VAO Validation Database Statistics and Issues

Michael Preciado

Tom McGlynn

VAO Validation Service

- Goal: Periodically validate all registered VO Services using standard validators, collect and maintain statistics on validation issues. We do not build or maintain the validators themselves.

- Currently:

  - All SIA and Cone search services being validated

  - SSA validation started but some concerns about validator

  - OAI and registry validation being incorporated now.

  - SLAP, TAP services will be validated if a validator is provided.

- Current validators define 'validator' protocol.

- Service available at:
  http://heasarc.gsfc.nasa.gov/vo/validation

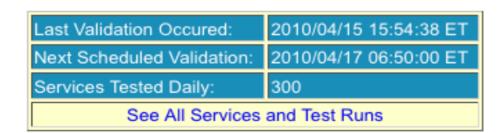- Does not test robustness of service under load.

Summary

Total Number of Services	6602
Never Validated	80
Validated	6522

Status	Total	Percent
Passed	3298	50.5
Failed	3189	48.8
Aborted	35	0.53

View Results by Center

View Results by Type

Query by identifier or string. Any identifiers containing this string will be matched.

Last Validation Occured:	2010/04/15 15:54:38 ET
Next Scheduled Validation:	2010/04/17 06:50:00 ET
Services Tested Daily:	300
See All Services and Test Runs	

Validation Results By Center

Center (*click for individual test results)	Total Services	Total Validated	not validated	passed	failed	aborted	skipped	Running List of Errors per Center
AGES	1	1	0	0	1	0	0	view list
AIASCR-STEL	2	0	2	0	0	0	2	view list
ANU	4	4	0	3	1	0	0	view list
Aladin	1	1	0	0	1	0	0	view list
Arecibo	2	2	0	0	2	0	0	view list
Astro-Wise	4	4	0	0	4	0	0	view list
Astrogrid	3	0	3	0	0	0	3	view list
Astronet	21	20	1	0	14	6	1	view list
CADC	11	11	0	0	11	0	0	view list
CDS	5226	5225	1	2832	2393	0	1	view list
CDS Simbad	1	1	0	1	0	0	0	view list
CNRS	1	0	1	0	0	0	1	view list
CXC	2	2	0	0	2	0	0	view list
CXC MIT	2	2	0	1	0	1	0	view list
Caltech	2	0	2	0	0	0	2	view list

- The table shows the number of services currently passing and failing validation.

Validation Results by Type

Center (*click for individual test results)	Total Services	Total Validated	not validated	passed	failed	aborted	skipped	Running List of Errors per Center
Cone	6335	6335	0	3229	3075	31	0	view list
SIA	187	187	0	69	114	4	0	view list
SSA	57	0	57	0	0	0	57	view list
Registry (Harvest)	15	0	15	0	0	0	15	view list
Registry (Search)	8	0	8	0	0	0	8	view list

- All Cone services have been validated. A little more than half are passing.

- All SIA services have been validated. Less than half are passing.

- SSA and Registry services have not yet been validated.

Statistics

- 6610 total services

- 3304 services are passing (50.5%) (as of May 10).

  - One year ago only 6.7% were passing.

  - VizieR cone search services dominate statistics.

- 3191 services are failing (48.8%)

- Just over 50% of Cone services are passing validation

- About 67 % of SIA services are passing validation

- 35 tests on services have aborted for unknown reasons. Connection issues were probably at fault.

- 80 services have not been validated. These include SSA and Registry services.

- 300 services are tested per day.

- A few dozen deleted services are improperly included in stats.

All SIA Errors

Current error types for center: sia:SimpleImageAccess		
Errorname	**Error Description**	**Frequency**
responseType	Apparent communication error produced an exception inside the validater:	112
comm	Apparent communication error produced an exception inside the validater:	93
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	42
4.2.4d/b	The POS_EQ_RA_MAIN column must have datatype='double'.	32
4.2.4e/b	The POS_EQ_DEC_MAIN column must have datatype='double'.	32
4.2.4i/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Format'.	21
validater	Internal Validater Error: No stylesheet configured for response type="v1.2" (Please report to validater service provider.)	16
4.2.4f/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Naxes'.	14
4.2.4h/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Scale'.	12
4.2.4x/b	The VOX:Image_FileSize column must have datatype='int'.	12
4.2.4e/a	The TABLE must contain exactly one FIELD or PARAM with ucd='POS_EQ_DEC_MAIN'.	11
4.2.4d/a	The TABLE must contain exactly one FIELD or PARAM with ucd='POS_EQ_RA_MAIN'.	11

- Communication Errors and malformed URLs appear frequently.

