

US National Virtual Observatory

Data Access Layer Working Group Working Group Report and Summary

Doug Tody
National Radio Astronomy Observatory



NVO Team Meeting, Pittsburgh July 26-27 2005

DAL Overview

SSA

- Spectrophotometric tabular data
 - photometry points – flux vs spectral coord, time
 - 1D spectra, time series, SEDs
- New query interface presented and discussed at Kyoto meeting
 - revisions, early implementations in progress (JHU)
 - target first stable version for Madrid interop
- Query interface at V0.9, data model at V0.83b

• SIA

- Regularly gridded image data
- Query interface at V1.0
 - predates data models, utype, etc.
 - update will be based on technology developed for SSA
 - target V1.1 interface draft for Madrid; stable early 2006

• Event, Visibility Data

- Only current activity is on instrumental data models
- Main strategy for VO is to render data into images, spectra, etc.



SSA Topics

- **Expanded query interface**
 - search on more than just sky region
- **Standard dataset metadata**
 - uniformly describe, characterize astronomical datasets
- **Service metadata**
 - service coverage, capabilities, compliance, extensions
- **Data model issues**
- **Spectral linelist access protocol (SLAP)**



SSA Query Interface

- **Types of query**
 - Parameter-based (parameter objects, 90/10 rule)
 - Syntax-based (query language based on attributes)
- **Basic query**
 - Spatial region (POS,SIZE), spectral bandpass, time bandpass
 - Allows query in all physical axes: spatial, temporal, spectral
 - Range syntax VAL1,VAL2 to express ranges
- **Other parameters**
 - TOP (ranked by best match)
 - OBJTYPE, COLLECTION, CREATORID, PUBID
 - APERTURE, SPECRES, TIMERES
 - SNR, REDSHIFT, TARGETCLASS



SSA Query Response

- **Metadata is grouped into component data models**
 - GROUP, UTYPE representation in VOTable
 - Same components used in query response and in dataset itself
- **Classes of Query Metadata Returned**
 - Query metadata Describes the query itself
 - Dataset metadata Describes data object; object-specific
 - Target metadata Astronomical target
 - Curation metadata External identification of dataset
 - Characterization Coverage, Accuracy, Frame, etc.
 - Instrument metadata Optional; service-defined
 - Access metadata Describes how to access the dataset

Most of this is generic dataset metadata and will apply equally to SIA as well.



SSA Data Model

- **General issues**
 - Finalize component models (characterization, identification)
 - How to handle "accuracy" information
 - Fuzzy concepts (resolution, sampling)
- **Spectrophotometric data**
 - Spectrum, TimeSeries, SED should all be top-level objects
 - Approach
 - subclass "Dataset", aggregate reusable component models
 - For SSA core model is a photometric sequence (segment)
 - Spectrum: one photometric sequence, plus dataset metadata
 - TimeSeries: same, but time varies instead of spectral coord
 - SED: multiple sequences, dataset metadata describes entire SED
- **What do we call the SSA data model?**
 - SED data model, spectral data model, spectrophotometric data model, etc.



Simple Line Access Protocol (SLAP)

- **Purpose**
 - Standard interface to access spectra line lists
 - Used for both calibration and analysis
- **Authors**
 - ML Dubernet, P.Osuna, et.al.
- **Consists of**
 - Detailed physical spectral linelist data model
 - Query interface
 - Line list returned as VOTable
- **Status**
 - Broad working group established
 - Draft specification in preparation (by Madrid)



SIA Upgrade

- **Goals**

- Query parameters
 - Add spectral and time bandpass to minimal required parameters
 - Query by dataset characteristics as for SSA
- Update and enhance query response metadata
 - Query response (dataset metadata) consistent with SSA
 - New metadata extension mechanism
 - VOTable currently; also needed for FITS

- **Some new functionality**

- Logical name associations, score, compression, etc.
- Improved service metadata query, as for SSA

- **Under Consideration**

- Asynchronous staging, VOStore integration, authentication, etc.
- Web Service interface
- 3D image support



Asynchronous Data Staging

- **Basic interface**

- queryImages data discovery, service negotiation
- accessImages asynchronous data staging, multiple datasets
- getData get a dataset (synchronous)

- **Staging interface**

- Approach
 - based on WSRF/SOAP (jobID, messaging infrastructure, etc.)
 - Information content defined by each access protocol
- Features
 - accessImages – start job to stage datasets based on query acref
 - WSRF messaging or polling may be used to query job status
 - data may be staged directly to a designated VOStore
 - events (messages) for image ready, job done, etc.
 - integrated SSO authentication capabilities



