To use Provenance DM for VizieR catalogues

-

G.Landais

F.Bonnarel (CDS), G.Landais (CDS), M.Louys (IceCube), M.Sanguillon (LUPM), M.Servillat (LUTH)

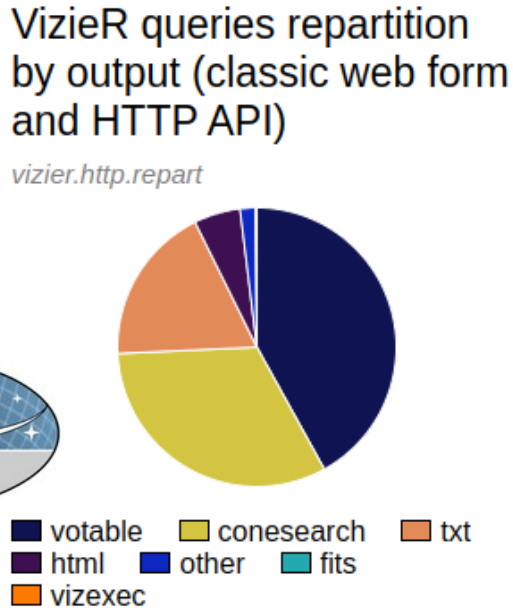
VizieR Provenance

VizieR contents statistics (~21,600 catalogues)

- >30K authors (~10K first authors)
- 10 main journals (+ ~30 other)
- Large tables and surveys coming from Space Agencies

Motivation

- Continue the FAIR-isation effort to be a Open Data Center
  - More transparency on workflow
  - Provide interoperable data in an interconnected network (=> citation, ..)
- Give explanation to final users/authors on workflows Operation executed on original data :
  - Join tables, crossmatch
  - (CDS) Photometry (fitler) assignment
- Lack of visibility on Data origin in the data provided through the Virtual Observatory
  (75 % queries generated throught the VO in VizieR)



Provenance applied to tables

Data origin (Provenance) is hidden in the Virtual Observatory

- Data Origin(Provenance) information are needed for the data understanding
  - author(s)
  - publication date, journal, article references,
  - internal & external identifiers, etc.
- Information **available in the VO registries**
- No standard in VOTable format to provide the Data Origin
  It is possible to add the data origin in the table description

```
<RESOURCE ID="yCat_51332464" name="J/AJ/133/2464">
  <DESCRIPTION>Parameters and abundances of nearby giants (Luck+, 2007)</DESCRIPTION>
```

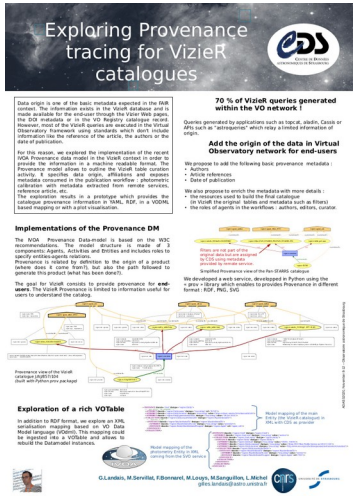
- Idea: a library (in Python for instance)

```
from astroquery.vizier import Vizier
catalog = Vizier.find_catalogs('II/349')
prov = Provenance(catalog)
print(prov)
prov.plot()
```

Implementation

The VizieR Provenance implementation

- Initiative following VOFrance 2020
- “french” group to define a Provenance output in conformance with the IVOA ProvDM model : F.Bonnarel (CDS), G.Landais (CDS), M.Louys (IceCube), M.Sanguillon (LUPM), M.Servillat (LUTH)



Idea: to give a quickview to users and authors about the VizieR table origin

- Contents : tables, associated data, meta-data
- Data origin: journal, author, publication year, ...
- Operation applied on Data:
  - Positions columns added from Simbad database or an other reference catalogue using a target name
  - Join and Fusion operation on tables
  - Metadata : CDS assigns filter on photometric columns

