# Pollux and Provenance

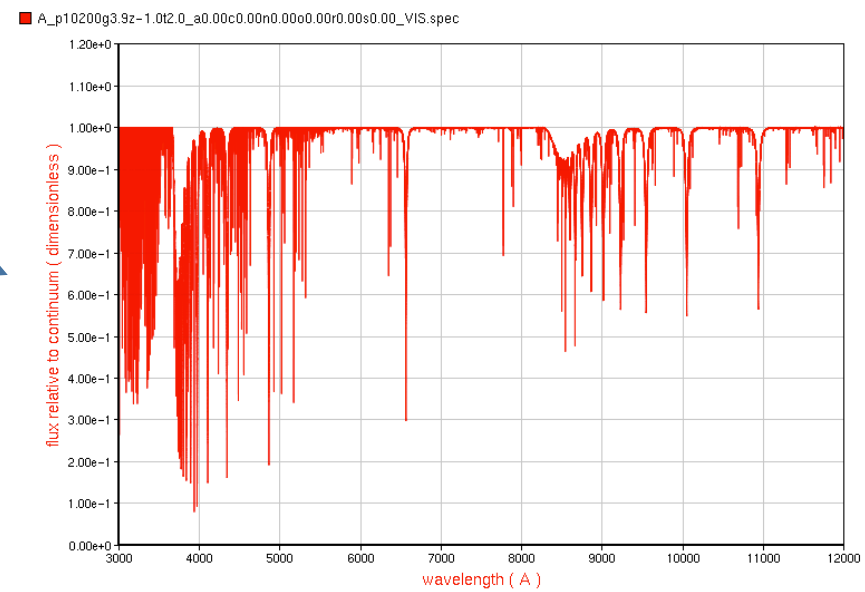
Scientists : Ana Palacios, Agnès Lèbre

Software engineer : Michèle Sanguillon

# Pollux Context

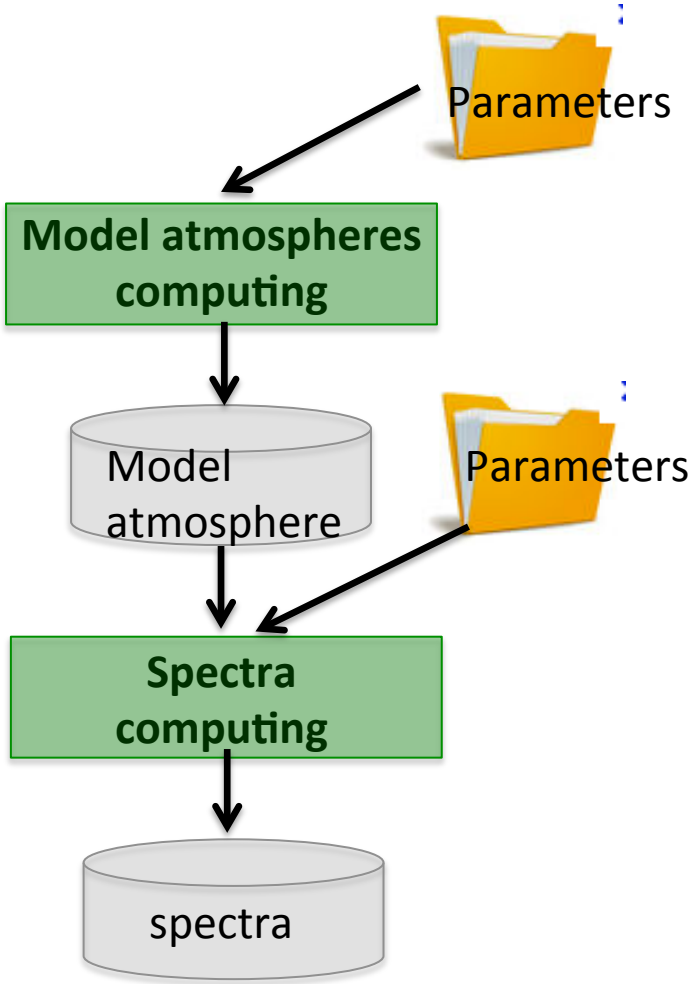
temperature = 5000°K
gravity = …
…

POLLUX database collects and presents synthetic spectra computed* at high resolution.




* The computing of the spectra is done by the producers and not done on the fly.

