VO Interoperability Meeting
Johns Hopkins University
Physics and Astronomy Building
Room 462

8:30am Introductions, Goals (Genova)

Would like to exchange information and ideas prior to adoption of standards. Would like to include more people, but meeting room limitations precluded this. Will begin using IVOA website to make documents widely available. Will try to plan for next meeting in spring, and again at ADASS '03 (Strasbourg).

8:45am Data models (McDowell/Louys)

Held meeting at SAO last week; compared existing models, discussed modeling process and how to reach standards. Began work on merging IDHA, AstroGrid, and NVO data models. Presentations available at http://hea-www.harvard.edu/~jcm/vo/dmtm.

Framework for compliant data models:

- White paper defining concepts
- UML (or equivalent) based model
- Convert UML to XMI or equivalent and verify that the model schema satisfies VO DM standards.
- Provide reference XSD schema and XML instance examples. IVOA or appropriate body can endorse these models.
- Conforming IVOA data models would be required to have:
    - URI
    - Description
    - Curation metadata
    - URL for white paper
    - Class descriptions
        - Name and description
        - Attributes
        - Properties/relationships
        - Methods
        - Abstract or concrete?

Bandpass methods and attributes

- Waveband
- Generic name
- Specific name
- Effective wavelength
- Effective minimum wavelength and maximum wavelength
- Transmission curves (theoretical, measured)
- Add and multiply bandpasses
- Quality object
- Uncertainties

Problems

- Red leaks
- Periodic filters
- Wedge filters, notch filters
- Multiple filters
- Atmospheric transmission

Attributes

- Name (filter ID), generic name, waveband name

- Central and end wavelengths
- Measured or theoretical

Wavelength or frequency-based model? Or have models that can be expressed in wavelength, frequency, or energy terms. Need to look at FITS Paper III for complete representations. Registry of bandpass names?

Much discussion—do we need names or numbers? VO service/directory of transmission curves?

What is an image?

- Sparse (with bad pixel map) is ok
- Regular pixelated array
- Pixels must have scalar simple data types (including complex)
- *n*-dimensions, *n* is arbitrary

This does not cover descriptions of photographic plates or other physical entities that do not exist in digital form.

Model comparisons

- IDHA has separate concepts for Raw and Processed data. What happens when an image is the result of combining others from different sources? How much of the processing history do you try to carry forward?
- "has a" vs. "is associated with"
- What is "data"? Hypermaps, images, tables, documents, plots, links, software,…

Data models will have to be partial, incomplete, given the great variety in data from different missions. What do you do when you don't recognize something? Most important thing is *an identifier* (i.e., a unique label).

What about links to publications? Could be in curation metadata for the data model.

Need to try some interoperability tests. A long way away from this right now.

Is this extensible to the Grid? The concept of "compute object" is in the AstroGrid data model.

What applications will make use of these data models?

- Make HST, ISO, and ESO… data structures look the same. Allows one piece of software to operate on all data.
- General exposure-time calculator
- Comparison of theoretical models to observed data

Next steps:

- Try process on concrete examples
- Improve integration of existing models
- Will communicate with dm@us-vo.org distribution list
    - majordomo@us-vo.org
    - subscribe dm (in body of message)

10:15am NVO Simple Image Access Protocol (Plante, Tody)

History/Motivation

- First-year NVO science prototypes
- Catalog and image access
- Cone search, simple image access

- Has been in discussion throughout the summer

Goals

- Uniform access to image data
- Service-oriented: distributed, heterogeneous
- Simple to implement and use (low buy-in)
- Both FITS and graphics image formats

Scope

- First step to provide uniform data access
- Implementation: URLs first, web services later
- Data models and metadata

Status

- V1.0 of specification now available at us-vo.org
- Various implementations in progress

Interface Overview

- QueryImages (pos, size, format, …) → VOTable
- GetImage (acref) → image/fits, image/jpeg, text/html, …
- AccessImages (clientID, images) → status [messages]
- Service registry
    - Resource metadata (coverage, etc.)
    - Service metadata (atlas, cutout, mosaic, etc.)
- Protocol independent
    - URL-based, web services based, etc.

Might get hundreds or thousands of images that match the query – may not be what users really want. How to limit? Interface seems asymmetric – greater detail on output side than on input side. Output parameters map to UCDs but input parameters do not. Service parameters are a bit different, though, than the metadata associated with the outputs. Commonality will be in types of parameters.

Concerned about keeping this simple for implementation. Service can provide additional metadata in response if it desires.

Needs authentication mechanism. V1.0 does not address this. Also does not address expiration of images (e.g., when staged or constructed on-demand). Will need to allow query to specify where to put the output – staged remotely, downloaded to client, etc. V1.0 assumes staging and provides URLs to images. Need to make it easy for data providers to make data available, and secondary services will be required to make optimal use of the data.