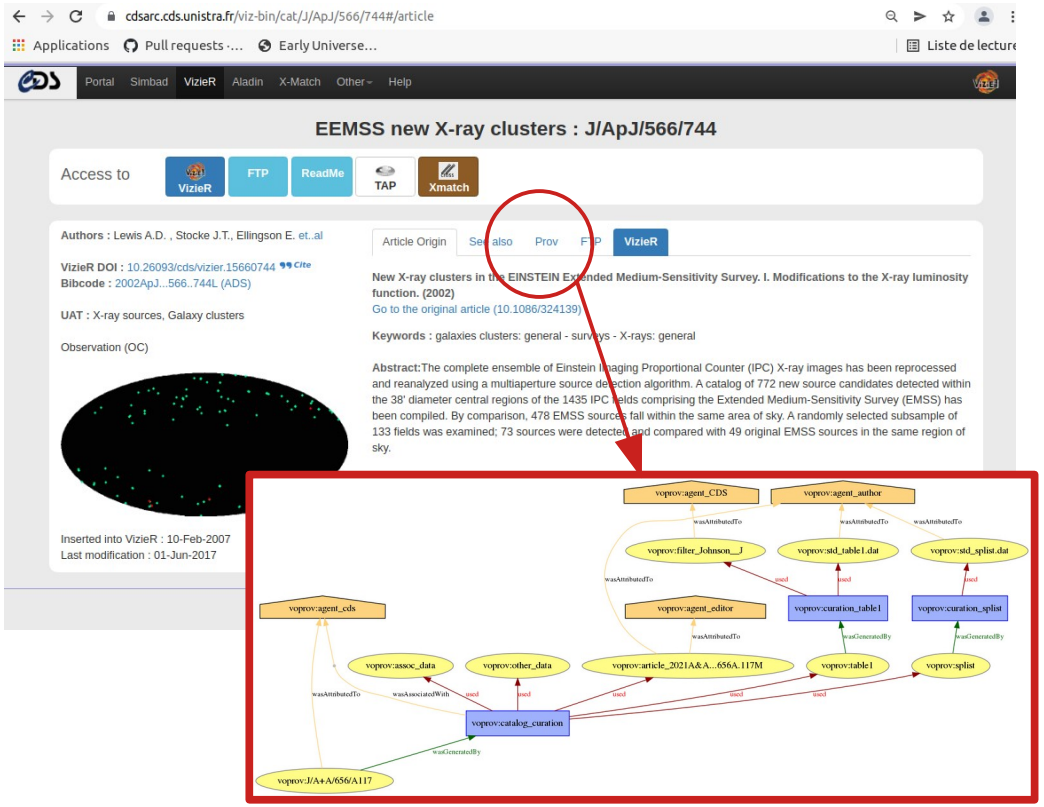
Implementation

Where to find Provenance ?

- Landing page
- Rich VOTable output (MangoDM)

Format available in output

- Prov:json, Prov:ttl
- Yaml
- PNG
- VOTable (Vo-instance dml)



