VIRTUAL ASTRONOMICAL OBSERVATORY

VAO SIAv2 Prototype

Ray Plante, NCSA

Data Cube Team:

Bruce Berriman (IPAC), Mark Cresitello-Dittmar (SAO), John Good (IPAC), Matthew Graham (Caltech), Gretchen Greene (STScI), Bob Hanisch (STScI), Tim Jeness (Cornell), Joe Lazio (JPL), Pat Norris (NOAO), Olga Pevunova (IPAC), Arnold Rots (SAO), Doug Tody (NRAO)

The VAO is operated by the VAO, LLC.

Contents

- Motivation
- VAO Approach to SIAv2
- The prototype results
- Feedback to standards development
- SIAv2 in the context of Cube Access

Motivation

- Heidelberg
  - Call for creation of prototypes various approaches to cube data access
  - Data Providers: First show discovery and downloading of whole datasets
  - VAO response
    - Discovery and static download first priority
    - Sub-cube access (e.g. cutouts): as time allows
- VAO Project Goals
  - Demonstrate to the IVOA what an SIAv2 service would look like
    - Ability to see under the hood as well.
  - Exercise key science use cases to assess the standards
    - Provide feedback that can constructively inform further development of the standards
    - Focused selection of data based on coverage

The VAO approach to SIAv2

- See IVOA Twiki page: SIA2VAOPrototype
  http://wiki.ivoa.net/twiki/bin/view/IVOA/SIA2VAOPrototype
- Development “phases”
  - Initial Planning and setting scope (white paper)
    - Priority: discovery and static download
  - Updating the relevant documents: ImageDM and SIAv2
  - Implementation based on DALServer toolkit
  - Testing and evaluation of deployed prototype

The VAO approach to SIAv2

- See IVOA Twiki page: SIA2VAOPrototype
  http://wiki.ivoa.net/twiki/bin/view/IVOA/SIA2VAOPrototype
- Development “phases”
  - Initial Planning and setting scope (white paper)
  - Updating the relevant documents: ImageDM and SIAv2
    “Froze” documents to *baseline* version:
    - Key for testing/evaluation: what are we testing against?
    - Used a change control process to allow only critical changes
    - IVOA versions continue to evolve; thus, baseline is a minor fork.
      - SIAv2: VAO-SIA-2.0-20130830.pdf
  - Implementation based on DALServer toolkit
  - Testing and evaluation of deployed prototype

The VAO approach to SIAv2

- See IVOA Twiki page: SIA2VAOPrototype
  http://wiki.ivoa.net/twiki/bin/view/IVOA/SIA2VAOPrototype
- Development "phases"
  - Initial Planning and setting scope (white paper)
  - Updating the relevant documents: ImageDM and SIAv2
  - Implementation based on DALServer toolkit
    - DALServer: a Java-based toolkit developed by NVO/VAO
    - Collected wide variety of 2-D and N-D images
      - ALMA, VLA, JVLA, Heracles, OSIRIS (Keck), Califa, JCMT, JWST, CARMA
      - Traditional 2D images, 3D radio cubes, moment maps, sparse images, intergral optical spectrograph
      - Extracted/transformed FITS metadata, loaded into a database
        - Noted where we needed to "make things up" (mainly identifiers)
    - Deployed to server at NRAO
  - Testing and evaluation of deployed prototype

The VAO approach to SIAv2

- See IVOA Twiki page: SIA2VAOPrototype
  http://wiki.ivoa.net/twiki/bin/view/IVOA/SIA2VAOPrototype
- Development “phases”
  - Initial Planning and setting scope (white paper)
  - Updating the relevant documents: ImageDM and SIAv2
  - Implementation based on DALServer toolkit
  - Testing and evaluation of deployed prototype
    - Developed a set of Test Queries based on spec documents
      - Independent of implementation
      - Science-driven, focused on coverage: spatial, spectral, temporal, & polarization
    - Assembled Test Plan that included tests of compliance against our baseline documents
    - Tested for building and deployment on Linux/Apache/MySQL/Tomcat platforms