# Pollux data and workflow

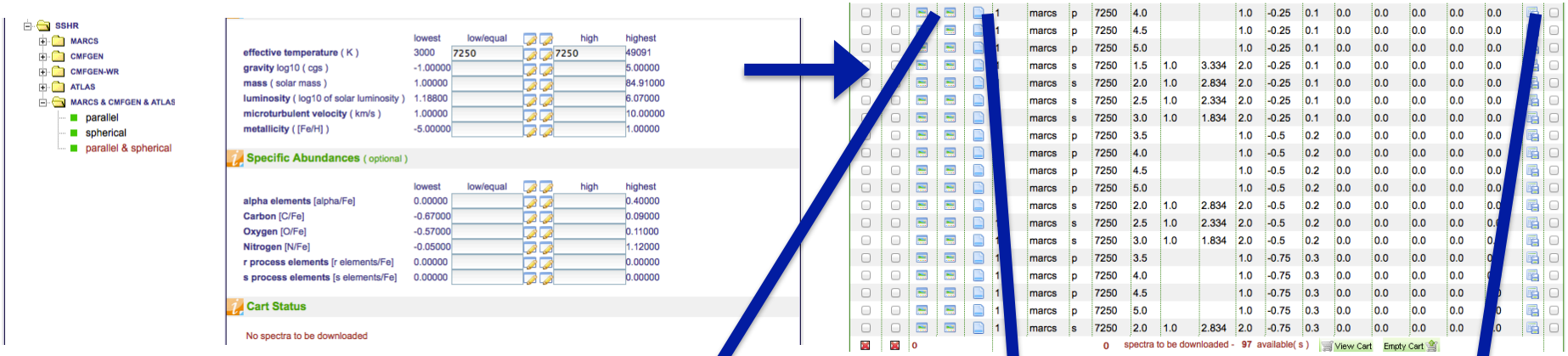



Only the spectra are available on the VO

# Pollux accessibility

- via Web
  - Spectra in flat, fits, xml, votable formats (header + data)
- via SSA protocol for OV users
  - Spectra in fits format (header + data)

