International Virtual Observatory Alliance

IVOA DataLink protocol Version 2.0

IVOA Working Draft May 12, 2012

This version:
http://www.ivoa.net/Documents/WD/DataLink-20120419.html

Latest version:
http://www.ivoa.net/Documents/latest/DataLink.html

Previous versions:
http://www.ivoa.net/Documents/WD/DataLink.html

Editor(s):
François Bonnarel

Authors:
François Bonnarel, Laurent Michel, Mireille Louys, TBC....

Abstract

This document describes the DataLink protocol. This protocol defines services allowing to link datasets with various resources such as other complementary related datasets, metadata or other services. The current document situates the protocol within the global VO architecture, describes in detail various use cases for Data Link services and specifies input parameters and query response of such services.

Status of this Document

This document is an IVOA Working Draft.

Acknowledgments

1 Introduction

This protocol defines a new category of services allowing to link datasets with various resources such as other related datasets, metadata or other services. The specifity of these services is to provide a binding mechanism and metadata structure necessary to describe connected datasets or secondary data for independant datasets discivered in previous VO operations.

1.1 DataLink within overall VO Architecture

DataLink is one of the acces protocols, relies on a DataLink model in close relationships with other protocols such as ObsTAP, SIAP, and SSAP as is illustrated by figure 1

1.2 DataLink within DAL architecture

DataLink offers a binding mechanism to represent logical links between data and metadata in the IVOA. It defines an access protocol compatible to DAL protocols and extends the possibilities to select and retrieve complex datasets. It may extend the capabilities of existing dedicated protocols such as SSA, SIA, ObsTAP etc... It can be implemented as an independant service linking informations to an identified Dataset. It can also easily be included as extension of ObsTAP or typed DAL services. It can be a corner stone for simulation data access where logical connections between datasets and exploratory data need to be explicit, in close relationship with SimDAL.

1.3 Use cases: Linking the Discovery results with various types of access.

Suppose we have queried a DAL service... A table of results comes backwith, each row in the result table describing a dataset... In the case of ObsTAP the description contains a field named "reference" which is there for direct retrieval. Obviously richer types of "links" could be provided. Here we list various use cases and types of links to cover:

