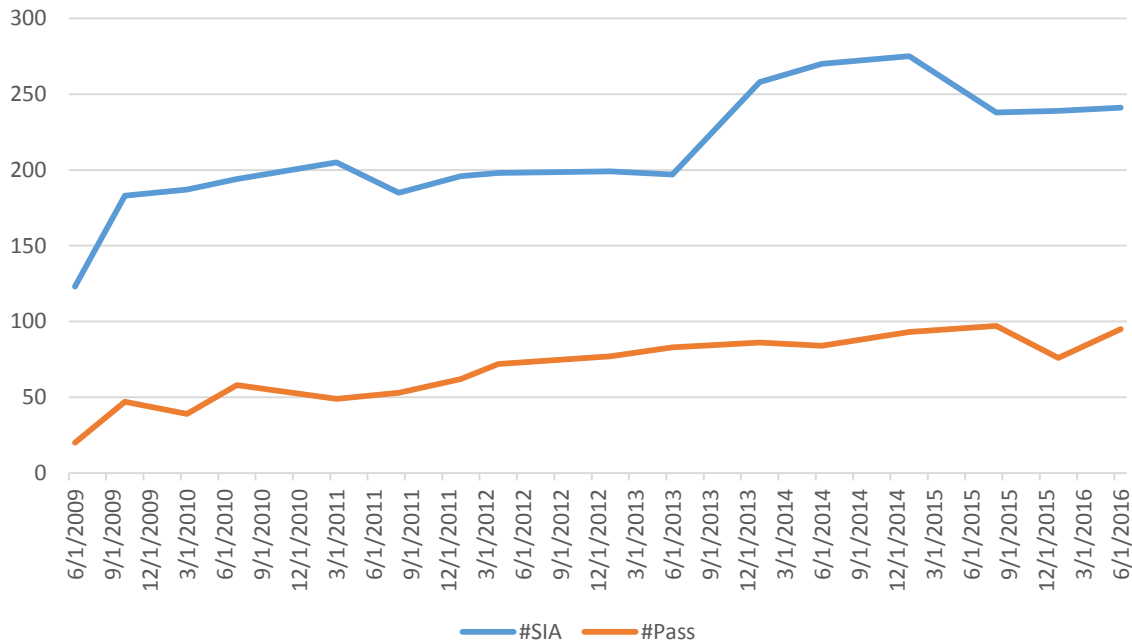# Comparing DAL Validators
A personal tale of error and repentance
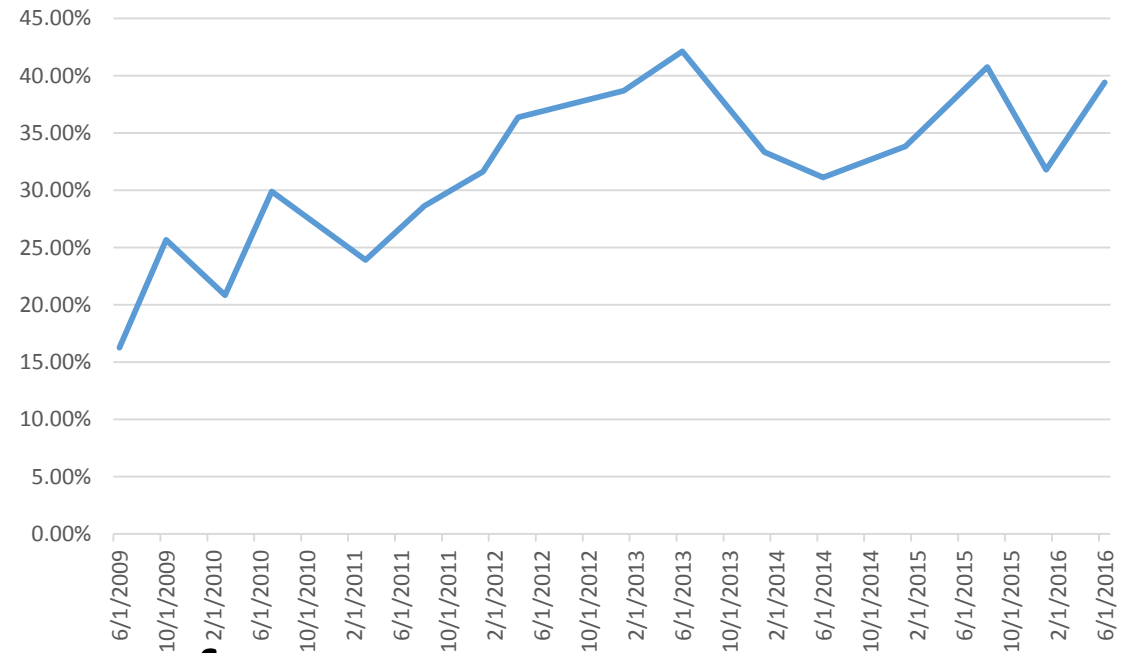
Tom McGlynn

NASA/HEASARC

Validation History (using NCSA/HEASARC SIA validator)

Use all SIA's since cone search stat's essentially measure VizieR.

