



DOI in the IVOA: the VizieR implementation example

G.Landais (CDS), A. Schaaff (CDS), G.Muench (AAS),
R.D'Abrusco (Center for Astrophysics)
.. and all DCP- IVOA WG

VizieR team: E.Perret, P.Vannier, M.Brouty, P.Ocvirk, C.Fix, G.Monari, FX.
Pineau, T.Boch

Editors: A&A, AAS and ADS

□ What is VizieR ?



VizieR provides the most complete library of published astronomical catalogues --tables and associated data-- with verified and enriched data, accessible via multiple interfaces. Query tools allow the user to select relevant data tables and to extract and format records matching given criteria. Currently, 20516 catalogues are available [more info](#)



A trusted repository for authors

- Open data repository, certified by the Core Trust Seal (CTS)
- Contents:
 - Tables from papers published in the major astronomical journals : AAS, A&A, MNRAS, ..
 - Reference catalogues & surveys e.g. Gaia, SDSS, 2MASS, UCAC, WISE
 - Logs of observations and incremental datasets updated periodically
 - Associated data: spectra, images, time-series

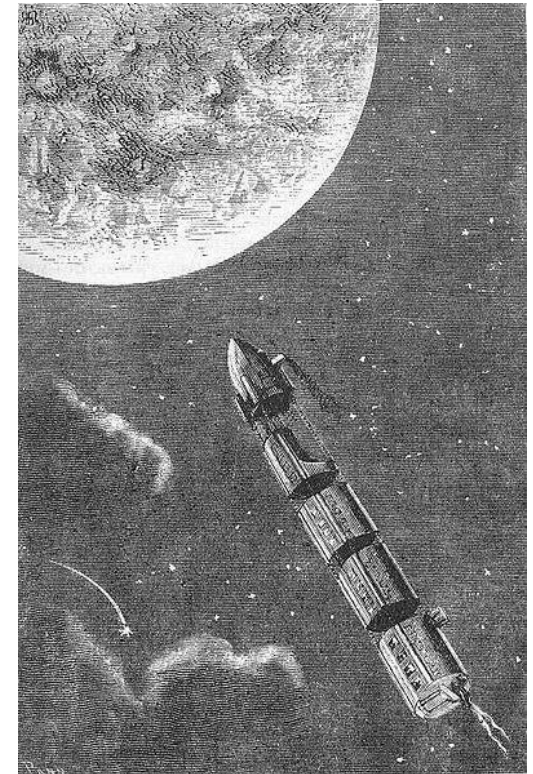
Free access to the data

Enriched data to be reused in the astronomical landscape through standards

- Identifiers: bibcode, ivooid, DOI
- Registered and available through the Virtual Observatory network



20,000 catalogues
under the sky



Henri de Montaut, an illustration from the novel *From the Earth to the Moon* by Jules Verne, 1868

□ DOI and FAIR principles



FAIR principles by RDA

Working group: FAIR Data Maturity Model WG

<https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>



Table 1 FAIR data maturity model indicators

- Provides a document to specify indicators for assessing compliance with FAIR principles
- DOI architecture meets a number of criteria : (if DOI metadata are well used)
 - **F:** Data is identified by a persistent identifier
 - **F:** Metadata is offered/published/exposed in such a way that it can be harvested and indexed
 - **A:** Data identifier resolves to a digital object
 - **A:** Metadata is guaranteed to remain available after data is no longer available
 - **I:** Data includes references to other data
 - **R:** Metadata includes information about the licence under which the data can be reused

FAIR	ID	Indicator	Priority
F1	RDA-F1-01M	Metadata is identified by a persistent identifier	●●● Essential
F1	RDA-F1-01D	Data is identified by a persistent identifier	●●● Essential
F1	RDA-F1-02M	Metadata is identified by a globally unique identifier	●●● Essential
F1	RDA-F1-02D	Data is identified by a globally unique identifier	●●● Essential
F2	RDA-F2-01M	Rich metadata is provided to allow discovery	●●● Essential
F3	RDA-F3-01M	Metadata includes the identifier for the data	●●● Essential
F4	RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	●●● Essential
A1	RDA-A1-01M	Metadata contains information to enable the user to get access to the data	●● Important
A1	RDA-A1-02M	Metadata can be accessed manually (i.e. with human intervention)	●●● Essential
A1	RDA-A1-02D	Data can be accessed manually (i.e. with human intervention)	●●● Essential
A1	RDA-A1-03M	Metadata identifier resolves to a metadata record	●●● Essential
A1	RDA-A1-03D	Data identifier resolves to a digital object	●●● Essential
A1	RDA-A1-04M	Metadata is accessed through standardised protocol	●●● Essential
A1	RDA-A1-04D	Data is accessible through standardised protocol	●●● Essential
A1	RDA-A1-05D	Data can be accessed automatically (i.e. by a computer program)	●● Important
A1.1	RDA-A1.1-01M	Metadata is accessible through a free access protocol	●●● Essential
A1.1	RDA-A1.1-01D	Data is accessible through a free access protocol	●● Important
A1.2	RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	● Useful
A2	RDA-A2-01M	Metadata is guaranteed to remain available after data is no longer available	●●● Essential
I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	●● Important
I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	●● Important
I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	●● Important
I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	●● Important
I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	●● Important
I2	RDA-I2-01D	Data uses FAIR-compliant vocabularies	● Useful
I3	RDA-I3-01M	Metadata includes references to other metadata	●● Important
I3	RDA-I3-01D	Data includes references to other data	● Useful
I3	RDA-I3-02M	Metadata includes references to other data	● Useful

