



---

# How to find the origin of VizieR tables in the VO?

G.Landais,  
and VizieR team !



# □ Providing Data origin



## GO Fair principle guideline (EOSC)



### Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

**F1. (Meta)data are assigned a globally unique and persistent identifier.**

**F2. Data are described with rich metadata (defined by R1 below).**

**F3. Metadata clearly and explicitly include the identifier of the data they describe.**

**F4. (Meta)data are registered or indexed in a searchable resource.**

### Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

**A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.**

**A1.1 The protocol is open, free, and universally implementable.**

**A1.2 The protocol allows for an authentication and authorisation procedure, where necessary.**

**A2. Metadata are accessible, even when the data are no longer available.**

### Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

**I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.**

**I2. (Meta)data use vocabularies that follow FAIR principles.**

**I3. (Meta)data include qualified references to other (meta)data.**

### Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

**R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.**

**R1.1. (Meta)data are released with a clear and accessible data usage license.**

**R1.2. (Meta)data are associated with detailed provenance.**

**R1.3. (Meta)data meet domain-relevant community standards.**



# □ Providing Data origin



## Importance of Data origin especially in a virtual network

- Needed in Open science
- Enough metadata to understand data :
  - To be understood by end-users
  - To be interoperable
  - To create an interconnected network using references
  - To be cited

## Origin meta-data for published resources

- Authors, curator, provider
- References:
  - Link to original data
  - Link to related resources  
eg: articles or any resources used
- Transformation/update (eg: selection of columns, added values)
- Years of publication, last release, etc.
- (Licenses ?)

The screenshot shows the VizieR interface. At the top, there is a search bar containing 'archivestsciedu-hlsp\_classy'. Below the search bar, the interface is divided into two main sections. On the left, there is a 'Tap Nodes' tree view showing a hierarchy of data sources. The tree includes nodes for 'archivestsciedu-hlsp\_classy', 'nedipac-tap', 'TAP\_SCHEMA', 'NEDTAP', 'vizier', 'tap\_schema', 'J\_PASJ', 'IX\_HE', 'II\_photometry', 'J\_other', 'VI\_misc', 'J\_BaltA', 'J/BaltA/11/153/table4', 'J/BaltA/11/441/table5', 'J/BaltA/11/153/table8', and 'J/BaltA/17/293/stars'. On the right, there is a 'Hide tree' button and a title bar for the selected table: 'vizier>J\_BaltA>J/BaltA/11/153/table4>761'. Below the title bar, there is a 'Show' dropdown menu set to '5' and the text 'entries'. A table with 5 rows and 4 columns is displayed. The columns are labeled 'recno', 'HIP', 'm\_HIP', and 'n\_HIP'. The rows contain the following data: (1, 110, , ), (2, 110, A, ), (3, 110, B, ), (4, 210, , ), and (5, 210, AB, ). A tooltip is visible over the table, containing the text: 'Individual values of proper motions ( Bartkevicius A., Gudas A.) Double click or drag and drop to display it'.

recno	HIP	m_HIP	n_HIP
1	110		
2	110	A	
3	110	B	
4	210		
5	210	AB	

# □ To find Data origin in the VO



## Where to provide origin metadata in the VO ?

- **VO registry**
  - Identifiers, authors, data origin (article)
  - References to other resources
- **ProvDM**
  - Ingestion Workflow
- **VO Table**
  - No “standard” provenance information : DOI/bibcodes, authors...
- **In Protocols?**
  - TAP ? Scs ? Etc.

