Reviewing VO Data Publishers

Michael Preciado and Tom McGlynn

NASA/NAVO

Overview

- Hourly monitoring of VO data providers (selected services)
- Monthly validation of VO services
- Biannual (?) review of VO Publishers

- Activities supported by NASA/NAVO

Uptime Percentages for VO Data Providers

Site	% up last year	% up prev. year
3CR Snapshots	99.1	97.4
ARCHES	96.3	-
ARVO	97.6	-
ASDC	99.9	98.8
Astronet	84.9	77.3
BSDC	98.5	78.5
CADC	97	95.9
CDPP	98.5	98.4
CDS	99.2	98.8
CSIRO	98.9	99.2
China VO	96.1	90.3
DAME	96	-
ESA VO	98.4	98.8
ESO ORG	99.2	99.6
GAVO	98.5	87.4
GRAAL-VO	97.8	98.8
HEASARC *	97.8	98.2
HESIOD	97.7	96.7
Harvard ADS *	96.6	96.2
Harvard CXC *	99.5	99.8

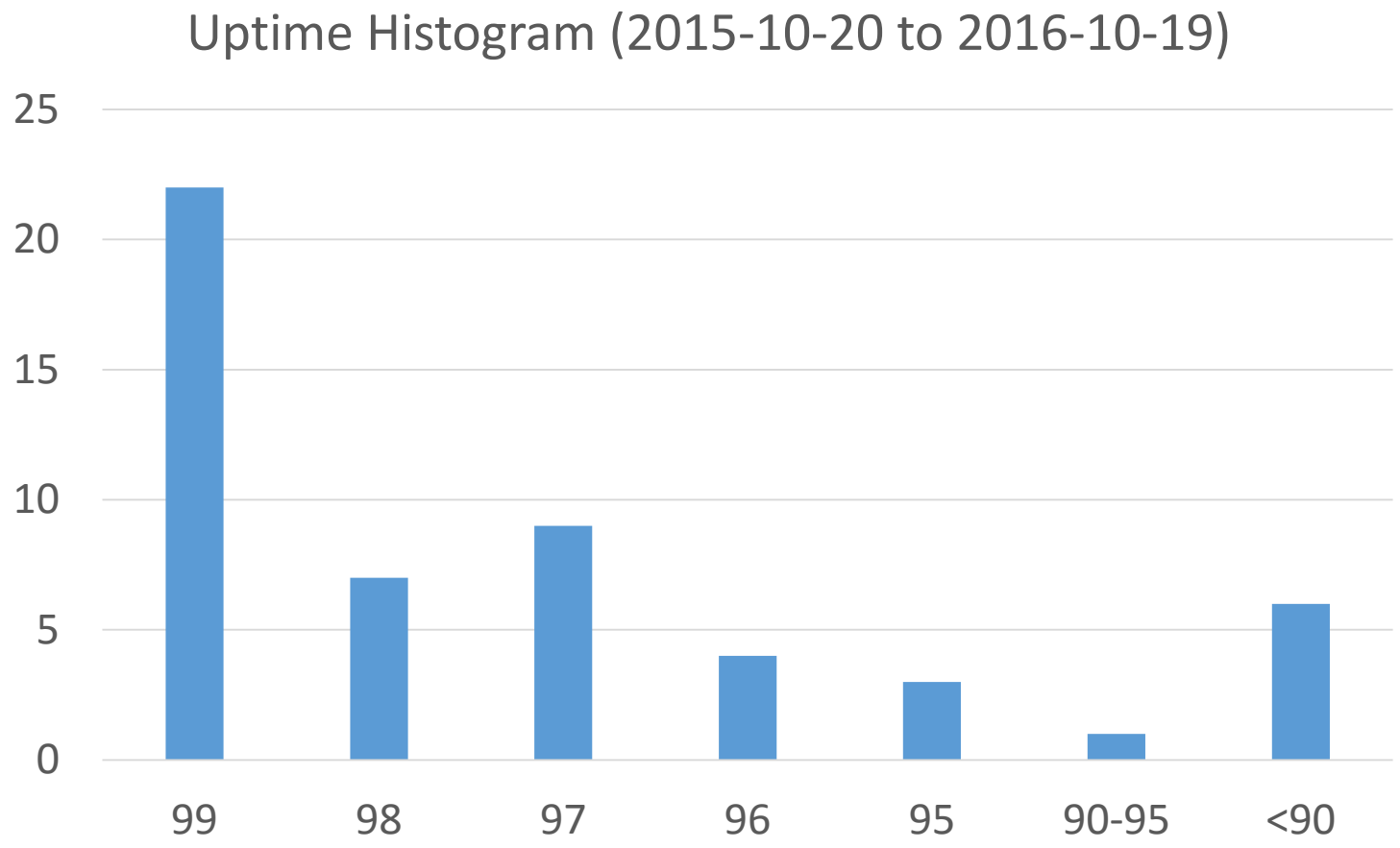
Site	% up last year	% up prev. year
Helio VO	95.1	98.9
Helio VO Mirror	81.2	99.8
IACVO	97	100
IDOC	28.8	-
INAF	97.5	98.6
IRSA *	99.7	99.5
IVOA	99.9	99.8
IVOA RofR	99.4	98.2
JVO	99.5	99.3
MAGIC	99.7	95.8
MAST *	99.6	99.7
MAST NAVO Registry	99.7	99.8
NAO	97.4	97
NCI	97.3	91.7
NED *	99.8	99.5
NEXSCI	99.9	99.8
NOAO	92.3	99.1
NOVA	89.4	91.8
NRAO	99.9	-
OVGSO	95.1	98