SIA Service Validation History

%Pass

Not much progress for a long time!

Automated DAL service validators

- Currently running at VO-Paris, ESA, HEASARC (from NCSA)
- Effort in last year to understand differences
- Comparison of tests (see Ops Wiki page) suggests that results should be similar, but in practice this is not the case
- Why?

- Take a look at the HEASARC web pages and see how they do…

A few weeks ago…

HEASARC Services Validation Statistics: March 2016										
Validator	Combined					Cone			Simple Image Access	
	Total	Pass	Fail	%	Total	Pass	Fail	Total	Pass	Fail
HEASARC	903	774	129	86%	852	748	104	50	30	20
VO-Paris	909	0	909	0%	859	0	859	50	0	50
ESA	899	0	899	0%	849	0	849	50	0	50

ESA: Loop 136
　　　Cone Search: 826 C, 11 C-, 12 C—
　　　SIA: 20 B, 30 C—
VO-Paris numbers are estimated (but no services passed!)

How do the results compare?

- Numbers of services similar.

- Doing well with HEASARC based tests
  - We'd tried to fix these in the past
    - May not be any easier than other validators, just the one we'd paid attention to.
  - But even here lots of errors

- Nothing fully validated by other testers

But now…

Validator	Combined				Cone			Simple Image Access		
	Total	Pass	Fail	%	Total	Pass	Fail	Total	Pass	Fail
HEASARC	903	794	109	88%	852	748	104	51	46	5
VO-Paris	911	750	161	82%	861	700	161	50	50	0
ESA	901	618	283	69%	851	618	233	50	0	50

ESA: Loop 138

    Cone Search: 618 A+, 214 C, 6 C-, 13 C—

    SIA: 20 B, 30 C—

HEASARC and VO-Paris from 5/4/2016.

Why did the HEASARC complain?

- Cone Search
  - Mostly explicit choice to break CS protocol: either no ID column or ID column not unique. Compatibility of results with non-VO interfaces deemed more important than detailed CS compliance.
  - Found a few tables where metadata was missing.
- Simple Image Access
  - All failures due to inability to find downloadable data
    - Bad test data. Need to specify position/size
    - Only HEASARC tests data download?

Why did VO-Paris complain

- Cone Search
  - Wrong VOTable format (1.0 or 1.1 required)
  - Errors in VOTable
- Simple Image Access
  - Errors in PARAMs: missing attributes
  - Nominal VOTable version (1.1) incompatible with organization (used INFO at end of RESOURCE which is only allowed starting in VOTable 1.2)

VOTable details matter.

Why did ESA complain?

- Cone Search
  - Test of erroneous request was done by not specifying required parameter. We allowed default so query succeeded.
    - Comparable tests at HEASARC/VOParis used erroneous input

- Simple Image Access
  - Still looking!

How do validators compare?

- HEASARC/NCSA focuses on DAL standard.
  - Uniqueness in ID column
  - Downloadability of SIA data.
- VO-Paris does more testing of the underlying VOTable
  - Checking of VOTable version.
  - Checking against DTD/Schema
- ESA does more checking of different request types

Strengths of Web Pages

- HEASARC gives nice summary of what errors are found for a given institution or protocol, with links to services showing any given issue
- VO-Paris supports broadest array of validators with SIAv2 coming soon
- ESA ADQL interface supports most complex queries

Can we put all of these features together somewhere?

Take-away

- DAL Sx validators tend to be minimalistic: typically one test per requirement.
  - Assume all failures are real! False negatives seem to be rare.
  - Good for continuous testing, but more exhaustive validation (e.g., TAPlint) is often desirable
- Need to worry about version issues. E.g., some protocols require specific versions of VOTable, UCDs, etc.
- VOTables need to be checked with VOTlint or similar
- Multiple validators are useful since validator authors may have different focus or understandings of standard
  - Should we combine?
  - How do we get developers/maintainers to notice?
- Still a long way from 100%.
- Aside: Need to review periodically and after software updates, cf. recent issues with VizieR.