Simple Spectral Access (SSA)

Data model issues and revised access protocol

IVOA Victoria, May 2006

Doug Tody (NRAO/NVO/IVOA)

DAL/DM Joint Session

- Spectral Access
  - SSA Overview (25m)
  - Jonathan McDowell – Spectrum DM (45m)
  - Francois Bonnarel – Discussion on UTYPEs (20m)
  - Doug Tody – SSA Protocol (45m)
  - Markus Dolensky – Ranking algorithm (20m)
- Discussion of SSA and Spectrum DM (25m)
- SSA roadmap update

SSA Status

- Activities
  - Major focus currently is on data model issues
  - Access protocol too, but this also involves data models
- Goals
  - Want reasonably simple, robust, consistent data models
  - Generic dataset metadata very important
    - Needed by all data access interfaces
    - Much of SSA query concerns generic dataset metadata
  - Data model-based data access
    - Spectrum/SED models must support actual data analysis
- Requirements
  - We must finalize V1.0 WD in a matter of weeks
  - Only requirement is that SSA be self-consistent
  - Compatibility with general data models desirable but not required

DAL Data Model Guidelines

- Architecture
  - Datasets form a class hierarchy
    - Generic Dataset class at the root
    - Catalog, Image, Spectrum, TimeSeries, SED, etc. subclasses
    - Instrument data collections are in effect subclasses of these
    - Generic dataset metadata and basic access methods inherited
  - Each class of Dataset is modeled by composition
    - Aggregate, associate component data models
    - Both standard and extension components are possible
    - Containers used to flexibly aggregate components
  - Components
    - Small enough to be reasonably simple
    - Re-usable via composition to model more complex things
    - Associations are made at a higher level

DAL Data Model Guidelines

- Formal Interfaces
  - A data model is an abstraction
    - Defined independently of implementation
    - Can/will be implemented using various languages, technologies
    - Specified as human-readable document plus UML
    - If have only XML, it is an implementation, not a data model
  - Parameterization
    - Need to be able to reduce data model to a set of parameters
    - For example, to map to a set of table fields, query params
    - Parameter representation is language and technology-neutral
  - UTYPE
    - Purpose is to identify elements of a data model
    - Allows data model to be reduced to a set of parameters
    - Utype identifies role of individual parameter in data model

Spectral Data Model Refactoring

- Approach
  - Spectrum, TimeSeries, SED become co-equal top level objects
  - Composed mainly from same components
  - Core Spectrum/SED data model remains much the same
- Advantages
  - Simplifies individual models, e.g., simple 1D spectrum
  - Enables custom object composition

