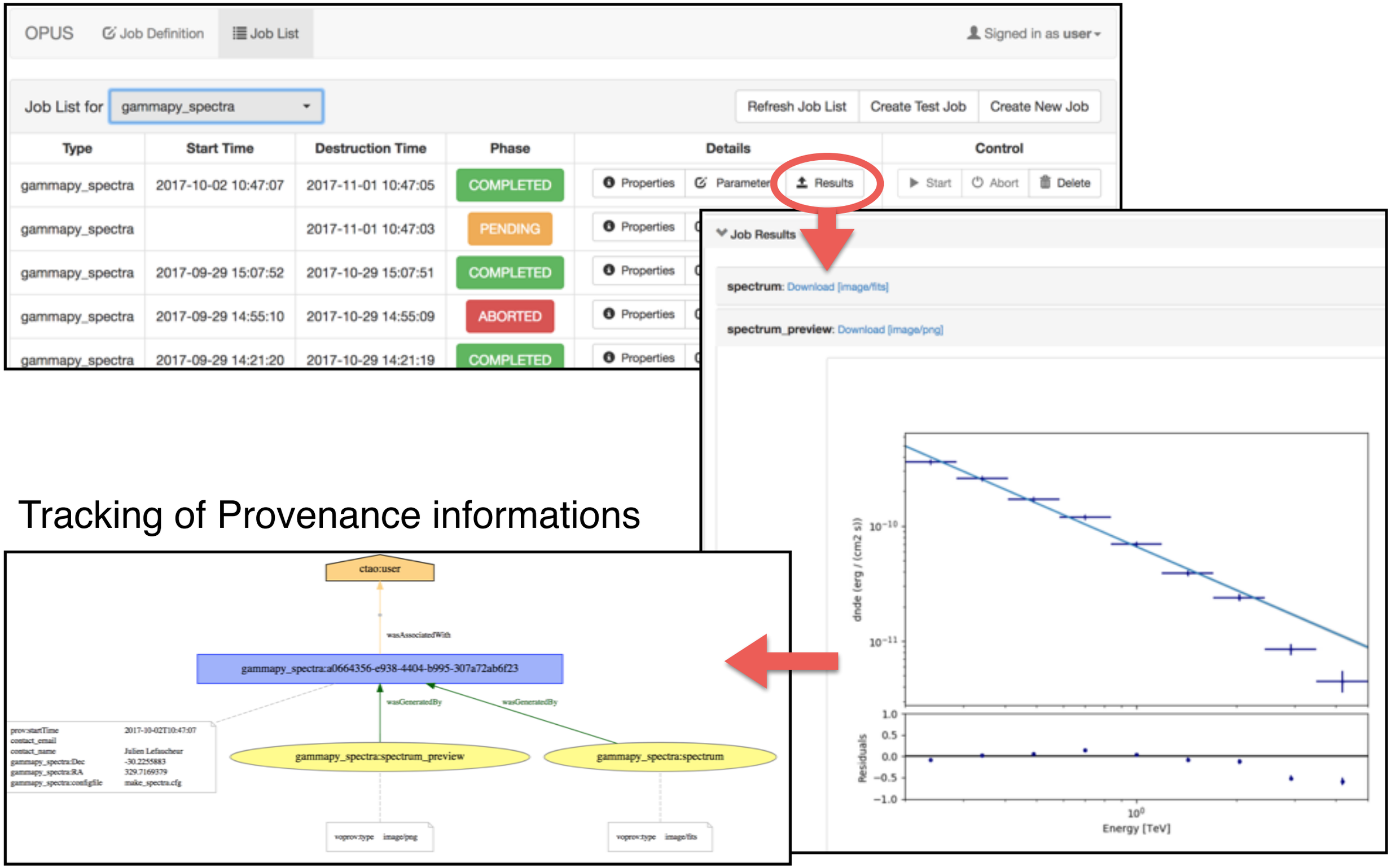
Client main features

- Javascript based + Python Flask
  - `UwsLib.js`: sends AJAX requests to the server
  - `uws_manager.js`: handles and displays responses
    - Integration with Bootstrap3
    - HTML pages with specified elements (id=joblist, parameters, results…)
- **Job definition editor**
  - Interface to create the ActivityDescription file
    - parameters, used and generated entities
    - bash execution script
- **Job manager**
  - list jobs, create jobs, control jobs, view results

Client - Job definition



Client - Job list



Tracking of Provenance informations

Serializations - W3C PROV formats

```
<prov:document xmlns:ctadata="ivo://vopdc.obspm/cta#" xmlns:ctajob
  <prov:activity prov:id="ctajobs:ctbin">
    <prov:startTime> 2016-03-13T23:44:46 </prov:startTime>
    <prov:endTime> 2016-03-13T23:44:56 </prov:endTime>
  </prov:activity>
  <prov:agent prov:id="cta:consortium">
    <prov:type xsi:type="xsd:string"> Organization </prov:type>
  </prov:agent>
  <prov:wasAssociatedWith>
    <prov:activity prov:ref="ctajobs:ctbin" />
    <prov:agent prov:ref="cta:consortium" />
  </prov:wasAssociatedWith>
  <prov:entity prov:id="uwsdata:parameters/inobs" />
  <prov:used>
    <prov:activity prov:ref="ctajobs:ctbin" />
    <prov:entity prov:ref="uwsdata:parameters/inobs" />
  </prov:used>
  <prov:entity prov:id="uwsdata:results/outcube" />
  <prov:wasGeneratedBy>
    <prov:entity prov:ref="uwsdata:results/outcube" />
    <prov:activity prov:ref="ctajobs:ctbin" />
  </prov:wasGeneratedBy>
  <prov:wasDerivedFrom>
    <prov:generatedEntity prov:ref="uwsdata:results/outcube" />
    <prov:usedEntity prov:ref="uwsdata:parameters/inobs" />
  </prov:wasDerivedFrom>
  <prov:entity prov:id="uwsdata:results/logfile" />
  <prov:wasGeneratedBy>
    <prov:entity prov:ref="uwsdata:results/logfile" />
    <prov:activity prov:ref="ctajobs:ctbin" />
  </prov:wasGeneratedBy>
  <prov:wasDerivedFrom>
    <prov:generatedEntity prov:ref="uwsdata:results/logfile" />
    <prov:usedEntity prov:ref="uwsdata:parameters/inobs" />
  </prov:wasDerivedFrom>
</prov:document>
```

```
{
  - wasAssociatedWith: {
      - _:id1: {
            prov:agent: "cta:consortium",
            prov:activity: "cta:anactools_v1.1"
        }
    },
  - agent: {
      - cta:consortium: {
            prov:type: "Organization"
        }
    },
  - entity: {
        uwsdata:results/fit_results: { },
        uwsdata:results/configfile: { },
        uwsdata:results/butterfly: { },
        uwsdata:results/spectrum_plot: { },
        uwsdata:results/spectrum: { }
    },
  - prefix: {
        uwsdata: "https://voparis-uws-test.obspm.fr/rest
        cta: "http://www.cta-observatory.org#",
        voprov: "http://www.ivoa.net/ns/voprov#"
    },
  - activity: {
      - cta:anactools_v1.1: {
            prov:startTime: "2016-04-07T00:26:00",
            prov:endTime: "2016-04-07T00:27:15"
        }
    },
  - wasGeneratedBy: {
      - _:id5: {
            prov:entity: "uwsdata:results/butterfly",
            prov:activity: "cta:anactools_v1.1"
        },
      - _:id4: {
            prov:entity: "uwsdata:results/fit_results",
            prov:activity: "cta:anactools_v1.1"
        },
```