□ And what about Data ?



According to scholexplorer :
9,673,560 dataset DOI (31,120,836 doi publication)

Scholeexplorer

Data providers who provide DOI in astronomy

CDS, Chandra, Exoplanets, IPAC, IRSA, KOA, Nasa, MAST, NED, ChineseVO, ESA, ...
but also Faisharing, zenodo...

Particularities in astronomy

- Precursor to share data
- The IVOA provides standards of interoperability :
a large network of interoperability with registries, softwares, services, databases...



Identifiers used in astronomy

- Bibcode: identifier dedicated for bibliography (ADS, VizieR, NED)
→ popular and used by authors for citation

ex: 2019A&A...628A..62M

- ivooid: identifier used in the Virtual Observatory framework
→ used in the VO registries to identify services

ex: ivo://CDS.VizieR/J/A+A/628/A62

□ Why a new identifier ?



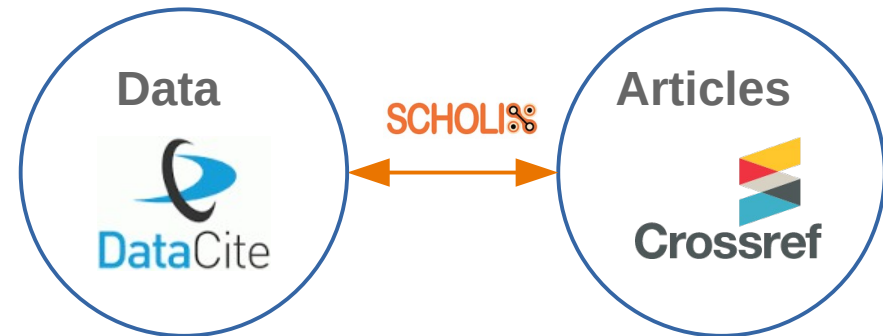
To take advantage of DOI

- Persistent identifier
- DOI is a standard commonly used by authors to cite articles / data
- Metadata included in the identifier : author, date, abstract, references, ...
- FAIR compatible

An architecture web designed

- Search engine (datacite: <https://www.datacite.org/>)
- The URL resolver links DOI to a “landing page”
- Cross references availability using identifier/URL..
- Rich semantics to specify interactions between resources

XML semantic compatibility



Identifier comparison

	bibcode	ivoid	DOI
Visibility/popularity			
Standard scope	Biblio. Astro.	IVOA	WWW
Citation			
Preservation guaranty			
URL resolution			
API available			
Metadata			
Cross references mechanism			

Comparison with VO registries



Identifier implementations

- Same harvesting technology (OAI-PMH)
- Metadata available in XML
→ Metadata +- compatible with some specializations !

Identifiers resolver adapted for the usage



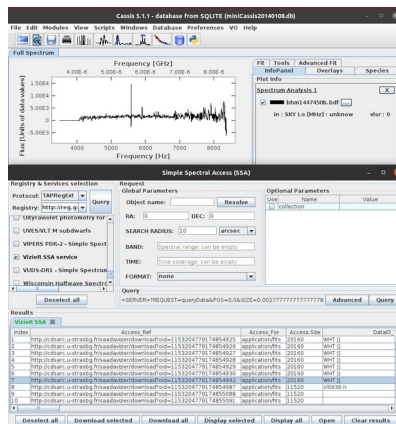
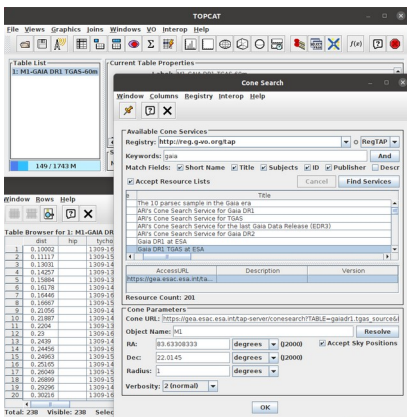
IVOA registry (ivoird)

