

Publication and Use of Validation Results
or
Validation — What is it good for?

Mark Taylor (Bristol)

IVOA Interop
Shanghai

19 May 2017

`$Id: validuse.tex,v 1.17 2017/05/19 01:51:59 mbt Exp $`

Outline

- What validation are we doing?
- Why are we doing it?
- Can/should we publish machine-readable bulk validation results?
 - Why?
 - ▷ To compare validation services?
 - ▷ For use of client applications and their users?
 - How?
 - ▷ Using existing Registry DM and services?
 - ▷ Simple Validation Access Protocol?
 - ▷ ValidationTAP?

Standards Validation Coverage

Known validators listed at [IvoaValidatorsSummary](#) wiki page

- Recent changes
 - ▷ Add moclint (MOC)
 - ▷ Add Ucidy (UCD)
 - ▷ Remove WS Basic Profile (obsolete)
- Missing
 - ▷ DALI? DataLink, STC-S, SimDAL, VTP, VOEvent, VOEventRegExt
- N/A?
 - ▷ SSO, DocStd, Data models?

Taplint to-do:

- RegTAP, EPN-TAP, ADQL standard functions, Units (Unity), UCD (Ucidy)
- Errata: COOSYS, TAP upload data types, Identifiers anyURI

Validation Overview

Validation activity

- There is a requirement for validators for Proposed Recommendations:
 - ▷ DocStd v2.0:

“When applicable, the Working Group should be able to demonstrate two interoperable implementations of each feature, and validation tools should be available.”
 - ▷ Operations IG charter:

“During the IVOA standards review process, assess the validation capabilities that are now required as part of new standards.”
- People are writing validation software
 - ▷ 20+ individual software items on wiki [IvoaValidatorsSummary](#) page
- Several data centers run bulk periodic validation of registered services
 - ▷ PADC, Euro-VO, HEASARC, others?
 - ▷ Results circulated in various ways: these meetings, web pages, by email

Questions:

1. What benefits does validation provide? Why do we require it?
2. Can we get more benefit from the validation work that we're doing?

Why Validate?

- To enable metrics on how well the VO is running
 - Useful internally to understand what is/isn't working well, especially what's getting better/worse
 - Useful externally to report to funders, community etc (*but use with care*)
- To assist in building (compatible, correct) implementations
 - Q: otherwise, how do you know you've got it right?
 - A: by writing your own test suites
 - Danger: implementors may end up coding to a validator rather than to the standard, or not bothering to write their own tests
- To improve the quality of standards
 - Writing a comprehensive validator is a proof by construction that the standard is comprehensible, unambiguous and (in some ways) implementable
 - Does a similar (complementary?) job to requiring implementations
 - Not producing validator+implementations while drawing up standard is more expensive in the long run than doing it ...
 - ... at least in most cases (*"When applicable ..."* — DocStd 2.0)

Bulk Validation Results

- Various organisations run “bulk” validation:
 - VO-Paris, EuroVO, HEASARC, more...?
 - Periodically test all registered services of particular types (e.g. SCS, SIA, SSA, TAP)
- Results available in various ad-hoc ways
 - Presented at these meetings
 - ▷ Useful to assess health of VO
 - ▷ Feeds into standards development
 - Passed to operators of problematic/failing services
 - ▷ Positive effect on service compliance
 - In some cases available on line
 - ▷ Service operators can go looking for test results for self-assessment
- Currently, no standard machine-readable access to validation results

Validation results publication

Standard validation result publication would be nice to have...

- (1) Comparing validation services would improve consistency and understanding of VO health
- (2) Knowing which services “work” could improve user experience significantly

... but tricky to do

- Service provision
 - ▷ Who publishes validation results?
 - ▷ How do clients access them?
- Characterising validity
 - ▷ What does a “validation result” look like?