How does client specify computations they want to be done on the data being requested? Requires different service/interface. Need data models to support this properly. This interface does not accommodate this. Downstream work.

Service types: image cutout, image mosaic, atlas image, pointed image archive.

Query parameters, region of interest: POS, SIZE

POS combines RA and Dec. Easier when dealing with list of regions, and with syntax that allows combination of CFRAME and POS. POS is a position object.

Image generation parameters: NAXIS, CFRAME, EQUINOX, CRPIX, CRVAL, CDELT, ROTANG, PROJ (simplified from FITS WCS)

Need to keep interface definitions like this simple. Higher level services and clients should provide multiple coordinate formats. Keep middleware thin.

Re/ POS: Should design XML first and then reverse engineer the interface definition.

Re/POS: Had same discussions in design of ASU six years ago.

Important thing is to agree – implementers can deal with it either way.

Image formats: FITS, PNG, JPEG. What about compressed FITS? Should be transparent to this service – negotiated at the web server level. Already a standard way to do this in HTTP; should be described in appendix to the SIAP document.

Where to next? NVO project is starting implementations? Would be interesting to have a variety of client applications, beyond NVO initial science demos. Document will need to be updated following testing experiences.

Are the various VO project teams interested in doing test implementations of the SIAP? Yes, CDS/Aladin. Others…? CVO can implement trivially on top of existing services. Will need something for IAU IVOA demonstrations. Need list of issues to address in subsequent versions. Must distinguish between data you can analyze and data that is for preview purposes.

How to communicate further on this? Set up data access sub-group. Data access layer: dal@us-vo.org.

12:45pm UCDs and ontology (Ochsenbein, Rixon, Szalay, Williams)

UCDS: designed for automatic cross-correlations; hierarchy of terms is not fundamental. UCDs are not accurate definitions, do not express physical dimensionality, are not tied to units. UCDs are different from names; the names are the choice of the catalog creator, and are usually unique within a table. Several columns in a table might have the same UCD. UCDs are different from units. FITS WCS Paper II gives conventions for units. The units domain depends on the UCD. Java code now available from CDS for unit conversions. UCDs are assigned based on column name, units, and explanation. Done semi-automatically with s/w tool.

Usage: searches in catalogs and databases, filtering, visualization.

Updating: Cannot split UCDs into groups assigned to different groups for maintenance, as portioning the hierarchy is not unique. Central control is also awkward politically.

Tested UCDs for use in radio astronomy, including interferometry, and on SDSS. Alex worked at CDS to map SDSS database attributes to UCDs. About 70% of the SDSS fields (1300 from 20 tables) mapped directly onto UCDs. Found errors in units fields in SDSS. Needed 10-12 additional UCDs (standard deviation, variance, covariance, …).Also needed UCDs to describe UCDs themselves.

Can browse UCDs at http://vizier.u-strasbg.fr/UCD/.

UCDs can also be used for quality assurance. Helps to identify inadequately described columns.

Data models will refer to UCDs, but are likely to require additional ones.

UCD problems:

- Too many UCDs without precise definitions

- INST_BANDPASS – what it is exactly?
    - ERROR?
- Too many gaps in UCD tree
    - **MODEL**_POS_EQ_RA?
    - PHOT_JHN_V_**MAIN**?
- UCDs for structures?
    - POS_EQ composed of POS_EQ_RA., etc.

Refactor into "atoms"

- Define atoms individually: energy-flux, wavelength, declination. Include modifiers: mean, median, etc.
- Allow application developers to combine atoms into UCDs
    - Energy-flux/Johnson-v/photometry
    - Energy-flux/uncertainty/Johnson-v/photometry
    - Wavelength/….

Advantages

- Tuples can be written alphabetically or in XML structure
- ~400 atoms for all ~1500 current UCDs
- Easier to relate UCDs from different namespaces in they share some atoms
- Can treat old UCDs as atoms; add new atoms to remove ambiguity

Need some amount of hierarchy, but not single one for everything. POS_RA makes sense, but RA_POS does not.

Combination of some atoms would lead to meaningless terms. Same is true for any language.

Should have preferred "molecules". How detailed does a registry need to be? _ERROR can be appended to many molecules. Need not specify all combinations. Begs for a formal grammar. How structured vs. how loose? Preferred orderings? Structure relates back to data models/relationships…

Relationship between data models and metadata: data model shows relationships, defines terms, and terms are encoded in a metadata standard. Model essential building blocks first, rather than try to construct one massive diagram that represents all relationships in the world. All interconnected through re-use of common atoms.

What does "/" mean? "Is a" or "has a"?

Continue discussion on dm@us-vo.org.

1:45pm Resource and service metadata (Hanisch)

Get notes from Francoise…

2:15pm VOTable V1.0 review and retrospective (Williams, Ochsenbein)

Complex objects in cells? Should this be changed?

Error conditions: is a message the same as a table? Should exceptions be expressed in some other way?