Site	% up last year	% up prev. year
Obs Lyon	97.6	95.8
ROE Astrogrid	87.6	97
SDSS	98.5	92.5
STECF	95	99.6
SVO	99.1	99.1
SkyBot	94.9	93.7
StarDB	99.5	98.8
Swinburne	99.9	99.1
TBL Narval	94.9	99.1
USNO	72.4	87.5
VAO Closeout	99.9	99.7
VO Paris	98.3	99.5
XAOVO	99.3	-
XCAT	99.1	98.3

21 blue (improving)
15 green (>99% in all years)
16 red (deteriorating)
4 with <99% in last year only.

http://heasarc.gsfc.nasa.gov/vo/monitor/stats.html

Number of sites versus reliability



Uptime overview

- About half of sites reaching 2 nines.
  - 99% means down 3 ½ days down per year.
- Over 90% of sites have >95% reliability
  - 95% is equivalent to being down one hour/day

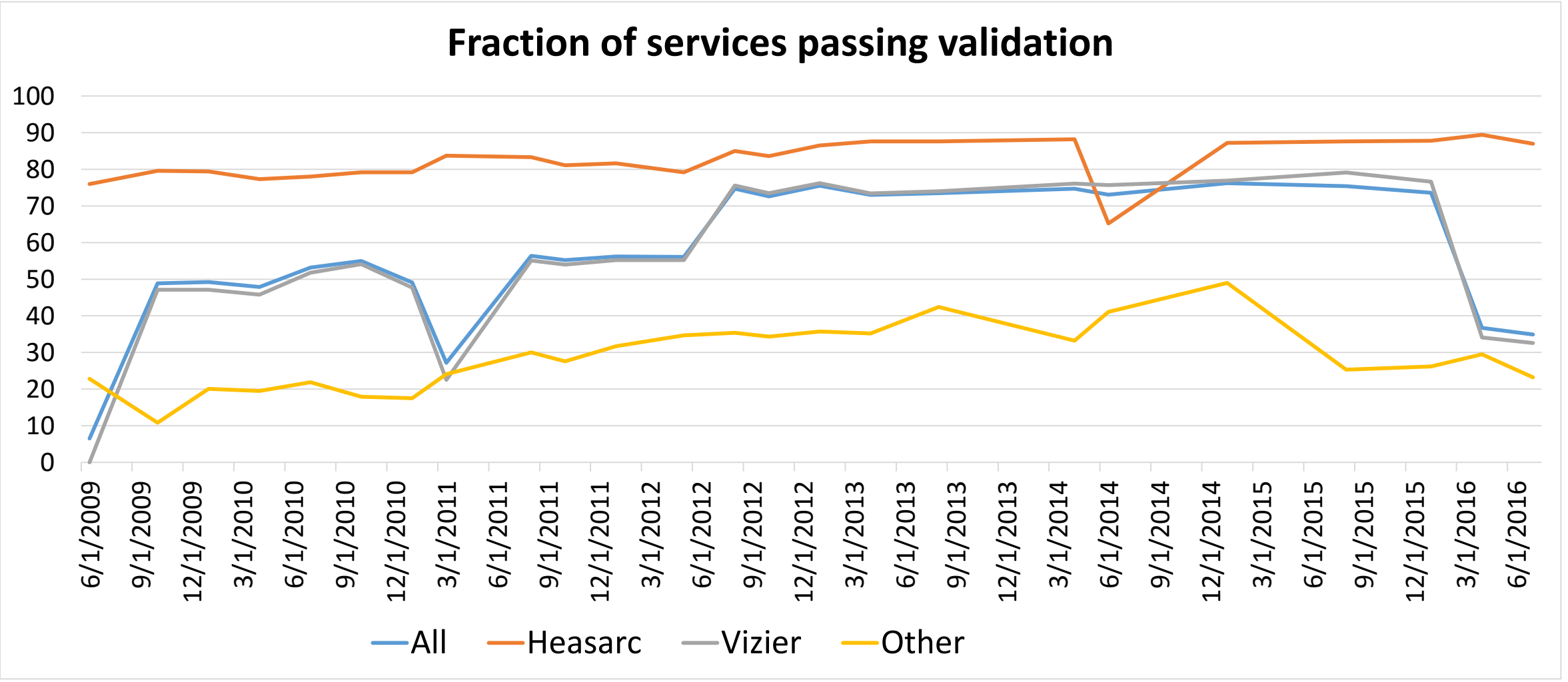
- Distribution of down times affects apparent reliability.

NAVO Validation Service

- Mostly don’t write validators:
  - NCSA Cone and SIAV1 (local tweaks)
  - TAPLint
  - VOParis SSA and SIAV2
- Supplement S[IS]A checks with data download request

- http://heasarc.gsfc.nasa.gov/vo/validation
  - Individual services
  - Service types
  - Institutions

Historical Results



Site	Validated	Passed	%Pass
3CRSnapshots	1	1	100.0
ARCHES	2	0	0.0
ARVO	1	0	0.0
ASDC	1	0	0.0
Astronet	27	0	0.0
BSDC VO	5	3	60.0
CADC	5	1	20.0
CDPP	1	0	0.0
CDS	1	0	0.0
CDS SIMBAD	2	1	50.0
CDS Vizier	16436	5502	33.5
CSIRO	7	2	28.6
China VO	4	4	100.0
DAME	1	0	0.0
ESAVO	13	5	38.5
ESO	1	1	100.0
GAVO DC	50	39	78.0
GRAAL-VO	1	0	0.0
HEASARC	909	811	89.2
HESIOD	3	2	66.7
Harvard ADS	1	1	100.0
Harvard CXC	7	5	71.4
Heidelberg	3	1	33.3
Helio Registry	2	0	0.0
Helio Registry Mirror	2	0	0.0
IA2 Inaf	36	27	75.0
IAP VOPDC	1	1	100.0
IDOC	2	0	0.0
IRSA	377	5	1.3
ISON	1	0	0.0
IVOA NET	1	0	0.0

