Astronomical Data Analysis Software and Systems XXVI
ASP Conference Series, Vol. 521
Marco Molinaro, Keith Shortridge, and Fabio Pasian, eds.

Organizing Standardisation of Astronomical Data Access: the DAL WG Current Experience

Marco Molinaro,[1] and François Bonnarel[2]

[1]*INAF – Osservatorio Astronomico di Trieste, via G.B. Tiepolo 11, 34143 Trieste, Italy;* marco.molinaro@inaf.it

[2]*CDS, Observatoire astronomique de Strasbourg, 11 Rue de l'Universitè, 67000 Strasbourg, France;* francois.bonnarel@astro.unistra.fr

Abstract. The tremendous amount of data available in astronomy at all wavelengths allows astronomers to make new science and to correlate an extremely wide range of phenomena. It is also a challenge for digital data management distribution and processing. Interoperable data access protocols as designed by the IVOA take a major place in this challenge. This contribution reviews the current trends of IVOA efforts in this context.

1. Introduction

In the last decades, astronomy has evolved to become a multi-approach, multi-wavelength discipline. Full understanding of the astronomical objects requires cross correlating data obtained all over the electromagnetic spectrum (and astroparticles and gravitional waves) and by exploring also polarization and time dimensions. This is only achievable through access to very specialized archives, dedicated to a specific observing technologies. At the same time the amount of available data has increased reaching now several petabytes, and still growing; the upcoming LSST or SKA surveys among others will provide several petabytes a day in the coming years. Designing standard protocols for finding, describing and accessing all these data in a user friendly and interoperable way is critical. Within the scope of the IVOA, the DAL (Data Access Layer) WG has developed a large set of protocols, with a lot of services already built implementing them. This makes a complex set of 14 specifications, some of them presenting several versions. Bonnarel et al. (2017) presented the historical development of this landscape. This paper aims at providing a view of who is doing what in this landscape and how it can evolve in the future.

2. DAL protocol properties and classification

DAL protocols can be considered under several aspects or *properties*: **types of the data** they are dealing with; **functionalities** they are performing on the data; **software design** of the protocol.

Types of data. ObsCore specification distinguished the seven following data types in astronomy: spectra, images, time series, cubes, event lists, visibility, measurements

Table 1. DAL protocol properties. Note that SIA-1.0 is very different from others

		Functionalities					
Data Type	**Design**	Cone Search Discovery	MultiD Discovery	Description	Simple Access	Access Processing	Link
Catalogues, Tables	sync	TAP, CS, ObsTAP	TAP, ObsTAP		TAP, CS		DataLink
	async	TAP, ObsTAP	TAP, ObsTAP		TAP		
	ADQL	TAP, ObsTAP	TAP, ObsTAP		TAP		
	PBL	CS			CS		DataLink
	DALI	TAP, ObsTAP	TAP, ObsTAP		TAP		DataLink
	no-DALI	CS			CS		
Spectra, Time Series	sync	SSA, ObsTAP	SSA, ObsTAP	SSA, ObsTAP	SSA	SSA	DataLink
	async	ObsTAP	ObsTAP	ObsTAP			
	ADQL	ObsTAP	ObsTAP	ObsTAP			
	PBL	SSA	SSA	SSA	SSA	SSA	DataLink
	DALI	ObsTAP	ObsTAP	ObsTAP			
	no-DALI	SSA	SSA	SSA	SSA	SSA	DataLink
Images, Cubes	sync	SIA–1.0, SIA–2.0, ObsTAP	SIA–2.0, ObsTAP	SIA–1.0, SIA–2.0, ObsTAP	SIA–1.0, SODA–1.0	SIA–1.0, SODA–1.1	DataLink
	async	SIA–2.0, ObsTAP	SIA–2.0, ObsTAP	SIA–2.0, ObsTAP	SODA–1.0	SODA–1.1	
	ADQL	ObsTAP	ObsTAP	ObsTAP			
	PBL	SIA–1.0, SIA–2.0	SIA–2.0	SIA–1.0, SIA–2.0	SIA–1.0, SODA–1.0	SIA–1.0, SODA–1.1	DataLink
	DALI	SIA–2.0, ObsTAP	SIA–2.0, ObsTAP	SIA–2.0, ObsTAP	SODA–1.0	SODA-1.1	DataLink
	no-DALI	SIA–1.0		SIA–1.0	SIA–1.0	SIA–1.0	
Raw data, Event lis,t Visibility	sync	ObsTAP	ObsTAP	ObsTAP			DataLink
	async	ObsTAP	ObsTAP	ObsTAP			
	ADQL	ObsTAP	ObsTAP	ObsTAP			
	PBL						DataLink
	DALI	ObsTAP	ObsTAP	ObsTAP			
	no-DALI						
Spectral Lines	sync	SLA	SLA	SLA			
	async						
	ADQL						
	PBL	SLA	SLA	SLA			
	DALI						
	no-DALI	SLA	SLA	SLA			
Theory Data	sync		SimDAL	SimDAL	SimDAL	SimDAL	SimDAL
	async						
	ADQL						
	PBL						
	DALI						
	no-DALI		SimDAL	SimDAL	SimDAL	SimDAL	SimDAL

Note that SIA1.0 is very different from others
SCS: Simple Cone Search. SLA: Simple Line Access. SSA: Simple Spectrum Access.
SIA: Simple Image Access. TAP: Table Access Protocol. ObsTAP: ObsCore TAP service.

(seen as tables). We can group them this way: catalogues/tables, spectra/time-series (equivalent to unsampled spatial axis), images/cubes (regular bitmaps where spatial axis is sampled), event lists/visibility data (and raw observational data in general). But the distinction may be more subtle. For example some catalogs or tables of measurements may be directly extracted from images or cubes.