The requirements may differ for the two use cases

Publication Use Case 1: Comparison

If bulk validation results were available in a standard form, we could:

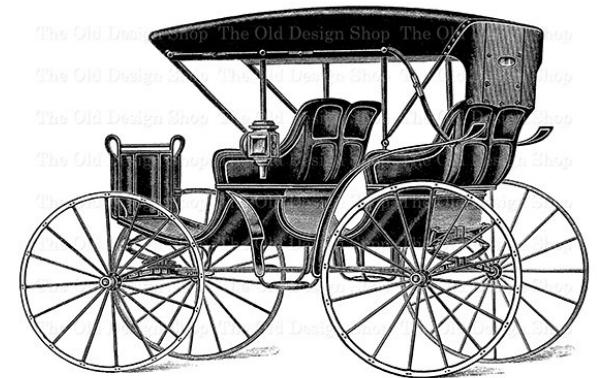
- compare them to understand validator behaviour (identify validator issues)
- aggregate them to improve results
- make it easier for service providers to see how they are doing

Requirement:

- Multiple result sets visible alongside each other
- Fairly detailed enumeration of results by protocol and test type

Publication Use Case 2: Applications

Publication Use Case 2: Applications



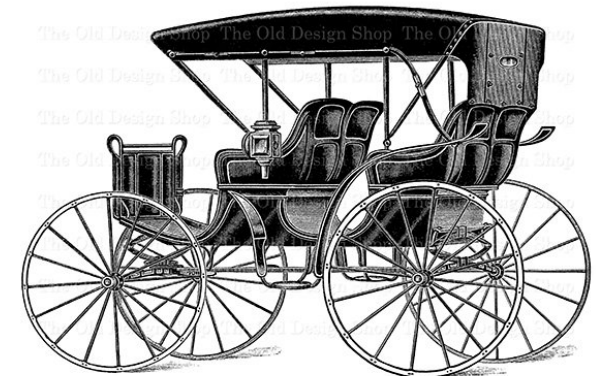
Publication Use Case 2: Applications

AstroGrid VODesktop application (circa 2008)

The screenshot displays the AstroGrid VODesktop application window. The title bar reads "Astroscope - 14 Cat. Object Services - 5 Image Services - 4 Spectra Services". The interface includes a menu bar (File, Edit, View, History, Result, Window, Help), a search panel on the left with filters for "Cat. Objects", "Images", "Spectra", and "Timed Data", and a main results table. The search criteria are set to "At Position (RA,Dec) or Object Name" with coordinates 187.750000, +12.400000 and a search radius of 0.010000 degrees. The results table lists 15 resources with columns for Status, Results, Capabilities, Title, and SubName. Below the table, there are tabs for "Results", "Information", "Table Metadata", and "XML". The "Results" tab is active, showing a list of files with columns for Icon, Name, Date Modified, Size, and Type. The "Process" panel on the left includes actions like "View", "Download...", "Send tables to Topcat", and "Send tables to Aladin".