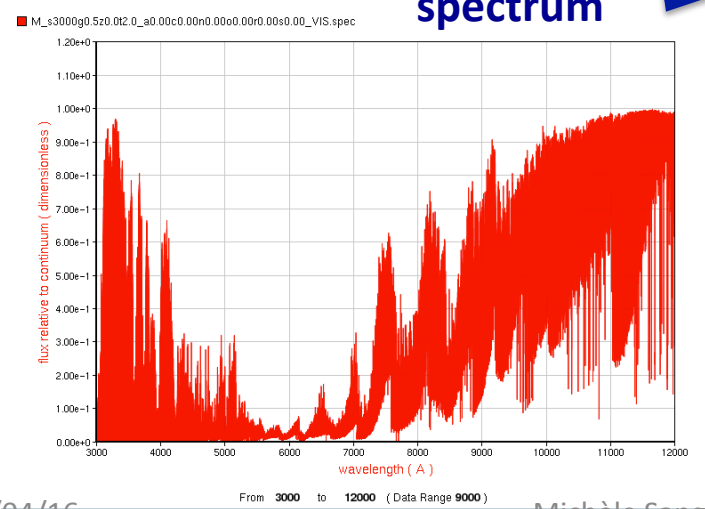
# Pollux accessibility via Web




**Display spectrum**

**Header**

**Download**



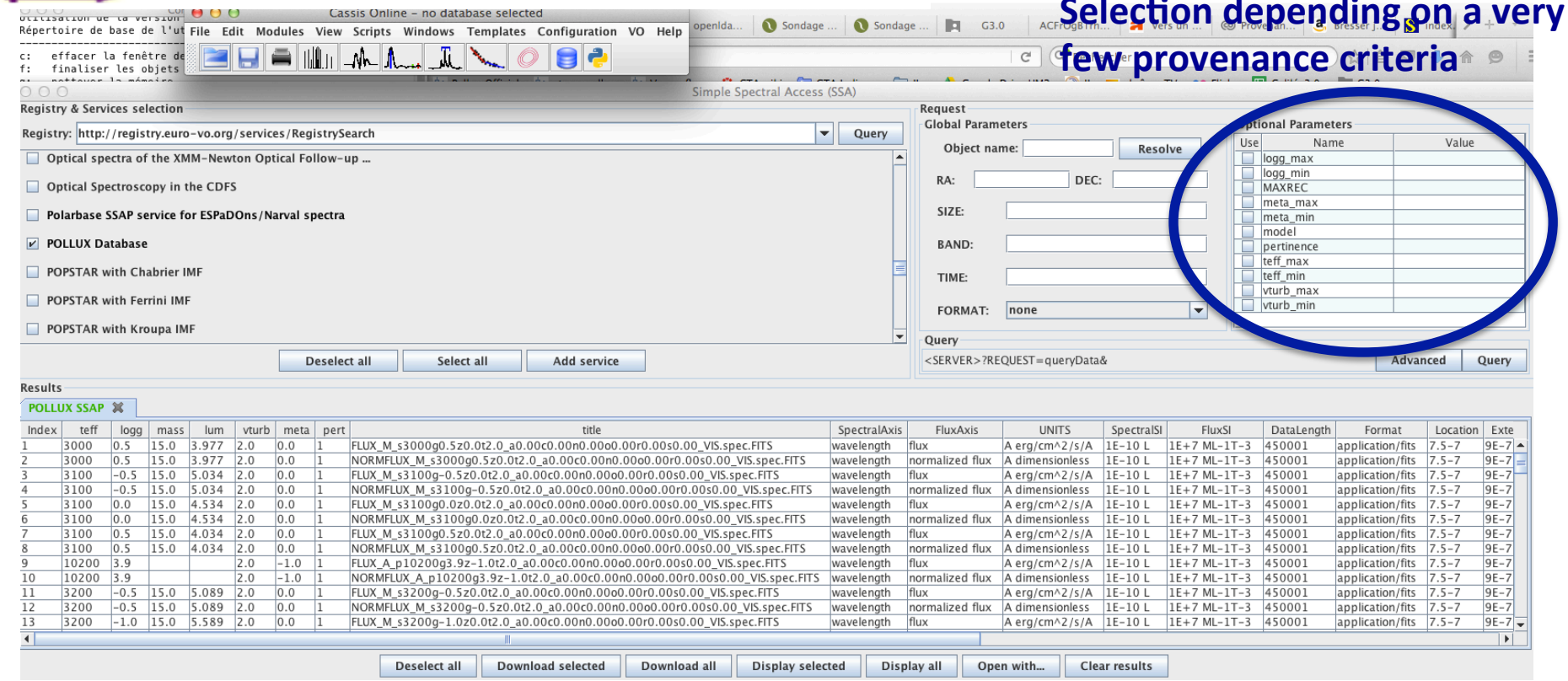

```
Spectrum: M_s3000g0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0

code1            = marcs35            code for atmosphere model
version1         = 2006.1             version of code for model
type             = s                  type of model atmosphere
filename         = 3000g0.5m15z0.00t2.0.mod model atmoshere filen
author_mod       = plez               model atmosphere creator

Teff             = 3000               effective temperature (K)
logg             =  0.5               log10(gravity) (cgs)
mass             = 15.0               mass (solar mass)
lum              =  3.977             log of luminosity (solar
turbvel          = 2.00               microturbulent velocity (

conv_alpha       = 1.5                convection parameter (con
conv_ny          = 8.0                convection parameter (con
conv_y           = 0.076              convection parameter (con
conv_beta        = 0.5                convection parameter (con
macroturbvel     = 0.0                macroturbulence parameter
macrobeta        = 0.0                macroturbulence parameter

Mdot             = irrelevant         log10(mass loss) (solar m
Vinfty           = irrelevant         terminal velocity (km/s)
beta             = irrelevant         velocity law parameter
finfty           = irrelevant         1st clumping law paramete
vcl              = irrelevant         2nd clumping law paramete

metallic_mod     = 0.000              metallicity ([Fe/H])
alpha_mod        = 0.000              [alpha/Fe]
r_process_mod    = 0.000              [r elements/Fe]
s_process_mod    = 0.000              [s elements/Fe]
```

**Provenance information**

# Pollux accessibility via a VO tool

**CASSIS**
**A free interactive spectrum analyser**

**Selection depending on a very few provenance criteria**




| Index | teff | logg | mass | lum | vturb | meta | pert | title | SpectralAxis | FluxAxis | UNITS | SpectralSI | FluxSI | DataLength | Format | Location | Exte |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3000 | 0.5 | 15.0 | 3.977 | 2.0 | 0.0 | 1 | FLUX_M_s3000g0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | flux | A erg/cm^2/s/A | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 2 | 3000 | 0.5 | 15.0 | 3.977 | 2.0 | 0.0 | 1 | NORMFLUX_M_s3000g0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | normalized flux | A dimensionless | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 3 | 3100 | -0.5 | 15.0 | 5.034 | 2.0 | 0.0 | 1 | FLUX_M_s3100g-0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | flux | A erg/cm^2/s/A | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 4 | 3100 | -0.5 | 15.0 | 5.034 | 2.0 | 0.0 | 1 | NORMFLUX_M_s3100g-0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | normalized flux | A dimensionless | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 5 | 3100 | 0.0 | 15.0 | 4.534 | 2.0 | 0.0 | 1 | FLUX_M_s3100g0.0z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | flux | A erg/cm^2/s/A | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 6 | 3100 | 0.0 | 15.0 | 4.534 | 2.0 | 0.0 | 1 | NORMFLUX_M_s3100g0.0z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | normalized flux | A dimensionless | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 7 | 3100 | 0.5 | 15.0 | 4.034 | 2.0 | 0.0 | 1 | FLUX_M_s3100g0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | flux | A erg/cm^2/s/A | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 8 | 3100 | 0.5 | 15.0 | 4.034 | 2.0 | 0.0 | 1 | NORMFLUX_M_s3100g0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | normalized flux | A dimensionless | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 9 | 10200 | 3.9 | | | 2.0 | -1.0 | 1 | FLUX_A_p10200g3.9z-1.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | flux | A erg/cm^2/s/A | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 10 | 10200 | 3.9 | | | 2.0 | -1.0 | 1 | NORMFLUX_A_p10200g3.9z-1.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | normalized flux | A dimensionless | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 11 | 3200 | -0.5 | 15.0 | 5.089 | 2.0 | 0.0 | 1 | FLUX_M_s3200g-0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | flux | A erg/cm^2/s/A | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 12 | 3200 | -0.5 | 15.0 | 5.089 | 2.0 | 0.0 | 1 | NORMFLUX_M_s3200g-0.5z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | normalized flux | A dimensionless | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |
| 13 | 3200 | -1.0 | 15.0 | 5.589 | 2.0 | 0.0 | 1 | FLUX_M_s3200g-1.0z0.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.FITS | wavelength | flux | A erg/cm^2/s/A | 1E-10 L | 1E+7 ML-1T-3 | 450001 | application/fits | 7.5-7 | 9E-7 |

