SIA Planning, Cube Access, Generic Dataset

IVOA Victoria, May 2006

Doug Tody (NRAO/US-NVO/IVOA)

DAL Sessions 3-4

- Spectral Data Access
  - Markus Dolensky – Ranking algorithm (20m)
- Spectral Line Access
  - Jesus Salgado, Pedro Osuna – SLAP specification (20m)
- SIA Upgrade Topics
  - Pat Dowler – Image indexing and searching (20m)
  - Silvia Dalia – Time-oriented Queries (20m)
  - Francois Bonnarel – CGPS radio data cube use case (15m)
  - Markus Dolensky – SINFONI IFU data cube use case (15m)
  - Doug Tody – Complex data, generic dataset, data cube access
- All – Discussion of SIA upgrade and related topics (30m)

SIA Upgrade Planning

- Main objectives
  - Upgrade metadata, query interface as for SSA
    - standard generic dataset metadata
    - much more powerful query interface
    - metadata extension mechanism
  - Precision image data access enhancements
    - e.g., cube data, image slicing, projection, filtering
    - (TBD whether this is folded into “basic SIA” or done as a separate service class)
  - Service capabilities
    - versioning, metadata query
    - asynchronous data staging, authentication, VOStore integration

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

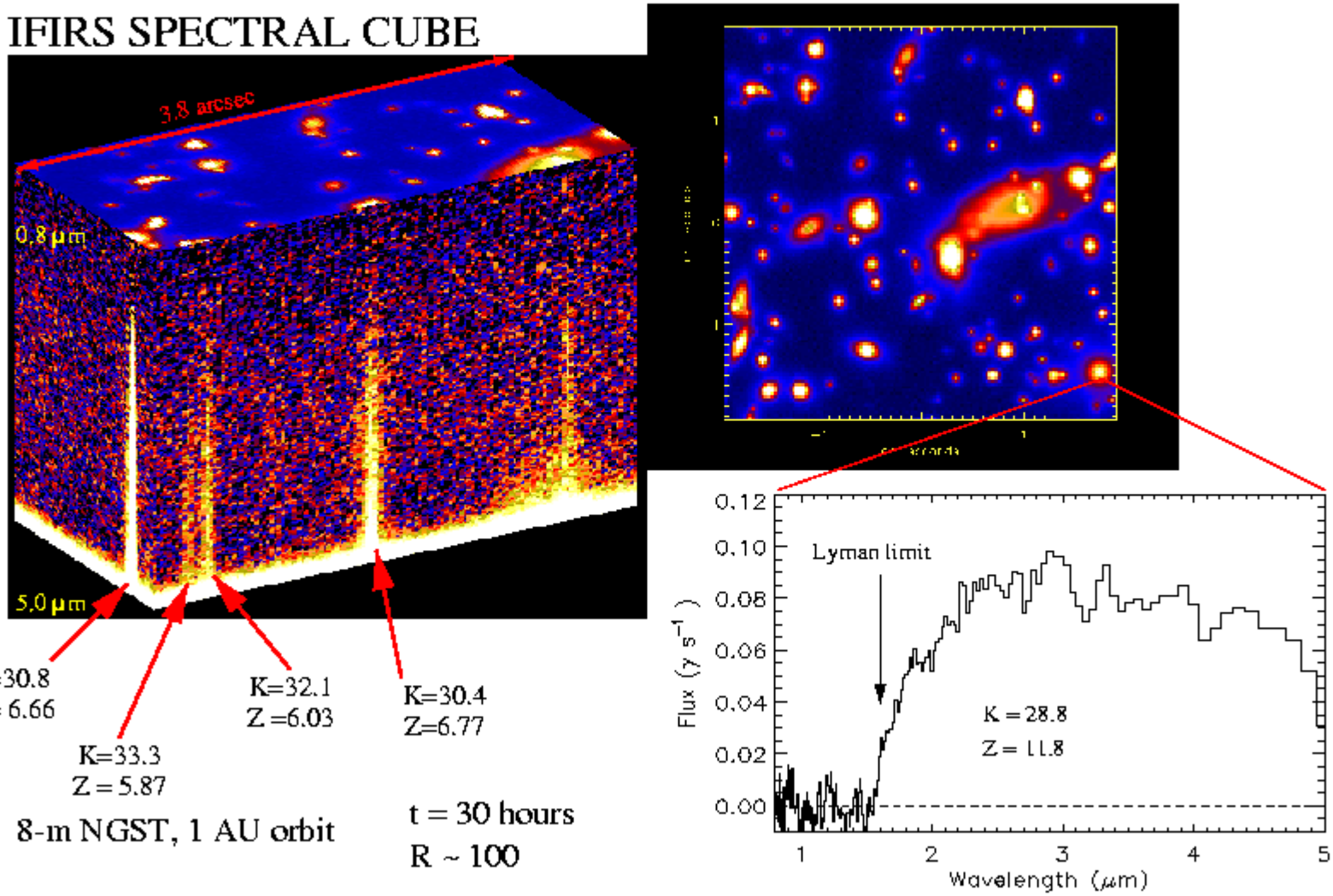
Recommended Query Parameters

#	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

Cube Data

- Overview
  - Motivated primarily by radio data surveys (CGPS, Arecibo)
  - Many O/IR integral field unit (IFU) instruments coming online as well
  - Datasets can be both large and complex
- Large datasets
  - Current data cubes are several hundred MB up to several GB
  - Future wide-field wide-band: 2048x2048x8192x4 = 128 GB
  - With polarization, multiple bands, could have 1/2 TB datasets!
- Complex datasets
  - e.g., CGPS: HI cube, CO cube, continuum, IQUV, IRAS same field
  - Multiple ways to view the same data
  - Multi-band surveys are a simpler example of this trend
- Use-Cases for recent study