- Consumed by softwares (machine readable)
- Hidden for users
- Can't be used as citation



DOI

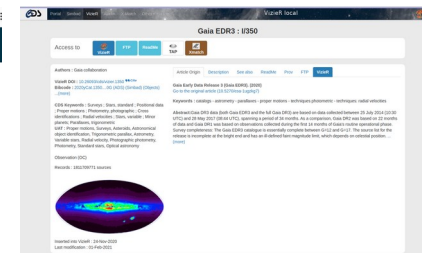
- Landing page resolver (human readable)
- Well adapted to be included in articles, web pages
- Used to cite data



DOI	https://doi.org/10.5270/esa-1vgfht?
Name	Gaia Early Data Release 3 (Gaia EDR3)
Mission	Gaia
Portal URL	https://www.cosmos.esa.int/web/guest/early-data-release-3
Content URL	Gaia EDR3 full data repository Gaia Archive at ESA
Version	1.0

CONTENTS OF GAIA EDR3

- The full astrometric solution (5 parameters) - positions on the sky (α , δ), parallaxes, and proper motions - for around 1.468 billion sources, with a limiting magnitude of about G=21 and a bright limit of about G=3. The astrometric solution is accompanied with error and quality indicators, the B/G/R/I, and source image descriptors.
- The full astrometric solution has been done as 5-parameter solution for 585 million sources and as 6-parameter solution for 885 million sources. In the 6-parameter solution, the additional third quantity is the so-called pseudo-colour that had to be included for sources without high quality colour information.
- In addition, two-parameters solutions - positions on the sky (α , δ) - for around 344 million additional sources.
- G magnitudes for around 1.806 billion sources.
- GBP and GRP magnitudes for around 1.842 billion and 1.835 billion sources, respectively.
- Please be aware that the photometric system for the G, GBP, and GRP bands in Gaia EDR3 is different from the photometric system for used in other TESS and Gaia DR1.



Comparison with V&O registries



IVOA registry

```
http://www.openarchives.org/OAI/2.0/ ... |
//www.openarchives.org/OAI/2.0/ ... |
05-25709:48:112//responseDate=
```

```
<request verb="GetRecord" identifier="ivo://CDS.VizieR/J/AJ/159/9" metadataPrefix="ivo_vor" http://cdweb.u-strasbg.fr/registry/beta/</request>
</request>
<header>
<identifier ivo://CDS.VizieR/J/AJ/159/9/</identifier>
<datestamp 2021-05-19T12:38:14Z/</datestamp>
<setSpec ivo managed /setSpec>
<setSpec J.AJ /setSpec>
</header>
<metadata>
```

```
<ri:Resource xmlns:iv="http://www.ivoa.net/xml/VOResource/v1.0/ ... >
<title HD 150382 lithium-rich star at the early-AGB stage/</title>
<shortName J/AJ/159/9/</shortName>
<identifier ivo://CDS.VizieR/J/AJ/159/9/</identifier>
<altIdentifier doi:10.26903/cds/vizier.51590009/</altIdentifier>
<curator>
```

```
<publisher ivo-id="ivo://CDS">CDS</publisher>
<creator name="Holanda N."/</creator>
<creator name="Drake N.A."/</creator>
<creator name="Pereira C.B."/</creator>
<date role="Updated">2020-02-20T07:59:08Z/</date>
<date role="Created">2020-02-20T07:59:08Z/</date>
<contact>
<name CDS support team/</name>
<address CDS, Observatoire de Strasbourg, 11 rue de l'Universite, F-67000 Strasbourg, France/</address>
<email cds-question@unistra.fr/</email>
</contact>
</curator>
```

authors

Pub.

keywords

```
<subject Chemical abundances/</subject>
<subject Effective temperature/</subject>
<subject Giant stars/</subject>
<description>
We report the discovery of a lithium-rich giant, HD 150382, a post-red-giant-branch clump star. The atmospheric parameters, the chemical abundances for 17 elements, and the isotopic ratio 12C/13C were determined using the ...
</description>
<source format="bibcode">2020AJ...159....9H</source>
<referenceURL http://cdsarc.unistra.fr/cgi-bin/cat/J/AJ/159/9/</referenceURL>
<type Catalog/</type>
<contentLevel Research/</contentLevel>
```