DataLink

Browser Based
Apps

Desktop Apps

Script Based
Apps

Utypes

DataLink

Resource Metadata

R
E
G
I
S
T
R

VOResource

UCD

Semantics

VO Query
Languages

CharDM

D
A
P

SIA

A   O

SSAP

Data
Models

A   O
C   C
C   O
E   L
S   S
S

SimpleDALRegExt

VOSI

Formats

Resource Identifier

VOTable

Storage
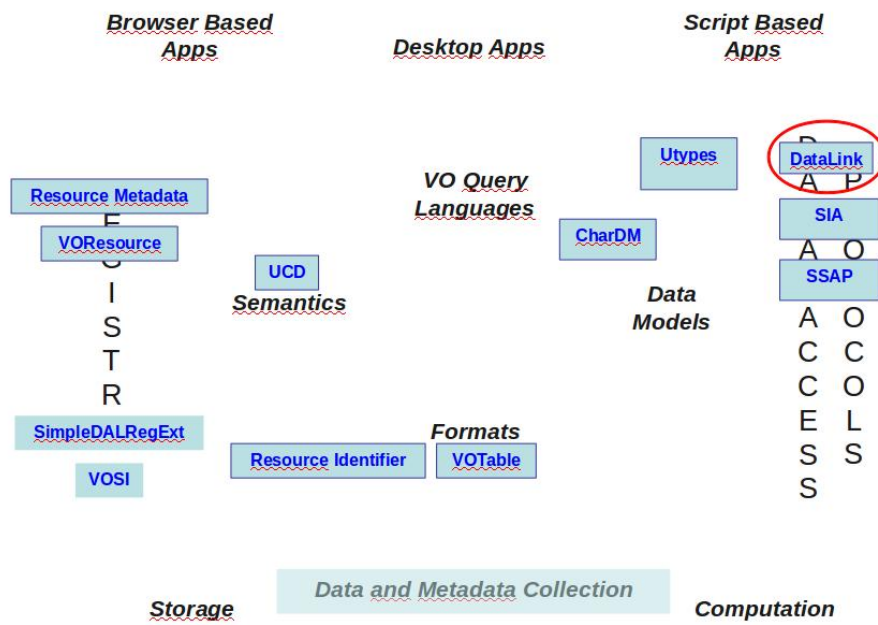
Data and Metadata Collection

Computation

Figure 1: Data Link in the global VO architecture

3

- Retrieving other data formats: ObsTAP provides reference for retrieval of the dataset in one single format. Datalink can provide other formats (eg, fits instead of VOTABLE for spectra, tar instead of Zip for archives, etc ...)
- Whatever the nature of the dataset we want to provide acces to some kind of "preview" of it such as 2D-preview of the field , an undersampled cube , a graphical and rough plot of the spectrum profile.
- Many datasets exposed with ObsTAP are actually complex data sets: bunch of spectra for echelle spectrograph or IFUs, ancillary data associated with science data in the same set, various exposures and intruments for the same field in X ray data[1], etc .... One typical use case there could be to explode the data set components ... It will be appreciated to obtain the list of the content of the dataset and to be able to extract some subpart of the dataset.
- Each dataset in the response could be more extensively characterised by extra metadata. These metadata could be either standardized by IVOA (full characterisation, Provenance) or proprietary... A specific mime type has to be given for each dataset...
- Link to another IVOA service with some fixed parameters such as PublisherDId, position, collectionID, etc... It can offer an access to complementary data/metadata via another IVOA service like ObsTAP or any kind of TAP service (for access to sources obtained by extraction from the current image),or simple access protocols such as SIA,SSA, etc... The specialized services can be used In query mode providing more detailed metadata or in AccessData mode with virtual data generation (cutout, resampling, etc....). Other UWS Web services for reprocessing could be pointed out too.... It could describe the ivorn of the service in the registry, and provide an URL template fixing some of the parameters dependant from the considered dataset if it makes sense.

Of course each dataset described in the Obstap or DAL service query response may have several such links and the nature of the links has to be described somehow.. The reference itself (generally a web access) has to be described by its URL, its format and size like in SSA services or ObsTAP services query responses but in addition the structure of the file has also to be given ... All the access rely on a key that uniquely identify the row:

---

[1]The pipeline of XMM-Newton produces source lists with a lot of scientific products attached to individual detections (preview, spectra, time series ...). There is no easy way to retrieve datasets linked to the sources selected from a VO resource yet. That can only be done by using various source detection parameters to state very specific queries on other VO services with a risk of mismatch for people without a fine expertise in XMM data.
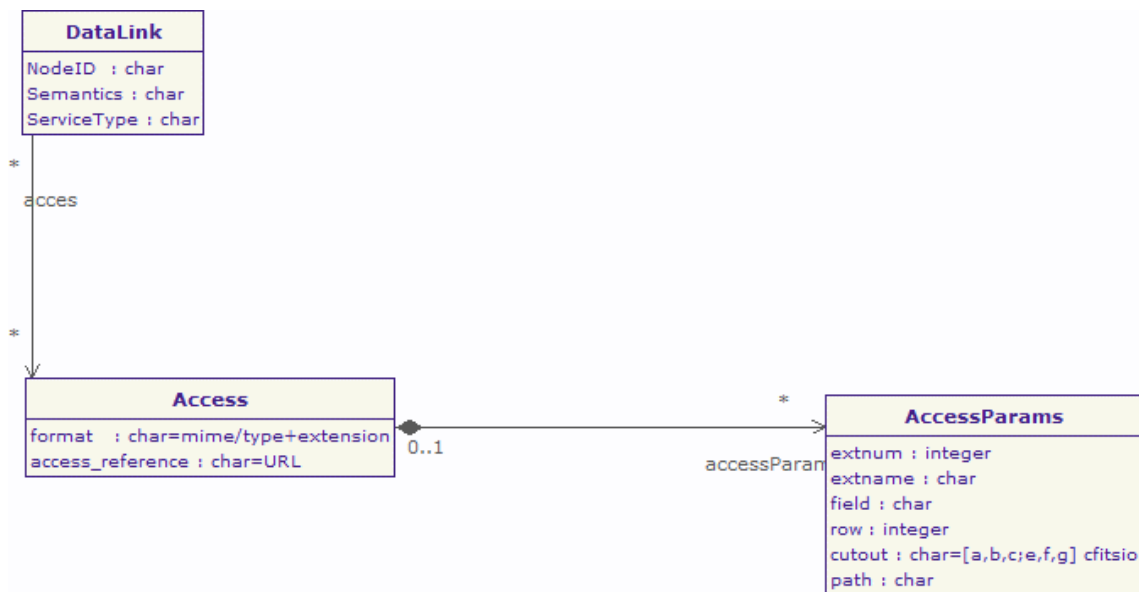
Figure 2: DataLink class UML diagram

