



*International
Virtual
Observatory
Alliance*

IVOA DataLink Implementation note Version 1.0

IVOA Note 2024-01-19

Working Group

DAL

This version

<https://www.ivoa.net/documents/DataLinkImp/20240119>

Latest version

<https://www.ivoa.net/documents/DataLinkImp>

Previous versions

This is the first public release

Author(s)

François Bonnarel, Markus Demleitner, Patrick Dowler, Laurent Michel, Mark Taylor

Editor(s)

François Bonnarel

Abstract

This note gives some hints on how the DataLink specification should be used. The DataLink specification document (Dowler and Bonnarel et al., 2023) describes the linking of data discovery metadata to access to the data itself, further detailed metadata, related resources, and to services that perform operations on the data.

Status of this document

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

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

Contents

1	Introduction	2
2	DataLink {links} endpoint recognition mechanisms	3
2.1	DataLink and SIAP-2.0 or ObsTAP services	3
2.1.1	DataLink discovery via format and reference columns .	3
2.1.2	SIAP-2.0, ObsTAP and service descriptors	4
2.2	DataLink in the context of other dataset discovery methods .	4
2.3	DataLink outside Data discovery context	5
2.4	Various recognition solutions for use cases introduced in section 3 and 4	6
3	Datalink {links} endpoint response FIELDS usage	6
4	Changes	8
	References	8

Acknowledgments

The authors would like to thank all the participants in DAL-WG discussions for their ideas, critical reviews, and contributions to this document.

Conformance-related definitions

The words “MUST”, “SHALL”, “SHOULD”, “MAY”, “RECOMMENDED”, and “OPTIONAL” (in upper or lower case) used in this document are to be interpreted as described in IETF standard RFC2119 (Bradner, 1997).

The *Virtual Observatory (VO)* is a general term for a collection of federated resources that can be used to conduct astronomical research, education, and outreach. The *International Virtual Observatory Alliance (IVOA)* is a global collaboration of separately funded projects to develop standards and infrastructure that enable VO applications.

1 Introduction

The DataLink specification (Dowler and Bonnarel et al., 2023) defines mechanisms for connecting data items discovered via one service to related data products, and web services that can act upon the data.

The {links} web service capability is a web service capability for drilling down from a discovered data item such as a dataset identifier, a source in a

catalog or any other data item. In the first case (typically an IVOA publisher dataset identifier) it allows to find details about the data files that can be downloaded, alternate representations of the data that are available, and services that can act upon the data (usually without having to download the entire dataset). In the latter cases it allows to associate metadata, images, cubes, spectra, timeseries or ancillary data to a source in the sky or other kind of item. The expected usage is for DAL (Data Access Layer) data discovery services (e.g. a TAP service (Dowler and Rixon et al., 2010) with the ObsCore (Louys and Tody et al., 2017) data model or one of the simple DAL services) to provide an identifier that can be used to query the associated DataLink capability. The DataLink capability will respond with a list of links that can be used to access the data.

The current document is first an implementation note for {links} endpoint recognition in various contexts (section 2). We are reviewing three methods by which DataLink documents can be associated to resource entries in service responses and suggest possible choices according to the context.

In addition it documents how to fill the {links} endpoint response FIELDS (section 3).

2 DataLink {links} endpoint recognition mechanisms

2.1 DataLink and SIAP-2.0 or ObsTAP services

SIAP-2.0 (Dowler and Bonnarel et al., 2015) services are queryable by setting constraints on the four "axes" of the data (2D-space, spectral, time and polarization) and their properties. Other archive metadata or data details are also queryable (target, collection, facilities, etc...). ObsTAP services are TAP services delivering the `ivoa.ObsCore` table queryable via ADQL. The successful response is a VOTable containing a mandatory result TABLE and optional service descriptor resource(s).

2.1.1 DataLink discovery via format and reference columns

The result table describes the dataset characterization on the different axes and give additional operative features. The various columns are mapped from the ObsCore-1.1 data model (Louys and Tody et al., 2017). The `accessReference` and `AccessFormat` fields give different solutions to reach the dataset for download or access. This can be done in two ways:

- The data discovery step has yielded a result with the `AccessFormat` value (media type): `application/x-votable+xml;content=datalink`
To the client, this indicates that what is given in the access reference

(e.g. the `access_url` column in ObsTAP or SIAP-2.0) is a DataLink document (see below for DataLink functionalities).