**Currently no possibility to see the provenance information,**
**but a DATALINK could be added to download the header file in the « flat » format.**

# Our motivations for Provenance

- **1) A Data Model** to describe our data (spectra and provenance) => use of utypes for every piece of information**:**
  - Spectrum DM 1.1 / Spectral DM 2.0:
    - adequate for spectra
    - with a ObsConfig part
    - provenance information not included
  - SimDM:
    - adequate for simulations
    - don't describe the spectrum
  - Spectral DM + ProvDM or SimDM?
    - Will describe spectra and the way they were produced

# Our motivations for Provenance

**2)** A way to **display** all our provenance information in VO spectra tools, easy readable and **normalized**

## Flat header

## (Web + VO users in the future)

```
header_name_SSHR = 'M_p7250g5.0z1.00t1.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec.txt'
short_name_SSHR  = 'M_p7250g5.0z1.00t1.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec'
Key_SSHR         = 'M_p7250g5.0z1.00t1.0_a0.00c0.00n0.00o0.00r0.00s0.00_conva1.5convny8.0convy0.076convb0.5mt(

code1          = 'marcs'                    / code for atmosphere model
version1       = '2008.5'                   / version of code for model atmosphere
type           = 'p'                        / type of model atmosphere (Spherical/Parallel)
filename       = 'p7250_g+5.0_m0.0_t01_st_z+1.00_a+0.00_c+0.00_n+0.00_o+0.00_r+0.00_s+0.00.mod'  / model atm(
author_mod     = 'Marcs-team'               / model atmosphere creator name

Teff           = '7250'                     / effective temperature (K)     - model atmosphere data
logg           = ' 5.0'                      / log10(gravity) (cgs)          - model atmosphere data
mass           = 'irrelevant'               / mass (solar mass)              - model atmosphere data
lum            = 'irrelevant'               / luminosity (solar luminosity)  - model atmosphere data
turbvel        = '1.00'                     / microturbulent velocity (km/s) - model atmosphere data

conv_alpha     = '1.5'                      / convection parameter (conva)   - model atmosphere data
conv_ny        = '8.0'                      / convection parameter (convny)  - model atmosphere data
conv_y         = '0.076'                    / convection parameter (convy)   - model atmosphere data
conv_beta      = '0.5'                      / convection parameter (convb)   - model atmosphere data
macroturbvel   = '0.0'                      / macroturbulence parameter (mt) - model atmosphere data
macrobeta      = '0.0'                      / macroturbulence parameter (mb) - model atmosphere data

Mdot           = 'irrelevant'               / log10(mass loss) (solar mass/year) - model atmosphere data
Vinfty         = 'irrelevant'               / terminal velocity (km/s)           - model atmosphere data
beta           = 'irrelevant'               / velocity law parameter             - model atmosphere data
finfty         = 'irrelevant'               / 1st clumping law parameter         - model atmosphere data
vcl            = 'irrelevant'               / 2nd clumping law parameter (km/s)  - model atmosphere data

metallic_mod   = '1.00'                    / metallicity ([Fe/H])
alpha_mod      = '0.000'                    / [alpha/Fe]
r_process_mod  = '0.000'                    / [r elements/Fe]
s_process_mod  = '0.000'                    / [s elements/Fe]
```

Easy to read and to understand

Keywords and format not normalized

No UCD, no utypes

## Fits header

## (OV)

```
<PARAM arraysize="60" datatype="char" name="short_name_SSHR" value="M_p7250g5.0z1.00t1.0_a0.00c0.00n0.00o
<PARAM arraysize="106" datatype="char" name="Key_SSHR" value="M_p7250g5.0z1.00t1.0_a0.00c0.00n0.00o0.00r0
<PARAM arraysize="5" datatype="char" name="code1" ucd="meta.id;meta.code" value="marcs">
<DESCRIPTION>code for atmosphere model</DESCRIPTION>
</PARAM>
<PARAM datatype="float" name="version1" ucd="meta.code;meta.version" value="2008.5">
<DESCRIPTION>version of code for model atmosphere</DESCRIPTION>
</PARAM>
<PARAM arraysize="1" datatype="char" name="type" ucd="meta.code.class" value="p">
<DESCRIPTION>type of model atmosphere (Spherical/Parallel)</DESCRIPTION>
</PARAM>
<PARAM arraysize="76" datatype="char" name="filename" ucd="meta.id;meta.file" value="p7250_g+5.0_m0.0_t01
<DESCRIPTION>model atmosphere filename</DESCRIPTION>
</PARAM>
<PARAM arraysize="10" datatype="char" name="author_mod" ucd="meta.bib.author" value="Marcs-team">
<DESCRIPTION>model atmosphere creator name</DESCRIPTION>
</PARAM>
<PARAM datatype="int" name="Teff" ucd="phys.temperature.effective" unit="K" value="7250">
<DESCRIPTION>effective temperature (K)     - model atmosphere data</DESCRIPTION>
</PARAM>
<PARAM datatype="float" name="logg" ucd="phys.gravity;arith.zp" unit="log(cm/s2)" value="5.0">
<DESCRIPTION>log10(gravity) (cgs)          - model atmosphere data</DESCRIPTION>
</PARAM>
<PARAM arraysize="10" datatype="char" name="mass" ucd="phys.mass" unit="M_sun" value="irrelevant">
<DESCRIPTION>mass (solar mass)              - model atmosphere data</DESCRIPTION>
</PARAM>
<PARAM arraysize="10" datatype="char" name="lum" ucd="phys.luminosity" unit="L_sun" value="irrelevant">
<DESCRIPTION>luminosity (solar luminosity)  - model atmosphere data</DESCRIPTION>
</PARAM>
```