- POS RA Main columns with incorrect datatype value are also common.

All Cone Errors

	Current error types for center: cs:ConeSearch	
Errorname	**Error Description**	**Frequency**
2.2.1a	The TABLE must contain exactly one FIELD with ucd='ID_MAIN'.	2650
3a	Error message must appear in an INFO or PARAM with name='Error' in required location.	1548
2.2.3b	The POS_EQ_DEC_MAIN column must have datatype='double'.	965
2.2.2b	The POS_EQ_RA_MAIN column must have datatype='double'.	963
2.2.1c	The ID_MAIN column must have a string-appropriate arraysize (recommend '*').	275
2.2.1b	The ID_MAIN column must have datatype='char'.	260
2.2.2a	The TABLE must contain exactly one FIELD with ucd='POS_EQ_RA_MAIN'.	73
2.2.3a	The TABLE must contain exactly one FIELD with ucd='POS_EQ_DEC_MAIN'.	71
2.2.1d	Values in the ID_MAIN column must be unique.	56
comm	Apparent communication error produced an exception inside the validater:	53
2.1a	VOTable must contain only one RESOURCE.	31
error parsing page	cannot connect to web page or response could not be parsed	31
2.1b	The RESOURCE element must contain exactly one TABLE.	31
validater	Internal Validater Error: No stylesheet configured for response type="VOTable-1.1.xsd" (Please report to validater service provider.)	21
responseType	Apparent communication error produced an exception inside the validater:	15
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	1

- Tables missing ID_MAIN. Some have two MAIN columns.
E.g.
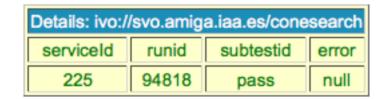<FIELD name="Tile" ucd="ID_MAIN:1" datatype="unsignedByte" width="1">
<FIELD name="ID" ucd="ID_MAIN:2" datatype="short" width="4">

- Errors need to be in right location

- DEC and MAIN columns need to have right datatype values

Validation history example:
SVO Amiga

Current error types for center: SVO Amiga										
serviceId	shortname	status	time ↑	serviceURL	runtest	type	ivoid	test_ra	test_dec	radius
225	AMIGACS	pass	2010/03/31 06:01:19	url	run	Cone	ivo://svo.amiga.iaa.es/conesearch	83.8221	-5.3911	0.3

Details: ivo://svo.amiga.iaa.es/conesearch			
serviceId	runid	subtestid	error
225	94818	pass	null

View All Tests

- Current Errors (top) and Details page for this test (bottom). Test currently passes validation.

- Click on View All Tests for history

SVO Amiga History (cont.)

serviceId	runid	status	type	time ↑
			Old Tests	
225	94818	pass	Cone	2010/03/31 06:01:19
225	88164	pass	Cone	2010/03/06 06:01:50
225	81682	pass	Cone	2010/02/13 06:01:40
225	74992	pass	Cone	2010/01/22 06:03:11
225	68391	pass	Cone	2009/12/31 08:26:28
225	61984	pass	Cone	2009/12/10 06:03:42
225	55556	pass	Cone	2009/11/17 16:52:48
225	51013	pass	Cone	2009/10/30 06:02:04
225	44640	pass	Cone	2009/10/08 11:28:57
225	39936	fail	Cone	2009/09/23 06:57:42
225	33644	fail	Cone	2009/09/02 06:57:02
225	27040	fail	Cone	2009/08/13 06:57:39
225	20417	fail	Cone	2009/07/24 06:56:36
225	14011	fail	Cone	2009/07/08 17:49:49
225	13057	fail	Cone	2009/07/06 07:00:39
225	6459	fail	Cone	2009/05/26 09:47:47
225	98	fail	Cone	2009/05/11 15:06:51

- Service was failing on 2009/09/23
- Click on the "fail" status to see the failure on that day

SVO Amiga History (cont.)

Details: ivo://svo.amiga.iaa.es/conesearch			
serviceId	runid	subtestid	error
225	39936	3a	Error message must appear in an INFO or PARAM with name='Error' in required location.

- Test shows that Service used to respond with an Error that was not properly contained within the VOTable. This was the only error for this service in 2009.

- Providers acknowledged on 10/01/2009 that the issue was fixed. Validating the service again has confirmed the fix.

