

VO INVENTORY SERVICE

Virtual Observatory Inventory Service

John Good
VAO System Engineer
IPAC / Caltech

What is the VO, anyway?

A personal viewpoint:

The VO is fundamentally about distributed data and services.

- Finding datasets by characteristics
- Finding subsets of data by attribute
- Retrieving data through standard interfaces

Uniform access to all data and metadata.

- Levels the playing field

An example: Image metadata

Basic service:

- Find all the image datasets with data covering a location or locations or overlapping a region
- For a given image dataset, find the list of (URL-downloadable) images or cutouts for a location, locations, or region

And do all this fast.

VOInventory implementation

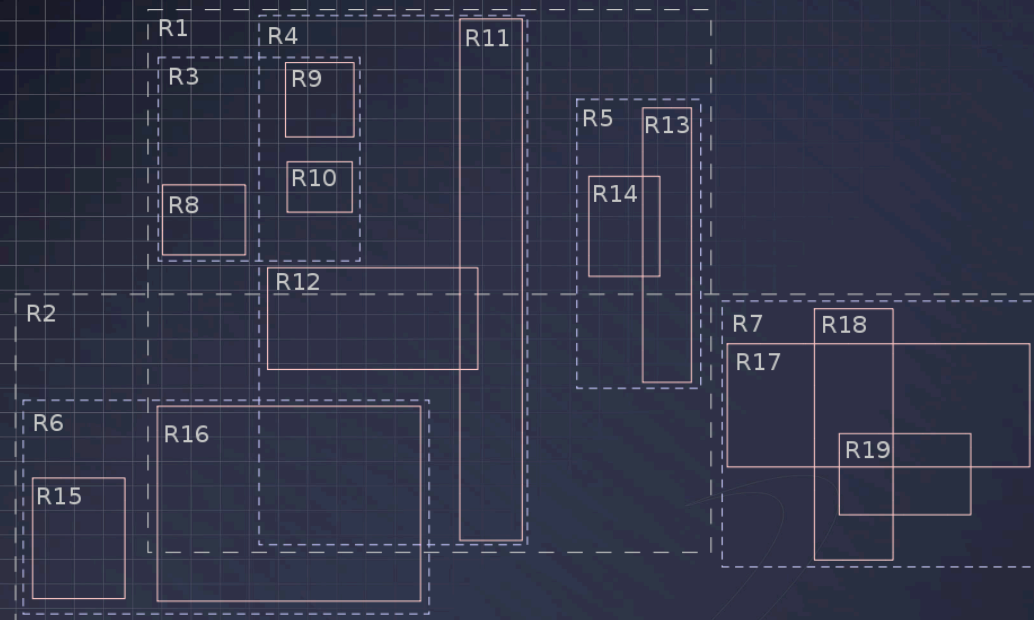
Implementation choices

- R-Tree indexing for $\log(N)$ searching
- Memory-mapped files
- Parallelization / cluster processing
- Adjunct SQL filtering

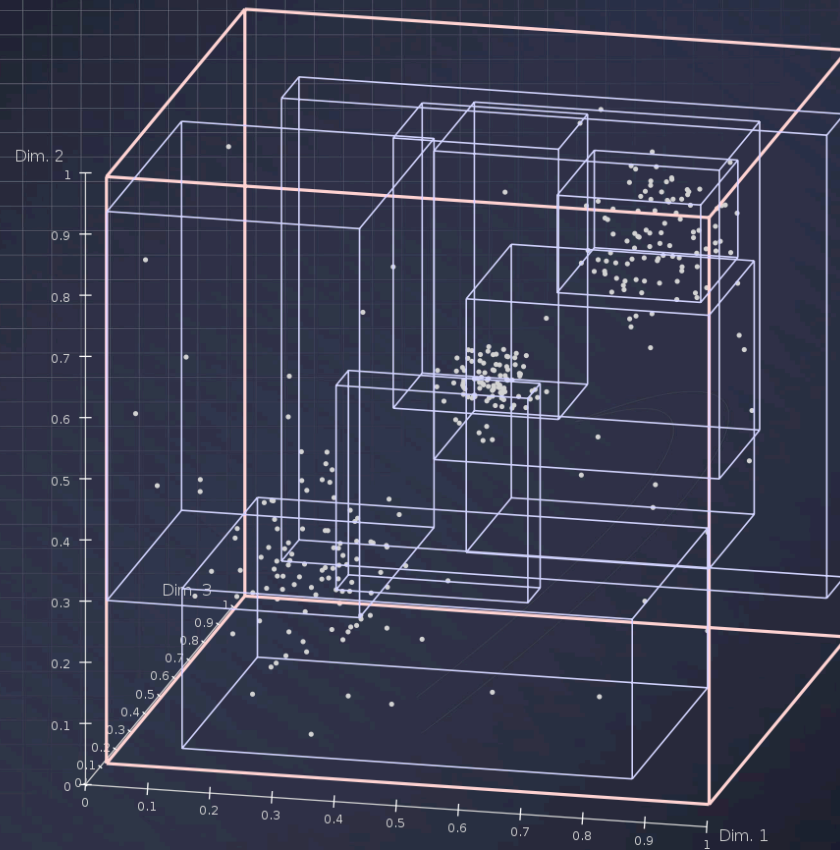
Access

- REST-based web services (*coming; older HTTP service for now*)
- HTML / AJAX NVO-specific interface

R-Tree basics



R-Tree basics (2)



Memory-mapped files

- A segment of virtual memory is assigned a byte-for-byte correlation with a portion of a file.
- Applications can then treat the file contents as if it were in primary memory.
- R-Tree structures contain complicated memory mappings and references and are slow to build. Putting them in memory-mapped files means they can be built “in memory” but saved and moved around as files.