Not very easy to read and to understand

VOTable (normalized)

UCD, no utypes

# Our motivations for Provenance

**3) Select** spectra **on provenance criteria** as…

- **Use case 1:**

  Show me a list of synthetic spectra satisfying :
  - domain of wavelength = visible
  - domain of effective temperature = [4000, 5000]

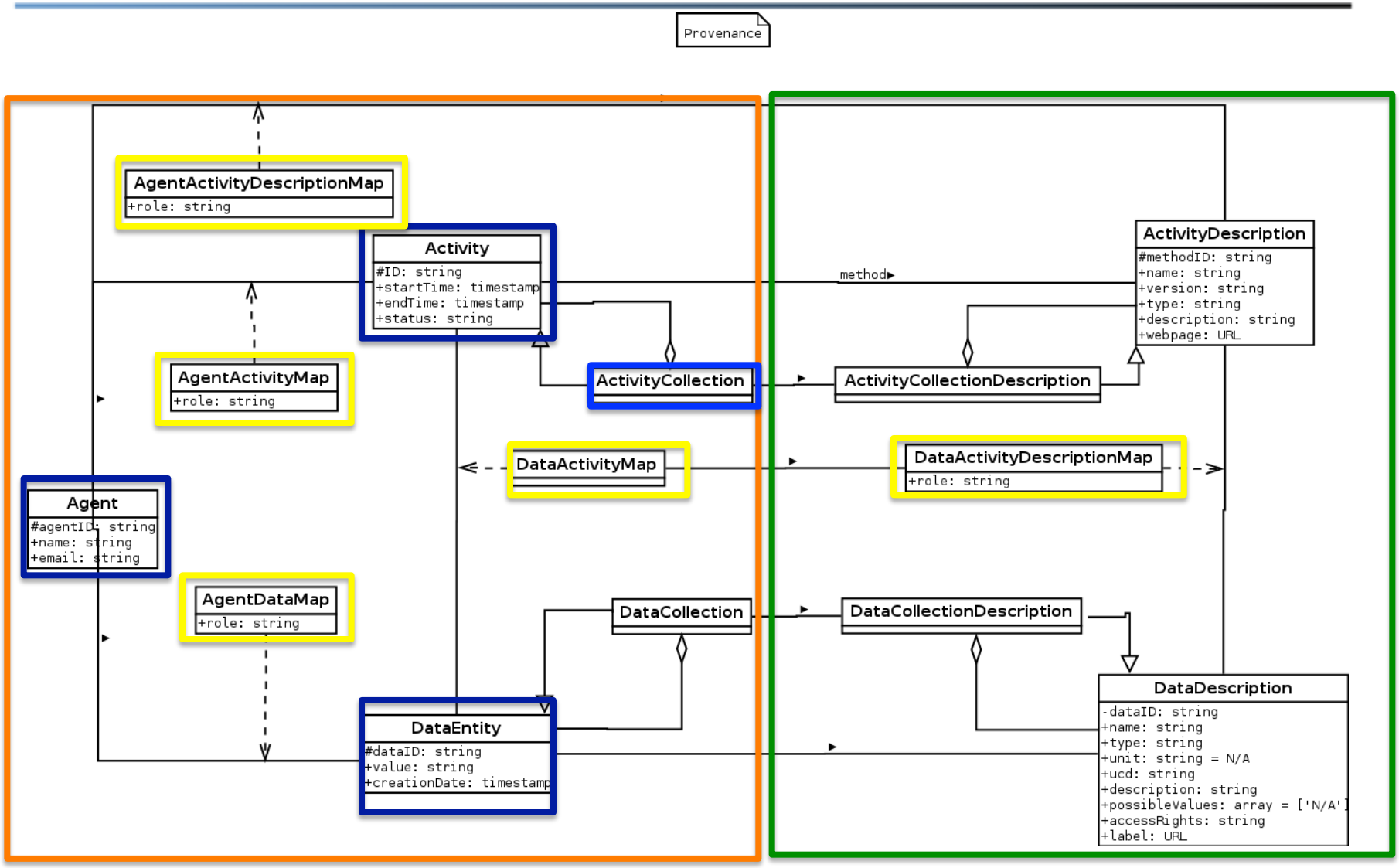
- **Use case 2:**

  Show me a list of synthetic spectra satisfying :
  - code for model atmosphere = MARCS
  - type of model atmosphere = spherical

