



---

## VizieR use case

### ProvenanceDM to specify Data origin

Contributors:

G.Landais, M.Servillat, F.Bonnarel, M.Louys,  
M.Sanguillon, L.Michel



CENTRE DE DONNÉES  
ASTRONOMIQUES DE STRASBOURG



# To get provenance information in VizierR



## Open data context asks for more transparency

- How to cite ?
- How to get data origin ? (article, date..)
- How to know authors, organisation/people involved in the data that I use ?
- Transparency sometimes hidden by the VO framework
  - A lot of resources (VizieR 21K catalogues, 45K tables..)
  - Metadata sometimes hidden in the output
- Data origin required in the Core Trust Seal certification .. sometimes required in a Data Management Plans (DMP)



# □ VizieR workflows



## Catalogue content

- Tables available in different format : original, CDS- standardized (MRT), binary, database..
- Associated data (spectra, images)
- Metadata : built from ReadMe and article and CDS tools



## 2 types of workflows

- Initiated by the CDS (eg: AAS, space agencies)
- Initiated by authors (eg : A&A )



## Repository / copy

- original tables/data are in the CDS (A&A)
- Original tables/data are outside (eg: gaia, AAS tables)



## CDS transformation

copy / added values / aggregation / paste ..







# How to get provenance information ?



- The **ReadMe** file  
<https://cdsarc.unistra.fr/viz-bin/cat/J/ApJ/728/38>

- The **DOI** :  
→ Metadata in DOI: authors, article, links,...

- The **ivoid** :  
<source format="bibcode">  
  2017ApJ...841...33M  
</source>

- The **VOTable** : no standard way !??

→ VizieR particularities:

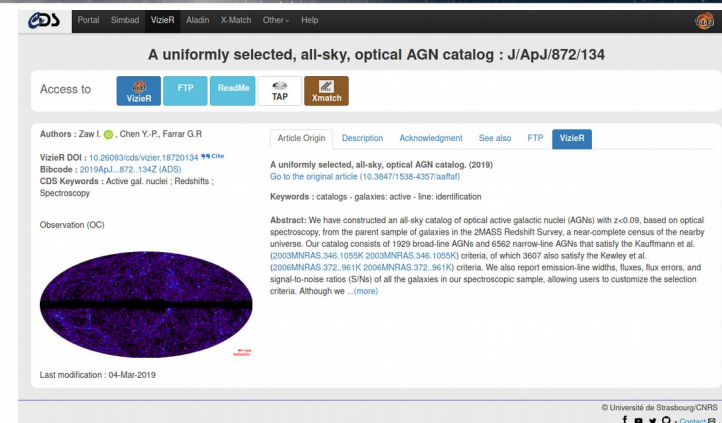
Author, publication year are added in tables description

<DESCRIPTION>

  AEGIS: demographics of X-ray and optical AGN (Yan+, 2011)

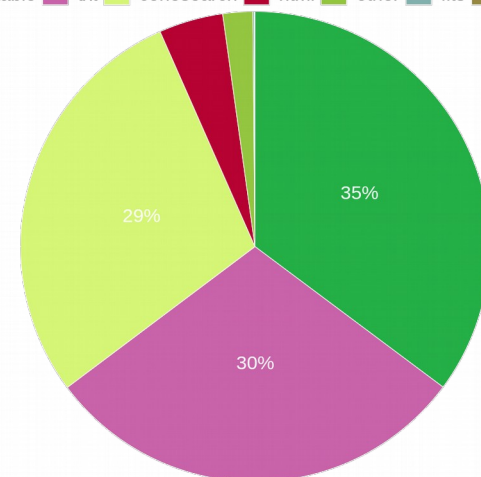
</DESCRIPTION>

- **DatasetDM** (Workinfg draft): describe resources but no way to link dataset together



VizieR queries repartition by output (classic web form and HTTP)

votable txt conesearch html other fits vizexec



# □ Information expected



	ReadMe / landing page	DOI metadata	Ivoid metadata	VOTable
Basic origin metadata: article ref.	x	x	x	
Basic origin metadata: authors	x	x	x	~
Basic origin metadata: pub. year	x	x	~	~
Rich metadata (resource relations / workflow)	x	x	~	
Human readable	x			
Machine readable		x	x	x
astroquery/topcat				x

## Provenance information expected

- Basics provenance information :
  - article ref (doi/bibcode, editor, link)
  - Publication date
  - author (creator)
- Workflow information:
  - resources used: tables + external metadata
  - specify the roles of agents in the workflow



