



Data Origin in VO result

Contributors:

G.Landais, M.Demleitner, R.Savalle, G. Muench

and A.Accomazzi, R.D'Abrusco

OV FRANCE





Décrire l'origine != décrire les données

Définition : Un ensemble d'information sur l'origine des données distribuées, qui peuvent avoir été soumises à des sélections, des modifications résultantes de flux de données ou d'une requête utilisateur.

Exemple d'information :

Dublin Core



- Identifiant
- Auteurs
- Licence
- ...

Méta-data de reproductibilités

- Centre de données
- URL + paramètres
- Date d'exécution
- ...

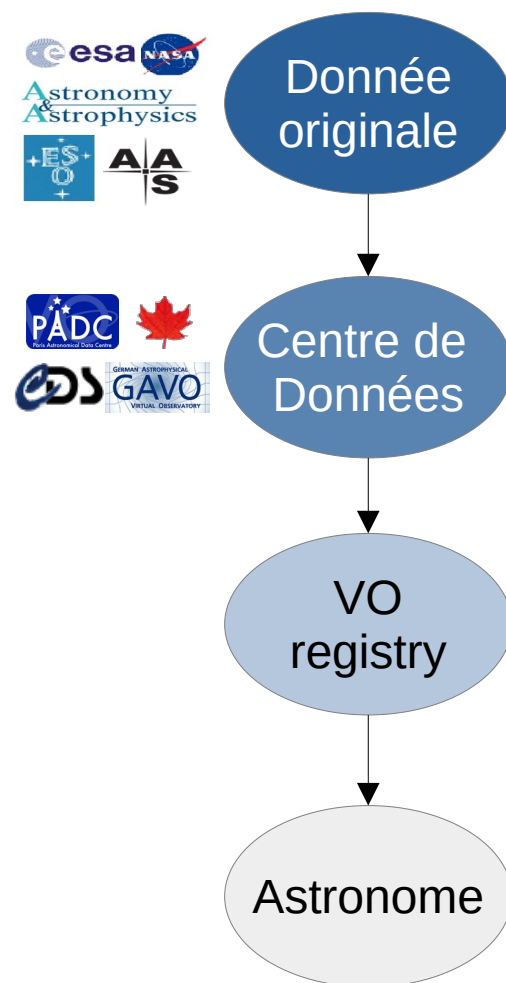
Motivations



- Amélioration de la compréhension des données
- Reproductibilité
- Citation

➔ Améliorer la chaîne « centre de données → utilisateurs »

- **Multiplication de la donnée (transformation/sélection)**
Plusieurs versions d'un même dataset disponible dans différents Centre de Données
Ex : Gaia (Gavo, CDS, ESA, ...)
- **Obtenir les informations d'origine**
ex :
 - construire un bibtex
 - Contacter le centre de donnéesAujourd'hui, il faut :
 - Accès aux « landing page »
 - Accès aux méta-données dépendant de l'implémentation



Où trouve t'on l'information dans le VO ?



- **VO registry**
inclus le DublinCore, +- Datacite compatible
 - Identifiants, auteurs, date de publication, droits, ...
 - Références vers d'autres sources

- **Data Models**
 - ProvDM
 - LastSetpProvenance (non standard)
 - DatasetDM (non standard)

- **VOTable**
 - Pas de standard pour mettre DOI, bibcode, auteurs, ...
 - Mivot (en cours de standardisation)

- **Protocoles ?**
 - DALI: information minimal: QUERY_STRING
 - TAP ? SCS ? Etc.

Méta-données proposées



<https://github.com/gilleslandais/ivoa-dcp-data-origin>

- Liste des méta-données :
 - Informations de reproductibilité
 - Méta-données d'origine (extension du Dublin Core)
- Méta-données dans les registres VO
- Sérialisation VOTable



Data Origin in the VO

Version 1.0

IVOA Note 2022-10-30

Working Group

DCP

This version

<https://www.ivoa.net/documents/data-origin/20221030>

Latest version

<https://www.ivoa.net/documents/data-origin>

Previous versions

This is the first public release

Author(s)

G.Landais, G.Muench, M.Demleitner, R.Savalle, looking for contributors

Editor(s)

G.Landais

Abstract

The goal of the document is to make the Data Origin more visible in the query results executed in the Virtual Observatory. The document lists metadata required to provide sufficient traceability to end-users in order to improve the understanding of the resultsets and enabling its reuse and its citation.