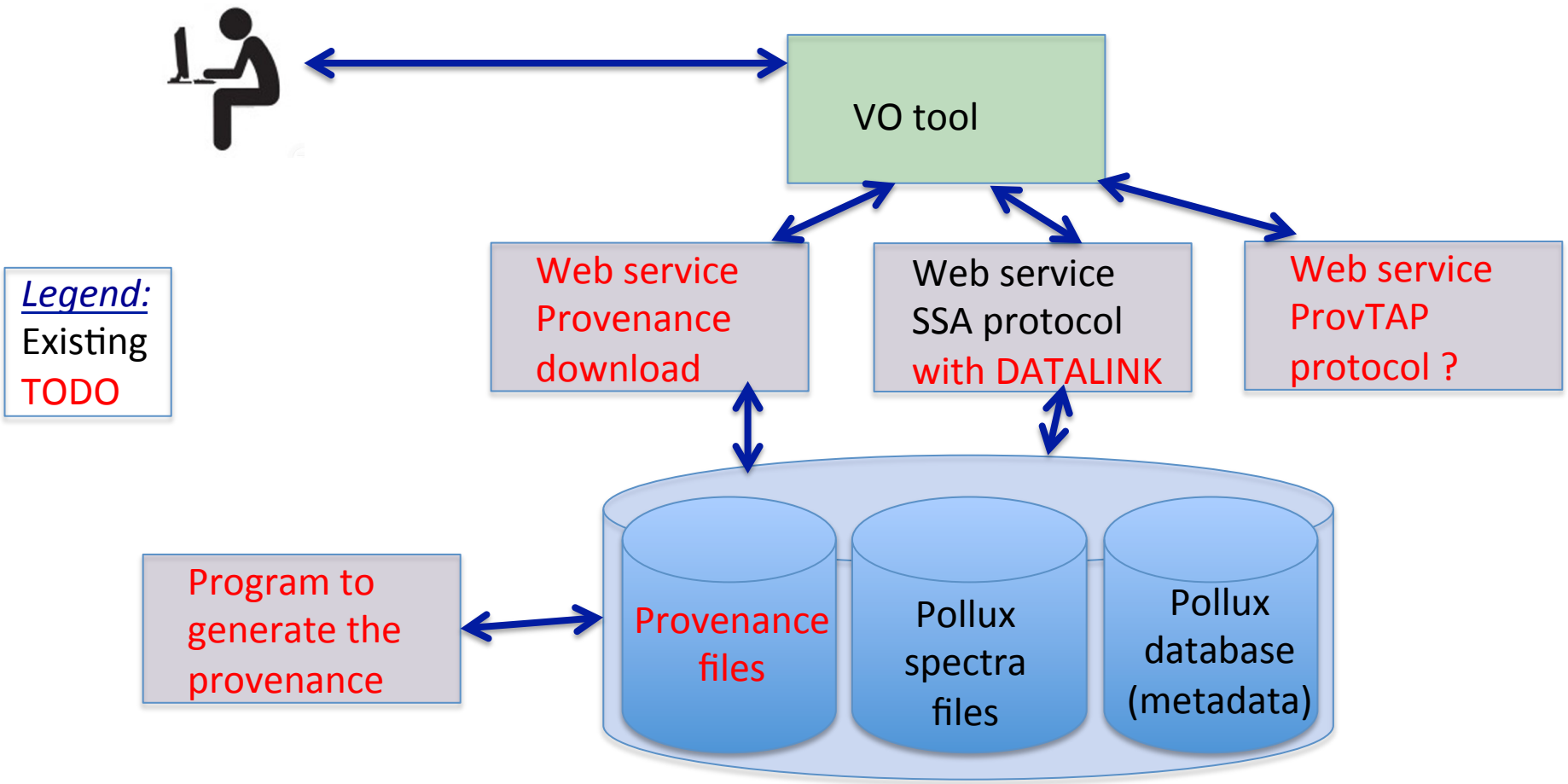
- **Use case 3:**

  Show me a list of synthetic spectra satisfying :
  - code for spectral synthesis = turbospectrum
  - version of this code = 2008.1

# Data Model used



# Global use


VO tool
Web service
Provenance
download
Web service
SSA protocol
with DATALINK
Web service
ProvTAP
protocol ?
Legend:
Existing
TODO
Program to
generate the
provenance
Provenance
files
Pollux
spectra
files
Pollux
database
(metadata)

# Creation of provenance files

- Program to generate for each spectrum the provenance files (prov 1.4.0 python package)

- The provenance files will be stored on the server

- Different formats:
  - Serialized formats : PROVN, JSON, VOTABLE
  - Graphic formats : PNG, SVG, PDF