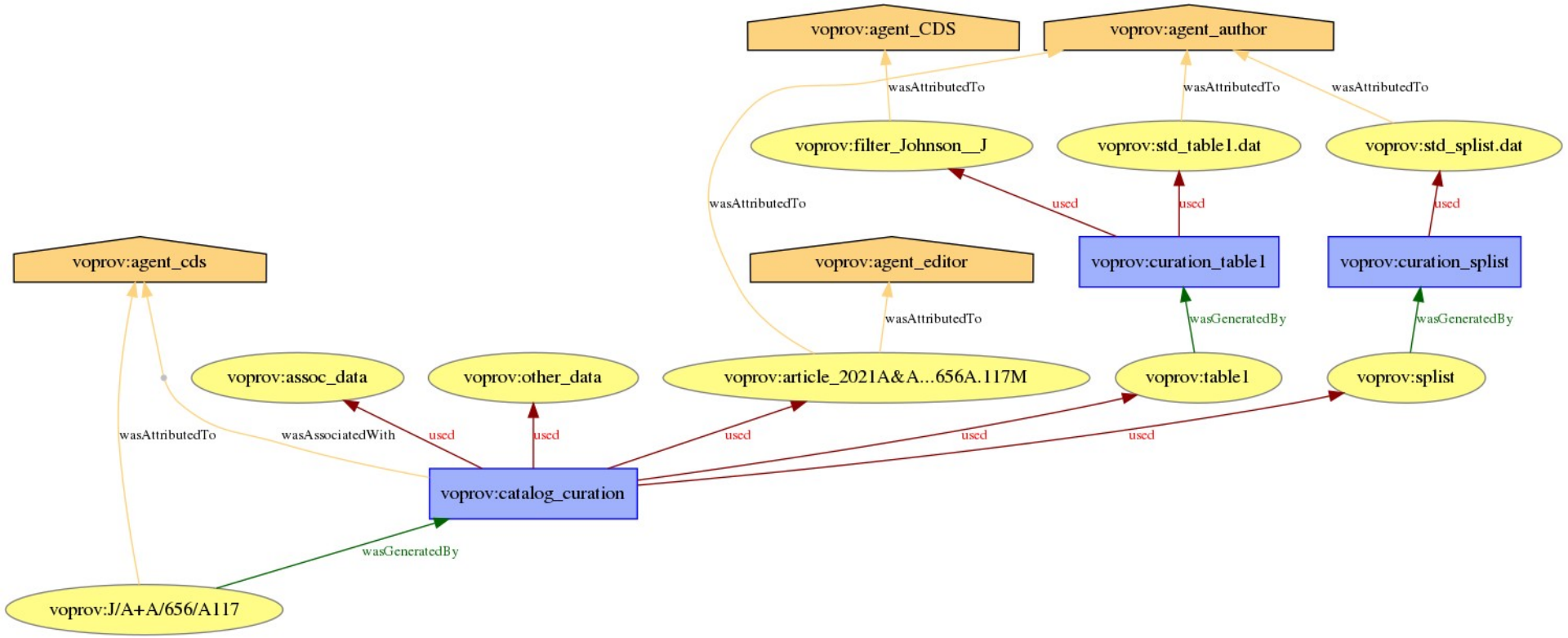
Implementation details

- Language : Python prov (=>? voprov)
- Contents: Prov W3C mainly for VizieR (but could be extended)

