



VIRTUAL ASTRONOMICAL OBSERVATORY

# Guiding Queries with Data Summaries

Tom Donaldson, STScI



The VAO is operated by the VAO, LLC.



# The Challenge

Suppose you have some data tables.

- ObsCore
- Source Catalog
- Real Estate Listings
- ...

... and your users want to find interesting entries in your table.



# Ultimately Just a Database Query

For example:

```
select * from obsdata where waveband = 'Optical'
```

# User Interfaces for Generic Queries

Enter query directly.

- May have assistance with column names and operators.

```
11  
12 select * from obsdata where waveband = 'Optical'  
13
```

Query name

Output format

**Run**

Quickview

Reset



# User Interfaces for Generic Queries

Enter filter values into traditional web form.

[Archive Status](#)

## SID Search

[Help](#)  
[Field Descriptions](#)

[Standard Form](#)
[File Upload Form](#)

---

**Retrieval Type**  
 Package  File

**Target Name**

**User-specified field 1**    **Field Descriptions**  
Retrieval Type

**Test Name**

**Instrument**

**User-specified field 2**    **Field Descriptions**  
Retrieval Type

**Test Number**

**Data Mode**

# User Interfaces for Generic Queries

More advanced interfaces provide more assistance.

The screenshot shows the TAPHandle web interface. At the top, there is a logo and the text 'TAPHandle' followed by a search box containing 'caom2' and a 'Node Selector @ b' link. Below this is a breadcrumb trail: 'caom2 > caom2 > Observation'. A 'Tan Nodes' section is visible below the breadcrumb. The main interface has a navigation bar with buttons for 'Select What', 'Where', 'Plain Text Query', and 'Job Control'. On the left, there is a 'Result Limit' input field set to '100'. The central part of the interface features a table of fields with their data types and units, and a 'Kw Filter' section. On the right, there are input fields for 'Coord/Name', 'Radius(arcmin)' (set to 1), and 'System' (set to ICRS). Below these is a 'List of Active Constraints' section with a dropdown menu showing 'caom2.Observation.collection ='. At the bottom right, there is a 'QL stmt' field containing 'IS NOT NULL' and a sorting dropdown set to 'desc'.

algorithm_name	VARCHAR		
collection	VARCHAR		
environment_ambientTemp	DOUBLE	C	
environment_elevation	DOUBLE	deg	
environment_humidity	DOUBLE		
environment_photometric	INTEGER		
environment_seeing	DOUBLE	arcsec	
environment_tau	DOUBLE		
environment_wavelengthTau	DOUBLE	m	

Kw Filter: ra [ ] dec [ ]  
caom2. [ ]

Coord/Name [ ] SIMBAD System ICRS [ ]  
Radius(arcmin) 1 [ ] Upload Position List [ ]  
List of Active Constraints [ ]  
caom2.Observation.collection = [ ]  
QL stmt IS NOT NULL [ ] desc [ ] asc [ ]



# User Interfaces for Generic Queries

```

11
12 select * from obsdata where waveband = 'Optical'
13

```

Query name: ta **Archive Status** **SID Search** [Help](#) [Field Descriptions](#)

[Quickview](#)

**Standard Form** **File Upload Form**

Search

Retrieval Type  
 Package  File

Target Name

User-specified field 1 Field Description  
 Retrieval Type

**TAPHandle**  Node Selector @ [b](#)

Meta

Tan Nodes

Select What Where Plain Text Query Job Control

Result Limit

algorithm_name	VARCHAR		
collection	VARCHAR		
environment_ambientTemp	DOUBLE	C	
environment_elevation	DOUBLE	deg	
environment_humidity	DOUBLE		
environment_photometric	INTEGER		
environment_seeing	DOUBLE	arcsec	
environment_tau	DOUBLE		
environment_wavelengthTau	DOUBLE	m	

Coord/Name  [YIMBAD](#) System

Radius(arcmin)  [Upload Position List](#)

List of Active Constraints

Kw Filter

QL stmt: IS NOT NULL

In common:

- Column names can be provided to user.
- User must know something about the possible values.
- User doesn't know how many records will be returned, so may be overwhelmed by the results.



# Guidance on Column Contents

Borrow paradigm from VAO/MAST Data Discovery Tool (and actually lots of other web pages):

- User interface generically provides a filter for each column.
- More importantly, filters provide a summary of the column contents.
- Summary is essentially a histogram of the column values.





