



HAL
open science

Build FAIR workflow for astronomical catalogues

Gilles Landais, Pierre Ocvirk, Mark G. Allen, François Bonnarel, Emmanuelle Perret, Patricia Vannier, Marianne Brouty, Coralie Fix, Giacomo Monari

► **To cite this version:**

Gilles Landais, Pierre Ocvirk, Mark G. Allen, François Bonnarel, Emmanuelle Perret, et al.. Build FAIR workflow for astronomical catalogues. *Astronomical Data Analysis Software and Systems XXXI*, Oct 2021, Cape Town, South Africa. pp.295. <hal-04733919>

HAL Id: hal-04733919

<https://hal.science/hal-04733919v1>

Submitted on 16 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Build FAIR workflow for astronomical catalogues

G.Landais¹, P.Ocvirk¹, M.Allen¹, F.Bonnarel¹, E.Perret¹, P.Vannier¹,
M.Brouty¹, C.Fix¹, and G.Monari¹

¹*CDS, Observatoire Astronomique de Strasbourg, CNRS UMR 7550, France*

Abstract. Documented workflows are an important aspect of CDS operations, and are the basis for the CoreTrustSeal certification of the VizieR catalogue service (Ochsenbein et al. 2000). The workflow processes include metadata assignment in conformance with the standards and the needs of the astronomy research community. Among the standards, the Virtual Observatory promotes data models and access protocols to provide reusable data. In parallel, other standards and concepts are emerging with the rise of open data, for example the widespread use of DOIs. The FAIR principles and the increasing importance of open data for science, overlap and complete the IVOA capabilities. While IVOA standards enable a very high level of interoperability allowing data mining, workflows and software usage, the DOI enables the connection to a network of references resources well adapted for citation. We will describe FAIR-workflows developed in the VizieR catalogue service of the CDS and present how the data benefit from IVOA standards and DOIs.

1. FAIR understanding

With the growing of Open Data, it becomes essential for Data Centers to provide data in open access through world wide framework according to the FAIR principles (Findable-Accessible-Interoperable-Reusable). The FAIR Data Maturity Model working group in RDA (Research Data Alliance) provides a list of indicators to check FAIR status. In the document, Identifiers and metadata are the keystones of the FAIR concept.

Interoperability means for RDA, an interconnected framework which enables to link the data to each other (for instance data tables linked to an article). The resulting framework enables data connections in web pages and is adapted for citation.

For the IVOA, the interoperability concept is more focused on a framework which enables comparison and combination of data coming from different Data Centers.

The reusable concept is more important for RDA which focused the topic on license usage. License information in IVOA registration is possible but not often provided. One is to note that regarding the international laws, data without license are not considered to be reusable.

2. Implement documented workflow

2.1. The Core Trust Seal (CTS)

The CTS is an Open Data certification dedicated for scientific Data Centers - It responds to Open Data requirements applied on scientific results publicly funded. The CTS is a synthesis of the TRUST (Transparency, Responsibility, User focus, Sustainability and Technology) and FAIR Concepts.

The CTS asks Data Center to be an Open Archive : it promotes the OAIS (Open Archive Information System) norm which covers the workflows from the deposit to the dissemination, the archival procedures and the governance. For instance, the criteria R6 asks for an Expert guidance "The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant)."

The Importance of metadata is illustrated with the CTS criteria which asks the Data Center to have "[..]appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations." (criteria R11)

The Virtual Observatory community, composed by members and experts coming from different institutes in the world, promotes standards covering a wide range of topics in astronomy. Implementations of IVOA standards in the CDS catalogue Service "VizieR" were crucial to get the CTS certification (Landais et al. 2018).

2.2. Taking advantage of the IVOA standards and implementation

The IVOA provides standards -based on Data models building blocks and specification of access methods- which enables data reuse in a large audience.

Data and services registered into VO registries (Dower & Demleitner 2018) benefit to search engines used by VO software and API. The data accessibility is guaranteed when VO access protocols are implemented. And finally, a good description of the data (eg: UCD for tables) and relevant standard metadata (DataModels) improve data reuse.

IVOA implementations open the data to be used through a wide list of softwares (Aladin, Topcat, Cassis, etc). The IVOA standards can be implemented with the needs of the Data Centers. For instance the Simple Cone Search (SCS) Implies a limited number of standardized features and metadata while other like the Simple Spectral Access Protocol (SSA) require more metadata.

2.3. The CDS Catalogue Workflow

In VizieR, the metadata generation is the result of technologies and human expertise.

The VizieR metadata assigned in the workflow are compatible with the VO standards. The VO Data Models can be used to organize the archival of the data or can be mapped on top of internal archive storage.

For instance, the ObsCore Data Model is used to archive the image/spectra metadata. The workflow makes a mapping with FITS (Landais et al. 2019) header to populate an ObsCore table.

In the other hand, models like the ProvenanceDM or the PhotoDM are mapped on to CDS metadata (Landais et al. TBD) - It results in the CDS photometry viewer (Allen

et al. 2014) which exploits a selection of tables with photometry published in VizieR.

In parallel, the metadata extraction enables to publish the catalogues in the DOI network. This workflow is in action since 2019 for catalogues coming from A&A and AAS. In particular, VizieR exploits the relation capability of the DOI architecture to link the catalogues with dataset or article origins.