Site	Validated	Passed	%Pass
JVO	26	2	7.7
MAGIC	1	0	0.0
MAST	47	45	95.7
MAST II	3	1	33.3
MSSL	2	0	0.0
NAO	3	3	100.0
NCI	4	1	25.0
NEA	1	0	0.0
NED	3	0	0.0
NOAO	4	0	0.0
NOVA	5	4	80.0
NRAO	1	0	0.0
OVGSO	1	0	0.0
Obs Haute Provence	1	1	100.0
Obs Lyon	2	1	50.0
ROE Astrogrid	222	163	73.4
SDSS	24	0	0.0
STScI Opo	1	0	0.0
STScI Registry	40	31	77.5
SVO CAB	37	8	21.6
SVO IFCA	1	0	0.0
StarDB	1	0	0.0
Swinburne	1	0	0.0
TBL Narval	1	0	0.0
USNO	3	0	0.0
VO Paris	23	2	8.7
XAOVO	5	2	40.0
XCAT	12	4	33.3

What happens?

- Sites generally start out with very low numbers (including HEASARC before the start of this chart).
- Once they are informed of failures most substantially improve often in discrete phases.
- Bugs may cause one or more tests (typically in edge cases) to fail and bring down existing services.
- New centers may come online, or existing centers may release new suites of services, bringing down overall numbers (see first point!)

Sites should validate:

- On initial deployment of services
- After any changes
- Regular checks

But since providers don’t always do it…

HEASARC Review of All VO Data Providers

- Detailed summary of all issues seen at one site created
  - Problems
  - Suggested solutions
- Emailed to all centers
  - Did not send out to centers with no problems – should we give gold stars at IVOA?
- About 20-25% responded usually indicating that they would work on problems.
- Sometimes catches issues in validators (e.g., too strict TAPlint constraint)

****
Center: HEASARC

++++++++++++++++++++++++++
+ Validator Error Summary +
++++++++++++++++++++++++++

(Cone)
A.Values in the ID_MAIN column must be unique.
B.The TABLE must contain exactly one FIELD with ucd='ID_MAIN'.
C.The ID_MAIN column must have datatype='char'.
D.The ID_MAIN column must have a string-appropriate arraysize (recommend '*').
E.The POS_EQ_DEC_MAIN column must have datatype='double'.
F.The POS_EQ_RA_MAIN column must have datatype='double'.
G.The TABLE must contain exactly one FIELD with ucd='POS_EQ_RA_MAIN'.
H.The TABLE must contain exactly one FIELD with ucd='POS_EQ_DEC_MAIN'.
I.The RESOURCE element must contain exactly one TABLE.

(SIAP)
A. Could not find image url
B. Fits Test Exception
…

All problems seen on any service at site.

Explanations, Examples and Suggested Solutions

…
++++++++++++++++++++++++++++++++
+Explanations and Mitigations +
++++++++++++++++++++++++++++++++
(Cone)
A. Your service:
http://heasarc.gsfc.nasa.gov/cgi-bin/vo/cone/coneGet.pl?table=a2lcscan&RA=136.9&DEC=-82&SR=1.0

...returns a VOTable where the ID Main values are not unique.

E.g.
<TD>1H0922-810</TD>
<TD>1H0922-810</TD>

The values in the ID main column must be unique.

B.-D. Your service:
http://heasarc.gsfc.nasa.gov/cgi-bin/vo/cone/coneGet.pl?table=isolog&RA=136.9&DEC=-85&SR=.06

...returns a VOTable with 2 ID_MAIN columns.

…

Is this useful?

- How frequently?
- Are we providing sufficient information?
  - Sufficient explanation of error?
  - Adequate examples?
- Are we providing too much information?
- Should we follow up with institutions?
  - ~75% did not acknowledge receipt.
  - Did we send it to right person?
  - How do we encourage feedback?
- Data from other validation services?
- Should we make reports public?