Versioning Issues

- **Concern about impact on community**
 - Current services, applications are affected by a change to a major VO interface such as SIA
- **Versioning essential**
 - describe both in registry and at service level
 - find services based on version of interface supported
 - a single service can support multiple interfaces
- **Plan to keep old version around for at least one cycle**
 - user wants to defer upgrading until time is right



DAL/ADQL integration strategy

- **Architecture**

- ADQL/SkyNode is an alternative query method
 - e.g. two *query/images* methods
 - Parameter-based interface will also be retained
- Query response (VOTable) is unaffected
 - except for optional field selection/projection
- Data staging is unaffected
- Data access (getData) is unaffected

- **Major interfaces required**

- Client query interface
- Data collection metadata in RDBMS
- Specification of virtual data product



Framework Stuff

- **Service framework**
 - SkyNode/ADQL technology for syntax-based queries (optional)
 - Scripted component framework for data processing
 - Can be used for data analysis and pipelines as well as services
- **VO-Client**
 - Client side of DAL interfaces
 - Data discovery capabilities (virtual directories)
 - Cache data locally (expose VOStore interface)
 - Efficient local object APIs for data analysis
- **VO-Services**
 - Web service interface
 - CEA/UWS approach





NVO Team Meeting, Pittsburgh July 26-27 2005

ADQL/DAL Integration

- **Agree on Architecture**
 - ADQL/SkyNode is an alternative query interface
 - Parameter-based interface will be retained
 - Query response is unaffected
 - Data staging is unaffected
 - Data access (getData) is unaffected
- **Major Interfaces**
 - Client query interface
 - Data collection metadata in RDBMS
 - Specification of virtual data product (for getData)



SSA – Service Levels, Compliance

- **Minimally-Compliant Service**
 - Provide at least the parameter-based GET interface
 - WS interface will be required in the future for advanced capabilities
 - Implement "MUST provide" query parameters
 - Implement "MUST provide" query response metadata
 - Implement at least one SSA DM-compliant output format
 - Implement service metadata query method
- **Query-Compliant Service**
 - Implements minimal query, query response
 - Returns foreign/external dataset
 - Some support for understanding contents of external dataset
- **Fully-Compliant Service**
 - Implements everything in minimal service
 - Implements all "SHOULD provide" query parameters and query response metadata



SSA Query Interface

- **Basic Query**
 - Add spectral and time bandpass to minimal required parameters
 - POS, SIZE, BANDPASS, TIME, FORMAT
 - Allows query in all physical axes: spatial, temporal, spectral
 - Range syntax VAL1, VAL2 to express ranges
- **Other Changes**
 - Add TIMERES at same level as SPECRES (for time series)
 - Drop VERBOSITY – ADQL will provide a better way to provide this
 - Compression applies to query response as well as data retrieval
- **Other New Parameters**
 - TOP, APERTURE, SPECRES, TIMERES
 - COLLECTION, CREATORID, PUBID
 - SNR, REDSHIFT, TARGETCLASS



Data Model Integration

- **Data model-based data access**
 - Basic capability has been achieved
- **Concepts**
 - Primary data object types are all top-level Datasets
 - Spectrum, TimeSeries, SED
 - Data model name should apply to a class of data
 - Working name: Spectral data model
- **Other issues**
 - Good progress on Characterization
 - Role of dimensional analysis still under discussion



SIA V1.1 Upgrade

- **Goals**

- Query parameters
 - Add spectral and time bandpass to minimal required parameters
 - Query by dataset characteristics as for SSA
- Update and enhance query response metadata
 - Query response consistent with SSA
 - New metadata extension mechanism
- Some new functionality
 - Logical name associations, score, compression, etc.
 - Improved service metadata query

- **Probably Defer**

- Web Services interface
- Asynchronous i/o, staging to VOSTore, authentication, etc.
 - possibly we can include this as an experimental feature



DAL Roadmap Update

- **SSA**
 - V0.9 WD Discussed in Kyoto
 - V0.9+ WD Summer 05 updated draft, trials
 - V1.0 PR Madrid (Oct 05) – ready for general implementation?
- **SIA**
 - V1.0 WD Current version
 - V1.1 WD Discuss in Madrid (Oct 05), trials to follow
- **ADQL/DAL integration**
 - Lower priority for DAL than data access protocols
 - Work with VOQL group on SkyNode/DAL interface in Fall 2005
- **Reference Implementations**
 - Service framework (still in concept/design phase)
 - VO-Client (still in concept/design phase)