# Origin in the registry



## Example of VizieR catalogue in VO registry

```
<title>ROXA: multi-frequency large sample of blazars</title>
<shortName>J/A+A/472/699</shortName>
<identifier>ivo://CDS.VizieR/J/A+A/472/699</identifier>
<altIdentifier>doi:10.26093/cds/vizieR.34720699</altIdentifier>
```

Metadata available in pyVO !

```
<curator>
  <publisher ivo-id="ivo://CDS">CDS</publisher>
  <creator>
    <name>Turriziani S.</name>
  </creator>
  <creator>
    <name>Cavazzuti E.</name>
  </creator>
  <creator>
    <name>Giommi P.</name>
  </creator>
  <date role="Updated">2017-10-13T15:41:40Z</date>
  <date role="Created">2007-11-11T15:44:53Z</date>
```

authors, pub. date

```
<subject>Active galactic nuclei</subject>
<subject>BL Lacertae objects</subject>
<subject>Quasars</subject>
<description>Although Blazars are a small fraction of the overall AGN population they are expected to be the dominant population of extragalactic sources in the hard X-ray and gamma-ray bands and have been shown to be the largest contaminant of CMB fluctuation maps. So far the number of known blazars is of the order of several hundreds, but the forthcoming AGILE, GLAST and Planck space observatories will detect several thousand of objects of this type. In preparation for these missions it is necessary to identify new samples of blazars to study their discovery of new High Energy Peaked BL Lac (HBLs).Our catalog therefore includes many new potential targets for GeV-TeV observations.</description>