Resource links

```
<relationship type="IsServedBy"/>
<relatedResource ivo-id="ivo://CDS.VizieR/TAP">TAP VizieR generic service</relatedResource>
</relationship>
<relationship type="IsServedBy"/>
<relatedResource id="Conesearch service"</relatedResource>
</relationship>
<relationship type="related-to"/>
<relatedResource ivo-id="ivo://CDS.VizieR/J/A/364/674">Li-rich giants atomic lines (Casti)
<relatedResource ivo-id="ivo://CDS.VizieR/J/A/45/142/279">Stellar activity of late-type stars
</relationship>
```

Resource links

Rights

```
</content>
</rights-public/</rights>
```

```
<capability>
<interface xsi:type="cs:ParamHTTP">
<accessURL use="base" http://vizier.unistra.fr/viz-bin/votable?source=J/AJ/159/9/</accessURL>
<mirrorURL title="VizieR at IUCAA, Pune, India" https://vizier.iucaa.in/viz-bin/votable?source=J/AJ/159/9/</mirrorURL>
<mirrorURL title="VizieR at SAO, South Africa" http://vizieridsa.sao.ac.za/viz-bin/votable?source=J/AJ/159/9/</mirrorURL>
<queryType GET/</queryType>
<resultType text/xml/votable/</resultType>
</interface>
</capability>
<capability standardID="ivo://ivoa.net/std/TAPmax">
...
</capability>
<capability standardID="ivo://ivoa.net/std/Conesearch" xsi:type="cs:Conesearch">
<description Cone search capability for table J/AJ/159/9/table5</description>
<interface role="std" xsi:type="cs:ParamHTTP">
<accessURL use="base" http://vizier.unistra.fr/viz-bin/conesearch/J/AJ/159/9/table5?</accessURL>
<mirrorURL title="VizieR at IUCAA, Pune, India" https://vizier.iucaa.in/viz-bin/conesearch/J/AJ/159/9/table5?</mirrorURL>
<mirrorURL title="VizieR at SAO, South Africa" http://vizieridsa.sao.ac.za/viz-bin/conesearch/J/AJ/159/9/table5?</mirrorURL>
<queryType GET/</queryType>
<resultType text/xml/votable/</resultType>
</interface>
</capability>
<coverage>
<footprint ivo-id="ivo://oci1v0d" http://alasky.unistra.fr/footprints/cats/vizier/J/AJ/159/9/product=HOC&inside=512/</footprint>
</coverage>
<tableset>
<ichem>
<name default/</name>
<table>
<name J/AJ/159/9/table5/</name>
<description Properties of the Li-rich giants adopted for comparison/</description>
<column name="recno" name="description" record number .../</column>
<column name="Name" name="description" Star name/</column>
<column name="f" name="description" Flag on Name/</column>
<column name="Teff" name="description" Effective temperature/</column>
<column name="logg" name="description" Surface gravity/</column>
<column name="[Fe/H]" name="description" Metallicity/</column>
<column name="A(Li)" name="description" Lithium abundance, log(epsilon)(Li)/</column>
...
</table>
</tableset>
</tableset>
</ri:Resource>
</metadata>
</record>
</GetRecord>
</OAI-PPH>
```

? Interoperability desc.

doi DOI metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns="http://datacite.org/schema/kernel-4" ... >
<identifier identifierType="DOI">10.26093/CDS/VIZIER.51590009</identifier>
<creators>
```

```
<creator>
<creatorName Holanda N.</creatorName>
<nameIdentifier nameIdentifierScheme="ORCID" schemeURI="http://orcid.org/">0000-0002-8504-6248</nameIdentifier>
</creator>
<creator>
<creatorName Drake N.A.</creatorName>
</creator>
<creator>
<creatorName Pereira C.B.</creatorName>
</creator>
</creators>
```

authors

```
<titles>
<title xml:lang="en">HD 150382 lithium-rich star at the early-AGB stage</title>
</titles>
<publisher>Centre de Donnees Strasbourg (CDS)</publisher>
<publicationYear>2020</publicationYear>
```

Pub.

```
<subjects>
<subject schemeURI="https://cdsarc.unistra.fr/doc/ADCKwds.htm" subjectScheme="ADC Keywords"> Stars, giant
<subject schemeURI="https://cdsarc.unistra.fr/doc/ADCKwds.htm" subjectScheme="ADC Keywords"> Effective temperatures
<subject schemeURI="https://cdsarc.unistra.fr/doc/ADCKwds.htm" subjectScheme="ADC Keywords"> Abundances
<subject schemeURI="https://cdsarc.unistra.fr/doc/ADCKwds.htm" subjectScheme="ADC Keywords"> Rotational velocities
<subject schemeURI="https://cdsarc.unistra.fr/doc/ADCKwds.htm" subjectScheme="ADC Keywords"> Equivalent widths
</subjects>
```

keywords

```
</subjects>
<dates>
<date dateType="Created">20-Feb-2020</date>
</dates>
```

Resource links