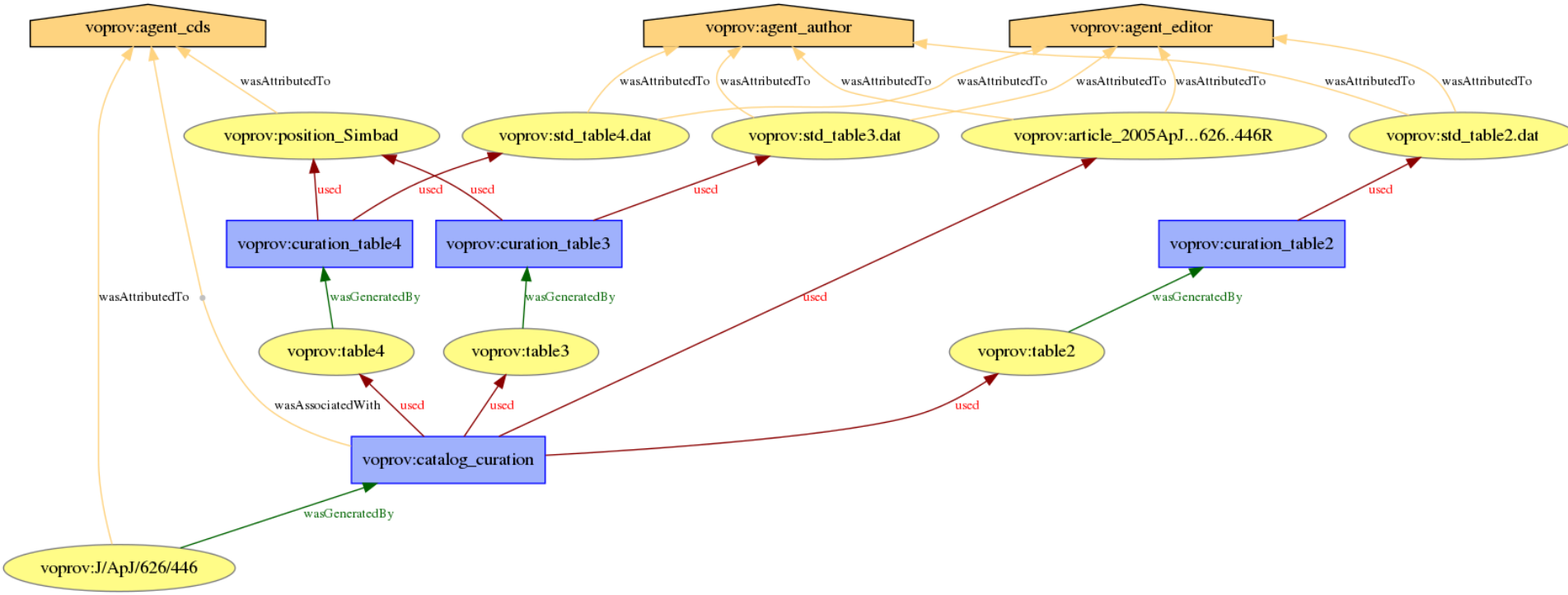
```
<GLOBALS>
 <INSTANCE id="agent_cds" dmtype="voprov:Agent">
    <ATTRIBUTE dmrole="Agent.name" dmtype="ivoa:string" value="CDS"/>
    <ATTRIBUTE dmrole="Agent.type" dmtype="ivoa:string" value="organisation"/>
    <ATTRIBUTE dmrole="Agent.url" dmtype="ivoa:string" value="https://cds.unistra.fr/"/>
  </INSTANCE>
 <INSTANCE id="agent_author" dmtype="voprov:Agent">
    <ATTRIBUTE dmrole="Agent.name" dmtype="ivoa:string" value="Page M.J."/>
    <ATTRIBUTE dmrole="Agent.type" dmtype="ivoa:string" value="person"/>
    <ATTRIBUTE dmrole="Agent.url" dmtype="ivoa:string" value=""/>
  </INSTANCE>
 <INSTANCE id="agent_editor" dmtype="voprov:Agent">
    <ATTRIBUTE dmrole="Agent.name" dmtype="ivoa:string" value="Monthly Notices of the Royal Astronomical Society"/>
    <ATTRIBUTE dmrole="Agent.type" dmtype="ivoa:string" value="organisation"/>
    <ATTRIBUTE dmrole="Agent.url" dmtype="ivoa:string" value=""/>
  </INSTANCE>
 <INSTANCE id="agent_CDS" dmtype="voprov:Agent">
    <ATTRIBUTE dmrole="Agent.name" dmtype="ivoa:string" value="agent_CDS"/>
    <ATTRIBUTE dmrole="Agent.type" dmtype="ivoa:string" value="organisation"/>
    <ATTRIBUTE dmrole="Agent.url" dmtype="ivoa:string" value=""/>
  </INSTANCE>
 <INSTANCE id="filter_XMM_OT__UVW2" dmtype="voprov:Entity">
    <ATTRIBUTE dmrole="voprov:Entity.comment" dmtype="ivoa:string" value="Filter added by CDS (not part of original data)"/>
    <ATTRIBUTE dmrole="voprov:Entity.location" dmtype="ivoa:string" value="http://xmm.esac.esa.int/"/>
   <COLLECTION dmrole="voprov:Entity.wasAttributedTo">
     <INSTANCE dmrole="voprov:WasAttributedTo" dmtype="voprov:WasAttributedTo">
        <ATTRIBUTE dmrole="voprov:WasAttributedTo.role" dmtype="ivoa:string" value="service"/>
        <INSTANCE dmrole="voprov:WasAttributedTo.agent" dmtype="voprov.Agent" ref="agent_CDS"/>
      </INSTANCE>
    </COLLECTION>
  </INSTANCE>
 <INSTANCE id="filter_XMM_OT__UVM2" dmtype="voprov:Entity">
    <ATTRIBUTE dmrole="voprov:Entity.comment" dmtype="ivoa:string" value="Filter added by CDS (not part of original data)"/>
    <ATTRIBUTE dmrole="voprov:Entity.location" dmtype="ivoa:string" value="http://xmm.esac.esa.int/"/>
   <COLLECTION dmrole="voprov:Entity.wasAttributedTo">
     <INSTANCE dmrole="voprov:WasAttributedTo" dmtype="voprov:WasAttributedTo">
        <ATTRIBUTE dmrole="voprov:WasAttributedTo.role" dmtype="ivoa:string" value="service"/>
        <INSTANCE dmrole="voprov:WasAttributedTo.agent" dmtype="voprov.Agent" ref="agent_CDS"/>
      </INSTANCE>
    </COLLECTION>
  </INSTANCE>
 <INSTANCE id="filter_XMM_OT__UVW1" dmtype="voprov:Entity">
    <ATTRIBUTE dmrole="voprov:Entity.comment" dmtype="ivoa:string" value="Filter added by CDS (not part of original data)"/>
    <ATTRIBUTE dmrole="voprov:Entity.location" dmtype="ivoa:string" value="http://xmm.esac.esa.int/"/>
   <COLLECTION dmrole="voprov:Entity.wasAttributedTo">
     <INSTANCE dmrole="voprov:WasAttributedTo" dmtype="voprov:WasAttributedTo">
        <ATTRIBUTE dmrole="voprov:WasAttributedTo.role" dmtype="ivoa:string" value="service"/>
        <INSTANCE dmrole="voprov:WasAttributedTo.agent" dmtype="voprov.Agent" ref="agent_CDS"/>
      </INSTANCE>
    </COLLECTION>
  </INSTANCE>
```