```

```
<source format="bibcode">2007A&A...472..699T</source>
<referenceURL>https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/472/699</referenceURL>
```

Reference to article

References given by authors in the article and which are in VizieR database

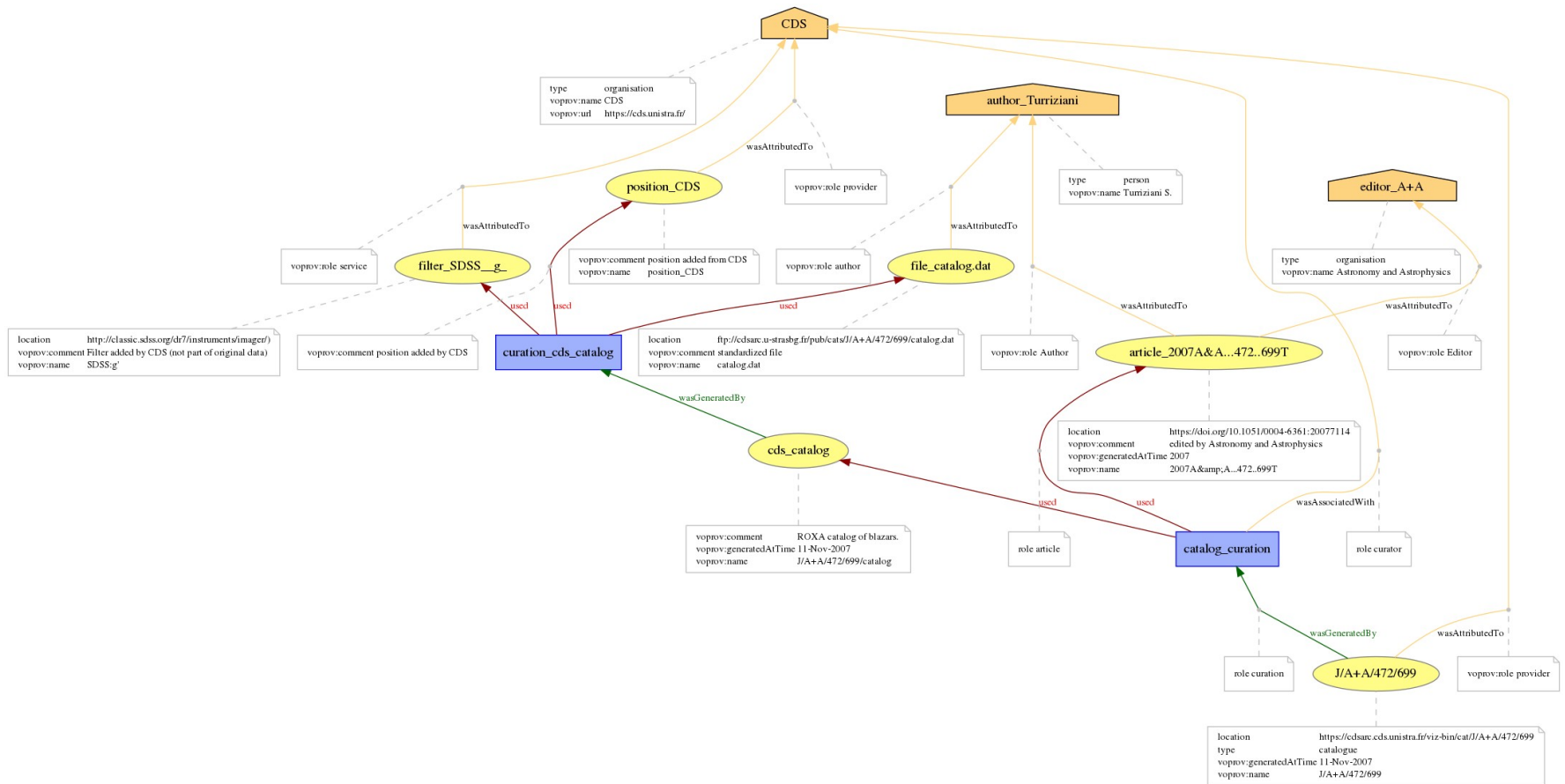
```
<relationship>
  <relationshipType>related-to</relationshipType>
  <relatedResource ivo-id="ivo://CDS.VizieR/VIII/65">VIII/65 : 1.4GHz NRAO VLA Sky Survey (NVSS) (Condon+ 1998)</relatedResource>
  <relatedResource ivo-id="ivo://CDS.VizieR/IX/29">IX/29 : ROSAT All-Sky Survey Faint Source Catalog (Voges+ 2000)</relatedResource>
  <relatedResource ivo-id="ivo://CDS.VizieR/IX/10">IX/10 : ROSAT All-Sky Bright Source Catalogue (1RXS) (Voges+ 1999)</relatedResource>
  <relatedResource ivo-id="ivo://CDS.VizieR/II/276">II/276 : The SDSS Photometric Catalog, Release 5 (Adelman-McCarthy + 2007)</relatedResource>
  <relatedResource ivo-id="ivo://CDS.VizieR/VII/226">VII/226 : The 2dF Galaxy Redshift Survey 100k Data Release (2dFGRS Team, 2001)</relatedResource>
  <relatedResource ivo-id="ivo://CDS.VizieR/VII/241">VII/241 : The 2dF QSO Redshift Survey (Croom+ 2004)</relatedResource>
  <relatedResource>ATCAPMN : http://www.atnf.csiro.au/resources/catalogues/pmn_atca</relatedResource>
</relationship>
```