SOAP services and VOTable – how to integrate?

How are references among VOTable elements best handled?

VOTable parsers

- Perl — Eric Winter, GSFC
- Java-lite — Andre Schaaff, CDS
- Java — Roy Williams, CACR
- C++ — VO India
- C# — Antonio Volpicelli

Standardization process? Follow FITS process (but faster)? Follow W3C RFC process? Discuss further in IVOA.

3:15pm Web services (Szalay, Rixon)

Grid services = web services + 1 or more standard ports + optional security features
Open Grid Services Architecture (OGSA): http://www.globus.org/ogsa/
Basis of future Globus toolkit

Web services: stateless, can't relocate, created by sysadmin, no metadata provision, no standard security, identified by URL

Grid services: can remember state, can relocate, created by user, GridService port, built-in security, located by Grid Service Handle (GSH)

Standard ports

- GridServices: emit service metadata, manage service instances, resources
- Factory: make service instances
- Registry: record service instances
- HandleMap: locate instances
- Notification{Source|Sink}: progress reporting

Use grid services for:

- Recording user-defined temporary state, e.g., caching results
- Providing service metadata in a standard way, e.g., image query service
- Pre-selecting service parameters, e.g., accrefs from image query service
- Access control on services

Web services at JHU (implemented in C#/.NET):

- Photometric redshift pipeline
- Galaxy clustering analysis
- Density maps from SQL query
- Bandpass service (w/ J. McDowell), filter transmission curves
- C# wrapper class for FITSIO
- Wrappers of legacy statistical tools in C (w/ A. Moore)

Other web service environments: Mono at http://go-mono.com

Starlink wrapping applications into web services

- Needs work on interfaces
- Keen to support IVOA service definitions
- Requires data to be co-located with service; can use GridFTP to access remote (i.e., local to user) data

CDS planning to make some SIMBAD functions available as web service. Please do name resolver first.

4:00pm General discussion, plans for future workshops

IVOA activities and milestones. See http://ivoa.net/.

Set up VO code repository in SourceForge?

Schedule for future discussions: After ADASS '03, but need something in between. ~April '03.

4:30pm Adjourn

Participants:

Françoise Genova	CDS	genova@astro.u-strasbg.fr
Bob Hanisch	STScI	hanisch@stsci.edu
George Fekete	JHU	gfekete@pha.jhu.edu
Jonathan McDowell	CfA	jcm@cfa.harvard.edu
François Ochsenbein	CDS	francois@astro.u-strasbg.fr
Arnold Rots	CfA	arots@head-cfa.harvard.edu
Vincent McIntyre	ATNF	vincent.mcintyre@csiro.au
Markus Dolensky	ESO	mdolensk@eso.org
Niall Gaffney	STScI	gaffney@stsci.edu
Tamas Budavari	JHU	budavari@jhu.edu
Riccardo Smareglia	INAF	smareglia@ts.astro.it
Ray Plante	NCSA	rplante@ncsa.uiuc.edu
David Giaretta	AstroGrid/Starlink	d.giaretta@rl.ac.uk
Alex Szalay	JHU	szalay@jhu.edu
Sebastien Derriere	CDS	derriere@astro.u-strasbg.fr
Alberto Micol	ST-ECF	Alberto.Micol@eso.org
André Schaaff	CDS	schaaff@astro.u-strasbg.fr
Pierre Fernique	CDS	fernique@astro.u-strasbg.fr
Guy Rixon	AstroGrid	gtr@ast.cam.ac.uk
Roy Williams	NVO/Caltech	roy@caltech.edu
Doug Tody	NRAO	dtody@nrao.edu
Patrick Dowler	CADC/NRC	patrick.dowler@nrc.ca
Roberta Allsman	NOAO	robyn@noao.edu
Steve Lubow	STScI	lubow@stsci.edu
Ryusuke Ogasawara	Subaru/NAOJ	ryu@naoj.org
Jonas Haase	ST-ECF	jonas.haase@eso.org
Francesco Pierfederici	ST-ECF	fpierfed@eso.org
Olga Pevunova	IPAC/Caltech	olga@ipac.caltech.edu
Erik Deul	Leiden Observatory	deul@strw.leidenuniv.nl
Mark Allen	CDS	allen@astro.u-strasbg.fr
William O'Mullane	STScI	womullan@yahoo.co.uk
Benoit Pirenne	ESO	bpirenne@eso.org
Gretchen Greene	STScI	greene@stsci.edu
Christopher Arviset	ESA	christophe.arviset@esa.int
Jose Hernandez	ESA	Jose.Hernandez@esa.int
Patricio Ortiz	AstroGrid	pfo@star.le.ac.uk
Clive Page	AstroGrid	cgp@star.le.ac.uk
Tom McGlynn	NASA/GSFC	tam@lheapop@gsfc.nasa.gov
Mario Nieto-Santisteban	JHU	nieto@pha.jhu.edu