PublisherDID is a good candidate, but any other key should be accepted as long as it is common to the main table and the DataLink response.

2   The DataLink model: links parameters

The remaining part of the document is there to ilustrate a possible implementaion of the DataLink concept. It may evoluate significanty in the future

We define here a DataLink class. See the UML class diagram in figure 2. The data Link node contains the following elements:

- the nodeID attribute: PublisherDID is a good candidate.
- an attribute giving the meaning (or semantics) of the link (Calibration file or SIA DESCRIPTION or catalogue part in a complex dataset -archive -)
- the IVOA type of the link: simple retrieval, IVOA service as other ObsTAP, SIA, SSA service, with either query or DataAccess method, UWS service, etc....
- an Access class detailing the structure of the link (see [1] for more complete description of this Access package.... It contains:
    - an URL or URI
    - a mime/type eg image/fits...
    - an estimated size for the response

- – a subtype : table, votable, mef, archive
- – a set of access parameters
    - * path
    - * cutout
    - * row
    - * field
    - * extnum
    - * extname

The Access Parameters are really important to provide localisation of a data product inside a complex data structure...

- The "Path" attribute allows to describe the file path and name inside an "archive" dataset
- The "array" attribute defines a cutout in a n-dimensional array image using the cfitsio syntax: [50:100,70:200] being the extracted subimage from pixel 50 to 100 in x and 70 to 200 in y.
- The "extnum" or "extname" attribute designates the extension number or name of a data subset in a multi extension FITS file. Extname can also be used here to designate a RESOURCE or TABLE name in a complex VOTABLE document ...
- "FIELD" and "ROW" have current significations in a FITS or VOTABLE table...

All these attributes are optional but must be used following this order : $[path][extnum|extname][field][row][array]$ It should be possible to locate any kind of significant structure in archives or datasets containing the most commun astronomy standards for files ...

3   Query parameters

A DataLink service is a simple DAL service providing data links for 1 to n Datasets identified by their dataset IDs

The specific input parameters are :

- nodeID: The value for example is a PublisherDiD or any other form of valid IVOA identifier) .
- nodeIDfile: The value is a set of PublisherDids stored in a file .

This is an example of a service query:
http://aaa.bbbb.fr/dal-services/datalink?nodeID="ivoa://xxx.yyy.edu/123345"

where ivoa://xxx.yyy.edu/12345 is an IVOA identifier (publisherDID or whatever) of a specific observation which could have been provided by eg an Obstap Query, or any other exploration.

4   Response

The result is presented as a VOTABLE with one field per attribute (attributes described above)... Table 1 gives a detailed description of the response TABLE field attributes
Appendix B gives an example of a query response.

Table 1: DataLink response fields

name	datatype	ucd	utype
nodeID	char	meta.ref.ivorn	dl:DataID.PublisherDiD
semantics	char	meta	dl:Semantics
serviceType	char	meta	dl:VOtype
reference	char	meta.ref.url	dl:Access.Reference
format	char	meta.code.mime	dl:Access.Format
size	int	meta.number	dl:Access.Size
subtype	char	meta.code	dl:Access.subtype
path	char	meta	dl:Access.AccessParams.path
extnum	char	meta	dl:Access.AccessParams.extnum
extname	char	meta	dl:Access.AccessParams.extname
field	char	meta	dl:Access.AccessParams.field
row	char	meta	dl:Access.AccessParams.row
array	char	meta	dl:Access.AccessParams.cutout

5   Extensions of ObsTAP for dataLinking

In any case an ObsTAp service is a TAP service and may accept several query languages...([2]) The mandatory ADQL interface can be usefully completed by PQL for example. In the case we use a Obstap service with a PQL interface the standard doesn't require the single table response ... This allows to add DAL extensions (additional tables) - see SSA recommendation for a definition of DAL extension mechanism - to the main standard Obstap table. Adding a specific DataLink response to the main Obstap table becomes then possible.

