The VizieR catalogues system certified by the "Data Seal of Approval"

IVOA
Santiago, 2017

Gilles Landais,
Francoise Genova, JY Hangouet
CDS team
VizieR team: P.Ocvirk, T.Boch, FX.Pineau, M.Brouty, E.Perret, S.Guehenneux, P.Vannier, T.Pouvreau

Acknowledgments: F.Ochsenbein

CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG

Open Data, what does it means for data center ?

Open data is a very active concept, demanded by authorities for publicly funded, especially for science data.

- The first mission consists of data **preservation** and to give **free access**

  In returns users have to cite the origin (author, article) with persistent identifier

2009A&A...501..539U

IVO-ID

- The second mission consists of providing **useful data**
  FAIR : **F**undable, **A**ccessible, **I**nteroperable, **R**eusable

  - The **basic meta-data**
    e.g. to provide table means the tables, column description, unit and everything needed for the data understanding
  - **Rich meta-data** which can be reused by softwares
    → interoperability : protocols, formats (IVOA)

DSA certification

DSA: Data Seal of Approval
http://datasealofapproval.org/en/

- The first edition of the DSA : 2008
- Dedicated to science data repositories
- A trusted-based certification simple to get than ISO

To get the DSA certification

The CDS (VizieR) has been certified by the DSA since 2014

- 16 criteria to document
- Certification has to be renewed every 3 years

The DSA requirements

New set of requirements in 2017

- WDS (World Data System) and DSA announce unified requirements
  → Common board and same criteria issued from work in RDA
- Importance of **data reusability**

DSA requirements based on the following five criteria:

- The data can be found on the Internet
- The data are accessible (clear rights and licences)
- **The data are in a usable format**
- The data are reliable
- The data are identified in a unique and persistent way so that they can be referred to

DSA evaluation

1. Answer the criteria and make the documentation
   - Auto-evaluation to indicate the compliance levels for each criteria
     - Not applicable
       e.g.: crit. confidentiality/ethic : VizieR is not subject to disclosure risk
     - 4 levels: Not considered it → implemented
2. DSA Peer review evaluation
   - 1 DSA + 1 WDS reviewer evaluate the responses and write a report
   - The Board examines the report and decides whether to allow the Seal to the data center
3. The Data center certification seal

   The answers will be public and available in the DSA website and the DSA logo can be displayed in the data center website

DSA criteria

- *Crit. 1: Mission/Scope : The repository has an explicit mission to provide access to and preserve data in its domain.*
- *Crit. 2:Licences : The repository maintains all applicable licenses covering data access and use and monitors compliance.*
- *Crit. 3 : Continuity of access : The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.*
- *Crit. 4: Confidentiality/Ethics : The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.*
- *Crit. 5 : Organizational infrastructure : The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.*
- *Crit. 6: Expert guidance: The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).*
- *Crit. 7: Data integrity and authenticity: The repository guarantees the integrity and authenticity of the data.*
- *Crit. 8: Appraisal : The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.*
- *Crit. 9 : Documented storage procedures : The repository applies documented processes and procedures in managing archival storage of the data.*
- *Crit. 10: Preservation plan : The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.*
- *Crit.11: Data quality : The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.*
- *Crit. 12: Workflows : Archiving takes place according to defined workflows from ingest to dissemination.*
- *Crit. 13 : Data discovery and identification : The repository enables users to discover the data and refer to them in a persistent way through proper citation.*
- *Crit. 14: Data reuse : The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.*
- *Crit. 15: Technical infrastructure : The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.*
- *Crit. 16: Security : The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.*