Functionalities. The basic functionality is discovery. This is done by a search against some criteria. One can distinguish discovery with spatial criterium (*cone search* like) and discovery with multidimensional search. Close to that is *data description*: standardized metadata describe in detail the datasets which may be retrieved by the service. Access to the data presents a large spectrum of aspects from simple retrieval to various serverside processing of the data. *Cutout* facilities are standing somewhat in the middle. The possible linkage (*data link*) of the result of a query to other associated resources allow a wide extension of functionalities.

Software design. The services can be distinguished by their synchronous or asynchronous nature. Some of them (older or most basic ones) are only synchronous, while the most recent and advanced are both synchronous and asynchronous (asynchronous services are based on the UWS standard (Harrison & Rixon 2010). Interfaces will also distinguish strongly between those accepting a relational-oriented query language such as ADQL (Osuna et al. 2008) and those only accepting a Parameter Based Language (PBL) which can be mapped to a datamodel more easily. At some point the DAL WG has recognized some best practice rules which may be shared by all the DAL specifications. These definitions define common interface features and are designed in the DALI specification (Dowler et al. 2013). Compliance to DALI is also a classification criterium for DAL protocols.

2.1. DAL protocol classification

Table 1 provides a tentative classification of the DAL protocols according to the three properties defined above. Some specifications are definitely outside the classification. DALI itself, because it gathers common definitions to all protocols; ADQL, being a generic language which may be used by other protocols. None of these two define services by themselves. Table 1 covers most of the DAL protocols used in IVOA services. Acronyms for the specifications and the documents themselves can be found on the IVOA Documents page[1]. The table shows clearly that some protocols are rather isolated and specific in the landscape: SLA, DataLink and SimDAL. TAP is fully adapted to catalogues but is also the basis for ObsTAP services dedicated to discovery of all types of data (including some lists of measurements themselves). Older protocols are generally PBL based and not compliant with DALI (except SimDAL which is significantly diverging from DALI due to a specific architecture more relevant for theoretical data). VTP is actually a very special protocol which doesn't belong to the same landscape of *access* but deals with real time distribution of information.

3. Future evolution

To imagine future evolution of the DAL protocols it can be useful to inspect what could be the evolution of our protocol *properties* in the future. As for types of the data, it is obvious that the time dimension has been poorly explored. The time domain importance will be dramatically enhanced in a near future with surveys like LSST. Beside this, the tremendous increase of the data volume is the main driver for adding new features for our *functionality* and *design* properties. The current *discovery* approach is based on a two step process (three step if we include discovery of the services themselves in the registry): first we query a database of metadata and then we choose datasets of interest and access them. In contrast to that, progressive fine tuning of the data discovery will be more and more necessary to explore the huge new collections of data. The HiPS specification (Hierarchical Progressive Surveys, Fernique et al. (2016)) is a

[1] `http://www.ivoa.net/documents/`

first approach to do this efficiently in the spatial domain. Distinct additional innovations may be needed in the future to tackle other data axes. The data access will also require more and more server side data processing to allow significant excerpts of the data to be delivered to the end user. This serverside processing requirement will drive several consequences on the protocol properties: simple description made by ObsCore will be insufficient; full metadata retrieval (including *mapping* information) will allow to prepare more sophisticated queries to access services. The algorithms necessary to achieve this processing will be more sophisticated and would hardly be standardized in all respects: DAL protocols have to enhance the possibility of defining *custom services* which should be nevertheless easily integrated in the IVOA interoperable backbone. Eventually the end user may want to test her own software on data without downloading large amounts of data: porting code close to the data and execute it remotely is then a promising perspective. These changes would reflect on the language. Is a parameter language (with 3 factor semantics - see Bonnarel et al. (2016)) sufficient for *access data* protocols designed in this context? Would PDL (Parameter Description Language) be a good basis for building a real generic PBL? What would be the role of datamodels in this new syntax definition? JSON format interfaces may also be added both for querying and responses for richer flexibility. The growing datasets represented in catalogue tables poses also a new view on the TAP protocol, whose relational approach in the primary query language and metadata structure presents a limiting constraint to non relational database back ends that start to be used for astrophysical resources.

4. Conclusion

DAL efforts have permitted an interoperable backbone of services for all kind of data. It allows easy access to these data with VO tools. The upcoming data avalanche is a big challenge for the evolution of this DAL backbone.

Acknowledgments. The authors acknowledge their work within the *ASTERICS* project, Horizon 2020 grant agreement n. 653477.

References

Bonnarel, F., et al. 2016, Ivoa server-side operations for data access version 1.0, IVOA Proposed Recommendation. URL `http://www.ivoa.net/documents/SODA`

Bonnarel, F., et al. 2017, in ADASS XXV, edited by N. P. F. Lorente, K. Shortridge, & R. Wayth (San Francisco: ASP), vol. 512 of ASP Conf. Ser., 549

Dowler, P., et al. 2013, Data Access Layer Interface Version 1.0, IVOA Recommendation 29 November 2013. `1402.4750`

Fernique, P., et al. 2016, A&A. URL `http://www.ivoa.net/documents/HIPS/20160623`

Harrison, P., & Rixon, G. 2010, Universal Worker Service Pattern Version 1.0, IVOA Recommendation 10 October 2010. `1110.0510`

Osuna, P., et al. 2008, IVOA Astronomical Data Query Language Version 2.00, IVOA Recommendation 30 October 2008. `1110.0503`, URL `http://www.ivoa.net/documents/latest/ADQL.html`