The PublisherDId FIELD which is common to the two tables allows to relate records in the main obstap table to records in the concatenated DataLink response, and can be used as a reference key. This approach avoids the necessity for the clients to extract the URL for the dataLink service from the main table and to start a new query on another service... In addition a future version of ObsTAp could benefit defining an additional FIELD with utype "dataLink" which will be pointing to a DataLink service (a PARAM could also be sufficient, giving the service root URL), using the PublisherDid value of each record as the main parameter for the query.

Appendix A: Background and motivation of the DataLink concept within the DAL landscape

Discussion in the IVOA has shown that "DataLink" is a usefull concept within the scope of the Generic data set protocols such as ObsTAP (1.2.2 below) . See [3] for first description of the GDS concept and [4] for first mention of the DataLink concept.

Within the scope of DAL protocols, the Generic dataset concept illustrate the need for a type of services valid for any kind of data , as a counterpart of typed interfaces such as SSA and SIA ... which can describe only a single category of data, but can do it in finer details, with a data model specific to the data.

As a conterpart to their generality generic dataset services can only describe data sets or files as they are stored in some archive, whereas the typed interfaces (SIA, SSA ...) can describe, and provide access to both static archival datasets as well as virtual data built on the fly from these static datasets (see [5]).

Finally, since generic dataset services such as ObsTAP are valid for any type of data they can also encompass all sorts of complex data.

A complex observation consisting of several related data products can be described globally via generic dataset query mechanisms such as ObsTAP query. For example we might have a survey field consisting of a spectral data cube, some 2-D projections of the cube (integrated flux, maps for a given wavelength, velocity/position maps, etc ...), a source catalog for the field computed from the 2-D continuum, and possibly some extracted spectra of objects in the field. Hence there is a need for a flexible frame where all related info can be gathered and made accessible.

It is the work of Client applications to deal with the data sets.

Therefore clients supporting simple access protocol will manage selection

and retrieval of existing archive datasets while more sophisticated applications can expose connected resources via DataLinks.

Identifying partnership between files require 3 basic properties for a link to be described : a dataset ID (PublisherDID or whatever) used as a key, a URL and a link type and semantic.

A dataset can have a number of such links.

Discovery with Obstap

ObsTAP is actually playing the role of a Generic dataset discovery service..

Built on the work done on data models ([6]) and TAP ([2]), the ObsTAP standard service protocol ([7]) allows to expose standard metadata describing available datasets stemming from observation collections, including complex datasets delivered as archive files. The goal of ObsTAP is to provide such a capability based upon an essential subset of the general observational data model.

Specifically, this effort defined a database table to describe astronomical datasets (data products) stored in archives that can be queried directly with the TAP protocol. This is very useful for global data discovery as any type of data can be described in a straightforward and uniform way.

The described datasets can be either directly downloaded, or linked to IVOA Data Access Layer (DAL) protocols such as for accessing images (SIA) or spectra (SSA) or any appropriate service. These links can potentially be used to perform more advanced data access operations on the referenced datasets.

Actually this is what is behind the "DataLink" concept which is described in the main part of this document.

Appendix B: response example

In the following response example, where the main dataset is discovered as a tar archive, the first record links to the full retrieval of the whole archive, the second record links to a FITS image cccc.fits in the directory image of the tar file.... The last record links to the query method of a SIAP service which will answer by description of images sharing the "ivoa://xxx.yyy.edu/123345" Obsid. Figure 3 shows a visualization of this query response with TOPcat.