VAO Implementation

- Based on VAO baselines of ImageDM and SIAv2 documents
- Further limits on scope:
  - Synchronous version of service (**`/sync`**)
  - Focus is on **`REQUEST=queryData`**
  - Assume **`INSTERSECT=OVERLAPS`** (default)
  - Assume **`MODE=archival`**; i.e. return static images
- Supported Input parameters
  - Required parameters: **`POS`**, **`SIZE`**, **`BAND`**, **`TIME`**, **`POL`**, **`FORMAT`**
  - Some additional optional parameters support, but not tested
- Output columns:
  - Required and recommended columns supported
  - Note use of VOTable GROUPs to associate related columns

Results

- Deployed prototype service with base URL:
  http://vaosa-vm1.aoc.nrao.edu/ivoa-dal/siapv2/
- Report on IVOA TWiki page: SIA2VAOPrototype
  http://wiki.ivoa.net/twiki/bin/view/IVOA/SIA2VAOPrototype
  - Includes clickable test queries
- Source code available:
  - Browse on-line:
    http://dev.usvao.org/vao/browser/vao/prototype/dalserver/branches/siav2aproto
  - Download:
    http://dev.usvao.org/siav2/dalserver-siav2proto.1.tar.gz
  - "Friendly IVOAers" can build and deploy on own platform
    - Requires Apache, Java, Tomcat, MySQL, Python
    - See prototype/README

Feedback to standards development

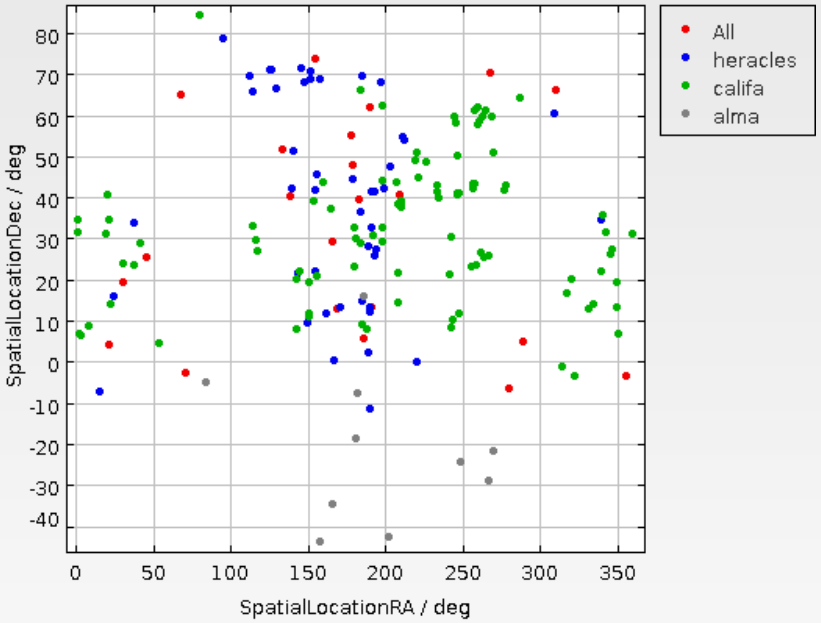
- General Assessments:
  - Successful demonstration of a service in the mode of SSA
  - Simple queries are compact, readable, and easily editable directly
  - Much improved metadata returned over SIAv1:
    - Can readily see what axes are found and in what order
    - Description of coverage is independent of the axes
      - E.g. Can see spectral coverage even if image does not formally have a spectral access
      - Does not depend on axis order, direction
    - Can determine what pixels measure (unit and UCD)
    - Can determine if dataset contains multiple image HDUs (Dataset.Image.Nsubarrays)
      - Browsing HDU metadata via Dataset.Image.Dataref not demonstrated
  - Flexible to complex, leading-edge N-D datasets
    - Proof of concept: study: applicable to x-ray event list