16 criteria grouped into 3 topics

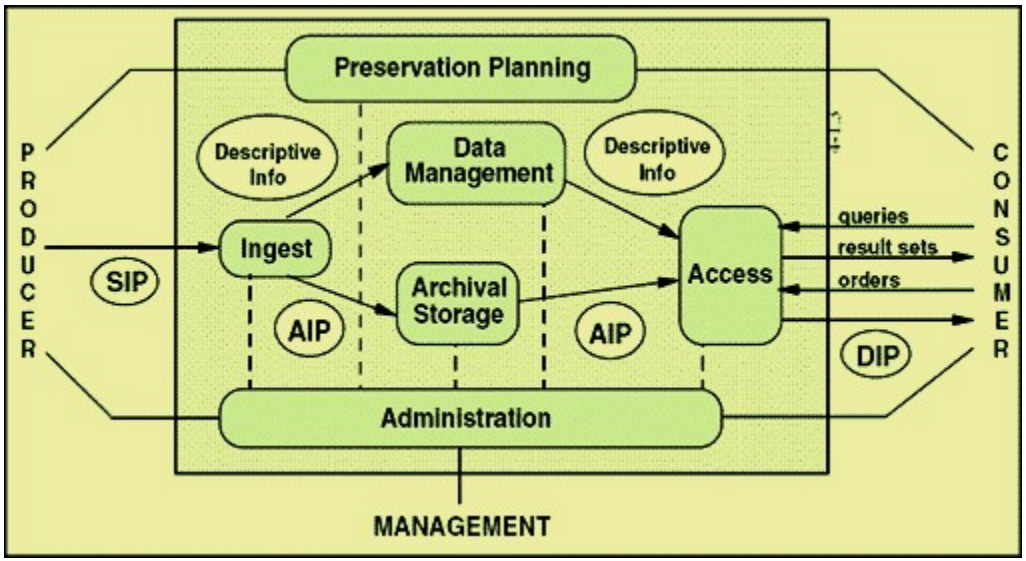
- **Organization** (4 criteria)
  - Mission
  - Continuity of access
  - contracts ...
- **Data management** (9 criteria)
  - Licences
  - Integrity and authenticity checking
  - identification
  - meta-data assignment
  - SI description (OAIS)
  - scientific expertise
- **Technology & security** (3 items)
  monitoring, softwares, backup, ...

Describing the Information System

OAIS: Open Archive Information System (norme OAIS, ISO 14721)

Defines a reference model and an editing guideline to describe an Information System (including the system, architecture, staff).

- The packaging information describes data together with meta-data
  - **SIP** : Submit Information Package
  - **AIP** : Archival Information Package
  - **DIP** : Diffusion Information Package
- The OAIS **entities**
  - the worflow description from the input to the dissemination
  - the preservation strategy for the long term
  - The data management
  - The archivage storage
  - The administration



The VO in the Data preservation context

VO responses to DSA criteria

- Open archive concept (from OAIS)
  - dissemination, free access
  - a source of proposals, outer from the Data Center with a scientific expertise
- Data acces :
  - **Reusability/Interoperability** with standardized format, protocols to access data
  - **Dicovery**: the IVO-ID, the registry
- **Sandardized meta-data** that can be applied on OAIS-packages

Criteria for which the VO doesn't provide solution

- Aspects which are in the charge of the Data Center
  - Local Organization and political aspects
  - Technology
  - Preservation strategy
    - Security/Backups
    - Long term management
  - ...
- Data curation
  - Licence management
  - Integrity & Authenticity checking