Site Summaries

CDS

Current error types for center: CDS		
Errorname	Error Description	Frequency
2.2.1a	The TABLE must contain exactly one FIELD with ucd='ID_MAIN'.	2508
3a	Error message must appear in an INFO or PARAM with name='Error' in required location.	1524
2.1b	The RESOURCE element must contain exactly one TABLE.	8
2.1a	VOTable must contain only one RESOURCE.	8
2.2.1d	Values in the ID_MAIN column must be unique.	4

- A little over half of all 5226 CDS services have passed validation completely

- 2832 have passed, 2393 have failed and 1 registry service has not yet been validated

- ID_MAIN missing in some cases

- some columns have 2 ID_MAIN UCDs.

- Error message must be inside INFO element

HEASARC

Current error types for center: HEASARC		
Errorname	Error Description	Frequency
2.2.1a	The TABLE must contain exactly one FIELD with ucd='ID_MAIN'.	80
2.2.1d	Values in the ID_MAIN column must be unique.	45
2.2.1b	The ID_MAIN column must have datatype='char'.	34
2.2.1c	The ID_MAIN column must have a string-appropriate arraysize (recommend '*').	34
2.2.3a	The TABLE must contain exactly one FIELD with ucd='POS_EQ_DEC_MAIN'.	14
2.2.2a	The TABLE must contain exactly one FIELD with ucd='POS_EQ_RA_MAIN'.	14
error parsing page	cannot connect to web page or response could not be parsed	8

• 425/536 HEASARC services passing. Only Cone services failing

• ID_MAIN column missing

• ID_MAIN contains wrong datatype value and arraysize atribute is missing

   -E.g. <FIELD name="sequence_number" datatype="int" ucd="ID_MAIN">

•POS RA and DEC missing from some tables

LEDAS

Current error types for center: LEDAS		
Errorname	**Error Description**	**Frequency**
2.2.3b	The POS_EQ_DEC_MAIN column must have datatype='double'.	954
2.2.2b	The POS_EQ_RA_MAIN column must have datatype='double'.	952
2.2.1b	The ID_MAIN column must have datatype='char'.	214
2.2.1c	The ID_MAIN column must have a string-appropriate arraysize (recommend '*').	214
2.1a	VOTable must contain only one RESOURCE.	10
2.1b	The RESOURCE element must contain exactly one TABLE.	10
2.2.1d	Values in the ID_MAIN column must be unique.	4
2.2.1a	The TABLE must contain exactly one FIELD with ucd='ID_MAIN'.	2
2.2.2a	The TABLE must contain exactly one FIELD with ucd='POS_EQ_RA_MAIN'.	2

- There are no LEDAS services passing validation as of yet.

- POS EQ RA and DEC do not have the required datatype value

-E.g. <FIELD name="decl" ID="rasscals.decl" ucd="POS_EQ_DEC_MAIN" datatype="float" unit="deg"/>

- ID_MAIN has the datatype value set to "int" instead of "char"

-E.g. (ivo://uk.ac.le.star.tmpledas/ledas/ledas/rassob)

MAST

Current error types for center: Mast		
Errorname	Error Description	Frequency
cannot find image	cannot find an ImageURL	7
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	3

• Cannot find an image at a given RA,DEC. This can be fixed.

• Malformed service URL recorded in Registry. The service URL must end with either a "?" or &.

-E.g. ivo://mast.stsci/siap/hst.appp is missing the "&".

-The url in the registry is:

  http://archive.stsci.edu/siap/search.php?id=appp

NRAO

Current error types for center: NRAO		
Errorname	Error Description	Frequency
7.1a	Metadata query produced an ERROR result.	7
4.1	Legal query produced an ERROR result.	7
error parsing page	cannot connect to web page or response could not be parsed	2
3a	Error message must appear in an INFO or PARAM with name='Error' in required location.	1
2.2.1a	The TABLE must contain exactly one FIELD with ucd='ID_MAIN'.	1
comm	Apparent communication error produced an exception inside the validater:	1
2.2.1d	Values in the ID_MAIN column must be unique.	1
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	1

- When FORMAT=METADATA is used, the response is an error message. There should only be a metadata response.

- When FORMAT=ALL (a legal query) is used, the result is an error.

- ID_MAIN missing when SR=0.

