International
Virtual
Observatory
Alliance

Data Origin in the VO

Version 1.0

UNKNOWN DOCUMENT (fix DOCTYPE) ???

Working group
DCP
This version
https://www.ivoa.net/documents/NOTE-ivoa-data-origin/???
Latest version
https://www.ivoa.net/documents/NOTE-ivoa-data-origin
Previous versions
This is the first public release
Author(s)
G.Landais, G.Muench, 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 meta-data 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

UNKNOWN DOCUMENT (fix DOCTYPE)

A list of current IVOA Recommendations and other technical documents can be found at https://www.ivoa.net/documents/.

Contents

1 Introduction **2**
1.1 Role within the VO Architecture 2

2 Use cases **2**

3 State of the art **3**

4 Expected Data Origin **3**

5 Implementation tracks **3**

A Changes from Previous Versions **3**

Acknowledgments

Conformance-related definitions

1 Introduction

Data origin is required for end users to understand data, for citation and for reusability. The provenance is cited as a mandatory criterion in the EOSC or in RDA FAIR definition.

The virtual observatory provides an advanced framework to search and consume data provided by Data Centers or Space Agencies who apply curation in different level. In this context, Data Origin in output includes meta-data from the data producer (author, space agency) and the Data Center which hosts the resource. Depending of the implementation, the users can find the origin information in the Data center web pages (landing pages) or in the Registries of the Virtual Observatory. For citation, ADS (Astrophysics Data System, Nasa) provides citation capabilities with bibtex output. There are no VO standards to get the information easily yet. For instance, the origin meta-data are not included neither in output format, nor in protocols used to access the data. A list of basics meta-data added in strategic location (as result output or resource listing) would give easier the authors search who is looking for how to cite VO resources. Tracing data origin, from the producer to the final query enables also to report to end users the different agents implied in the data preservation (authors, data center, space agencies, journal)- especially when data can be subject to a curation which depends of the different agents. We propose to list the meta-data which responds to the need of Provenance and methods available today for their implementations.

1.1 Role within the VO Architecture

2 Use cases

- To get basic provenance information in VO output(author, pulication date, article, DOI,...)
- To trace data origin: query, resuources used to compute the result...
- To homogenize the Origin-metadata information in VO output.
  Example: Query the Gaia catalogue using VO services (for instance with topcat or any other VO-software). The registry lists Data Center (eg: Gavo, VizieR, ESA) which provides Gaia tables using TAP. The results returns VOTable having information in the header. However, the information depends of the implementation.
  To get citation information require the user to query the providers web sites or to use ADS.
- Relevant meta-data for final users to **cite** resources
- Fill the AAS citation template.
  Example : "we searched optical astrometric data of these sources from the Gaia (Gaia Collaboration et al. 2016) Early Data Release 3 (Gaia Collaboration et al. 2021) via the Gaia archive (Gaia Collaboration 2020)."
- Relevant meta-data for final users to understand data origin.
  Table provided by a Data Center can be a copy of an existing resource. For instance, a table published in a journal or by a Space Agency is also hosted in a Data Center like CDS, GAVO, etc. The data curation depends of the Data Center which can add associated data, enrich meta-data (eg: add filter for magnitude) or make a sub-selection of columns.
- Give me a bibliography of everything I've used in the workflow"

The basics meta-data should contain the data origin (space agency or authors, article references), the data center providing the resource, the date of publication ...

3 State of the art

4 Expected Data Origin

5 Implementation tracks

- mapping metadata with Provenance, Datacite, ivo-registry ?...
- Serialization using existing standard ? (ProvDM, mivot, VOTable, registry?)

A Changes from Previous Versions

No previous versions yet.