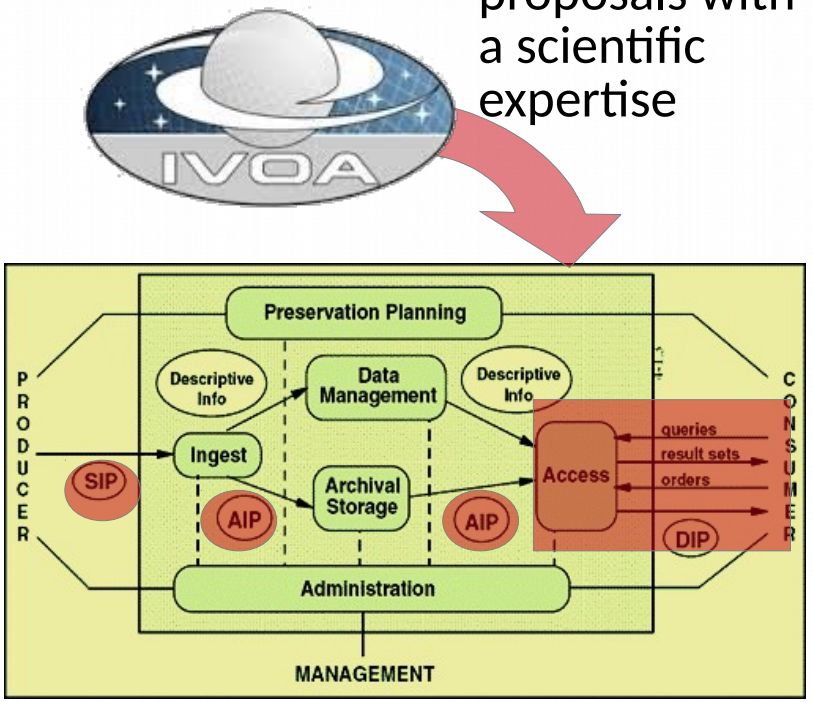
- *Crit. 1: Mission/Scope : The repository has an explicit mission to provide access to and preserve data in its domain.*
- *Crit. 2:Licences : The repository maintains all applicable licenses covering data access and use and monitors compliance.*
- *Crit. 3 : Continuity of access : The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.*
- *Crit. 4: Confidentiality/Ethics : The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.*
- *Crit. 5 : Organizational infrastructure : The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.*
- *Crit. 6: Expert guidance: The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).*
- *Crit. 7: Data integrity and authenticity: The repository guarantees the integrity and authenticity of the data.*
- *Crit. 8: Appraisal : The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.*
- *Crit. 9 : Documented storage procedures : The repository applies documented processes and procedures in managing archival storage of the data.*
- *Crit. 10: Preservation plan : The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.*
- *Crit.11: Data quality : The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.*
- *Crit. 12: Workflows : Archiving takes place according to defined workflows from ingest to dissemination.*
- *Crit. 13 : Data discovery and identification : The repository enables users to discover the data and refer to them in a persistent way through proper citation.*
- *Crit. 14: Data reuse : The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.*
- *Crit. 15: Technical infrastructure : The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.*
- *Crit. 16: Security : The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.*

Criteria for which the VO provides solutions

Criteria for which the VO provides partial solutions

Not the business of the VO

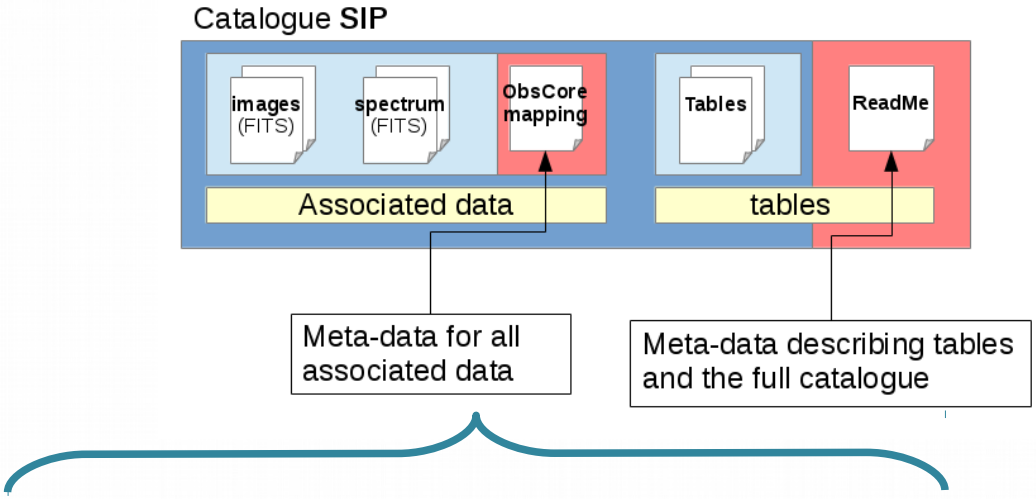
proposals with a scientific expertise



Example of standards usefull for DSA

- Meta-data & semantic
  - UCD
  - VO Data Model :
    ObsCore, Provenance build on W3C standards, ..
  - …
- Format
  - Votable based on XML
- Data access / protocols
  - TAP based on SQL
  - Access protocols on HTTP: SIA, SSA, SCS ….
  - Registry based on OAI-PMH