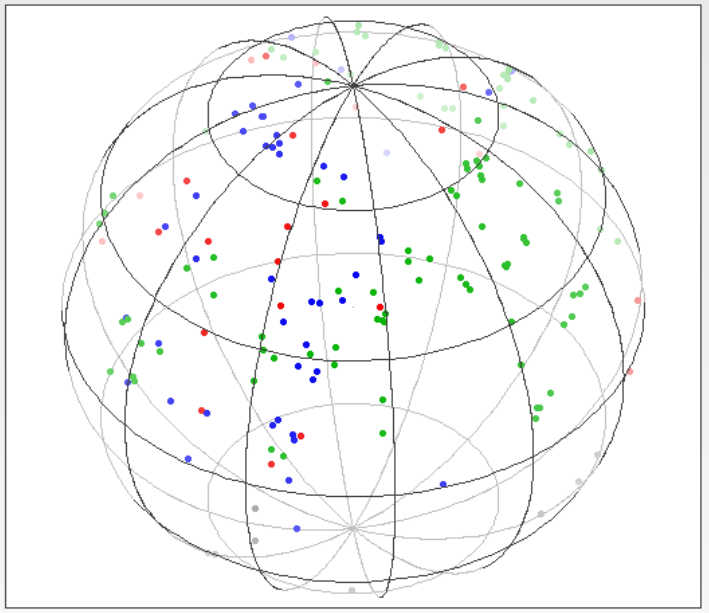
Feedback to standards development

- Various minor issues
  - Alternate Spectral coverage information (freq., energy, velocity) is useful to know
  - Make returning of Char.SpatialAxis.Coverage.Support.Area as “recommended”
    - STC-S description of spatial coverage
    - ImageDM needs clarification of UType definition
  - Metadata do not adequately identify compressed images
  - Consider search constraint on spectral line species
  - Make returning of Char.FluxAxis.Unit as “recommended”
  - Consider search constraint that tests pixel content
    - Against Char.FluxAxis.UCD? Char.FluxAxis.Unit?
    - Is SUBTYPE relevant?
    - Important for theory data: find theoretical optical depth images

Feedback to standards development

Distribution of test data collection





Returning spatial positions

- Current spec calls for positions to be an RA-Dec pair of numbers in single column
  - Image spatial center, lower & upper limits
- Cannot be interpreted by common tools
  - Requires tool understand DM to “know” pair is a position
- Better to render RA/Dec in separate columns
  - Need UTypes defined to identify components of position
  - Use GROUPs to associate RA and Dec components

Feedback to standards development

Handling unknown data in queries (NULLs)

- Consider query requesting data for time domain analysis: **/sync?REQUEST=querydata&TIME=1992-01-01/1997-01-01**
  - Current spec: if metadata for constraint is not known, the image should be returned in results
    - Thus, result contains (many) images that may not actually be within time range
    - Prototype: 4 images known to be in range, 290 returned
  - Rationale: don't let client miss potentially useful images
    - let client further filter on the client side
    - However, if time coverage is not known, client will be not be able to determine applicability post-query either
- Possible Solution
  - Extra parameter that provides a switch controlling whether to return images where information is unknown
  - Consider carefully what default behavior should be

SIAv2 in context of Cube Access

- What distinguishes SIAv2 from SIAv1
  - Full coverage querying; not just by position
  - Richer optional parameters that help refine queries:
    - Identifiers: Isolating specific images, collections, of source derivation
    - Processing levels, targets, sensitivity
  - Metadata model that is better suited to browsing N-d cubes
  - Sub-cube access (cutouts, subsetting, reprojections, …)
    - Client-directed “cutouts” separated explicitly from the discovery
    - Client uses the data returned in the discovery response to craft inputs into AccessData query
    - Allows for requesting richer transformations
  - Staging data on the server
    - Not currently specified

SIAv2 in context of Cube Access

- Potential for take-up
  - By clients
    - Parameter approach demonstrated to be convenient for human and agent clients alike
      - See PyVO Focus Demo
    - Small set of core discovery parameters
    - Common queries are compact, readable, editable
  - By data providers
    - SIAv1 has demonstrated large take-up
    - Small set of core metadata:
      - 6 required, others “strongly recommended” depending on the coverage of the data.
    - Aided if our approach is to adapt onto the metadata already maintained by data provider
      - As opposed to in effect requiring separately maintained database tables.