```
<resourceType resourceTypeGeneral="Dataset">Dataset</resourceType>
<alternateIdentifiers>
<alternateIdentifier alternateIdentifierType="internal ID">J/AJ/159/9/</alternateIdentifier>
<alternateIdentifier alternateIdentifierType="void">ivo://CDS.VizieR/j/aj/159/9/</alternateIdentifier>
</alternateIdentifiers>
<relatedIdentifiers>
<relatedIdentifier relatedIdentifierType="bibcode" relationType="IsSupplementTo">2020AJ...159....9H</relatedIdentifier>
<relatedIdentifier relatedIdentifierType="DOI" relationType="IsSupplementTo">10.3847/1538-3881/ab5528</relatedIdentifier>
<relatedIdentifier relatedIdentifierType="DOI" relationType="Cites">10.26093/cds/vizier</relatedIdentifier>
</relatedIdentifiers>
</rightsList>
```

Rights

```
<rights rightsURI="https://cds.unistra.fr/vizier-org/licences_vizier.html">Refer to CDS usage/</rights>
</rightsList>
<description descriptionType="Other">
VizieR online Data Catalogue associated with article published in journal Astronomical Journal (AAJ) with title 'HD 150382: a lithium-rich star at the early-AGB stage' (bibcode: 2020AJ...159....9H)
</description>
```

Pub.