ProvDM + Mango ?

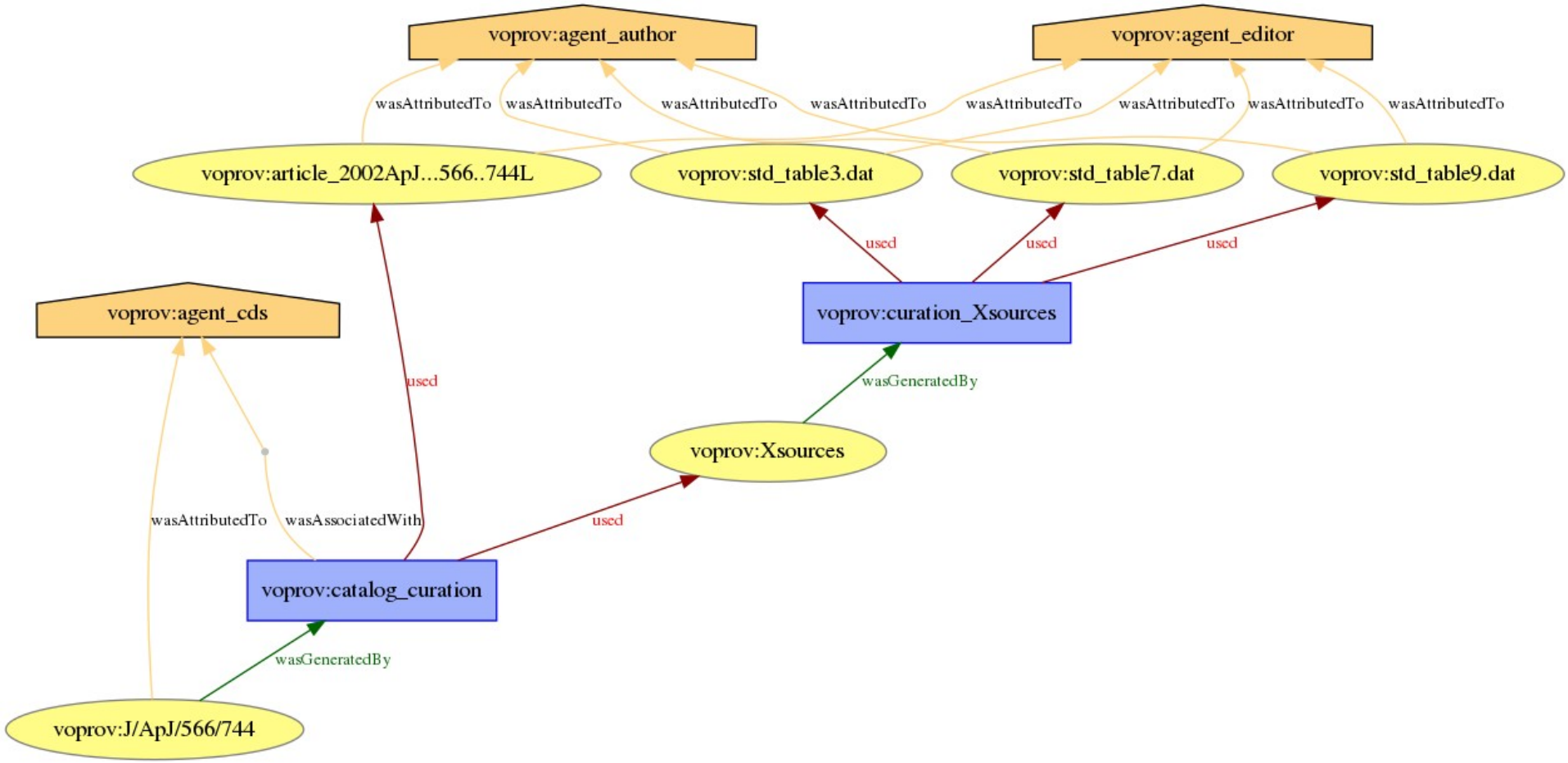
Exemple de Provenance (catalogue J/A+A/656/117)
(données associées + filtres)



