provenance data model: new use cases. Restart the effort.

F.Bonnarel (CDS),
And DM Working group

Motivation: trace the history of the « dataset »

- Some kind of quality assesment

    – What were the ambiant conditions ?

    – How was the telescope configured ?

    – What kind of processing has been done?

- Describe the previous steps and acces to « progenitors » if reprocessing of the data is needed.
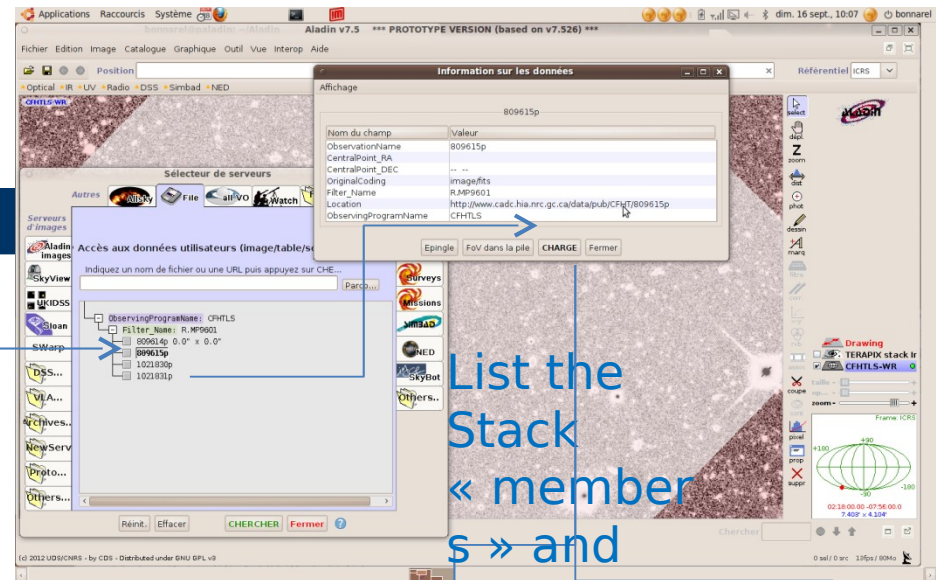
Use cases

- Radio data:

    - ALMA use case: Juan de Dios Santander et al (ALMA data distribution SPEC, 2010): Observing features such as ambiant conditions, link to raw data, calibration data