```
<descriptions>
<geoLocations>
<geoLocation>
<geoLocationPlace Strasbourg astronomical Observatory, France/</geoLocationPlace>
</geoLocations>
</descriptions>
</resource>
```

□ DOI in IVOA ?



➔ DOI are not a component in IVOA standards

Action in IVOA Data Curation and Preservation Interest Group (DCP)

- DOI Working group
 - DOI Survey : existing DOI in astronomy
 - expertise and feedback exchange
 - State of the art of DOI in IVOA
 - Eventually, DOI implementation proposals in IVOA standards
 - Eventually Proposal/ best practices for data-center



□ DOI implementation



Example of implementation



- **Chandra** catalogue: advanced DOI architecture that links resources at different granularity (catalogue/observation..) using the relatedIdentifier semantic



- **CADC**: self-serve DOI service deployed in the CANFAR Science platform. Cloud service with authentication and DOI hosting.



- **VizieR**: post publication DOI generation. Automated DOI Assignment for catalogues and link external resources (article, ivoid..)



- **China-VO**: data repository dedicated for authors with DOI reservation mechanism available before publication and published after article acceptance



- **VAMDC** (Atomic and Molecular Data) Query Store : allows users to reserve/assign DOI using zenodo for queries executed in the service.



- **MAST** : generate DOI for large datasets. DOI portal for users to search/generate DOI for a custom collection of one or more observations from Hubble, Kepler, GALEX, IUE,

□ Share expertise



DOI naming

A DOI should be opaque and not refer to a catalogue name, an astronomical object, etc. Nor should it bring up a hierarchy.

This is a guarantee of durability. Once registered, a DOI name can no longer be changed.

DOI sustainability

Do not modify value – use checksum!

If the dataset hosting is transferred to another data center, the DOI will not change, the resolving must connect to the new Landing Page which will give a link to the new dataset location. If this new data center will not manage DOIs a Landing Page must be maintained in any case.

Best practices

Updated dataset

It is possible that the data evolves for various reasons (error detection, generation with a new algorithm, new HPC infrastructure, etc.) In this case it is better to assign a new DOI.

These two DOIs will coexist and will be permanent references to different datasets

Case of multiple locations

A data provider could manage mirrors of the same data. In this case it is better to use a single DOI but to be sure that the datasets are the same and to resolve to the same Landing Page.