Object Composition Examples

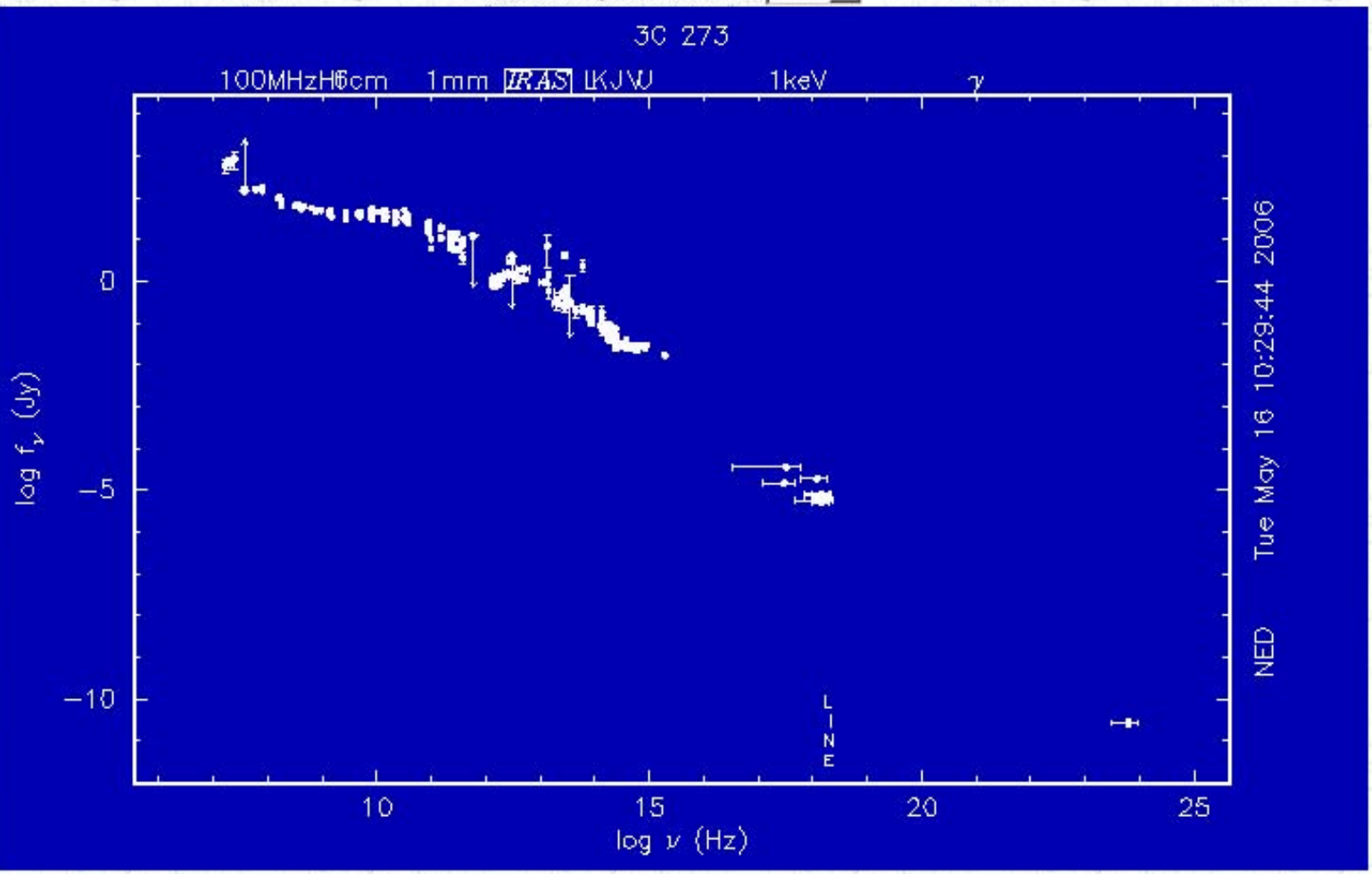
- Spectrum
  - Generic dataset metadata (target, char, id, curation, etc.)
  - Data (spectral coord vector, flux, bkg, accuracy)

- SED (proposed)
  - Generic dataset metadata (target, char, id, curation, etc.)
  - Homogenized SED segment data (as in NED for example)
  - Individual Segment objects
    - Spectrum, photometry points, time series
    - Data could be included or pointed to
  - Optional additional metadata on how SED was generated,
    - e.g., image cutout for a photometry point

Spectral Energy Distribution (SED) for 3C 273: 296 Data Points

New! Click to zoom in at a position of interest (uses a separate window).

Select zoom factor: 4X



Error Bars

No Point Labels

X=log(Freq.)(Hz)

Y=log(Fnu)(Jy)

Autoscale data range

Plot Again

New opti

Postscript version of the SED plot New!