- Namespaces:
  - prov: http://www.w3.org/ns/prov#
  - voprov: http://www.ivoa.net/documents/dm/provdm/voprov/
  - pollux: http://pollux.graal.univ-montp2.fr/documents/dm/provdm/std/
  - polluxData: http://pollux.graal.univ-montp2.fr/ssaserver/provenance?Id=

# Creation of provenance files
# Problems encountered and questions

- Serialized format:
  - JSON: 3906 lines, not very human readable
  - PROVN : 613 lines, not very human readable
  - VOTable : ? Lines; but not very human readable


```
"entity": {
  "polluxData:2702_line_list_1_name": {
    "voprov:ucd": "",
    "prov:value": "Hlinedata",
    "voprov:utype": "voprov:DataEntity",
    "voprov:type": "string",
    "voprov:description": "line list_1 : name",
    "voprov:unit": "",
    "voprov:name": "pollux:line_list_name"
  },
  "polluxData:2702_short_name": {
```



```
  entity(polluxData:2702_type, [voprov:utype="voprov:DataEntity", prov:value="s", voprov:unit="", voprov:ucd="meta.code.class
", voprov:description="type of model atmosphere (Spherical/Parallel)", voprov:type="char", voprov:name="pollux:type"])
  used(polluxData:marcs35, polluxData:2702_type, -)
  entity(polluxData:2702_filename, [voprov:utype="voprov:DataEntity", prov:value="3000g0.5m15z0.00t2.0.mod", voprov:unit="irr
elevant", voprov:ucd="meta.id;meta.file", voprov:description="model atmosphere filename", voprov:type="file", voprov:name="pol
```


- **Questions:**

  **Will these formats only be used by programs? Which ones ?**

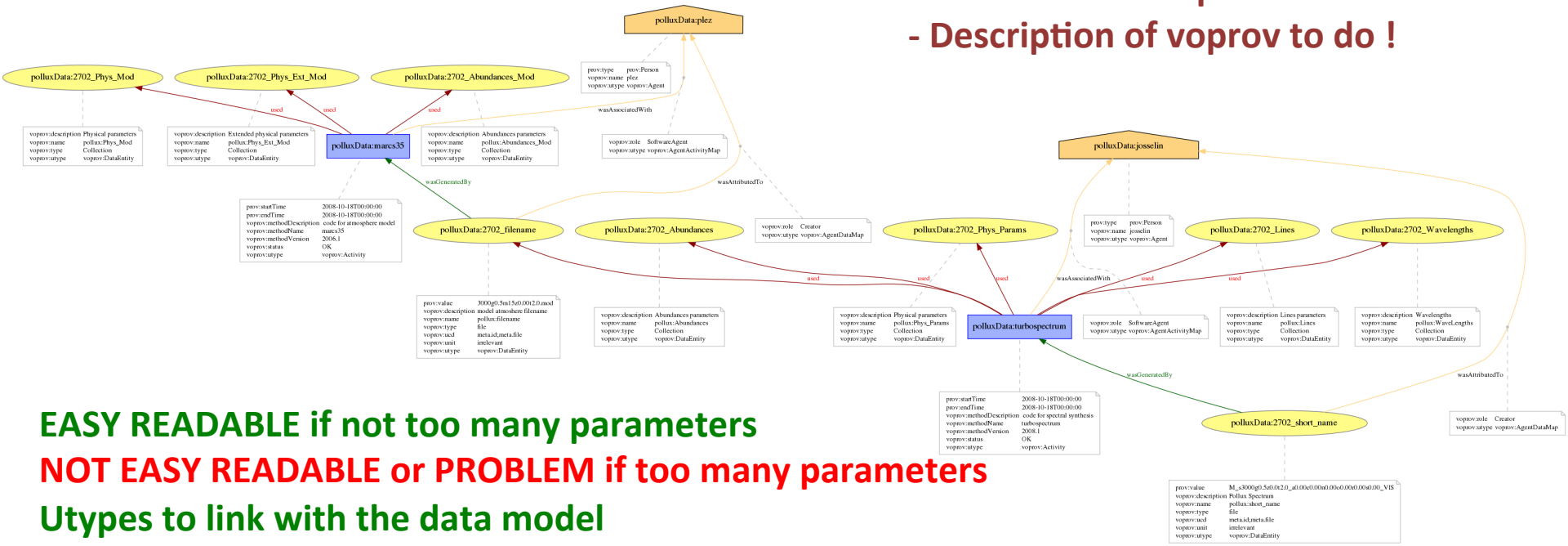
# Creation of provenance files
# Problems encountered and questions

- Graphic format (pdf, png, svg):
  - SVG: you can click on each declaration
  - PNG : only a picture
  - PDF : currently a picture