# Guidance on Column Contents

Select Collection: All MAST Observations Search: m51

Portal v1.6... [Examples: M101, 14 03 12.6 +54 20 56.7 r=0.2d, more...](#)

Start Page MAST: m51

3400 Total Rows MESSIER 051 Footprints: All

Filters Table View

Clear Filters Edit Facets... Help... Filter All Record Fields

Mission

Order Values by Count

- FUSE (3 of 3)
- GALEX (363 of 363)
- HST (2598 of 2598)
- HUT (1 of 1)
- IUE (25 of 25)
- SWIFT (410 of 410)

Instrument

Order Values by Count

- ACS (180 of 180)
- ACS/HRC (406 of 406)
- ACS/WFC (260 of 260)
- FOC/96 (8 of 8)
- FOS/RD (26 of 26)
- FUV (4 of 4)
- GALEX (363 of 363)
- LWP (11 of 11)
- LWR (4 of 4)
- NICMOS (56 of 56)
- NICMOS/NIC1 (8 of 8)
- NICMOS/NIC2 (100 of 100)
- NICMOS/NIC3 (932 of 932)
- STIS/CCD (62 of 62)
- SWP (10 of 10)
- UVOT (410 of 410)
- WFC3/IR (172 of 172)
- WFC3/UVIS (74 of 74)
- WFPC/PC (14 of 14)
- WFPC/WFC (4 of 4)

	Actions	Preview	Target Name	Instrument
1			M51	UVOT
2			SNinM51	UVOT
3			SNinM51	UVOT
4			SNinM51	UVOT
5			SNinM51	UVOT

AstroView

13:28:40.185 +46:57:31.018 RA DEC  
13:29:56.201 +47:13:50.016 hhmms/deg



# Guidance on Column Contents

Select Collection: All MAST Observations Search: m51  
 Portal v1.6... Examples: [M101, 14 03 12.6 +54 20 56.7 r=0.2d, more...](#)

Start Page MAST: m51

Displaying 2598 of 3400 Total Rows

MESSIER 051 Footprints: All

**Filters**

Clear Filters Edit Facets... Help...  
 Filter All Record Fields

**Mission**

Order Values by Count

- FUSE (0 of 3)
- GALEX (0 of 363)
- HST (2598 of 2598)
- HUT (0 of 1)
- IUE (0 of 25)
- SWIFT (0 of 410)

**Instrument**

Order Values by Count

- ACS (180 of 180)
- ACS/HRC (406 of 406)
- ACS/WFC (260 of 260)
- FOC/96 (8 of 8)
- FOS/RD (26 of 26)
- FUV (0 of 4)
- GALEX (0 of 363)
- LWP (0 of 11)
- LWR (0 of 4)
- NICMOS (56 of 56)
- NICMOS/NIC1 (8 of 8)
- NICMOS/NIC2 (100 of 100)
- NICMOS/NIC3 (932 of 932)
- STIS/CCD (62 of 62)
- SWP (0 of 10)
- UVOT (0 of 410)
- WFC3/IR (172 of 172)
- WFC3/LVIS (74 of 74)
- WFC3/PC (14 of 14)
- WFC3/WFC (4 of 4)

Actions	Preview	Target Name	Instrument
1		M51-WFPC2-POS3	WFPC2
2		M51-WFPC2-POS3	WFPC2
3		M-51-V06	WFC3/IR
4		M-51-V06	WFC3/IR
5		M-51-V06	WFC3/IR

**AstroView**

13:30:45.009 +47:04:35.976 RA DEC  
 13:29:56.201 +47:13:50.016 hhmss/deg



# Guidance on Column Contents

Select Collection: All MAST Observations Search: m51  
 Portal v1.6... Examples: [M101, 14 03 12.6 +54 20 56.7 r=0.2d, more...](#)

Start Page MAST: m51  
 Displaying 322 of 3400 Total Rows MESSIER 051 Footprints: All

Filters  
 Clear Filters Edit Facets... Help...  
 Filter All Record Fields

Mission  
 Order Values by Count

- FUSE (0 of 3)
- GALEX (0 of 363)
- HST (322 of 2598)
- HUT (0 of 1)
- IUE (0 of 25)
- SWIFT (0 of 410)

Instrument  
 Order Values by Count

- ACS (0 of 180)
- ACS/HRC (0 of 406)
- ACS/WFC (260 of 260)
- FOC/96 (0 of 8)
- FOS/RD (0 of 26)
- FUV (0 of 4)
- GALEX (0 of 363)
- LWP (0 of 11)
- LWR (0 of 4)
- NICMOS (0 of 56)
- NICMOS/NIC1 (0 of 8)
- NICMOS/NIC2 (0 of 100)
- NICMOS/NIC3 (0 of 932)
- STIS/CCD (62 of 62)
- SWP (0 of 10)
- UVOT (0 of 410)
- WFC3/IR (0 of 172)
- WFC3/UVIS (0 of 74)
- WFPC/PC (0 of 14)
- WFPC/WFC (0 of 4)

Table View	Actions	Preview	Target Name	Instrument
1	  		M51-POS3	ACS/WF
2	  		M51-POS3	ACS/WF
3	  		M51-POS3	ACS/WF
4	  		M51-POS3	ACS/WF
5	  		M51-POS3	ACS/WF

AstroView  
 13:30:41.547 +47:00:24.766 RA DEC  
 13:29:56.201 +47:13:50.016 hhmms/deg



# Extending the Concept

- That example was on a small subset of a table, where all the data and filtering were in the client.
- Try applying the same approach to a whole table
  - Compute summary histograms ahead of time
  - Use database queries to apply filters
- In this example, explore a JWST instrument test database.