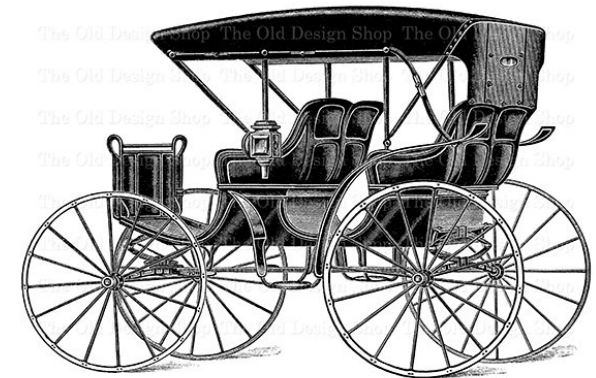
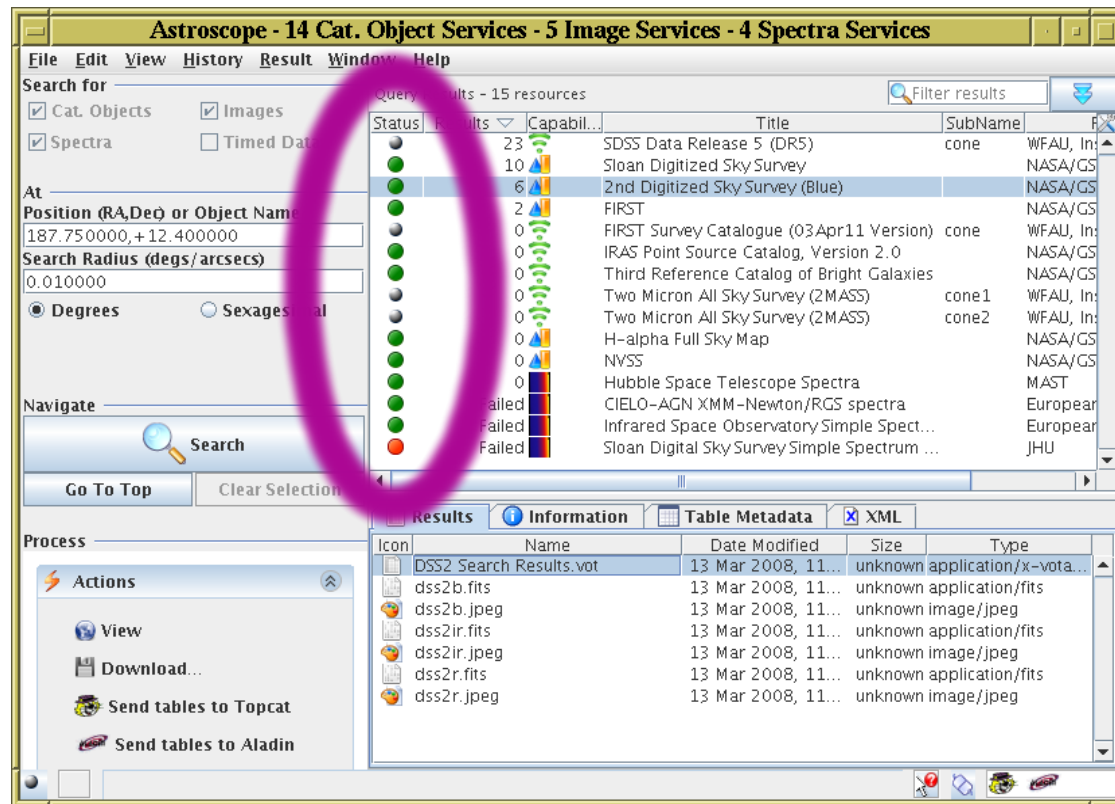
Status	Results	Capabil...	Title	SubName
●	23	●	SDSS Data Release 5 (DR5)	cone WFAU, In:
●	10	●	Sloan Digitized Sky Survey	NASA/GS
●	6	●	2nd Digitized Sky Survey (Blue)	NASA/GS
●	2	●	FIRST	NASA/GS
●	0	●	FIRST Survey Catalogue (03Apr11 Version)	cone WFAU, In:
●	0	●	IRAS Point Source Catalog, Version 2.0	NASA/GS
●	0	●	Third Reference Catalog of Bright Galaxies	NASA/GS
●	0	●	Two Micron All Sky Survey (2MASS)	cone1 WFAU, In:
●	0	●	Two Micron All Sky Survey (2MASS)	cone2 WFAU, In:
●	0	●	H-alpha Full Sky Map	NASA/GS
●	0	●	NVSS	NASA/GS
●	0	●	Hubble Space Telescope Spectra	MAST
●	Failed	●	CIELO-AGN XMM-Newton/RGS spectra	Europear
●	Failed	●	Infrared Space Observatory Simple Spect...	Europear
●	Failed	●	Sloan Digital Sky Survey Simple Spectrum ...	JHU