  - CGPS, SGPS, GALFA (Arecibo), SINFONI (ESO IFU)

IM & STOCKMAN NGST DEEP FIELD SIMULATION



Cube Data

- Data access considerations
  - Network download of large cubes can be impractical
  - VO-style virtual data access to remote data is required
    - subsetting, filtering (spectral or time regions), transformations (projections, spectrum extraction)
- Typical access modes
  - Whole image
  - Spectrum extraction
  - Cutout 2D planes
  - Cutout 3D sub-cube (permits local full 3D analysis)
  - 2D projection along one axis
  - 3D projection (general 3D transformation)
  - 2D slice through 3D cube at arbitrary 3D pos,orientation

Cube Data

- Typical access scenario
  - Discovery query to discover data, get access metadata
  - Access query to set up virtual data access (WCS based)
  - Data access, dynamically generating virtual data
  - Repeat for a different region or view
- Example: Compute 2D projection with spectral filtering
  - View 2D preview or projection, e.g., continuum
  - Extract 1D spectra in sky regions (SSA with synthetic aperture)
  - Analyze sky spectrum to determine night sky lines (SLAP)
  - Compute 2D projection of cube excluding sky emission, absorption
- Other examples
  - Extract 3D sub-cube for full 3D analysis locally
  - 2D slice at arbitrary position and orientation

Cube Examples

- Extract 2-D plane from cube, same orientation
  - queryData
    - PubID=<desired cube dataset>
    - POS=<center of 2-D plane>
    - SIZE=<spatial extent of 2-D plane>
      - (cutout of smaller region also possible here)
    - BAND=<spectral-coord of desired plane>
    - NAXES=2
    - FORMAT=FITS

Cube Examples

- 2-D Projection with spectral filtering
  - queryData
    - PubID=<desired cube dataset>
    - POS=<center of 2-D plane>
    - SIZE=<spatial extent of 2-D plane>
      - (cutout of smaller region also possible here)
    - BAND=<range-list of “good” spectral regions>
    - NAXES=2
    - FORMAT=FITS

(in SINFONI case original cube is in Euro-3D format)