hotometric Data --- Published and Homogenized [Frequency, Flux Density] Units

	Published Units				Homogenized Units [Frequency, Flux Density]				
No.	Observed Passband	Measurement	Uncertainty	Units	Freq (Hz)	Measurement	Uncertainty	Units	
	EGRET (0.1-5 GeV)	2.7200E-11	+/- 4.4200E-12	Jy	6.17E+23	2.72E-11	+/- 4.42E-12	Jy	19
2	6.4 keV Fe K{alpha}	30.3	...	ergs cm^-2^ s^-1^	1.55E+18	1.95E-08	...	Jy	20
3	F_2-10_ keV	8.3	+/- 10 %	10^-11^ erg/s/cm^2^	1.45E+18	5.72E-06	+/- 5.72E-07	Jy	19
4	2-10 keV (XMM)	7.87E-11	...	erg cm^-2^ s^-1^	1.45E+18	5.42E-06	...	Jy	20
5	3-9 keV (BeppoSAX)	1.17E-13	...	W m^-2^	1.45E+18	8.06E-06	...	Jy	20
6	3-9 keV (BeppoSAX)	1.12E-13	...	W m^-2^	1.45E+18	7.72E-06	...	Jy	20
7	3-9 keV (BeppoSAX)	1.08E-13	...	W m^-2^	1.45E+18	7.44E-06	...	Jy	20
8	3-9 keV (BeppoSAX)	1.03E-13	...	W m^-2^	1.45E+18	7.10E-06	...	Jy	20
9	4-8 keV (BeppoSAX)	7.01E-14	...	W m^-2^	1.45E+18	4.83E-06	...	Jy	20
10	4-8 keV (BeppoSAX)	1.08E-13	...	W m^-2^	1.45E+18	7.44E-06	...	Jy	20
11	4-8 keV (BeppoSAX)	1.16E-13	...	W m^-2^	1.45E+18	7.99E-06	...	Jy	20
12	4-8 keV (BeppoSAX)	8.67E-14	...	W m^-2^	1.45E+18	5.98E-06	...	Jy	20
13	4-8 keV (BeppoSAX)	1.01E-13	...	W m^-2^	1.45E+18	6.96E-06	...	Jy	20
14	0.5-10 keV (ASCA)	230	...	ergs cm^-2^ s^-1^	1.21E+18	1.90E-05	...	Jy	20
15	ROSAT (0.1-2.4 keV)	1.17E-10	+/- 2.16E-12	ergs sec^-1^ cm^-2^	3.25E+17	3.60E-05	+/- 6.65E-07	Jy	19
16	0.5-2 keV (XMM)	4.38E-11	...	erg cm^-2^ s^-1^	3.02E+17	1.45E-05	...	Jy	20
17	1549 A	2.110E-13	...	erg/cm^2^/s/A	1.94E+15	1.69E-02	...	Jy	20
18	log nu(Hz) 14.95	1.46	+/- 0.02	log f_nu (milliJy)	8.91E+14	2.88E-02	+/- 1.36E-03	Jy	19
19	U (Johnson)	27.11	+/- 0.79	milliJy	8.19E+14	2.71E-02	+/- 7.90E-04	Jy	19
20	U (Johnson)	24.92	+/- 0.60	milliJy	8.19E+14	2.49E-02	+/- 6.00E-04	Jy	19

Spectral Data Associations

- Motivation
  - Model multi-segment 1D spectral data
  - Echelle, Multi-object spectrograph (MOS), etc.
- Approach
  - Core object is a 1D spectrum
  - Complex object is a logical association of simple 1D spectra
  - Complex view is optional; simple 1D case remains simple

Spectral Data Associations

- Examples
  - MOS is an assocation of 1D spectra with common metadata
  - Echelle could be overall combined spectrum plus individual orders
- Access Model
  - Query describes all spectral elements and defines associaton
  - All data access is to individual 1D spectra
    - common metadata will be duplicated
    - access is uniform with common components
  - Client sophistication is optional
    - can "see" logical association or only separate 1D spectra

SSA Overview

- Summary
  - Initial focus on access protocol, 1D Spectrum, related data models
  - More complex data collections can be modeled by association
  - Refactored model allows SED to come along after Spectrum
  - TimeSeries is essentially the same as Spectrum (we think)
- Keeping it Simple
  - Core concept for SSA is simple 1D Spectrum
  - More complex cases can be built on top of this core concept
    - e.g., Echelle, MOS, SED
- Basic query interface and generic dataset metadata is global
  - Used everywhere, is inherited, hence can largely be done once and factored out of the problem of a specific data type

SSA Protocol

Basic Service Elements

- Version negotiation
- Numeric values
- Parameter ordering and case
- Range-list parameters
- Services with multiple operations

Version negotiation

- Registry
  - Service protocol version is recorded in registry
  - Service may support multiple protocol versions (also in registry)
  - Client queries registry to find desired version of services
- Service
  - Versioning also supported at level of client-server
  - Dynamic verification that client, server versions agree
    - Explicit VERSION parameter in GET call required
  - Client proceeds with call to queryData
    - With priory registry query, versions should match
    - If not, fallback to version negotiation is possible
- Version negotiation
  - Client calls getCapabilities operation to negotiate version
  - Request, response until match is found or negotiation fails

Numeric Values

- Integer and Real
  - Integer value specified as in XML Schema Datatypes spec
  - Real numbers specified in double as per XML Schema spec
  - All parameters permit real values unless otherwise specified
- Time Values
  - Try to limit these to ISO 8601 (UTC or TT) and MJD
- Formatting
  - Real defines integer, decimal, and exponential representations
  - Sexagesimal is not permitted except where explicitly specified

Parameter ordering and case

- Case
  - Parameter names are case-insensitive
  - Reserved parameter names are generally printed in upper case,
    - but this is no required
- Order
  - Parameters may appear in any order
  - If a parameter appears multiple times the result is undefined
- Unknown or service-defined parameters
  - A service should ignore unknown parameters
  - Services are allowed to define custom parameters
  - Custom parameters are described in the service metadata