# Origin and Provenance



## ProvDM

- An advanced mode dedicated for workflow
- Explain relations, roles between resources and agents



How to find the origin of VizieR tables in the VO? (IVOA april 2022)



# Origin in VOTable

- VOTable returned by a conesearch (SCS)

```
<VOTABLE xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.ivoa.net/xml/VOTable/v1.1" version="1.1"
xsi:schemaLocation="http://www.ivoa.net/xml/VOTable/v1.1
http://www.ivoa.net/xml/VOTable/v1.1">
  <DESCRIPTION> Vizier Astronomical Server vizier.u-strasbg.fr Date: 2022-04-21T16:49:34
  [V1.99+ (14-Oct-2013)] Explanations and Statistics of UCDS: See LINK below In case of
  problem, please report to: cds-question@unistra.fr In this version, NULL integer columns
  are written as an empty string <TD></TD>, explicitey possible from VOTable-1.3
  </DESCRIPTION>
  <!-- VOTable description at http://www.ivoa.net/Documents/latest/VOT.html -->
  <DEFINITIONS>
    <COOSYS ID="J2000" system="eq_FK5" equinox="J2000"/>
  </DEFINITIONS>
  <INFO ID="VERSION" name="votable-version" value="1.99+ (14-Oct-2013)"/>
  <INFO ID="Ref" name="-ref" value="VOTx4074306"/>
  <INFO name="MaxTuples" value="50000"/>
  <RESOURCE ID="yCat_34720699" name="J/A+A/472/699">
    <DESCRIPTION>ROXA: multi-frequency large sample of blazars (Turriziani+, 2007)
    </DESCRIPTION>
    <TABLE ID="J_A_A_472_699_catalog" name="J/A+A/472/699/catalog">
      <DESCRIPTION>ROXA catalog of blazars.</DESCRIPTION>
      <!-- Definitions of GROUPS and FIELDS -->
      <FIELD name="recno" ucd="meta.record" datatype="int" width="8">
        <!-- ucd="RECORD" -->
        <DESCRIPTION>Record number assigned by the Vizier team. Should Not be used for
        identification.</DESCRIPTION>
      </FIELD>
      <FIELD name="ROXA" ucd="ID_MAIN" datatype="char" arraysize="18">
        <!-- ucd="meta.id;meta.main" -->
        <DESCRIPTION>Source name (JHHMSS.s+DDMMSS.s) [datatype=char]</DESCRIPTION>
      </FIELD>
      <FIELD name="z" ucd="src.redshift" datatype="float" width="5" precision="2">
        <!-- ucd="REDSHIFT_HC" -->
        <DESCRIPTION>Redshift</DESCRIPTION>
      </FIELD>
      <FIELD name="Bjmag" ucd="phot.mag;em.opt.B" datatype="float" width="4" precision="1"
      unit="mag">
        <!-- ucd="PHOT_PHG_BJ" -->
        <DESCRIPTION>? Bj magnitude (from 2dF surveys)</DESCRIPTION>
        <VALUES null="NaN"/>
      </FIELD>
      <FIELD name="g'mag" ucd="phot.mag;em.opt.B" datatype="float" width="4" precision="1"
      unit="mag">
```

NO DOI ,  
NO bicode

Metadata  
(author, pub. Year)  
available in text, but  
not using a standard

No Metadata  
concerning the filter  
used

# Origin in TAP



Brown dwarf query executed in TOPcat

Window TAP Registry Edit Interop Help

VizieR

Select Service Use Service Resume Job Running Jobs

Metadata

Find: brown dwarf

Name  Descrip Or

TAPVizieR (30/49664)

J\_AA (6/16004)

- J/A+A/497/497/dwarfs\*
- J/A+A/574/A126/Dwarf\*
- J/A+A/620/A165/dwarf\*
- J/A+A/620/A165/dwarfbck\*
- J/A+A/640/A2/dwarfs\*
- J/A+A/640/A2/m\_dwarfs\*

J\_Aj (4/6949)

- J/AJ/133/439/dwarfs\*
- J/AJ/142/138/Mdwarfs\*
- J/AJ/145/102/Mdwarfs\*
- J/AJ/145/102/Mdwarfs\*

Service Schema Table Columns FKeys Hints

Name: "J/A+A/497/497/dwarfs"

Columns: 24

Foreign Keys: 0

Description: Dwarfs photometric data and reddenings used in the paper ( Gonzalez Hernandez J.I., Bonifacio P.)

Non-Standard Table Metadata:

Non-Standard Column Metadata:

Window TAP Registry Edit Interop Help

GAVO GERMAN ASTROPHYSICAL VIRTUAL OBSERVATORY

Select Service Use Service Resume Job Running Jobs

Metadata

Find: brown dwarf

Name  Descrip Or

GAVO DC TAP (1/206)

- browndwarfs (1/1)
  - browndwarfs.cat

Service Schema Table Columns FKeys Hints

Name: browndwarfs.cat

Columns: 23

Foreign Keys: 0

Description: A catalogue of brown dwarfs produced by Gelino et al. The database reflects the state of http://www.dwarfArchives.org on 2015-09-29.