Icon	Name	Date Modified	Size	Type
📄	DSS2 Search Results.vot	13 Mar 2008, 11...	unknown	application/x-vota...
📄	dss2b.fits	13 Mar 2008, 11...	unknown	application/fits
📄	dss2b.jpeg	13 Mar 2008, 11...	unknown	image/jpeg
📄	dss2ir.fits	13 Mar 2008, 11...	unknown	application/fits
📄	dss2ir.jpeg	13 Mar 2008, 11...	unknown	image/jpeg
📄	dss2r.fits	13 Mar 2008, 11...	unknown	application/fits
📄	dss2r.jpeg	13 Mar 2008, 11...	unknown	image/jpeg



Publication Use Case 2: Applications

AstroGrid VODesktop application (circa 2008)



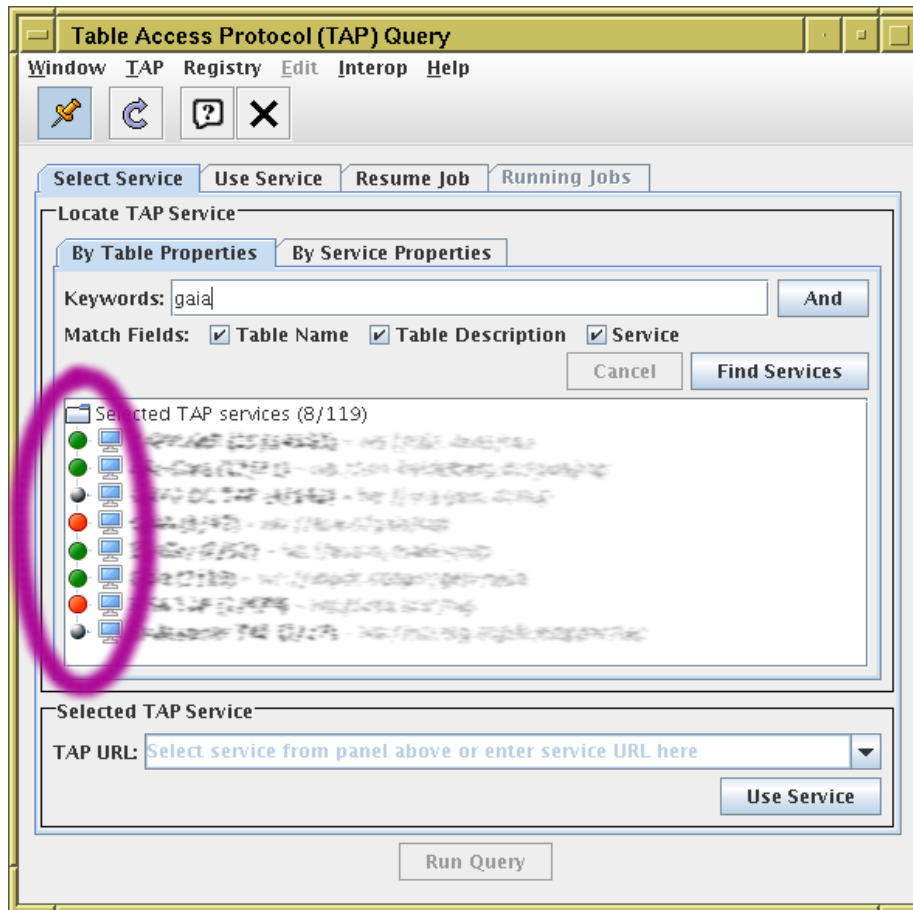
- **Status** column for each VO service

- Options ● Good / ● Bad / ● Unknown
- Nice indication to user of health/usability/correctness(?) of services
- Required central **“VoMon”** monitoring service to supply values
- Feature withdrawn in later VODesktop versions

Publication Use Case 2: Applications



Publication Use Case 2: Applications



Could we provide something similar in future applications?