Range-list parameters

- List-structured parameters
  - Comma-delimited list with no embedded whitespace
  - Null entries permitted (,,)
- Range specification
  - Specified as <*lower*>'/'<*upper*>['/'<*step*>]
  - Forward slash is used as ':' is valid in ISO 8601 (WMS does the same)
  - SSA does not use *step* currently, but it is in the syntax
  - Lower/upper value is included in the range
  - Open ranges are permitted
- Range-list
  - Is a list of scalar values or ranges
- Qualifiers
  - A range-list may be qualified as ';'<*qualifier*>
  - Qualifier applies to entire range-list
- Example
  - "1E-7/1.5E-7,3.3E-7/3.8E-7;source"

Multiple Service Operations

- Terminology
  - Service defines an *interface*
  - Interface may define multiple *operations* (methods)
- Examples
  - queryData, getData, stageData, getCapabilities
- Usage with HTTP GET
  - New REQUEST parameter specifies operation
  - Base-URL may be common or distinct; up to service provider
  - Typically all operations will share the same base-URL
  - Base-URL does not include VERSION or REQUEST
- Example
  - http://myvo.org/ssap?version=1.0&request=queryData&...

SSA Concepts

- Data Origin
  - Concept
    - Says something about how the data was originally generated
  - Classification
    - Atlas — Uniform coverage of parameter space (survey)
    - Pointed — Observations targeting specific objects
    - Modeled — Theory data, e.g., synthetic spectra
- Data Derivation
  - Concept
    - Says something about how the current dataset was produced
  - Classification
    - Observed — Single, direct instrumental observation
    - Composite — Composite of several observations
    - Simulated — Simulated observation, e.g., spectral extraction

SSA Concepts

- Service Type
  - Concept
    - General classification of service capability
  - Classification
    - Static — Returns whole, non-subsetted archive datasets
    - Cutout — Subsets data but does not resample pixels
    - Resampling — Resamples data, may reproject, cutout, etc.
  - Notes
    - A resampling service may also return cutouts or whole datasets
    - A cutout service may also return whole datasets
- Polymorphic Services
  - We can probably forbid this to simplify things
  - The generic Dataset is a better way to support polymorphism

Query Interface

- Operation
  - queryData (mandatory)
- Input Parameters
  - Mandatory
  - Recommended

Mandatory Query Parameters

#	Parameter	Sample value	Physical unit	Datatype	Utype
1	POS	52,-27.8	dec. deg., ICRS	double(2)	SSA.Position
2	SIZE	0.05	dec. deg.	double	SSA.RegionSize
3	BAND	0.1/2.7E-7 (=10cm-2700Å)	m	char(*)	SSA.SpectralBandpass
4	TIME	1998-05-21/1999	ISO 8601 UTC	char(*)	SSA.TimeBandpass
5	FORMAT	votable	-	char(*)	SSA.OutputFormat

Recommended Query Parameters

#	Parameter	Sample value	Unit	Datatype	Utype
1	APERTURE	0.00028 (=1″)	dec. deg.	double	SSA.Aperture
2	SPECRES	5D-10 (=5Å)	m	double	SSA.SpectralResolution
3	SPATRES	Xxxx	Dec. deg	double	SSA.SpatialResolution
4	TIMERES	xxxx	s	double	SSA.TimeResolution
5	SNR	5.0	-	double	SSA.MinSNR
6	REDSHIFT	-0.1,2.0	1	double(2)	SSA.Redshift
8	TARGETNAME	Mars	-	char(*)	SSA.TargetName
7	TARGETCLASS	star	-	char(*)	SSA.TargetClass

Recommended Query Parameters

#	Parameter	Sample value	Unit	Datatype	Utype
8	PUBID	ADS/col#R5983	-	char(*)	SSA.PublisherCreatorID
9	CREATORID	Ivo://auth/col#R1234	-	char(*)	SSA.CreatorCreatorID
10	COLLECTION	DSS2	-	char(*)	SSA.Collection
11	TOP	20	-	int	SSA.MaxTopRankedRecords
12	INDEX	50	Rows	Int	SSA.QueryIndex
13	COMPRESS	(TRUE)	-	boolean	SSA.Compress
14	SINCE	1995-04-15	ISO 8601	char(*)	SSA.Since
15	RUNID	<string>	-	char(*)	SSA.RunID