Non-Standard Table Metadata: sourceRD: brownDwarfs/bd

Non-Standard Column Metadata: sourceRD

Window Parameters Display Help

Table Parameters for 2: TAP\_2 J\_A+A\_497\_497\_dwarfs

Name	Value	Description
Name	result_s1650556909679	Table name
Column Count	24	Number of columns
Row Count	555	Number of rows
QUERY_STATUS	OK	
PROVIDER	CDS	TAPVizieR
QUERY	SELECT TOP 1000 * FROM "J/A+A/497/497/dwarfs"	

TOPCAT(1): Table Parameters

Window Parameters Display Help

Table Parameters for 1: TAP\_1\_browndwarfs.cat

Name	Value	Description
Name	cat	Table name
Column Count	23	Number of columns
Row Count	1000	Number of rows
CoordFlavor	SPHERICAL	
coord_axes	3	
CoordRefFrame	ICRS	
URI	http://www.ivoa.net/xml/STC/stc-v1.30.xsd	
server	http://dc.zah.uni-heidelberg.de	
sql_query	SELECT browndwarfs.cat.designation, browndwarfs.cat.raj2000, ...	
query	SELECT TOP 1000 * FROM browndwarfs.cat	
src_res	Contains traces from resource brownDwarfs/bd	A catalogue of brown dwar
copyright	brownDwarfs/bd copyright or license	If this table has been usef
src_table	Contains traces from table browndwarfs.cat	
QUERY_STATUS	OK	Query successful
citation	http://www.dwarfArchives.org	This resource contains dat
citation	http://dc.zah.uni-heidelberg.de/tableinfo/browndwarfs.cat	For advice on how to cite t
legal	If this table has been useful to you, please acknowledge: Thi...	

- Origin (authors only) in the description
- VOTable header more verbose in GAVO:
  - Citation
  - Legal
  - Copyright

But no authors/ data origin or references for the both !

How to find the origin of VizieR tables in the VO? (IVOA april 2022)



# □ Providing origin in rich output



## Adding origin to metadata

Extract of metadata in VOTable using MiVOT or MangoDM

Filter definition linked to the columns

```
▼<INSTANCE ID="SDSS_g'" dmrole="mango:stcextend.PhotFilter" dmttype="mango:stcextend.PhotFilter">
  <ATTRIBUTE dmrole="mango:stcextend.PhotFilter.name" dmttype="ivoa:string" value="SDSS/g'"/>
  <ATTRIBUTE dmrole="mango:stcextend.PhotFilter.zeroPointFlux" dmttype="ivoa:real" value="3643.0"/>
  <ATTRIBUTE dmrole="mango:stcextend.PhotFilter.magnitudeSystem" dmttype="ivoa:string" value="NotSet"/>
  <ATTRIBUTE dmrole="mango:stcextend.PhotFilter.effectiveWavelength" dmttype="ivoa:real" value="0.482"/>
  <ATTRIBUTE dmrole="mango:stcextend.PhotFilter.unit" dmttype="ivoa:string" value="um"/>
  <ATTRIBUTE dmrole="mango:stcextend.PhotFilter.bandWidth" dmttype="ivoa:real" value="0.1245"/>
</INSTANCE>
</GI ORAI S>
```

Table Provenance (including filters assignment) provided through a URL

```
▼<INSTANCE dmrole="mango:Source.associatedDataDock" dmttype="mango:WebEndPpoint">
  <ATTRIBUTE dmrole="mango:AssociatedData.semantic" dmttype="ivoa:string" value="computed"/>
  <ATTRIBUTE dmrole="mango:AssociatedData.dataType" dmttype="ivoa:string" value="provDM"/>
  <ATTRIBUTE dmrole="mango:AssociatedData.description" dmttype="ivoa:string" value="Complete VizieR catalogue Provenance"/>
  <ATTRIBUTE dmrole="mango:WebEndPpoint.ContentType" dmttype="ivoa:string" value="text/xml"/>
  <ATTRIBUTE dmrole="mango:WebEndPpoint.url" dmttype="ivoa:url" value="https://cdsarc.unistra.fr/viz-bin/provenance?cat=J/A+A/472/699&filter=true"/>
</INSTANCE>
```

# □ Conclusion



- The registry provides a schema that can be used to the needs of origins
- Does it concern only VizieR?
- Make a list of interesting metadata ?
- Where to put Provenance ?
  - In data discovery
    - Registry
    - Protocols (TAP)
  - In query output
    - VOTable ? `<info>`
    - Vodml ? (Link to Provenance / landing page)