Requirement:

- Per-service simple quality/fitness flag or score (*is this possible?*)
- Ideally a single source for results

Characterising Validity

Need some kind of Validity data model

- Generic items:
 - ▷ `ivoid = ivo://example.com/catalog`
 - ▷ `standard_id = ivo://ivoa.net/std/ConeSearch`
 - ▷ `validated_by = ivo://validator.com/`
- For comparing/aggregating results from different validators:
 - ▷ `test_name = CONE_BADMIME`
 - ▷ `test_status = fail`
 - ▷ `test_message = Bad MIME type for response document`
- For client applications to transmit to users:
 - ▷ `test_summary = good` (or some numeric score?)
 - ▷ `mean_availability = 0.98`
 - ▷ `latest_availability = true`

Questions

- Define some or all of the above groups?
- Attempt to standardise `test_names` between services (open or closed vocabulary)?
- Granularity of `test_names`?
- How to define `test_summary` values? (*very difficult*)

Service Provision

Who would provide validation result publication?

- Some centralised service or services needed
- Someone has to operate it/them
- Single service: Who claims/is given the authority?
- Multiple services: How do clients decide which to use?

⇒ Some rôle for the Registry

Implementation options:

- Store validation results in the Registry itself?
- Simple Validation Access Protocol?
- ValidityTAP?

Publication Option 1: Validation in Registry

Registry Metadata 1.12 section 4 has these items:

- ResourceValidationLevel
 - ▷ VOResource `vr:ValidationLevel`, RegTAP `rr.validation.val_level`
 - ▷ 0 (minimal) – 4 (expected to operate well as part of VO application)
 - ▷ Levels 3 and 4 by definition require human input
 - ▷ Wording seems to apply to resource registration, rather than resource content
- DataQuality
 - ▷ *not mentioned in VOResource or RegTAP?*
 - ▷ A (research quality) B (calibrated), C (uncalibrated), U (data quality unknown),
- ResourceValidatedBy
 - ▷ VOResource `vr:Validation/@validatedBy`,
RegTAP `rr.validation.validated_by`
 - ▷ IVOID of validating entity
 - ▷ Value is typically registry administrator (not resource owner)
 - ▷ But could be other publishing registries, operated by bulk validation services(?)

Can we use this for validation result publication?

- ▷ Kind of neat — no new standards!
- ▷ But perhaps the semantics aren't quite right, also no `DataQuality` in VOResource/RegTAP

Publication Option 2: SVAP

Define a new protocol for publishing validation results

- Standard Validation Access Protocol
- ValidationRegExt

Probably unnecessary/overkill

Publication Option 3: ValidationTAP

Could provide validation results from TAP services

- Analogous to RegTAP, ObsTAP
 - ▷ Data sits in existing or custom TAP service
 - ▷ Service declares implementation of standardised Validation DM
- Need to define data model:
 - ▷ simple DM for user consumption?
 - ▷ more detailed DM for validation aggregation/comparison?
 - with or without standardised test case vocabulary?
 - ▷ both?

Conclusions

- Writing validator software for standards:
 - It is important for standard definition and implementation quality
 - Keep doing it, and update [IvoaValidatorsSummary](#) page
 - The Ops IG will complain if you don't!
 - (but implementors should use validators *as well as* not *instead of* their own tests)
- Bulk validation:
 - Validating registered services and publishing results is useful
 - ▷ Thank you to services doing this, please keep it up
 - We could consider publishing results in a standard way
 - ▷ Useful for:
 - comparing or aggregating results from different validators
 - presentation to users in client applications
 - ▷ Data Model requirements are rather different for the two cases
 - ▷ Questions:
 - Is it worth it?
 - Is it possible to define a useful at-a-glance validity metric?
 - How best to do it? (*ValidityTAP?*)
 - Is anybody doing something like this already?
 - Discuss!