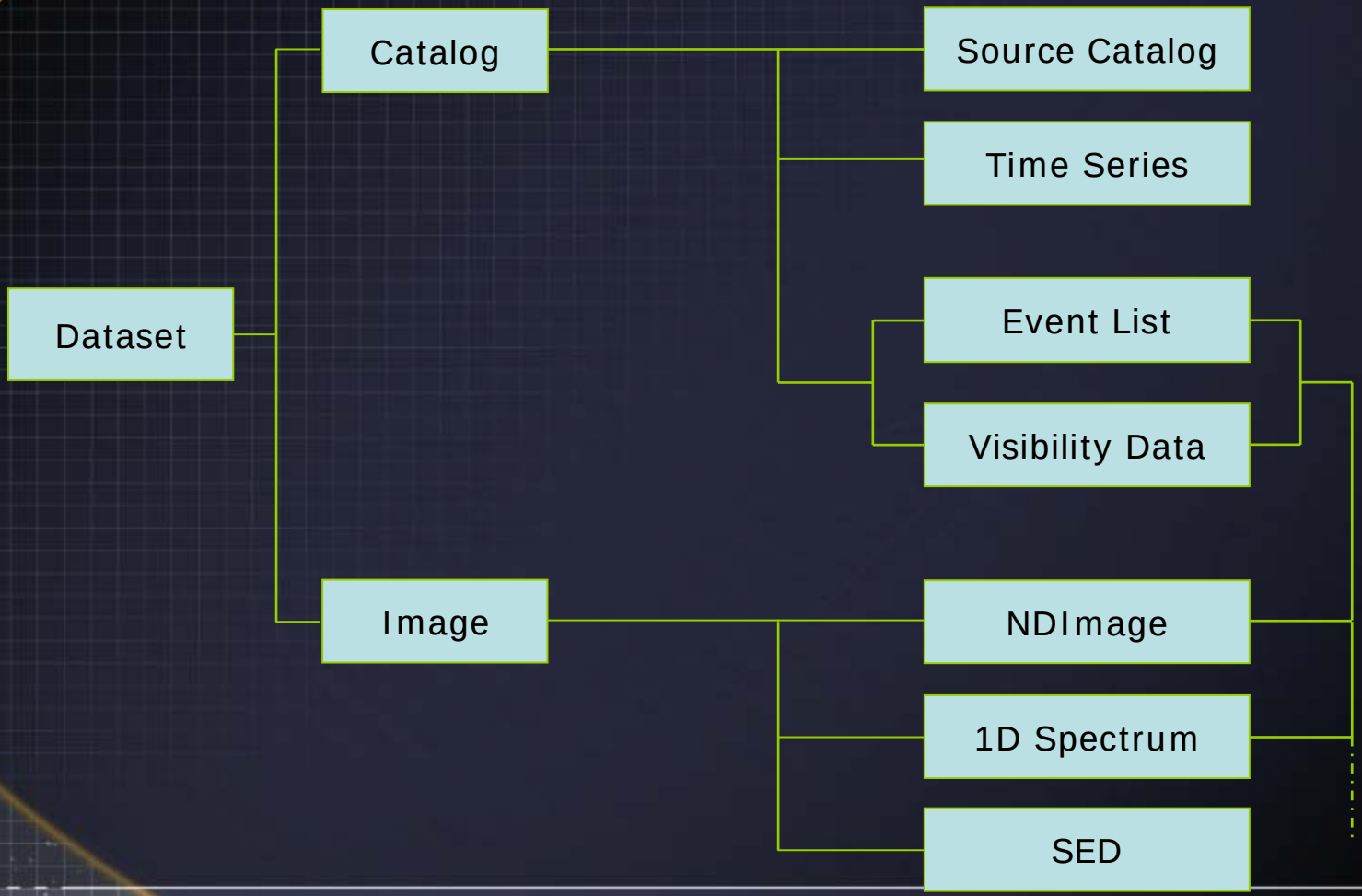
Cube Examples

- Extract 3-D Sub-Cube
  - queryData
    - PubID=<desired cube dataset>
    - POS=<spatial center of region>
    - SIZE=<spatial extent of sub-cube>
    - BAND=3.45E-7/8.76E-6
    - NAXES=3
    - FORMAT=FITS

Generic Dataset

- What is it?
  - Root Dataset class – applies to all types of data
  - Generic dataset metadata, generic Dataset query interface
  - SIA, SSA, etc. inherit from Dataset
- Needed to describe complex data
  - Aggregates of multiple data types (image, catalog, spectrum)
    - e.g., data cube survey use-cases
  - Multiple ways to view same data (image cutout, spectrum extraction)
- Powerful tool for data discovery
  - Can find any kind of data in a single query
  - Uniform description of all types of data
  - Ability to describe available access methods
- Would be natural place to introduce ADQL into DAL
  - Query is posed against generic dataset metadata table or tables
  - No virtual data access at this level

DAL Scope: Types of data (Cambridge 2003)



Generic Dataset

- Data discovery vs data access
  - Dataset query to discovery, understand complex data aggregates
  - SIA, SSA, etc. used for precision data access
    - Typed interfaces are still used for single-step virtual data discovery and access
  - SIA, SSA, etc. do not change at all
- Example Application: Dataset indexing
  - Google-type global index of all data, continually updated
  - Most powerful if it indexes all types of data
  - Want to index physical datasets, not virtual data
  - Never actually retrieve any data, just metadata
  - Hence, use generic dataset query interface
  - Query by modify time, large query response support required