Exemple de Provenance (catalogue J/ApJ/626/446)
(ajout de position a partir de la base Simbad)

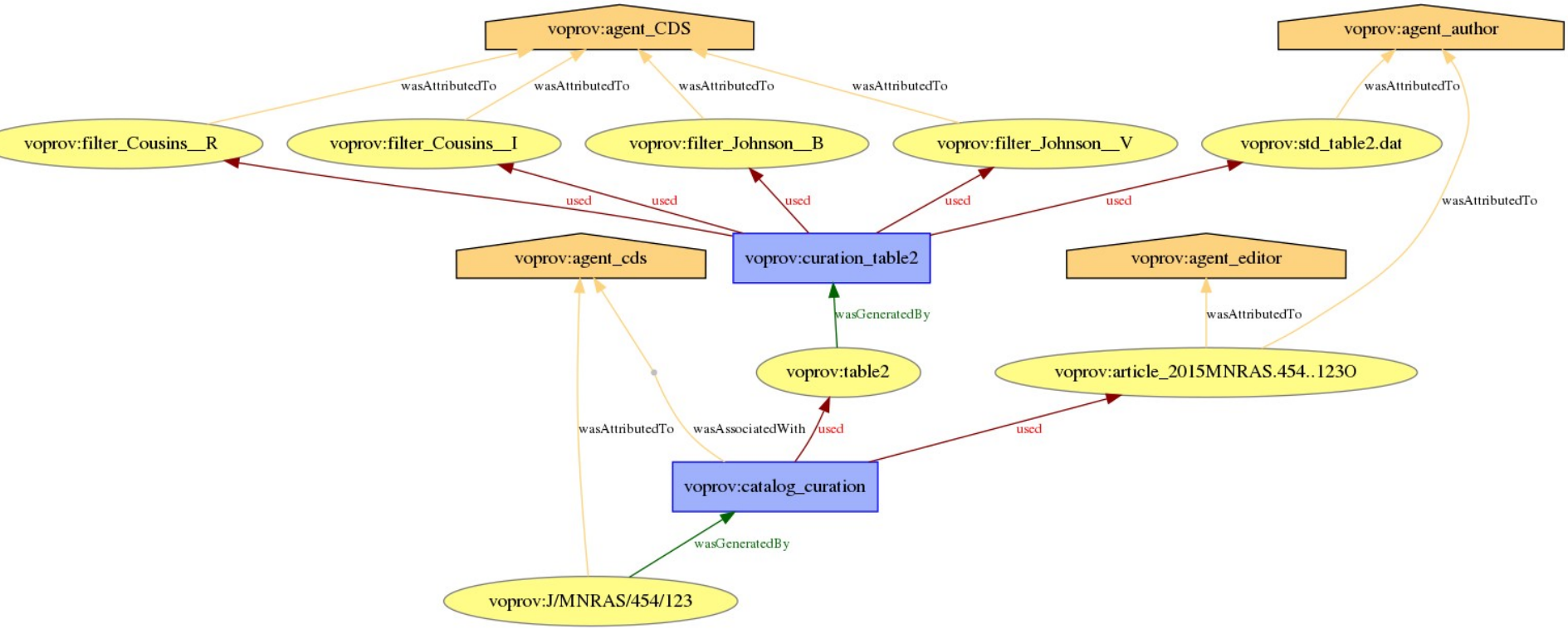


Conclusion – Provenance - Mango



Exemple de Provenance (catalogue J/ApJ/566/744)
operation de jointure

Exemple de Provenance (catalogue J/MNRAS/454/123)
Assignation de filtres



Possible evolution

(today) VizieR provenance uses (mainly) the W3C Provenance: Entity-Activity-Agent

- Other possibilities
  - Description (definition d'url)
  - Parameters to describe an activity