- In case the media type is different, the access reference is a direct link to the dataset download.

2.1.2 SIAP-2.0, ObsTAP and service descriptors

To enable additional DataLink functionalities, SIAP-2.0 or ObsTAP services can then add a service descriptor in the query response that indicates the availability of a DataLink service accompanying the DAL discovery service. It references one (or more) field(s) from the DAL response.

This is explained in detail in DataLink specification, section 4.4. The net result is that DataLink-enabled clients can find ancillary data and use associated services for data access or processing by virtue of being able to retrieve DataLink documents.

The service descriptor embedded in an ObsTAP query response is allowed because ObsTAP is a peculiar TAP service and TAP allows the TAP response to contain additional resources in addition to the "results" one (See section 2.9 of TAP-1.0 specification)¹. Hence the inclusion of a service descriptor service is valid.

2.2 DataLink in the context of other dataset discovery methods

Datasets can also be discovered in various other ways. Dataset "discovery" is defined as the discovery of the IVOA `publisher_id` of a dataset in standard services or by the ad hoc discovery of proprietary "dataset-ids" in combination with the a priori knowledge of a DataLink service "aware" of these dataset-ids.

Beside the most standard and modern discovery path via ObsTAP/SIAP-2.0, the following alternative scenarii are possible:

- The discovery can be done via an SIAP-1.0 or an SSA service. A DataLink service descriptor should have been added to the service query response to allow further guiding to additional resources
- The discovery can be done via a paper in a journal implementing publication of datasets associated with articles. The editor should describe or implement a DataLink service (eg via a service descriptor valid for the journal) for further usage and access to download DataAccess meta-data resources.

¹Apparently TAP-1.1 doesn't say explicitly that additional resources to the "results" one are allowed in the VOTable response

- Logs of observations are dataset descriptions belonging to the catalog category. They can be exposed in the VO using ConeSearch or TAP. They are in general not consistent with the ObsCore data model. However they allow some kind of "dataset discovery". A Service descriptor can be added to the VOTable containing the log file.
- HiPS is an IVOA standard for a global and hierarchical allsky access to pixel and catalogue data (reference HIPS). HiPS cells contain ids for original dataset which have been processed to build the Healpix maps. HiPS metadata file allows to generate a specific VOTable record for each cell of the HiPS collection including a "progenitor" URL. This can be a new way to discover datasets and start a DataLink session to find out additional associated resources.

2.3 DataLink outside Data discovery context

DataLink 1.1 provides description of resources to be attached to a dataset or to whatever item in a response table of a VO service. It is possible to use DataLink mechanisms in other VO contexts than Dataset discovery.

For example it could be useful to attach to a source in a catalog table such resources as

- data records for the same object contained in another catalog/database
- a service interface to an external database prepared to query around the object position.
- Datasets associated to the source (such as images, spectra, etc..)

In addition we could attach links to each measurement in a measurement list for an individual source formatted in VOTable. Light curves or radial velocity TimeSeries are good examples of such measurement lists. These links can attach:

- original dataset from which the measurement is extracted
- metadata describing the extraction method
- calibration information

We could also attach links to entities in an IVOA provenance metadata service response.

2.4 Various recognition solutions for use cases introduced in section 3 and 4

In such contexts, the {links} resource URL attached to a record may be given in various ways:

- A DataLink service descriptor (see DataLink 1.1 specification, section 4) could be added to the main table as long as the content of one of the table FIELDS may be used as the ID parameter in the {links} endpoint. An exemple is given in subsection 4.5 of the specification.
- In case the {links} endpoint URL is given in the main table for each row, a LINK element SHOULD be added to the FIELD containing this URL. The content-type of the LINK will be set to :

```
application/x-votable+xml;content=datalink
```

In the example below the content of the "Datalink" FIELD is used as an URL to the {links} endpoint for each row.

```
<FIELD name="DataLink" datatype="char" arraysize="*" ucd="meta.ref.url" >
  <LINK content-type="application/x-votable+xml;content=datalink" />
</FIELD>
```

- In case the main table contains an URL not systematically pointing to {links} endpoint, but may also point to responses with other content types (e.g. for single file download), the ObsCore utype "Access.reference" SHOULD be added to the FIELD and an additional FIELD with ObsCore utype "Access.format" SHOULD contain the "application/x-votable+xml;content=datalink" value when appropriate. A ref to the ID of the FIELD containing the URL SHOULD be added to avoid ambiguities.