Example of Submission Information Package in VizieR




```
-----------------------------------------------
image: liste (669)
              hdu: -1
      target_name: OBJECT
             s_ra: WCS.getCenter(1)
            s_dec: WCS.getCenter(2)
            s_fov: WCS.getFieldOfView()
         s_region: WCS.getWorldPixelRegion()
     s_resolution: WCS.getWorldPixelSize()
           system: WCS.getAstroFrame()
            t_min: DATE-OBS
            t_max:
       t_exptime:
     t_resolution:
           em_min: WCS.getMin(1)
           em_max: WCS.getMax(1)
     em_res_power:
          spcunit:
          em_band:
       pol_states: WCS.getStokes()
  instrument_name:
    facility_name: TELESCOP
```

IVOA Recommendation

International Virtual Observatory Alliance

IVOA Documents

Observation Data Model Core Components and its Implementation in the Table Access Protocol Version 1.1

IVOA Recommendation 09 May 2017

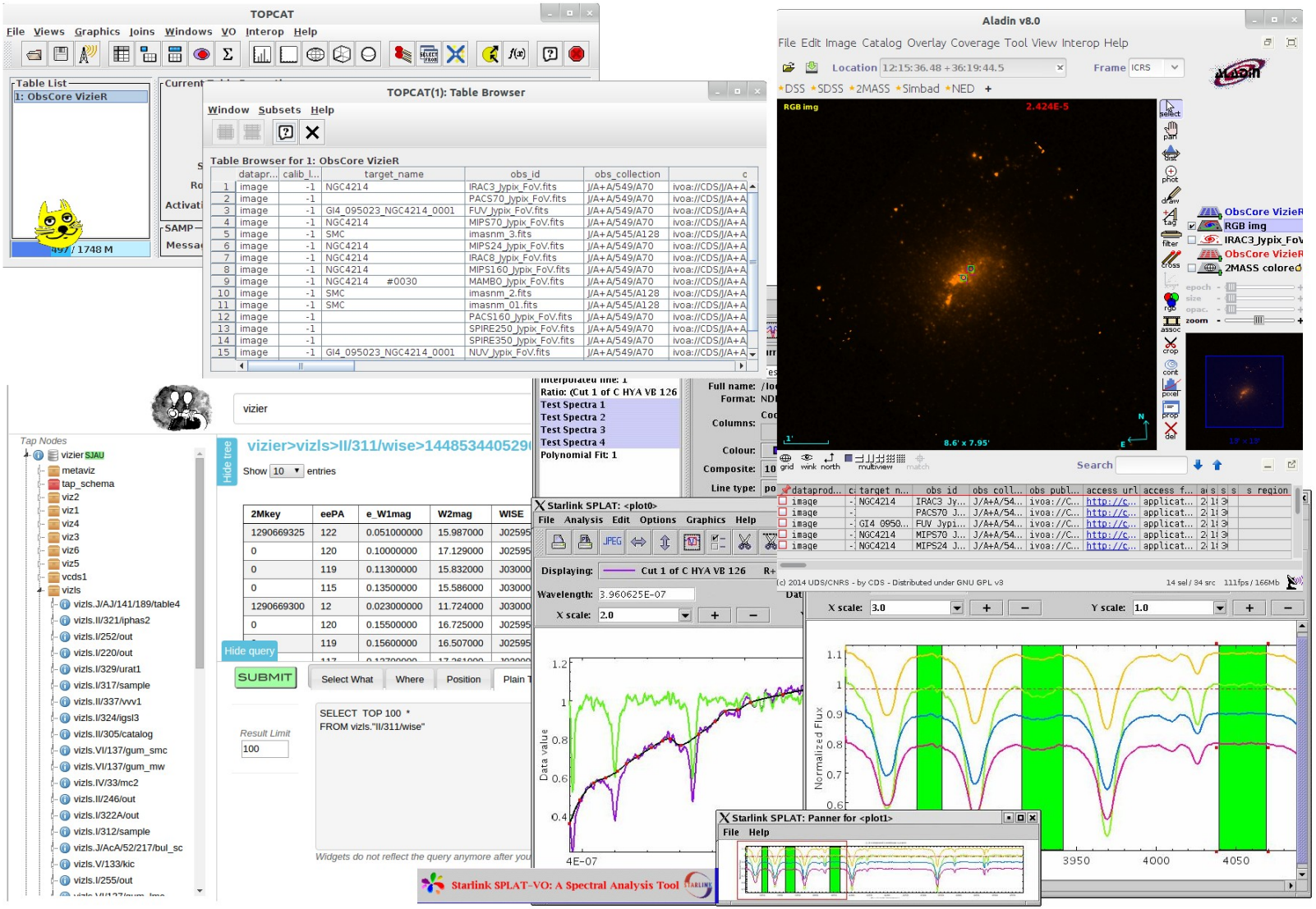
Providing useful data

Constitution of the Diffusion package (DIP)

