VO Data Access Layer
Working Group Summary

IVOA

Cambridge, UK 16 May 2003

Doug Tody, NRAO

IVOA DAL Working Group

- Christophe Arviset
- David Berry
- Francois Bonnarel
- Markus Dolensky (Minutes)
- Pat Dowler
- Pierre Fernique
- Mike Fitzpatrick
- Bob Hanisch
- Sang Chul Kim
- Young-Soo Kim
- Peter Lamb
- Tom McGlynn
- Pedro Osuna
- Clive Page
- Ray Plante
- Andre Schaaff
- Riccardo Smareglia
- Mike Taylor
- Doug Tody (Chair)
- Nicholas Walton
- Naoki Yasuda
- (our apologies if we didn't get everyone's name down here)

The above lists the DAL WG members who were present in Cambridge. The meeting was held at the IoA in Cambridge, UK on May 12, 2003.

Existing DAL Prototypes

- Cone search, SIA (NVO)
- Jodrell Bank – imaging online from Merlin data (Anita Richards)
- CVO – WFPC assoc, ROSAT, 2QZ (spectroscopic survey) – generic DA service for all data (query is separate); different service instances, obey same interface
- AUS/VO – HIPASS, ATCA archive, MACHO, SUMSS, 2DF GRSS – Web query i/f

Existing DAL Prototypes

- AVO
  - SIA, IDHA (all images)
- AstroGrid
  - HDX (Starlink) – mainly data model, not much data access yet
  - Also solar data, space physics data
- Others
  - We could surely continue this list further

Accomplishments

- IVOA DAL Working group agreements
  - Concept of DAL portal for unified client access to VO data
  - Agreed on DAL scope and high level architecture
  - Principal data types within the scope of the DAL
  - Mapping of data types to access services, e.g., images, spectra
  - Priorities for implementing the individual data access services
  - Roadmap and priorities for the next year
  - Preliminary list of target enhancements for simple image access