NOTE in work - template for a possible IVOA note.

Status of this document

Méta-données proposées



Informations de reproductibilité

Meta-data	Description	Mandatory
ivoid	ivoid identifier to link registry	yes
publisher	Data center that provides the VOTable	yes
version	Dataset version (or release date)	yes
service_protocol	Protocol access with version	
request	Request URL	
request_post	(POST Request) POST arguments	
request_date	Query execution date	
contact	email or URL contact	
landing_page	Dataset landing page	

Méta-données proposées



Méta-données d'origine

Metadata	Description	Level	Dublin Core	Registry
publication_id	Dataset identifier that can be used for citation	M	identifier	altIdentifier
curation_level	Controlled vocabulary (IVOA rdf, content_level)			contentLevel
resource_version	Dataset version or last release	R		version
rights	Licence URI		rights	rights
rights_type	Licence type (eg: CC-by, CC-0, ...)			
copyrights	Copyright text			rights
creator	The person or organization primarily responsible for creating the intellectual content of the resource.	R	creator	creator
editor	Editor name			
relation_type	An identifier of a second resource and its relationship to the present resource. Controlled vocabulary		relation	relationship
related_resource	Information about a second resource from which the present resource is derived.		source	relatedResource
publication_date	Date of publication	R		date
resource_date	Date of the original publication	R		

Sérialisation VOTable



- Simple sérialisation qui utilise les tags <INFO>
- Aucune modification des standards existants
- Bonne intégration des clients :
 - TOPcat
 - Fork Astropy (M.Demleitner)



TOPCAT(3): Table Parameters

Window Parameters Display Help

+ - ? X

Table Parameters for 3: J_AJ_161_36_table8.xml

Name	Value	Description
Name	//AJ/161/36/table8	Table name
Column Count	19	Number of columns
Row Count	6	Number of rows
Description	Planet candidate properties	
ivoid	ivo://cds.vizier/j/aj/161/36	
publisher	doi:10.26093/cds/vizier.51610036	
landing_page	https://cdsarc.cds.unistra.fr/viz-bin/cat//AJ/161/36	
publication_id	doi:10.26093/cds/vizier.51610036	
curation_level	Research	
resource_version	2022-10-07	
rights	https://cds.unistra.fr/vizier-org/licences_vizier.html	
creator	Bryson S.	
related_resource	2021AJ....161...36B	
editor	Astronomical Journal	
publication_date	2021-03-16	
resource_date	2021	
version	7.294	
protocol	Simple Cone Search 1.03	
request_date	2022-10-30T12:08:00	
request	https://vizier.cds.unistra.fr/viz-bin/conesearch//AJ/161/36/...	
contact	cds-question@unistra.fr	

Name: ivooid
Class: String
Shape:
Units:
Description:
UCD:
Utype:
Value: ivo://cds.vizier/j/aj/161/36

Exemple de sérialisation issue d'un SCS VizierR:

https://github.com/gilleslandais/ivoa-dcp-data-origin/blob/master/tests/J_AJ_161_36_table8.xml

Extraction d'un acknowledgment



Python code: <https://github.com/gilleslandais/ivoa-dcp-data-origin/tree/master/tests>

Template

We extract data published in <related_resource> (<creator>, <resource_date>),
via <publisher> services (ivoa resource=<ivoid>, <publication_date>)
using <service_protocol> (version <version>, executed at <request_date>)

Example

We extract data published in bibcode:2021AJ....161...36B (Bryson S., 2021),
via CDS services (ivoa resource=ivo://cds.vizier/j/aj/161/36, 2021-03-16)
using Simple Cone Search 1.03 (version 7.294, executed at 2022-10-30)