-E.g.http://archive.nrao.edu/vo/cone/getCatalog archive=NVSS&RA=270&DEC=13.0&SR=0

- ID_MAIN values not unique:

-E.g.
http://archive.nrao.edu/vo/cone/getCatalog?archive=NVSS&RA=3&DEC=3&SR=1

ROE

Current error types for center: ROE		
Errorname	Error Description	Frequency
4.2.4x/b	The VOX:Image_FileSize column must have datatype='int'.	12
error parsing page	cannot connect to web page or response could not be parsed	9
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	5
7.1a	Metadata query produced an ERROR result.	1

• VOX:Image_FileSize datatype values need to be integers. Currently set to "long".

• If VOTable responses for certain coordinates and SR values are large, the response time may exceed a predefined time. Changing the radius manually may fix this issue.

•METADATA query returns an error

-E.g. http://horus.roe.ac.uk:8080/ssa/SSS_SIAP?FORMAT=METADATA

Starlight

	Details: ivo://starlight.ufsc/synthesis.BS		
serviceId	runid	subtestid	error
221	100047	mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end
221	100047	3a	Error message must appear in an INFO or PARAM with name='Error' in required location.

- Center has one service

- There should be an error message in the required location when an erroneous query is entered.

- -E.g. http://www.starlight.ufsc.br/vo/conesearch.php?RA=as3f&DEC=13.0&SR=0.25

- URL is missing a "?" at end.

IRSA

Errorname	Error Description	Frequency
	Current error types for center: IRSA	
4.2.4d/b	The POS_EQ_RA_MAIN column must have datatype='double'.	28
4.2.4e/b	The POS_EQ_DEC_MAIN column must have datatype='double'.	28
responseType	Apparent communication error produced an exception inside the validater:	28
4.2.4m/c	The VOX:WCS_CoordRefPixel column must have arraysize='*'	4
4.2.4g/c	The VOX:Image_Naxis column must have arraysize='*'	4
4.2.4n/c	The VOX:WCS_CoordRefValue column must have arraysize='*'	4
4.2.4h/c	The VOX:Image_Scale column must have arraysize='*'	4
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	3
4.2.4h/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Scale'.	2
4.2.4a/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Title'.	2
4.3a	On error, RESOURCE must include one INFO element with name="QUERY_STATUS".	2
4.2.4f/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Naxes'.	2
4.2.4i/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Format'.	2
4.2.4d/a	The TABLE must contain exactly one FIELD or PARAM with ucd='POS_EQ_RA_MAIN'.	2
4.2.1b	There must be exactly one TABLE in the RESOURCE containing the results.	2
4.2.4g/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Naxis'.	2
4.3.1a	Service should report syntax errors (such as bad position formats)	2
4.2.4v/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_AccessReference'.	2
4.2.4e/a	The TABLE must contain exactly one FIELD or PARAM with ucd='POS_EQ_DEC_MAIN'.	2
3a	Error message must appear in an INFO or PARAM with name='Error' in required location.	1
error parsing page	cannot connect to web page or response could not be parsed	1
2.2.2b	The POS_EQ_RA_MAIN column must have datatype='double'.	1
2.1a	VOTable must contain only one RESOURCE.	1
2.2.3b	The POS_EQ_DEC_MAIN column must have datatype='double'.	1
2.1b	The RESOURCE element must contain exactly one TABLE.	1

IRSA

- With FORMAT=METADATA, RA/DEC datatype should be set to "double".

- VOX* Columns currently have integer arraysizes. The value should be "*".

-E.g. http://irsa.ipac.caltech.edu/cgi-bin/2MASS/IM/nph-im_sia?type=at&ds=asky&FORMAT=METADATA

- With FORMAT=METADATA, the query results in the error:

getDouble() imutil.c:809 could not convert value "1.0,1.0" to a double

The validator sees that many UCDs are missing and reports them.

-E.g. http://irsa.ipac.caltech.edu/cgi-bin/2MASS/IM/nph-im_sia
type=ql&ds=asky&POS=180.0,60.0&SIZE=1.0,1.0&FORMAT=ALL

- With an error in the positional information, the service returns an error that is outside of the RESOURCE element.