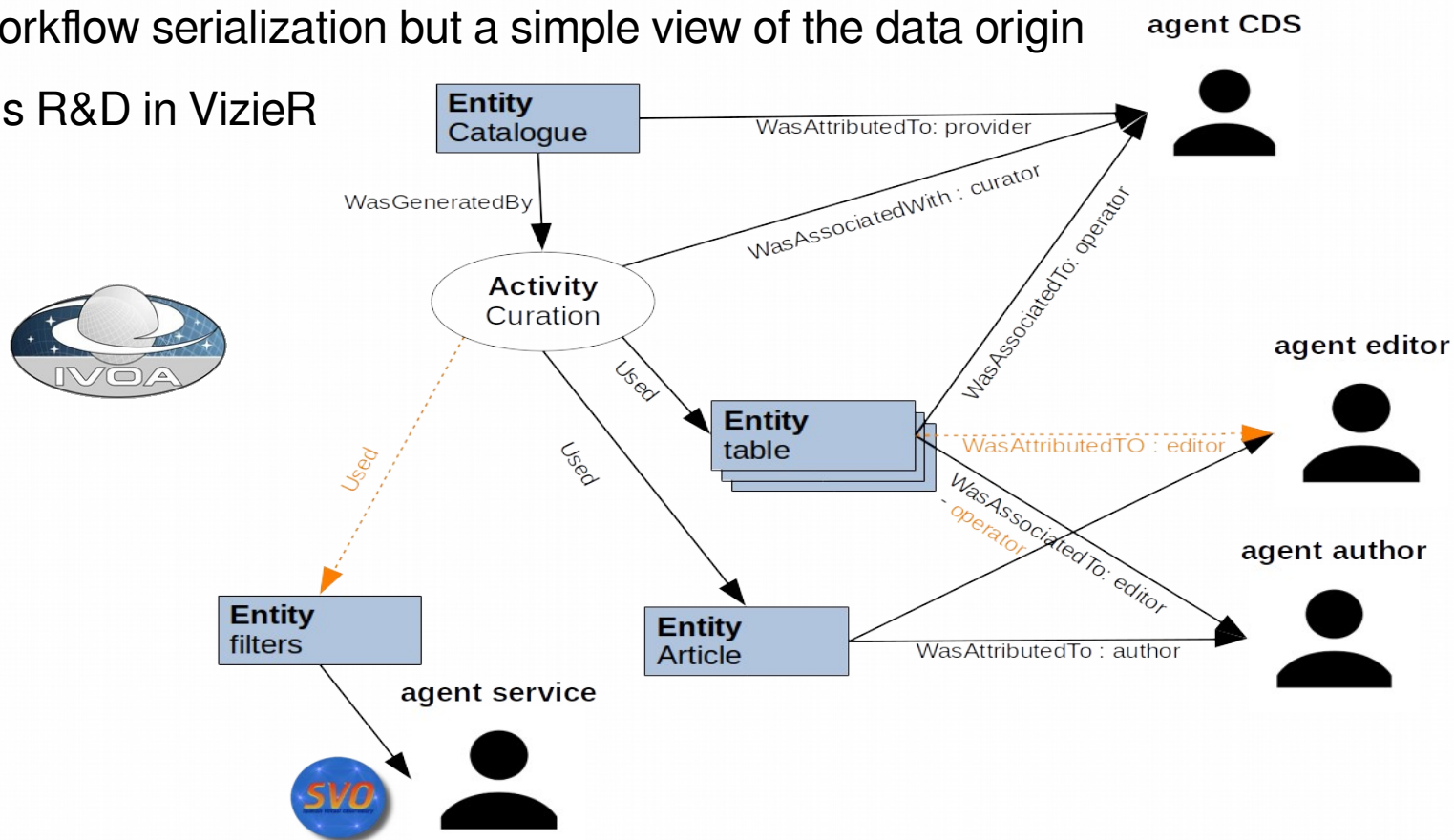
# To specify data origin in VizieR



## The Provenance DM draft to describe VizieR provenance

(G.Landais, M.Servillat, F.Bonnarel, M.Louys, M.Sanguillon, L.Michel)

- A **simple** workflow :  
not a workflow serialization but a simple view of the data origin
- In status R&D in VizieR





# □ Example



- Test url for vizier catalogues generated from existing metadata (mainly coming from the ReadMe file)
- Output format available in
  - VODMLITE
  - YAML (M.Servillat)
- Examples:
  - Gaia: 2 filters, editors doesn't interfere in the table  
<http://cdsarc.u-strasbg.fr/viz-bin/provenance?cat=I/345&out=yaml&filter=true>
  - J/AJ/159/8: AAS tables built by editors/authors  
<http://cdsarc.u-strasbg.fr/viz-bin/provenance?cat=J/AJ/159/8&out=yaml&filter=true>
  - J/A+A/401/959: A&A tables built by CDS/authors  
<http://cdsarc.u-strasbg.fr/viz-bin/provenance?cat=J/A%2BA/401/959&filter=true>

# □ Conclude ...



- This is (just a) test!  
vizier-provenance is not planned today to become official
  - Must be validated
  - Could be improved :
    - eg: CDS transformation (copy/added values/aggregation..) is not reported yet
    - Associated data not described
  - Information sometimes difficult to find...
- Serialization possible with YAML, VODMLite
- ProvDM is a standard based on W3C provenance  
→ interesting reflexion in the CTS context
- A verbose output may be not adapted in classic VOTable output yet.