3. To find and access data

Many search engines are available for science data. In astronomy, the *ivoa-registries* provides a data oriented framework for astronomical resources whereas the DOI architecture is privileged in the edition workflow and used in more global scientific contexts. Note that bridges exist which enable to link resources from one context to the other. For instance the European collaborative Data Infrastructure EUDAT provides the *b2find* search engine which is populated from the IVOA registries.

3.1. Similarities and overlap

IVOA registries as well as DOI architecture provide capabilities to find and access resources. Both provide identification with metadata based on *scholix* (<http://www.scholix.org/>) and include metadata like provenance information (eg: authors, publication, ...). The API provided is OAI-PMH which enables client to harvest the metadata.

3.2. Differences

IVOA and DOI registration have their particularities which respond to their usage. It impacts the harvesting architecture (see figure 1) but also the metadata contents.

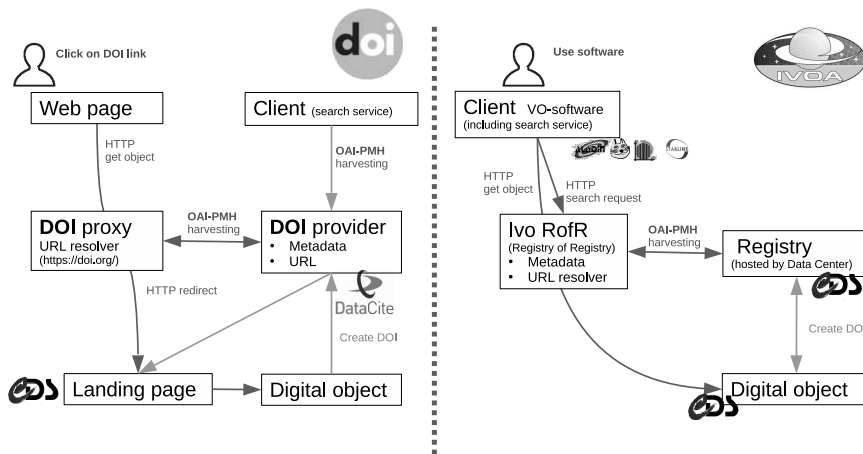


Figure 1. Comparison of the DOI and IVO registry architecture

The IVOA framework is a complex engineering enabling a high level of interoperability. The IVOA metadata requires an accurate data description including information

to access the Data. VO identifiers (ivoid) as well as protocols are hidden to the end-users - This framework is not adapted to cite the data.

In the other side, the DOI is visible for human. The DOI architecture provides an interconnected network, which enables to specify the role existing between linked resources. For instance the relation between the table and its article origin could be defined with the link verb "SupplementTo" (see figure 2) The DOI are built to be visible by users: the DOI are associated with landing pages and are well adapted to be included in articles, web pages. The DOI are commonly used for citation.

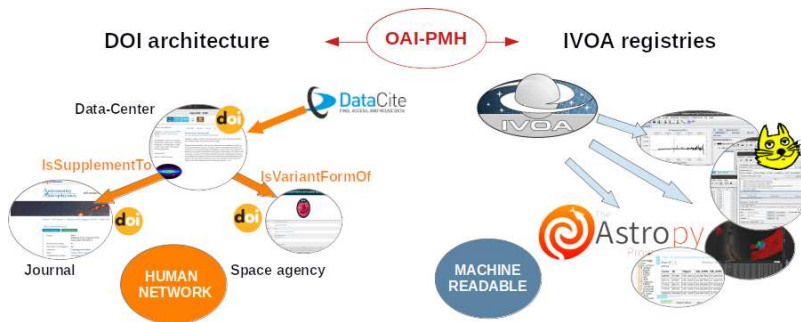


Figure 2. DOI and IVOA networks

4. Conclusion

With the rise of Open Data, concepts emerge and standards like DOI, or the needs of certification as the Core Trust Seal become essential. In astronomy, two Open Data worlds evolve: one in the editorial context promotes an interconnected network, the second is the Virtual Observatory dedicated for data. The both approach complete each other: one is human-readable enabling citation and human navigation in web pages, the second is machine-readable enabling data reuse by API and software

References

- Allen, M. G., Ochsenbein, F., Derriere, S., Boch, T., Fernique, P., & Landais, G. 2014, in *Astronomical Data Analysis Software and Systems XXIII*, edited by N. Manset, & P. Forshay, vol. 485 of *Astronomical Society of the Pacific Conference Series*, 219
- Dower, T., & Demleitner, M. e. a. 2018, Ivoa registry relational schema, version 1.1, IVOA standard. URL <https://www.ivoa.net/documents/RegistryInterface/>
- Landais, G., Genova, F., & Hanguët, J.-Y. 2018, in *European Physical Journal Web of Conferences*, vol. 186 of *European Physical Journal Web of Conferences*, 02003
- Landais, G., Michel, L., Derriere, S., Ocvirk, P., & Royer, F. 2019, in *Astronomical Data Analysis Software and Systems XXVI*, edited by M. Molinaro, K. Shortridge, & F. Pasian, vol. 521 of *Astronomical Society of the Pacific Conference Series*, 458
- Landais, G., Servillat, M., Bonnarel, F., Louys, M., Sanguillon, M., & Michel, L. TBD, in *Astronomical Data Analysis Software and Systems XXX*, edited by M. Molinaro, K. Shortridge, & F. Pasian, vol. TBD of *Astronomical Society of the Pacific Conference Series*, TBD
- Ochsenbein, F., et al. 2000, *A&AS*, 143, 230