New EuroVO registry

*

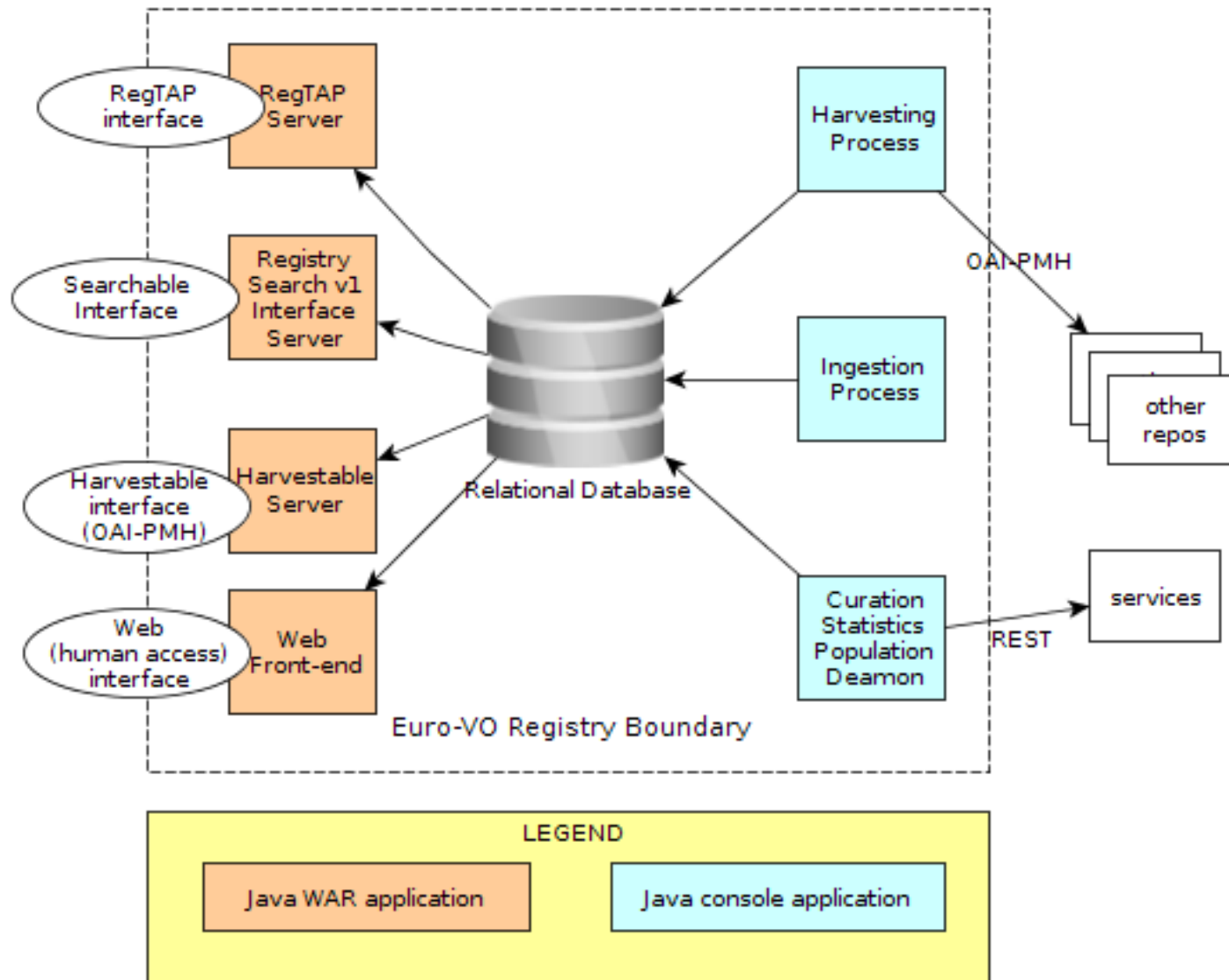## curation and service validation statistics

Menelaus Perdikeas, ESAC | Neuropublic

mperdikeas@sciops.esa.int

architecture

Curation library

- developed Java-based curation library for IVOA services (SCS, SIA and SSA); relying on `taplint` (M. Taylor) for TAP

- library does not require user to input values for service parameters (e.g. right ascension, declination or search radius) so its suitable for automated tests that lack a human to provide reasonable values.