**Questions/Remarks:**

**- How to create a pdf with links?**

**- Description of voprov to do !**




**EASY READABLE if not too many parameters**

**NOT EASY READABLE or PROBLEM if too many parameters**

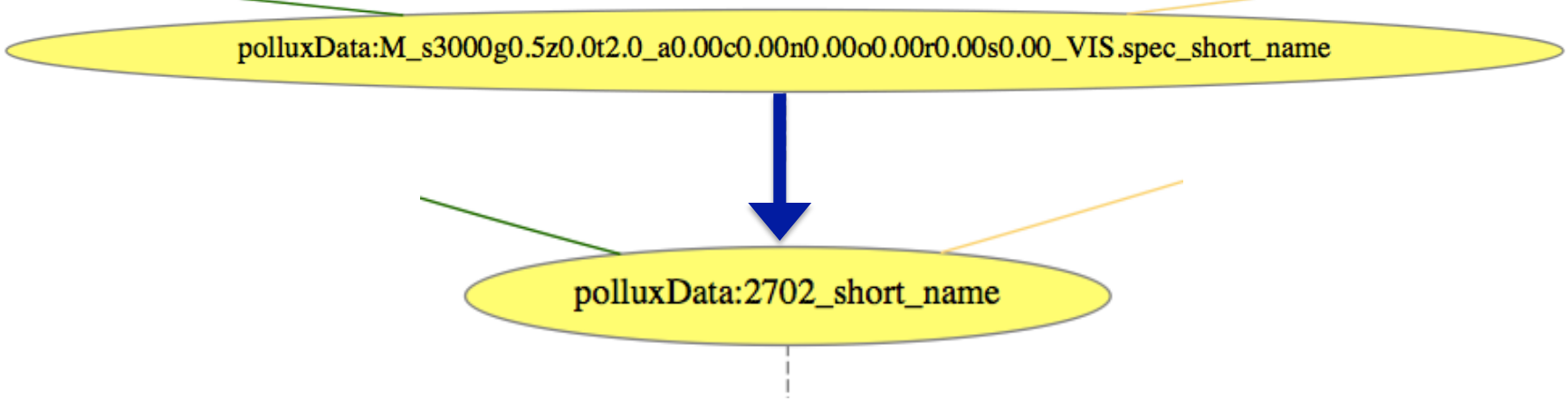
**Utypes to link with the data model**

**UCD to describe the data**

**Not yet included in CASSIS**

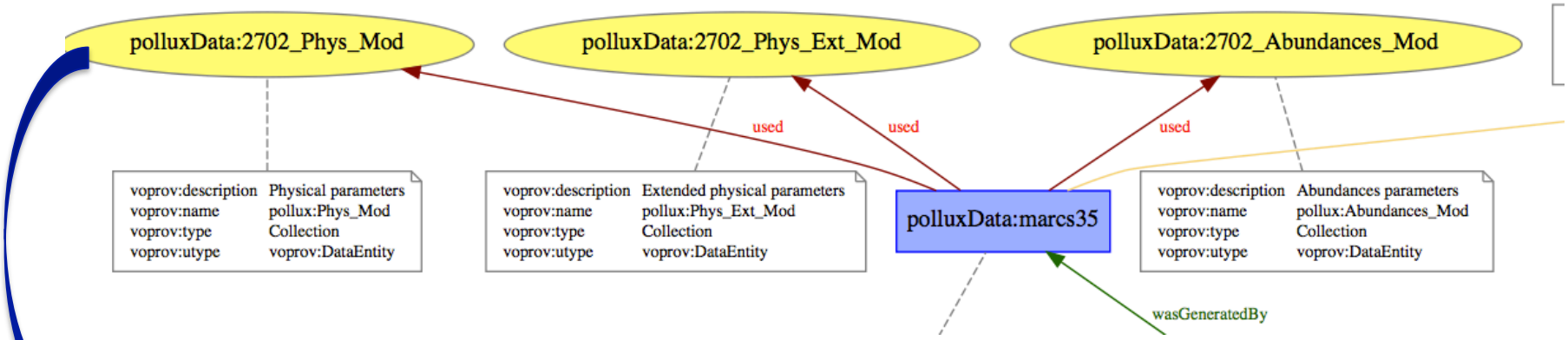
# Creation of provenance files
# Problems encountered and questions

- Name of instances too long
  => graphics not easy to read
  => add a column in the database to put a number (key too long)



# Creation of provenance files
# Problems encountered and questions

- A lot of input parameters
  => graphic too big
  => input parameters grouped




http://pollux.graal.univ-montp2.fr/ssaserver/provenance?Id=2702_Phys_Mod

- **Question : in which format do we give the details (svg, …)?**

# TO EXPLORE

- SimDM:
  - Resource : Experiment (=Activity) or Protocol (=ActivityDescription)
  - Contact (=Agent)
  - ParameterSetting, inputDataSet, outputDataSet (=Entity)
  - InputParameter,InputDataObjectType, OutputDataObjectType(=EntityDescription)
- SimDAL (3 composants):
  - Repository: contains projects definition (processes, parameters, …). When a project is selected, the links to the search and the data access services
  - search: gives the vues (one or serveral tables) of the project. The user can query to select datasets on project/process criteria
  - data access: downloads the dataset or a part of it (cutout)