Parallelization

- In general, fragmenting an index is a bad idea since $m \cdot \log(N)$ is substantially greater than $\log(mN)$.
- However, really large files start to have paging issues and it is often more operationally practical to manage data in subsets.
- The VOInventory uses a threaded front-end / multi-backend approach.

SQL filtering

- VOInventory is primarily a spatial search engine.
- However, many image metadata result tables need further filtering to be useful (*e.g.* isolating wavelength by file name filtering).
- VOInventory allows for “post” filtering of result tables using fully-functional SQL constraint “where” clauses.

Example

The NVO Inventory Service gives you a count of the number of entries, in each of many hundreds of catalogs and data collections, that are within a user-specified radius of a given position or list of positions on the sky. You can upload a list of sky positions and/or a list of catalogs and data collections of interest. You can examine individual datasets in detail, download the matched catalog entries for local analysis, or send results to other Data Discovery tools for further investigation.

Developed with the support of the [National Science Foundation](#) under Cooperative Agreement AST0122449 with the Johns Hopkins University. The NVO is a member of the [International Virtual Observatory Alliance](#). This NVO Application is hosted by [IRSA](#).

Table	Located At	Archive Subset	Table Record Count	Matched Sources
<input type="radio"/> HST Archived Exposures Catalog	HEASARC	HST	508518	80
<input type="radio"/> HST Planned and Archived Observations	HEASARC	HST	397638	45
<input type="radio"/> Extragalactic Radio Sources	HEASARC	GALAXY CATALOG	8603	40
<input type="radio"/> Revised and Updated Catalog of Quasi-stellar Objects <i>The catalogue</i>	VIZIER	Non-stellar	22484	14
<input type="radio"/> Master Radio Catalog	HEASARC	MASTER CATALOG	3581603	11
<input type="radio"/> HST Archived Exposures Catalog <i>First period from launch (1990Apr) to 1993Dec</i>	VIZIER	'External'	40008	8
<input type="radio"/> HST FOS spectral atlas <i>Observation Details</i>	VIZIER	Journal/ApJS	1357	5
<input type="radio"/> Master X-Ray Catalog	HEASARC	MASTER CATALOG	675218	4
<input type="radio"/> Dixon Master List of Radio Sources (Version 43)	HEASARC	RADIO CATALOG	84559	4
<input type="radio"/> SPECIND Catalog of radio continuum spectra <i>The parameters of the radio spectra</i>	VIZIER	Radio/Far-IR	375105	4
<input type="radio"/> The MAST Image Table/Spectra Scrapbook	IRSA	MAST	29291	4
<input type="radio"/> Master list of radio sources, updated 1978 <i>The Master List</i>	VIZIER	Non-stellar	79493	4
<input type="radio"/> Master Optical Catalog	HEASARC	MASTER CATALOG	4363156	4
<input type="radio"/> Einstein IPC Images	HEASARC	EINSTEIN	4132	3
<input type="radio"/> XMM-Newton XAssist Source List	HEASARC	XMM-NEWTON	67352	3
<input type="radio"/> Extragalactic Radio Source Identifications <i>The catalogue</i>	VIZIER	Non-stellar	14585	3
<input type="radio"/> Einstein Observatory IPC Parameters <i>Field Parameters for the IPC</i>	VIZIER	Astrometric	4295	3
<input type="radio"/> ROSAT All-Sky Bright Source Catalogue <i>1RXS Correlation to NED</i>	VIZIER	Astrometric	38670	3
<input type="radio"/> CRATES Flat-Spectrum Radio Source Catalog	HEASARC	RADIO CATALOG	14467	3
<input type="radio"/> Chandra Public Observations	HEASARC	CHANDRA	7191	2
<input type="radio"/> Chandra Observations	HEASARC	CHANDRA	9060	2
<input type="radio"/> Einstein Observatory IPC Parameters <i>Component Huts for IPC Sequences</i>	VIZIER	Astrometric	13265	2
<input type="radio"/> Chandra XAssist Source List	HEASARC	CHANDRA	73743	2
<input type="radio"/> Einstein Observation Log	HEASARC	EINSTEIN	5659	2
<input type="radio"/> INTEGRAL Observing Program	HEASARC	INTEGRAL	4491	2
<input type="radio"/> The Chandra Archive Log <i>The Chandra Log (2008-01-13)</i>	VIZIER	'External'	9022	2
<input type="radio"/> Einstein IPC Photon Event Data	HEASARC	EINSTEIN	4092	2
<input type="radio"/> Einstein IPC Unscreened Photon Event List	HEASARC	EINSTEIN	4028	2
<input type="radio"/> Master Observation Logs	HEASARC	MASTER CATALOG	44479	2
<input type="radio"/> ROSAT SIMBAD Identifications	HEASARC	ROSAT	639774	2

Applications

This service is currently being used by:

- The NVO Inventory service
- Spitzer Heritage Archive interface
- LSST image metadata search service prototype
- Montage on-demand mosaicking service