In the example below, the Image-Link contains the URL pointing to a {links} response or the dataset itself depending on the value of the Image-Format FIELD.

```
<FIELD ID="ILink" name="Image-Link" datatype="char" arraysize="*"
  ucd="meta.ref.url" utype="AccessReference" />
<FIELD name="Image-Format" datatype="char" arraysize="*"
  ucd="meta.code.mime" utype="AccessFormat" ref="ILink" />
```

3 Datalink {links} endpoint response FIELDS usage

This section lists the standard FIELDS in the response and trate how to use them when this seems required

- **ID** : The ID can be a `publisher_id` or any proprietary item id common to the main service and the Datalink service.
- **access_url** : The usage of this FIELD is straightforward from the specification. Just pay attention to the URL with fragments case. By using this feature several time for the same root URL in your `{links}` endpoint response you may distinguish different semantics, `content_qualifier`, and `local_semantics` inside the same linked target. For example this could allow to distinguish science data from calibration or auxiliary subsets in the same file
- **service_def** : This is intended to refer to a service described by the service descriptor embedded in the same VOTable document. A specific IVOA implementation note on service descriptors should be written in the future
- **error_message** : There is nothing to add to the specification text here
- **description** : This free text should be specific to each link for each specific ID value and is intended to be read by a user when looking at the response table of the `{links}` endpoint. It is not really useful for the user that data providers fill this field with a generic text valid for all links sharing the same semantics value
- **semantics** : This field is intended mainly to give information to the client software in order to help it to take some decision. However it is also an important information for the end user. It is qualifying the relationship between the main item identified by the ID value and the thing which is retrievable via the `access_url`. For example `#calibration` or `#auxiliary` are not intrinsic properties of this thing but relatively to the main item (probably a science dataset). The intrinsic thing type or quality is given by `content_qualifier` (in the case if a `#auxiliary` relationship the thing pointed by the `access_url` can be a table, -an image, a cube...). Although
- **content_type** : The usage of this Field is straightforward. Again it is intended to help the client software to manage the response of the `access_url`
- **content_length** : This field is useful to inform the user or the client software how long will the transfer last and how much memory is required to host it
- **content_qualifier** : The usage of this field is rather straightforward. At the time of writing of this note it is recommended to use the ObsCore vocabulary for dataproduct type instead of the ivoa external

vocabulary which is not stabilized. Other standard vocabularies could be used in the future according to the quality of the link response, for example it could be the identifier of some facility consistent with the emergent facility specification

- **local_semantics** : local_semantics is mainly intended to help the software to sort out the links and take decisions about what to do with them when semantics, content_type and content_qualifier do not allow to do it. This will contain a proprietary vocabulary useful for the end user but is not some totally free text like the one in description.
- **link_auth** : This field contains an information which relates to the status of the access_url and is not dependant of the user
- **link_authorized** : This field will be filled by the server after the user authentication process. It is dependant both from the link and the user

4 Changes

2024-01-18: upgrade to citations of most recent VO specifications and adding a section on usage of response FIELDS

2023-12608: This was the initial version of this document, porting on github a previous internal DAL note.

References

Bradner, S. (1997), ‘Key words for use in RFCs to indicate requirement levels’, RFC 2119.

<http://www.ietf.org/rfc/rfc2119.txt>

Dowler, P., Bonnarel, F., Michel, L., Demleitner, M. and Taylor, M. (2023), ‘IVOA DataLink Version 1.1’, IVOA Recommendation 15 December 2023, arXiv:1509.06152.

<http://doi.org/10.5479/ADS/bib/2015ivoa.spec.0617D>

Dowler, P., Bonnarel, F. and Tody, D. (2015), ‘IWOA Simple Image Access Version 2.0’, IWOA Recommendation 23 December 2015.

<http://doi.org/10.5479/ADS/bib/2015ivoa.spec.1223D>

Dowler, P., Rixon, G. and Tody, D. (2010), ‘Table Access Protocol Version 1.0’, IWOA Recommendation 27 March 2010, arXiv:1110.0497.

<http://doi.org/10.5479/ADS/bib/2010ivoa.spec.0327D>

Louys, M., Tody, D., Dowler, P., Durand, D., Michel, L., Bonnarel, F., Micol, A. and IVOA DataModel Working Group (2017), ‘Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.1’, IVOA Recommendation 09 May 2017.

<http://doi.org/10.5479/ADS/bib/2017ivoa.spec.0509L>