INTERNATIONAL VIRTUAL OBSERVATORY ALLIANCE
US National Virtual Observatory

IVOA Data Access Layer
TAP Use-Cases, Param Query (PQL) Update

D. Tody (NVO, NRAO)

Topics

- Key TAP Use-Cases (from USVAO planning)
  - Cross-Matching (large scale, small scale/scriptable)
  - Global Data Discovery, index tables, data linking
  - Implications for TAP evolution
- TAP Parameter Query Update (PQL V0.2)
  - Changes; support for above use cases

Distributed Cross-Matching (SkyQuery)

- Goals
  - Large scale (no size limit)
  - Distributed query capability (but also high performance)
  - Based on IVOA standards (TAP, VOSpace, etc.)

- Architecture
  - Client portal (one)
    - drives the cross match job
    - produces abstract XMatch definition
  - Large site(s)
    - turn XMatch definition into a workflow
    - take advantage of locality, e.g. table caching, native SQL, indexing
    - high performance link between large sites (VOPipe)
    - may take advantage of custom HPC protocols internally
  - Small data provider(s)
    - participate via IVOA standards (TAP etc.)
    - can locally perform simple spatial cross match with filtering

Cross-Match Requirements for TAP

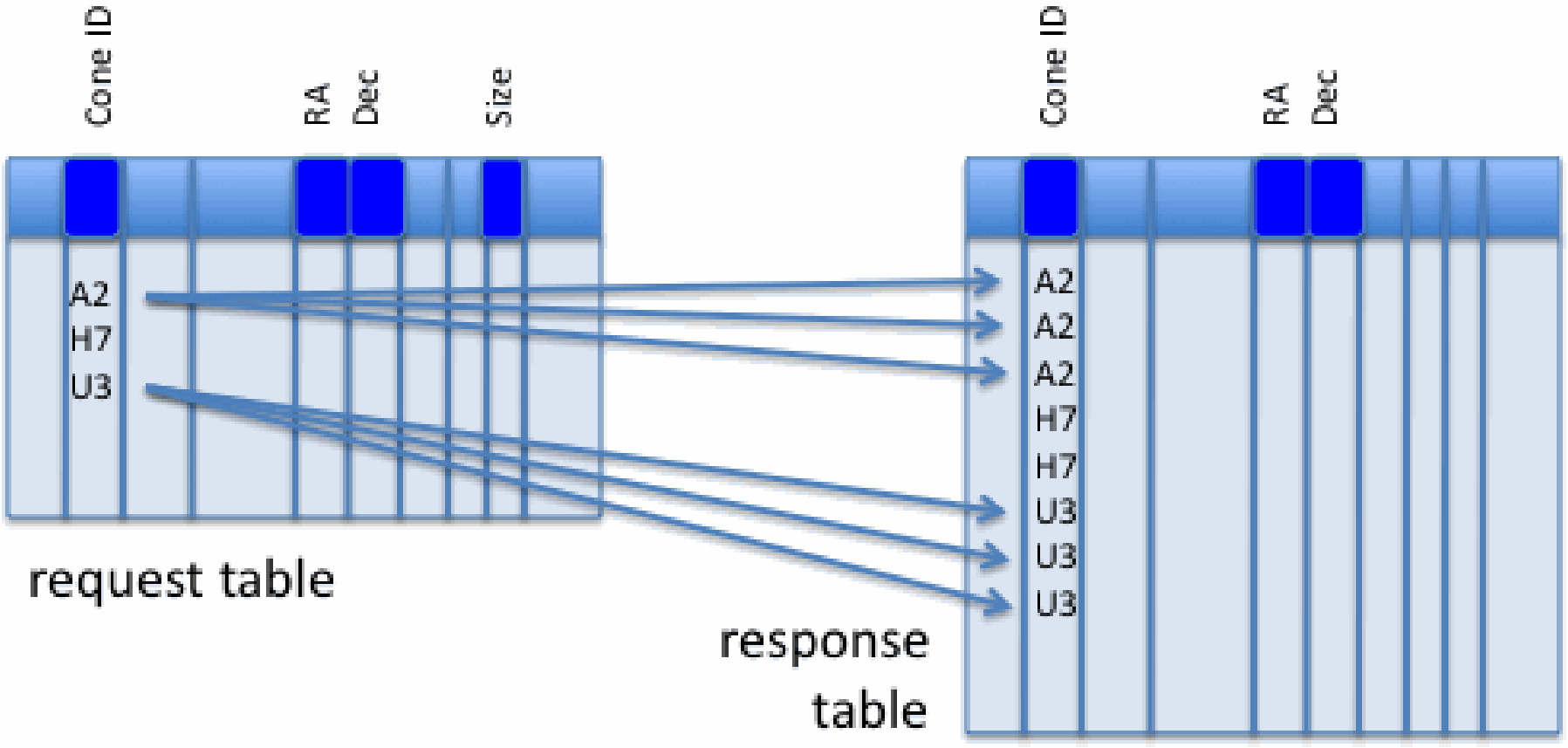
- Full-Up TAP Role
  - ADQL with region extensions, complex filters
  - VOSpace integration providing shared table access
  - Grid: VOPipe, SSO, UWS
- Minimal TAP Role
  - Multi-position (multicone) query (or ADQL+)
    - simple spatial cross match with filtering
  - Streaming output
  - Return filtered table to XMatch application

