

# **IVOA Data Access Layer (DAL) Work Package**

D. Tody, July 2003

## ***Working Group Meeting Agreements***

The kick-off meeting of the IVOA data access layer (DAL) working group was held in Cambridge, UK, on May 12 2003, with follow-on meetings the rest of the week. The goal of this first meeting was to agree on what the VO data access layer is, what we would ultimately like to produce, and what we would like to accomplish within the DAL working group over the next year.

Specific working group agreements were achieved in the following areas:

- Concept of DAL portal
- DAL scope and high level architecture
- Principal data types within the scope of the DAL
- Mapping of data types to access services
- Priorities for implementing the data access services
- Roadmap and priorities for the next year
- Enhancements to SIA V1.1

A more detailed summary of each agreement is given in the following sections.

### **Concept of DAL portal**

The DAL portal (the so-called generic client interface) provides unified client access to VO data. This portal is the primary interface between client data analysis applications and the VO. Developers use the DAL portal to build distributed multiwavelength data access and analysis applications. DAL client applications see mainly the portal interface and are largely isolated from the underlying VO architecture.

### **DAL scope and high level architecture**

The DAL portal consists of a consistent and integrated set of data access services, each for a particular type of data (image, spectrum, etc.), plus reference grade software (APIs, server frameworks) implementing key components of the data access layer. The DAL services are defined at the most fundamental level as implementation independent, over-the-wire protocols. Reference grade implementations help ensure that the protocols work on real data, while providing working software to demonstrate the practical application of the DAL protocols and services. Several levels of service are defined, including URL-based, Web Services-based, and Grid Services-based.

## Principal data types within the scope of the DAL

The principal classes of data to be supported by the VO data access layer include the following:

- **Source catalog.** This provides a simplified catalog query mechanism, e.g., for object catalog overlays on images, or for astrometric and photometric calibration of data. There was some question whether catalog access should be included in the DAL, but it was decided that a basic catalog access service is required for data analysis and integrated access should be provided. Large scale statistical analysis, catalog cross matches, etc., are handled elsewhere within the VO.
- **Image.** This includes 2D sky projections, spectral data cubes, long slit spectra, and ultimately sparsely sampled images such as IFU data. VO will emphasize calibrated data but it may be desirable to be able to use the VO framework and services to access raw data as well.
- **1D spectrum and SEDs.** This is a special case of the more general NDIImage, provided for ease-of-access to the most common spectral data.
- **Time series.** Light curves, variability data, etc. This category does not include synoptic image data, which is handled via the image interface instead.
- **Event and visibility data.** Although much access to radio and high energy data is via uniformly generated reference images, precise statistical analysis of event data, as well as imaging of both event and visibility data, requires access to (usually calibrated) event and visibility data.
- **Generic dataset.** Encompasses all types of data handled via the DAL. Used for basic data discovery, with subsequent access to data handled by the more specific DAL services.

## Mapping of data types to access services

Each principal type of data handled by the DAL has a corresponding data access service which is specific to and optimized for that particular class of data. Each type of data has a corresponding data model which is implemented by the service. Often the same data can be viewed via multiple services, e.g., synoptic or multiband imagery could be viewed as an image, as a spectrum or SED, or as a time series. Event and visibility data could be viewed as a table or as an image, spectrum, time series, and so forth, depending upon the capabilities provided by the service provider and the type of analysis being performed by the client.

Astronomical data falls into into two broad categories: catalog and image. The specific data classes include source catalog, time series, event list, visibility data, NDIImage (including the various image subclasses), 1D spectrum, and SED.

## Priorities for implementing the data access services

The highest priority services are image access, in particular 2D sky projections and spectral data cubes, 1D spectra and SEDs, and catalog access. Time series and general NDImage come next, followed by access to calibrated visibility and event data.

## Roadmap and priorities for the next year

In general the first year of IVOA DAL development emphasizes specification of the access protocols rather than reference implementations, which necessarily come later (although we have the science demos even in the first year or two). The priorities identified for the first 12-18 months of the IVOA standards effort are as follows:

- **Simple Image Access (SIA) V1.1** (target: summer 2003). Similar to the current version with selected enhancements.
- **Simple Spectra Access (SSA) V1.0** (target: fall 2003). Emphasize 1D spectra and SEDs.
- **SIA V2.0** (target: summer-fall 2004). General image model. Explore more general ways to structure metadata and attempt more complete data model-based representation of complex datasets.
- **Better integration with VO standards.** This includes registry integration, UCD normalization, development and use of standard data models, more comprehensive standards for data representation.
- **First steps for event and visibility data.** This includes a survey of existing resources, development of metadata standards for publishing raw data in archives, and initial steps towards open standards for calibrated and instrumental data.
- **Web Services versions of DAL services.** SOAP-based data access to complement the initial URL-based services. These will be alternative protocols for the same underlying DAL services. To be useful this also requires some experimentation with the client-side interface for data analysis.

## Enhancements to SIA V1.1

Simple image access (SIA) is currently at V1.0, the initial version released in the fall of 2002 and used in various demos in early 2003. Version 1.1 will be the first IVOA sponsored version of the SIA protocol. Due to the schedule, SIA V1.1 will

retain the same form as V1.0, with evolutionary enhancements. More extensive changes will be deferred to V2.0.

Enhancements were discussed in the Cambridge meeting as well as via the mail exploders and telecons, and in small meetings, e.g., at CDS and ESO. Planned enhancements include the first real registry support and integration, improved image characterization, e.g., to better define image provenance and identification, time of observation, and spatial resolution, further evolution towards formal data models, and normalization of UCDs including support for pointers into data models.

### ***Other Activities***

The DAL working group has only been active for about six months, and most activity during this period centered around the Cambridge meeting. Other activities included the following:

- Discussions of SIA enhancements. This was mainly via email and via various studies, most notably by CDS, NRAO, and SAO.
- D. Tody visited CDS for meetings with F. Bonnarel, F. Oschenbein, P. Fernique, and others in June to discuss further evolution of SIA as well as integration of SIA with the Aladin image visualization tool. Most of the discussions focused on normalization of UCDs, in particular the relationship of UCDs and data models and their use to model datasets in the DAL. A proposal to add a new parameter UTYPE to VOTable was produced and is currently undergoing broader discussion.
- A questionnaire was circulated on the IVOA exploders to survey existing spectral data archives and potential demonstration client analysis applications for 1D spectra and SEDs.
- M. Dolensky and D. Tody met at ESO in late June to review requirements for simple spectral access and start developing an initial interface specification.

### ***Near-Term Plans***

The top priorities for the remainder of 2003 are to release SIA V1.1 (for this to be worthwhile we need to evolve the underlying technology sufficiently first, e.g., the component data models), and to produce the initial version of simple spectral access. In terms of schedule, the principal driver for these developments are the various VO demos (IVOA, AVO, NVO, etc.) planned for early 2004. In addition we plan to continue research on scalable computational frameworks. A scalable data analysis framework is needed needed to integrate astronomical data analysis with VO, as well as to provide reference-grade framework software to implement scalable DAL services.