    - All projects : link to visibility data, link to velocity maps, 2D images, etc ....
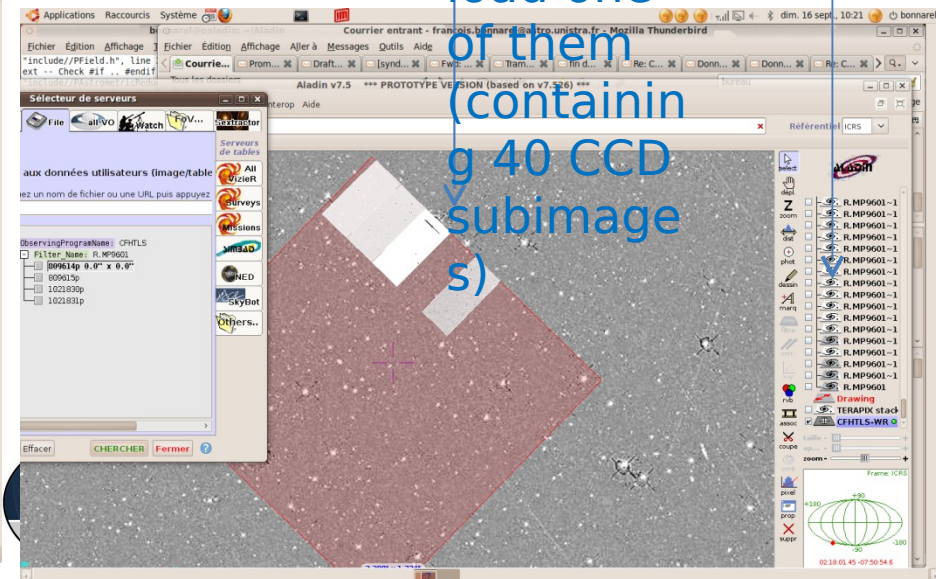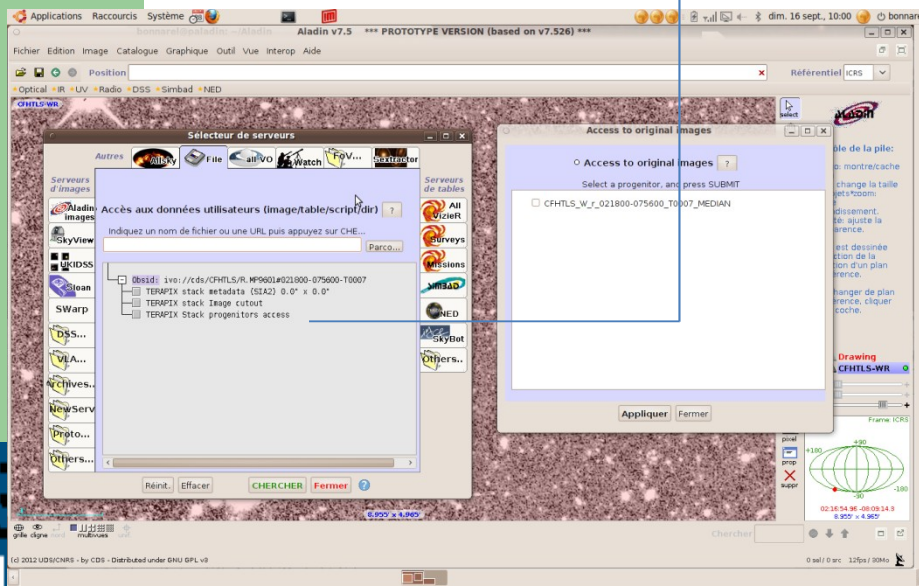
Use cases

Data processing:

- GWS VOSPACE and Workflow use case (VOSpace, M.Graham)

  - *The basic use case from the VOSpace point of view is that provenance information needs to be associated with data objects stored in a VOSpace and how this can be represented. This could be particularly important for temporary data objects or those used in a workflow.*

- Aladin and CFHTLS survey (to be released next week):

  - Access to progenitors CFH image of the TERAPIX Stack (mosaicking algorithm)

Use cases

List the Stack « members » and load one of them (containing 40 CCD subimages)

Use cases

- High energy data (eg HESS gamma ray observations)
  - Complex hierarchy of related products
    - Raw data
    - Event lists
    - Images and spectra
    - Object views
  - How to relate them in both directions with description of algorithms

Use cases

- Exposure time calculator (Nikos and LAM)
  - Parameters describing the observation features, ambiant conditions etc … allow to calculate exposure time
  - Express this in IVOA Observation Model standard words (input/output)
- Data curation and preservation use case (Arnold Rots):
  - Hetrogeneous repositories
  - Difference curation/provenance ?