Simple Client Cross-Match

- Scenario
  - User application drives remote TAP directly
  - Multi-pos (or ADQL+)
    - simple spatial cross match plus filtering
  - Second stage of cross-match done in client code
    - any user algorithm can be applied to table columns
    - a great deal can be done with this simple scheme
  - Simple, but scalable

- Minimal Requirements
  - Multi-pos for simplicity, upgradable to ADQL+regions
  - Streaming output or UWS with job download

Multi-Position Query



Input Table Metadata

Quantity	*utype*	*ucd*	*Nam*
Cone ID	src:Position.ID	meta.id	ConeI
First coordinate (e.g., Right Ascension), degrees	src:Position.Coord1	pos.eq.ra	RA
Second coordinate (e.g., Declination), degrees	src:Position.Coord2	pos.eq.dec	DEC
Diameter of search region	src:Position.Size	pos.angDistance	SIZE

Global Data Discovery

- Concept
  - Find archival datasets (static files) of any kind
    - Distinguished from data access where we compute virtual data
    - Can be truely global, or restricted to a single site
    - Can expose both VO data as well as instrumental data
  - Use generic VO tools to browse archives and download data
  - Integrate with VO services for image, spectrum, cube access etc.
  - Drive custom local services for pipeline reprocessing etc.

- Implementation
  - Table based – leverage TAP and TAP-based tools
  - Generic Dataset (GDS) model +custom models to describe data
  - Data linking to link to data products or services

Global Data Discovery

- Index Tables
  - Provide uniform (table) view of a *data collection*
    - instrument/survey, user-adhoc, global index of a site, etc.
    - "portfolio" concept in VOSpace is similar
  - Data model
    - custom, GDS, or a combination of the two
  - Generic Dataset (GDS) is a type of index table
    - more on this in DAL session 2

- Archives Provide:
  - Index catalogs describing data products
  - Data products (catalogs, images, spectra, instr. data, etc.)
  - Services to search, access, process data

Global Data Discovery

- Data Linking
  - Index table links to data product or other resources
    - data product, access service, HTML page, other table, etc.
    - Access reference is a special case of a data link
  - Generic tools
    - can browse tales, follow links, invoke services

- Associations
  - Associate related data products to model complex data
  - Can view complex dataset or individual data products
  - Data linking at level of individual datasets

Parameter Query (PQL V0.2) Update

- Data Model
  - More explicitly table oriented
    - fully specify how to access table data, metadata
  - Use GDS/Obs model for anything more complex than POS,SIZE
    - concepts like BAND, TIME not clear otherwise
    - GDS query vital for data discovery in DAL anyway

- New Advanced Topics Section
  - Fully specify semantics, parameter interactions
    - need more than individual param descriptions to define complex ops
  - Multi-position (multicone) query
  - Table metadata queries
  - Generic dataset queries (data discovery and access linkage)