# Extending the Concept

**MAST SID Test Interface**

Records Found: 207756 (Download Limit: 25000)

Search

Applied Filters

Columns: Defaults Hide All Clear All

Primary Columns:

- Retrieval Type
- Instrument
- File Name (p)
- Test Campaign
- Test Number
- Date (p)
- Test Title (p)
- Test Goal
- Optical Element
- Grating

Secondary Columns:

- File ID
- Package ID
- ASIC Part Number
- ASIC Temperature

Filters:

**Instrument**

Order Values by Count

<input type="checkbox"/> FGS	(6394 Total)
<input type="checkbox"/> NIRCAM	(28948 Total)
<input type="checkbox"/> NIRCam Optical Measurement Assembly (OMA)	(21056 Total)
<input type="checkbox"/> NIRSPEC	(57566 Total)
<input type="checkbox"/> TFI	(1234 Total)

**Date (p)**

1999-12-31 2013-04-03

**Test Goal**

Order Values Alphanumerically


<input type="checkbox"/> gain	(21813 Total)
<input type="checkbox"/> focus	(15217 Total)
<input type="checkbox"/> alignment	(14993 Total)
<input type="checkbox"/> dark	(14751 Total)
<input type="checkbox"/> flat	(13553 Total)
<input type="checkbox"/> external	(12118 Total)
<input type="checkbox"/> qe	(9633 Total)
<input type="checkbox"/> noise	(5799 Total)
<input type="checkbox"/> functional	(5613 Total)

**Optical Element**

Order Values by Count

<input type="checkbox"/> CLEAR	(8268 Total)
<input type="checkbox"/> clear	(238 Total)
<input type="checkbox"/> DHS2	(288 Total)
<input type="checkbox"/> F070W	(1295 Total)
<input type="checkbox"/> F090W	(502 Total)
<input type="checkbox"/> F115W	(232 Total)
<input type="checkbox"/> F140M	(3858 Total)
<input type="checkbox"/> F150W	(1840 Total)
<input type="checkbox"/> F150W2	(1113 Total)

# Extending the Concept



**MAST SID Test Interface**

Records Found: 502

**Search**

**Applied Filters**

Optical Element: F090W

**Columns**

Defaults Hide All Clear All

**Primary Columns**

- Retrieval Type
- Instrument
- File Name (p)
- Test Campaign
- Test Number
- Date (p)
- Test Title (p)
- Test Goal
- Optical Element
- Grating

**Secondary Columns**

- File ID
- Package ID
- ASIC Part Number
- ASIC Temperature


**Filters**

**Instrument**

Order Values by Count

<input type="checkbox"/> FGS	(6394 Total)
<input type="checkbox"/> NIRCAM	(28948 Total)
<input type="checkbox"/> NIRCAM Optical Measurement Assembly (OMA)	(21056 Total)
<input type="checkbox"/> NIRSPEC	(57566 Total)
<input type="checkbox"/> TFI	(1234 Total)

**Date (p)**



1999-12-31 2013-04-03

**Test Goal**

Order Values Alphanumerically

<input type="checkbox"/> gain	(21813 Total)
<input type="checkbox"/> focus	(15217 Total)
<input type="checkbox"/> alignment	(14993 Total)
<input type="checkbox"/> dark	(14751 Total)
<input type="checkbox"/> flat	(13553 Total)
<input type="checkbox"/> external	(12118 Total)
<input type="checkbox"/> qe	(9633 Total)
<input type="checkbox"/> noise	(5799 Total)
<input type="checkbox"/> functional	(5613 Total)

**Optical Element**

Order Values by Count

<input type="checkbox"/> CLEAR	(8268 Total)
<input type="checkbox"/> clear	(238 Total)
<input type="checkbox"/> DHS2	(288 Total)
<input type="checkbox"/> F070W	(1295 Total)
<input checked="" type="checkbox"/> F090W	(502 Total)
<input type="checkbox"/> F115W	(232 Total)
<input type="checkbox"/> F140M	(3858 Total)
<input type="checkbox"/> F150W	(1840 Total)
<input type="checkbox"/> F150W2	(1113 Total)

Tom Donaldson

September 28, 2013



# Soapbox Comment

Count-only queries are *\*very\** important.

- Very inefficient to spend time downloading a large dataset only to find out it's too much data for the client (and probably the user).



# Applicable to VO Use Cases?

TAP services already supply some column metadata.

- Summary histograms could become another piece of column metadata that could greatly enhance the utility of any TAP client.
  
- Are there other useful forms of summary data that we could supply “easily”?
  - MOC for spatial coverage
  - ???





# Notes/Limitations

- Computing histograms is fairly easy, but can take time for large datasets
- Histograms need to be recomputed when table is updated.
- Would be more useful if they were recomputed after each query.
  - Not fast enough in general.
  - Could pre-compute certain primary permutations.
  - Use estimates based on partial results?



# Thank You