Provenance UML Diagram (Preliminary)
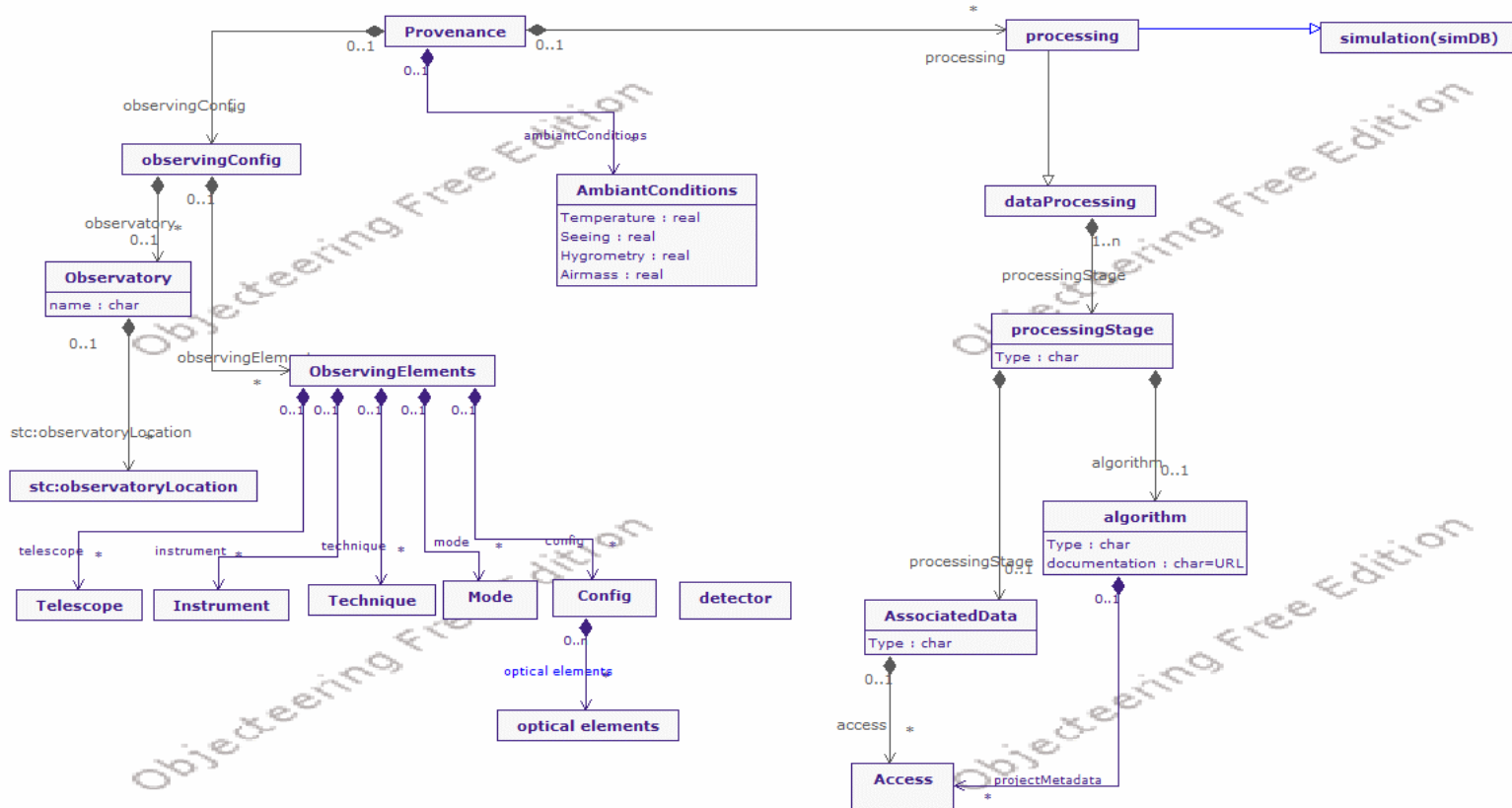
```xml
          <acref>http://project.org/data/exposure/exposure2.fits</acref>
          <observationMetadata>http://project.org/data/exposure/exposure2.xml</observationMetadata>
        </Access>
      − <Access>
          <type>proGenitor</type>
          <format>image/fits</format>
          <acref>http://project.org/data/exposure/exposure3.fits</acref>
          <observationMetadata>http://project.org/data/exposure/exposure3.xml</observationMetadata>
        </Access>
      − <Access>
          <type>proGenitor</type>
          <format>image/fits</format>
          <acref>http://project.org/data/exposure/exposure4.fits</acref>
          <observationMetadata>http://project.org/data/exposure/exposure4.xml</observationMetadata>
        </Access>
      </AssociatedData>
    </processingStage>
  − <processingStage>
      <type>confidenceMap</type>
    − <algorithm>
        <type>weight map</type>
      − <documentation>
          http://project.org/documentation/provenance/confidence.html
        </documentation>
      </algorithm>
    − <AssociatedData>
      − <Access>
          <type>confidenceMap</type>
          <format>image/fits</format>
        − <acref>
            http://project.org/data/exposure/thisimage_confidence.fits
```