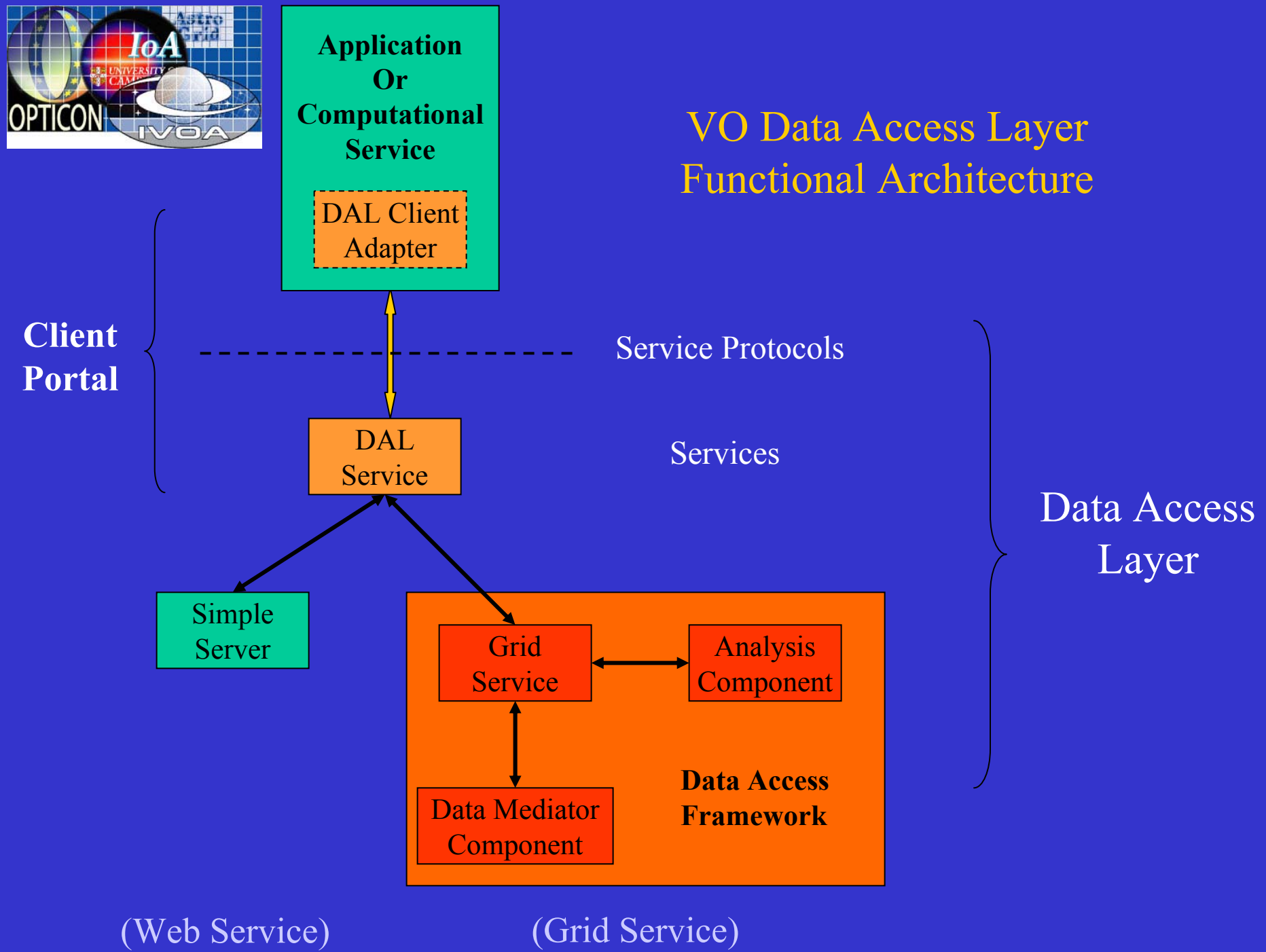
DAL Portal Concept

- Link between client (user) data analysis software and the VO
- Provides client with uniform access to VO data
- Enables distributed multiwavelength data access and analysis
- Key longer-term goals:
  - *Integrate everyday data analysis with VO*
  - *Achieve a scalable VO architecture (same software reused in multiple contexts)*

DAL Scope: Software Layers

- Client side integration
  - API, Web Services client, components, etc.
- Data access protocols
  - Catalog, image, spectrum, etc.
- Service frameworks
  - Virtual data generation and manipulation, grid computing
- Test cases and demonstrations
  - Required for verification and end-to-end testing

VO Data Access Layer Functional Architecture



VO Data Access Layer
Functional Architecture

Analysis components may be dynamically downloaded Java code, compiled legacy code wrapped as a component, etc.

Application Or Computational Service

DAL Client Adapter

Data Model

Grid Service

Analysis Component

Data Model

Data Access Framework

Data Mediator Component

Application framework may transform the data model seen by an application

Data mediators map external data into a VO data model, subsetting and filtering the data in the process.

External Dataset

External data may be in any format

(Application)

(Grid Service)

DAL Scope: Types of data

- Source catalog
  - Simplified catalog query included in DAL portal, e.g., for catalog overlays
  - More sophisticated catalog operations are provided elsewhere in VO
- Image
  - 2D sky images, spectral data cubes, long slit spectra
  - Sparsely sampled images including IFU data
  - Use VO framework to provide access to raw data as well
- 1D Spectrum, SED
  - Supported as a special case for ease-of-access to common spectral data
- Time series (later? - probably not year 1)
  - Light curves, variability data, etc. (does not include synoptic image data)
- Event and visibility data (later)
- Dataset (later)
  - Query for any type of data in a given ROI
  - Client then uses other data access services to drill-down to actual data

Data Access Services

- Multiple services
  - For data access we need one service for each view of the data
  - The “view” is the data model implemented by the service
  - At a high level, e.g., for a generic dataset, or for data mining, one may not need to differentiate the data
- View the same data via different services
  - e.g., and event list or visibility dataset may be viewed as either a table or an image
- Guidelines
  - Integrate common elements of services
  - Provide simple services for the most common cases, e.g., 2D skyImage, 1D spectrum