Example: the VOTable (XML) output of images meta-data for the catalogue *The RMS survey (J/A+A/501/539)*

```
<?xml version="1.0" encoding="UTF-8"?>
<VOTABLE version="1.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.ivoa.net/xml/VOTable/v1.2" xmlns="http://www.ivoa.net/xml/VOTable/v1.2">
    <RESOURCE type="results">
        <INFO name="PROVIDER" value="ObsTAPVizieR">ObsTAPVizieR TAP service (CDS)</INFO>
        <TABLE>
            <FIELD ID="dataproduct_type" name="dataproduct_type" datatype="char" arraysize="*" ucd="meta.id" utype="Obs.dataProductType">
                <DESCRIPTION>Data product</DESCRIPTION>
            </FIELD>
            <FIELD ID="calib_level" name="calib_level" datatype="int" arraysize="1" ucd="meta.code;obs.calib" utype="Obs.calibLevel">
                    <DESCRIPTION>Calibration level</DESCRIPTION>
            </FIELD>
            <FIELD ID="target_name" name="target_name" datatype="char" arraysize="*" ucd="meta.id;src" utype="Target.Name">
                <DESCRIPTION>Astronomical object observed</DESCRIPTION>
            </FIELD>
            <FIELD ID="access_url" name="access_url" datatype="char" arraysize="*" ucd="meta.ref.url" utype="Access.Reference">
                <DESCRIPTION>URL used to access (download) dataset</DESCRIPTION>
            </FIELD>
            <FIELD ID="access_format" name="access_format" datatype="char" arraysize="*" ucd="meta.code.mime" utype="Access.Format">
                <DESCRIPTION>File content format</DESCRIPTION>
            </FIELD>
            <FIELD ID="facility_name" name="facility_name" datatype="char" arraysize="*" ucd="meta.id;instr.tel" utype="Provenance.ObsConfig.facility.name">
                <DESCRIPTION>Name of the facility used for this observation</DESCRIPTION>
            </FIELD>
            <FIELD ID="instrument_name" name="instrument_name" datatype="char" arraysize="*" ucd="meta.id;instr" utype="Provenance.ObsConfig.instrument.name">
                <DESCRIPTION>Name of the instrument used for this observation</DESCRIPTION>
            </FIELD>
            ....
            <DATA>
                <TABLEDATA>
<TR><TD>image</TD><TD>-1</TD><TD>G010.002</TD><TD>G010.0653-02.0583.fits</TD><TD>J/A+A/501/539</TD><TD>ivo://CDS.VizieR/J/A+A/501/539?res=G010.0653-02.0583.fits</TD><TD>http://cdsarc.
u-strasbg.fr/saadavizier/download?oid=8649728439500144 65</TD><TD>application/fits</TD><TD>299520</TD><TD>273.903926494813</TD><TD>-21.2226327993674</TD><TD>0.0166714515062126</TD><TD>
Polygon ICRS  273.91286300929494 -21.21429684032974  273.9128691824609 -21.23096348895762  273.894988966411 -21.23096829183595  273.89498481723183 -21.214301639067266 </TD><TD>0.42426
4101508559</TD><TD>52904.0</TD><TD></TD><TD></TD><TD></TD><TD>0.0616844301873808</TD><TD>0.0616844301873808</TD><TD></TD><TD>obs.image</TD><TD>I</TD><TD>VLA</TD><TD></TD><TD>8649728433
950014465 </TD><TD>3</TD><TD></TD></TR>