E.g. http://irsa.ipac.caltech.edu/cgi-bin/2MASS/IM/nph-im_sia?type=ql&ds=asky&POS=as3f,13.0&SIZE=0.25

STScI

Current error types for center: STScI		
Errorname	**Error Description**	**Frequency**
comm	Apparent communication error produced an exception inside the validater:	57
2.2.2a	The TABLE must contain exactly one FIELD with ucd='POS_EQ_RA_MAIN'.	41
2.2.3a	The TABLE must contain exactly one FIELD with ucd='POS_EQ_DEC_MAIN'.	41
2.2.1a	The TABLE must contain exactly one FIELD with ucd='ID_MAIN'.	41
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	4
2.2.1c	The ID_MAIN column must have a string-appropriate arraysize (recommend '*').	3
error parsing page	cannot connect to web page or response could not be parsed	3
4.2.4a/c	The VOX:Image_Title column must have arraysize='*'	2
4.2.4s/b	The VOX:BandPass_HiLimit column must have datatype='double'.	2

- Only errors with the greatest frequency are displayed

- With bad positions, queries result in an SQL error:

   - E.g. http://archive.stsci.edu/fuse/search.php?RA=as3f&DEC=13.0&SR=0.25

- When SR=0, the response does not contain UCDs so the validator notes the missing POS EQ RA/DEC, and MAIN.

   - E.g. http://archive.stsci.edu/fuse/search.php?RA=270&DEC=13.0&SR=0

- Some responses do not contain the arraysize attribute for VOX* UCDs.

   - E.g.
http://galex.stsci.edu/gxWS/SIAP/gxSIAP.aspx?POS=180.0,60.0&SIZE=1.0,1.0&FORMAT=ALL

ESAVO

Current error types for center: ESAVO		
Errorname	Error Description	Frequency
4.1	Legal query produced an ERROR result.	3
4.2.4d/c	The POS_EQ_RA_MAIN column must be defined as a scalar--i.e. arraysize="1", arraysize="", or not provided (preferred).	2
4.2.4e/c	The POS_EQ_DEC_MAIN column must be defined as a scalar--i.e. arraysize="1", arraysize="", or not provided (preferred).	2
mal-formed URL	This service url is mal-formed. It is missing either a '?' or '&' at the end	1
comm	Apparent communication error produced an exception inside the validater:	1

• Optional parameters such as "INTERSECT" result in an error response, although the query is legal.

• With FORMAT=ALL, RA and DEC columns contain arraysize = "*". The value should be a scalar.

• With an RA/DEC error, response is a web page with an error, not a VOTable. A comm error results.

-E.g. http://isda.esac.esa.int:8080/aio/jsp/metadata.jsp?resource_class=ods&POS=as 3f,13.0&SIZE=0.25

ST-ECF

Current error types for center: ST-ECF		
Errorname	Error Description	Frequency
4.2.4m/c	The VOX:WCS_CoordRefPixel column must have arraysize='*'	3
4.2.4u/c	The VOX:Image_PixFlags column must have arraysize='*'	3
4.2.4n/c	The VOX:WCS_CoordRefValue column must have arraysize='*'	3
4.2.4f/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Naxes'.	3
4.2.4o/c	The VOX:WCS_CDMatrix column must have arraysize='*'	3
comm	Apparent communication error produced an exception inside the validater:	2

- VOX columns must have arraysize="*"

- VOX:Image_Naxes column must appear in a FIELD element

- WCS_CDMATRIX contains integer arraysize values. The arraysize should be set to "*".

- When FORMAT= METADATA, we get a communication error, specifically a code 500 error.

  - E.g. http://www.stecf.org/hst-vo/hst_sia?FORMAT=METADATA

Obs Paris/Paris Gepi

Current error types for center: Obs Paris		
Errorname	**Error Description**	**Frequency**
comm	Apparent communication error produced an exception inside the validater:	4
responseType	Apparent communication error produced an exception inside the validater:	3
2.2.1c	The ID_MAIN column must have a string-appropriate arraysize (recommend '*').	2
3a	Error message must appear in an INFO or PARAM with name='Error' in required location.	1

Current error types for center: VO Paris Gepi		
Errorname	**Error Description**	**Frequency**
4.2.4i/a	The TABLE must contain exactly one FIELD or PARAM with ucd='VOX:Image_Format'.	6
responseType	Apparent communication error produced an exception inside the validater:	2
4.2.1b	There must be exactly one TABLE in the RESOURCE containing the results.	2

• Comm errors. One or more service urls do not work (E.g. one response is a not found error)

• Arraysize attribute missing for some tables

• When FORMAT=METADATA, the VOX:Image_Format UCD and Table element are missing:

•E.g.

http://voparis-srv.obspm.fr/sia/sia_mama.php?SURVEY=POSSE&FORMAT=METADATA