Common Service Elements

- Registry
  - Cache resource and service data
  - Isolate service metadata from registry implementation
- Queries
  - Keywords for simple interface, AQL for complex queries
- Metadata standards
  - Add to UCD standard where necessary
- Data model standards
  - Spectral and spatial bandpass, resolution, observation time, etc.
- Data representation
  - e.g., VOTable
- Protocols
  - Web, Grid Services consistent across all services

Top Y1 DAL Service Priorities

- Simple Image Access Upgrade
  - Version to support IVOA (e.g., CDS/Aladin)
  - First upgrade since fall 2002
  - New IVOA version of SIA needed summer 2003 (e.g., July) to support Jan 2004 science prototypes
- Spectral Data Access (New)
  - Highest priority after basic catalog and image access
  - Highest priority re: spectra is for 1D spectra and SEDs (e.g., versus spectral data cubes, IFUs, slit spectra, etc.)

First cut at priorities for Y1

- SIA V1.1
  - Same as now with minor enhancements
- Simple spectral access
  - Emphasize 1D and SED
- SIA V2.0
  - Generalize image model (spectral data cubes etc.)
  - Explore ways to structure metadata (table vs hierarchical)
- Better integration with VO standards
  - Registries, queries, UCDs, data models, data representations
- First steps for event, visibility data
  - E.g., metadata standards to publish and document raw data

Spectral Data Access

- Goals / Use Cases
  - Discover, retrieve, display a 1D spectra for some object
  - Discover, retrieve, display a SED for some object
  - Retrieve a 1D science spectrum for analysis, e.g., classification
  - Multi-wavelength spectral combination
  - Generate spectra on the fly from grism/radio/HE data
    (this applies also to images, time series, etc.)

Spectral Data Access

- Priorities
  - 1D science spectra
  - SEDs
  - Spectral image cubes (2 spatial + 1 energy)
  - 2D science spectra (slit spectra)