<TR><TD>image</TD><TD>-1</TD><TD>G010.392</TD><TD>G010.3924+00.5390.fits</TD><TD>J/A+A/501/539</TD><TD>ivo://CDS.VizieR/J/A+A/501/539?res=G010.3924+00.5390.fits</TD><TD>http://cdsarc.
u-strasbg.fr/saadavizier/download?oid=8649728439500144 66</TD><TD>application/fits</TD><TD>308160</TD><TD>271.642678113325</TD><TD>-19.6848890416928</TD><TD>0.0166666849828754</TD><TD>
Polygon ICRS  271.6515282097792 -19.676555485396577  271.65152914924016 -19.693222153027072  271.63382709582254 -19.69322217037945  271.63382799844186 -19.676555502733045 </TD><TD>0.4
2426410150855 9</TD><TD>52986.0</TD><TD></TD><TD></TD><TD></TD><TD>0.0616844301873808</TD><TD>0.0616844301873808</TD><TD></TD><TD>obs.image</TD><TD>I</TD><TD>VLA</TD><TD></TD><TD>86497
2843950014466 </TD><TD>3</TD><TD></TD></TR>
<TR><TD>image</TD><TD>-1</TD><TD>G010.596</TD><TD>G010.5959-00.8734.fits</TD><TD>J/A+A/501/539</TD><TD>ivo://CDS.VizieR/J/A+A/501/539?res=G010.5959-00.8734.fits</TD><TD>http://cdsarc.
u-strasbg.fr/saadavizier/download?oid=8649728439500144 67</TD><TD>application/fits</TD><TD>299520</TD><TD>273.062207488219</TD><TD>-20.1903611167135</TD><TD>0.0166665963224313</TD><TD>
Polygon ICRS  273.07108598604833 -20.182027595679347  273.071086859846 -20.19869426334745  273.05332804064017 -20.198694192001778  273.0533290664026 -20.18202752439772 </TD><TD>0.4242
64101508559</TD><TD>52904.0</TD><TD></TD><TD></TD><TD></TD><TD></TD><TD>0.0616844301873808</TD><TD>0.0616844301873808</TD><TD></TD><TD>obs.image</TD><TD>I</TD><TD>VLA</TD><TD></TD><TD>8649728
3950014467 </TD><TD>3</TD><TD></TD></TR>
<TR><TD>image</TD><TD>-1</TD><TD>G010.588</TD><TD>G010.5971-00.3893.fits</TD><TD>J/A+A/501/539</TD><TD>ivo://CDS.VizieR/J/A+A/501/539?res=G010.5971-00.3893.fits</TD><TD>http://cdsarc.
u-strasbg.fr/saadavizier/download?oid=8649728439500144 68</TD><TD>application/fits</TD><TD>299520</TD><TD>272.611152632876</TD><TD>-19.956304634747</TD><TD>0.0166647107386544</TD><TD>P
olygon ICRS  272.6200188939921 -19.94797205929301  272.6200177504964 -19.96463872495216  272.6022854353009 -19.964636770031667  272.6022884533792 -19.947970106144684 </TD><TD>0.424264
101508559</TD><TD>52904.0</TD><TD></TD><TD></TD><TD></TD><TD></TD><TD>0.0616844301873808</TD><TD>0.0616844301873808</TD><TD></TD><TD>obs.image</TD><TD>I</TD><TD>VLA</TD><TD></TD><TD>8649728439
50014468 </TD><TD>3</TD><TD></TD></TR>
                ...
                </TABLEDATA>
            </DATA>
</VOTABLE>
```

Providing useful data with added values



Conclusion

- A good description of the workflows is an important part of the documentation required to get the DSA certification.
- Other important aspects -not developped in this talk- which are in charge of the Data repositories
  - Political aspects : the continuity of access & funding
  - Technical infrastructure : redundancy and data security level
  - Long term preservation strategy …
- The new criteria are adapted to VizieR in the astronomical context which has an advanced disciplinary data sharing framework with the Virtual Observatory that enables interoperability