It is also possible that multiple data providers host a copy of the same archive. In this case it is possible to create a DOI for providers in order to preserve the statistics of each one and to indicate in the metadata the existence of the other dataset(s) in order to globally generate statistics. It is obviously necessary to ensure that the dataset are identical.

DOI resolving and Landing Page

The Landing Page does not need to display all of the metadata but a summary of the dataset to give the user additional information to decide whether or not to download the dataset or to use a resource in case of a service. It includes a link to the dataset. The links must always be up to date and not lead to HTTP errors.

The maintenance is the responsibility of the data provider who has made a moral commitment to do so. Sometimes it can also be subject to a financial penalty if you decide, for example, to stop to host a Landing Page

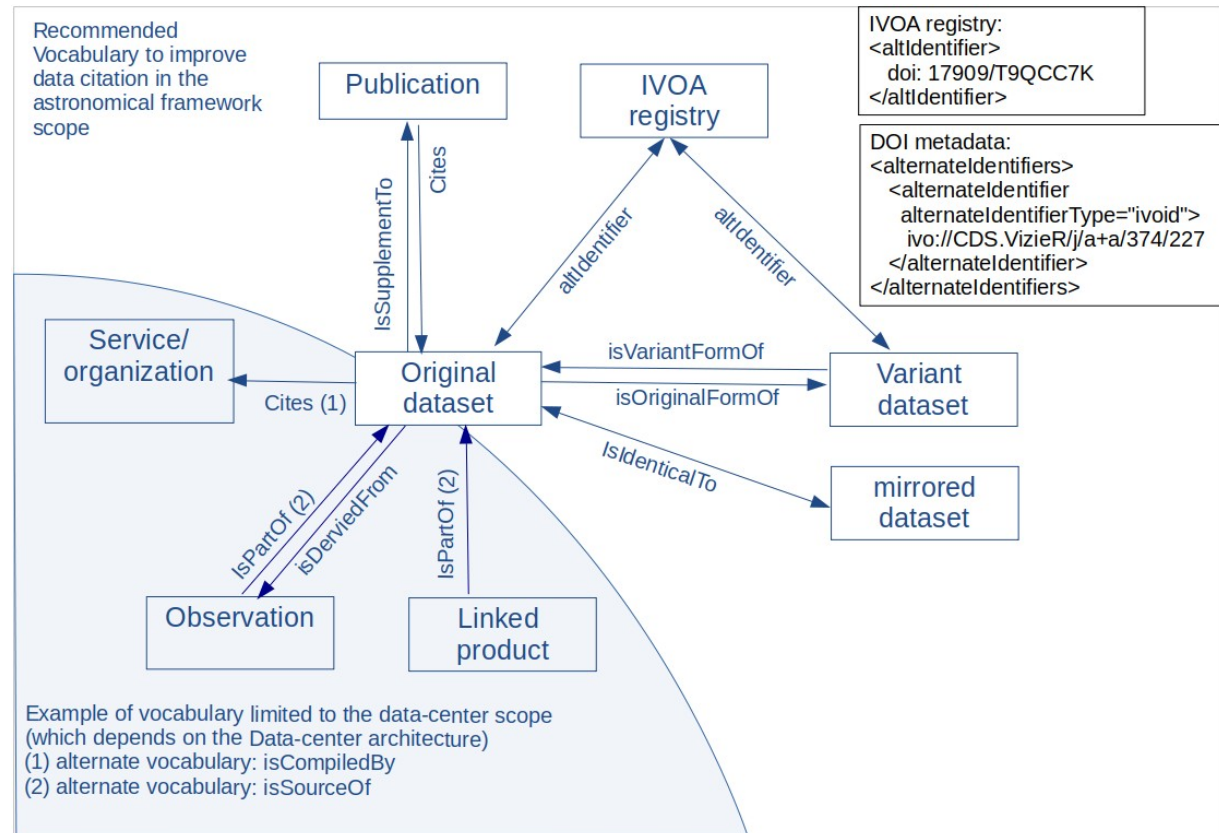
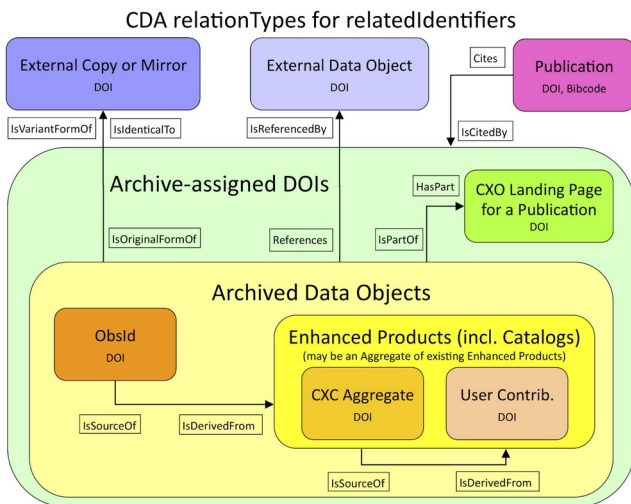
□ DOI recommendation



Linking resources using DOI meta-data

- Link the ivo-registry with DOI
- Exploit the DOI semantic “RealtedIdentifier” to link local and external resources

e.g.: Chandra DOI architecture



□ VizieR implementation



DOI in practice in VizieR

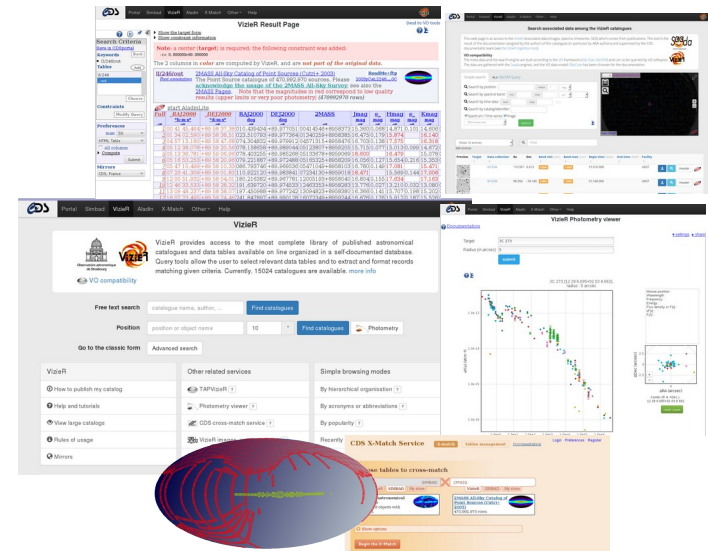
- ~14,000 DOI generated for A&A and AAS catalogues
- DOI registration organisation : Datacite
Only members can create DOI (in France, INIST (CNRS))
INIST: Institut de l'Information Scientifique et Technique
- A joint effort in CDS:
 - documentalist+develloppers
 - Editors feedbacks and acceptance



What is VizieR DOI ?

Granularity : the catalogue (!= table in the IVOA registry)

- The data : tables, associated data (FITS, spectra, images)
- The metadata (ReadMe) : coordinates, filters, time, ...
- Added values : links, plots....
- The services to access the data (VO, dedicated web pages)



□ DOI metadata



The VizieR DOI syntax

DataCite encourages DOI publishers to use a sufficiently opaque suffix
(=> Opaque suffix is preferable to avoid information in the identifier)

VizieR example : [10.26093/cds/vizier.34140699](https://doi.org/10.26093/cds/vizier.34140699)

DOI: [XXXX/XXXX](#)
Prefix (fixed by DataCite) Suffix (specific to each records)

Metadata subject to a particular attention in the VizieR context

- Remove abstract from DOI-metadata because abstract could be subject to licences
- Add journal origin in the title to improve its visibility
- Related identifiers :
 - Alternative identifiers : bibcode, IVOID
 - Link external identifiers with the relationship type :

DOI : the Datacite XML schema provides a rich grammar :
IsCitedBy, IsSupplementTo, IsSupplementedBy, IsContinuedBy, IsPartOf, IsReferencedBy, ...