Figure 3: DataLink response example visualized with TOPcat

```
<TABLE name="DataLinks">
  <FIELD name="Obsid" utype="dl:Dataid.ObservationID" datatype="char" arraysize="*"/>
  <FIELD name="Semantics" utype="dl:Semantics" datatype="char" arraysize="*"/>
  <FIELD name="Servicetype" utype="dl:Votype" datatype="char" arraysize="*"/>
  <FIELD name="reference" utype="dl:Access.Reference" datatype="char" arraysize="*"/>
  <FIELD name="format" utype="dl:Access.Format" datatype="char" arraysize="*"/>
  <FIELD name="size" utype="dl:Access.Size" datatype="char" arraysize="*"/>
  <FIELD name="subtype" utype="dl:Access.subtype" datatype="char" arraysize="*"/>
  <FIELD name="path" utype="dl:Access.AccessParams.path" datatype="char" arraysize="*"/>
  <FIELD name="extnum" utype="dl:Access.AccessParams.extnum" datatype="char" arraysize="*"/>
  <FIELD name="extname" utype="dl:Access.AccessParams.extname" datatype="char" arraysize="*"/>
  <FIELD name="field" utype="dl:Access.AccessParams.field" datatype="char" arraysize="*"/>
  <FIELD name="row" utype="dl:Access.AccessParams.row" datatype="char" arraysize="*"/>
  <FIELD name="array" utype="dl:Access.AccessParams.cutout" datatype="char" arraysize="*"/>
  <DATA>
    <TABLEDATA>
      <TR>
      <TD>ivoa://xxx.yyy.edu/123345</TD>
      <TD>full dataset</TD>
      <TD>retrieval</TD>
      <TD>http://xxx.yyy.de/archive.tar</TD>
      <TD>3.4Gb</TD>
      <TD>archive/tar</TD>
      <TD>none</TD>
      <TD>none</TD>
      <TD>none</TD>
      <TD>none</TD>
      <TD>none</TD>
      <TD>none</TD>
      <TD>none</TD>
      </TR>
      <TR>
      <TD>ivoa://xxx.yyy.edu/123345</TD>
      <TD>image</TD>
      <TD>retrieval</TD>
      <TD>http://xxx.yyy.de/archive.tar</TD>
      <TD>1Gb</TD>
      <TD>image/fits</TD>
      <TD>none</TD>
      <TD>image/cccc.fits</TD>
```

```
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        </TR>
        <TR>
        <TD>ivoa://xxx.yyy.edu/123345</TD>
        <TD>image metadata</TD>
        <TD>sia</TD>
        <TD>http://xxx.yyy.de/sinea?query&obsid="ivoa://xxx.yyy.edu/123345"</TD>
        <TD>1Kb</TD>
        <TD>application/xml+votable</TD>
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        <TD>none</TD>
        </TR>
    </TABLEDATA>
  </DATA>
</TABLE>
```

References

[1] François Bonnarel, Igor Chilingarian, and Mireille Louys. Ivoa working draft: Characterisation data model complements and new features. *http://www.ivoa.net/internal/IVOA/CharacterisationDataModel/NewChar2.pdf*, 2012.

[2] P. Dowler, G. Rixon, and D. Tody. IVOA Recommendation: Table Access Protocol Version 1.0. *ArXiv e-prints*, October 2011.

[3] Doug Tody, François Bonnarel, Markus Dolensky, and Jesus Salgado. Ivoa note: Ivoa data access layer service architecture and standard profile. *http://www.ivoa.net/internal/IVOA/SiaInterface/DAL2_Architecture.pdf*, 2008.

[4] Doug Tody. http://www.ivoa.net/internal/ivoa/200905dalsessions/siapv2-may09.pdf, 2009.

[5] François Bonnarel, Daniel Durand, Alberto Micol, Anita Richards, Jesus Salgado, and Doug Tody. Ivoa working draft: Simple image access protocol. version 2.0. *http://www.ivoa.net/internal/IVOA/SiaInterface/WD-SIAP-2.0-20091104.pdf*, 2009.

[6] M. Louys, A. Richards, F. Bonnarel, A. Micol, I. Chilingarian, J. McDowell, and the IVOA Data Model Working Group. IVOA Recommendation: Data Model for Astronomical DataSet Characterisation. *ArXiv e-prints*, November 2011.

[7] M. Louys, F. Bonnarel, D. Schade, P. Dowler, A. Micol, D. Durand, D. Tody, L. Michel, J. Salgado, I. Chilingarian, B. Rino, J. de Dios Santander, and P. Skoda. IVOA Recommendation: Observation Data Model Core Components and its Implementation in the Table Access Protocol Version 1.0. *ArXiv e-prints*, November 2011.