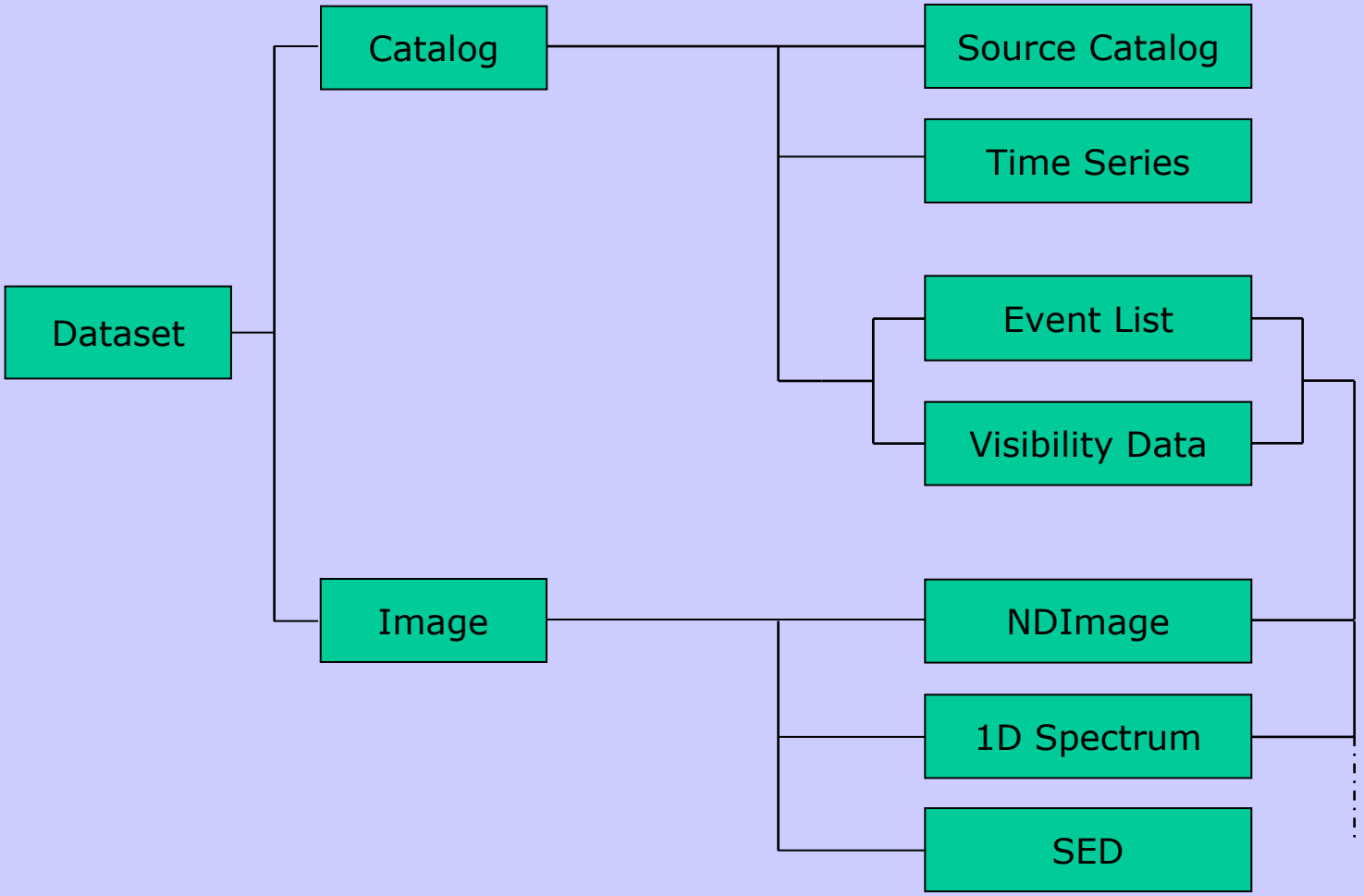
Spectral Data Access

- Key Issues
  - General 3/4D image model vs simple 1D/SED
    - Working group agrees that general ND model is needed to support full range of data.
    - However, the 1D/SED case should be supported via a simple, optimized interface since this is such a common case and is a higher priority.
  - Separate services for images and spectra vs combined service
    - Agreed: provide separate NDimage and 1D spectra+SED services
  - FITS vs XML (VOTable)
    - The following output formats are required for services: **FITS, Graphics, XML**
    - All should be available for 1D spectra and SEDs
    - We need a new FITS standard for 1D spectra to use FITS for VO 1D spectra, and should probably develop one along with the VOTable-based format.
    - 1D spectra via XML/VOTable will be first use of VOTable for data model-based data transfer.

SIA – Prioritized Enhancements

- Essential
  - Registry integration
  - Pixflags support for lossy compressed data (e.g., HCOMPRESS)
- Image Characterization
  - Image provenance and identification (collection ID, dataset ID, virtual data provenance, replica support)
  - Spectral bandpass (already present; may need tweaking for consistency)
  - Time of observation
  - Spatial resolution
  - Limiting flux (harder; may not make V1.1)
- Other
  - VO technology integration (normalize UCDs, data models, etc.)
  - Use of image attributes to refine query (e.g., band)
  - Default for case where there are multiple versions of same dataset
  - Spatial bandpass - 3
  - Image type (future- v2)
  - Logical hierarchies to describe complex metadata (as in IDHA – v2)

DAL Scope: Types of data



Data Discovery

Primary DAL Services

Second cut at priorities for Y1

- SIA V1.1 (Summer 2003)
  - IVOA release: similar to current version with selected enhancements
- Simple spectral access (Fall 2003)
  - Emphasize 1D and SED
- SIA V2.0 (Fall 2004)
  - General image model (spectral data cubes, IFUs, etc.)
  - Explore ways to structure metadata (flat table vs hierarchical)
- Better integration with VO standards
  - Query standards, registries, MD(UCDs), data model, data representations
- First steps for event, visibility data
  - E.g., metadata standards to publish raw data
- Web services versions of DAL services

Drivers/Priorities

- IAU demo July 2003
  - Add dynamic registry support
  - Add Aladin, other SIA services
  - Integrate SIA registry into clients (Aladin, DIS, etc.)
- AVO, AAS demos January 2004
  - AVO science case demo for 1D spectra and SEDs
  - Web services versions of services?

IVOA Data Access/Analysis Layer Agreements

IoA, Cambridge UK, May 2003