```
<relatedIdentifier relatedIdentifierType="DOI" relationType="IsSupplementTo">  
10.1051/0004-6361:20031671  
</relatedIdentifier>
```



- **Orcid** of the first author
→ populate DOI & Orcid in VizieR from ADS








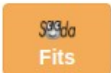
VizieR online Data Catalogue associated with article published in journal *Astronomical Journal* (AAS) with title 'K2-19b and c are in a 3:2 commensurability but out of resonance: a challenge to planet assembly by convergent migration.' (bibcode: 2020AJ....159....2P)


Link
catalogue
to its article




 Portal Simbad **VizieR** Aladin X-Match Other Help 

Spectra of CO and [C I] in protoplanetary disks : J/A+A/588/A108

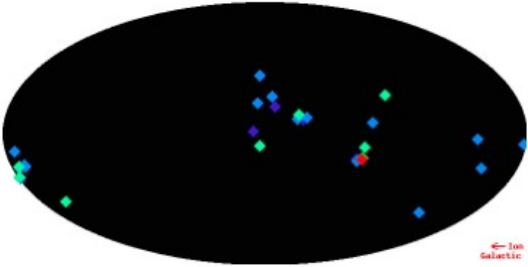
Access to      

Authors : Kama M.  , Bruderer S., Carney M. et..al

VizieR DOI : [10.26093/cds/vizieR.35880108](https://doi.org/10.26093/cds/vizieR.35880108)  [Cite](#)
Bibcode : 2016A&A...588A.108K (ADS)

CDS Keywords : YSOs ; Spectroscopy

Observation (OC)



Inserted into VizieR : 05-Apr-2016
Last modification : 13-Jun-2017

Article Origin **Description** FTP **VizieR**

Observations and modelling of CO and [C I] in protoplanetary disks. First detections of [C I] and constraints on the carbon abundance. (2016)
[Go to the original article \(10.1051/0004-6361/201526791\)](#)

Keywords : surveys - protoplanetary disks - submillimeter planetary systems

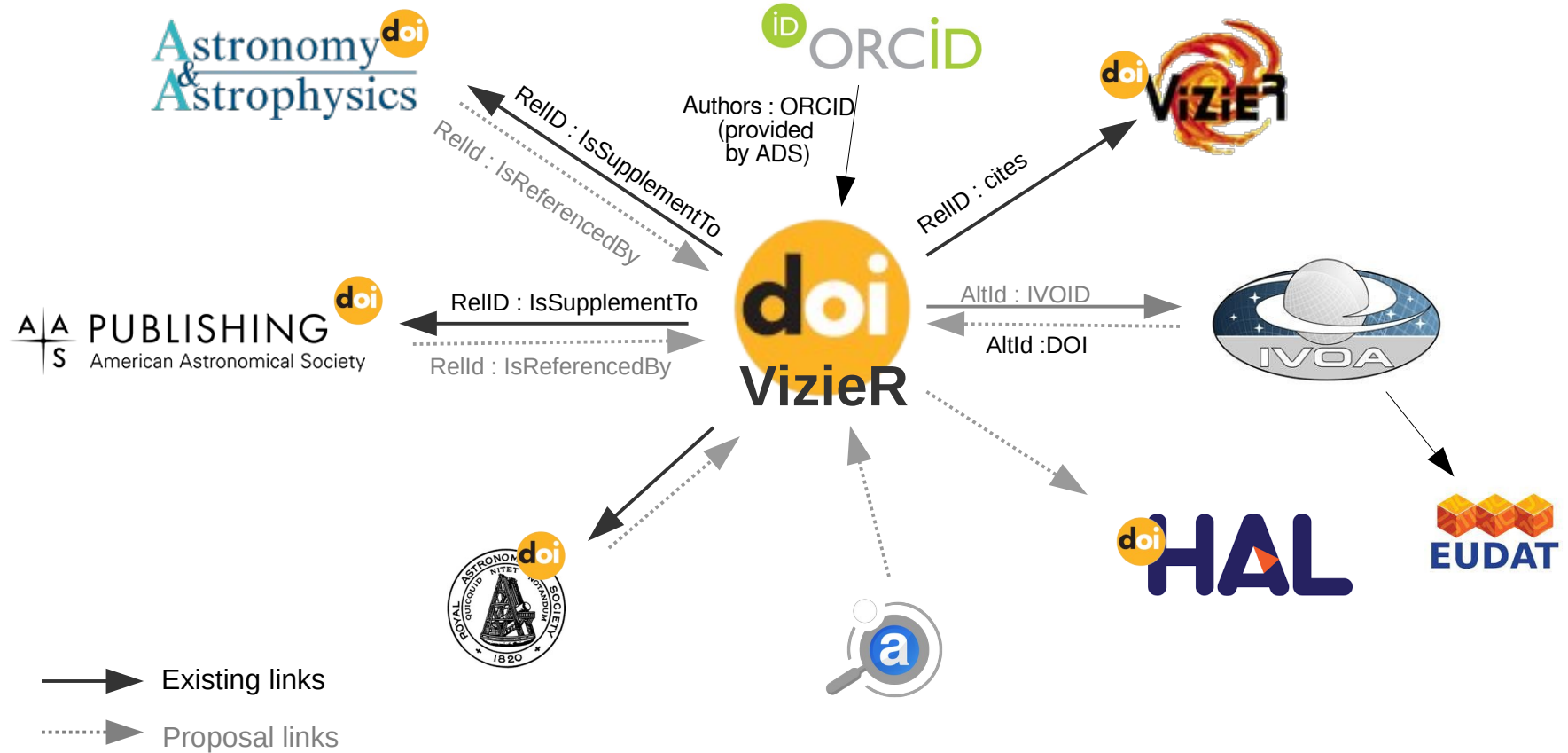
Abstract:The gas-solid budget of carbon in protoplanetary disks is related to the composition of the cores and atmospheres of the planets forming in them. The principal gas-phase carbon carriers CO, C⁰, and C⁺6 can now be observed regularly in disks. The gas-phase carbon abundance in disks has thus far not been well characterized observationally. We obtain new constraints on the [C]/[H] ratio in a large sample of disks, and compile an overview of the strength of [C I] and warm CO emission. We carried out a survey of the CO 6-5 line and the [C I] 1-0 and 2-1 lines towards 37 disks with the APEX telescope, and supplemented it with [C I] data from the literature. The data are interpreted using a ...(more)

<http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/588/A108>

VizieR DOI landscape



VizieR centric diagramm showing DOI links



□ Conclusion



- Comparison between ivo-registry and DOI publication
 - Not the same level of “interoperability”
→ metadata specialization
 - How to improve data-centers products links ?
 - How to improve data ↔ articles links ?
 - Howto include DOI in the IVOA standards ?
- VizieR implementations status :
 - Automated workflow available for AAS, A&A
→ how to manage post-publication DOI ?
 - Todo: include MNRAS in the workflow
 - Migrate ADC keywords used in VizieR DOI to UAT