- only requires type of service and access URL of service

Sample invocation

```
$ sval acc -t scs -u http://wfaudata.roe.ac.uk/twomass-dsa/DirectCone?DSACAT=TWOMASS%26DSATAB=twomass_psc%26


scs|http://wfaudata.roe.ac.uk/twomass-dsa/DirectCone?DSACAT=TWOMASS&DSATAB=twomass_psc&
Root name is: '{http://www.ivoa.net/xml/VOTable/v1.1}VOTABLE'
XTR :    x.ON --> true|http://wfaudata.roe.ac.uk/twomass-dsa/DirectCone?DSACAT=TWOMASS&DSATAB=twomass_psc&RA=0&DEC=0&SR=0.000|200
XTR : x.zero --> true|http://wfaudata.roe.ac.uk/twomass-dsa/DirectCone?DSACAT=TWOMASS&DSATAB=twomass_psc&RA=0&DEC=0&SR=0.000
XTR : x.zero-syntax --> false|org.xml.sax.SAXParseException|cvc-pattern-valid: Value '??' is not facet-valid with respect to pattern
'[A-Za-z0-9_.;\-]*' for type 'ucdType'.|false ( org.xml.sax.SAXParseException | line: 39 column: 66 | cvc-pattern-valid: Value '??'
is not facet-valid with respect to pattern '[A-Za-z0-9_.;\-]*' for type 'ucdType'. )|11169
XTR : x.some --> true|http://wfaudata.roe.ac.uk/twomass-dsa/DirectCone?DSACAT=TWOMASS&DSATAB=twomass_psc&RA=0&DEC=0&SR=0.176|9216
XTR : x.some-syntax --> false|org.xml.sax.SAXParseException|cvc-pattern-valid: Value '??' is not facet-valid with respect to pattern
'[A-Za-z0-9_.;\-]*' for type 'ucdType'.|false ( org.xml.sax.SAXParseException | line: 39 column: 66 | cvc-pattern-valid: Value '??'
is not facet-valid with respect to pattern '[A-Za-z0-9_.;\-]*' for type 'ucdType'. )|194812
REQ :    2.1 --> true|http://wfaudata.roe.ac.uk/twomass-dsa/DirectCone?DSACAT=TWOMASS&DSATAB=twomass_psc&RA=0&DEC=0&SR=0.000|200
REQ : 2.2a-i --> true
REQ : 2.2a-ii --> true
REQ :    2.2b --> false|org.xml.sax.SAXParseException|cvc-pattern-valid: Value '??' is not facet-valid with respect to pattern '[A-
Za-z0-9_.;\-]*' for type 'ucdType'.|http://wfaudata.roe.ac.uk/twomass-dsa/DirectCone?
DSACAT=TWOMASS&DSATAB=twomass_psc&RA=0&DEC=0&SR=0.176|false ( org.xml.sax.SAXParseException | line: 39 column: 66 | cvc-pattern-
valid: Value '??' is not facet-valid with respect to pattern '[A-Za-z0-9_.;\-]*' for type 'ucdType'. )|194812
REQ :    2.2c --> true
REQ :    2.2d --> true
REQ :    2.2e --> true
REQ :    2.2f --> true
REQ :    2.3a --> true
REQ :    2.3b --> true
REQ :    2.3c --> true
REQ :    2.3d --> true
REC : rec-2.2a --> true
REC : rec-2.2b --> true
REC : rec-2.3a --> true
REC : rec-2.3b --> true
```

This XML file does not appear to have any style information associated with it. The document tree is shown below.
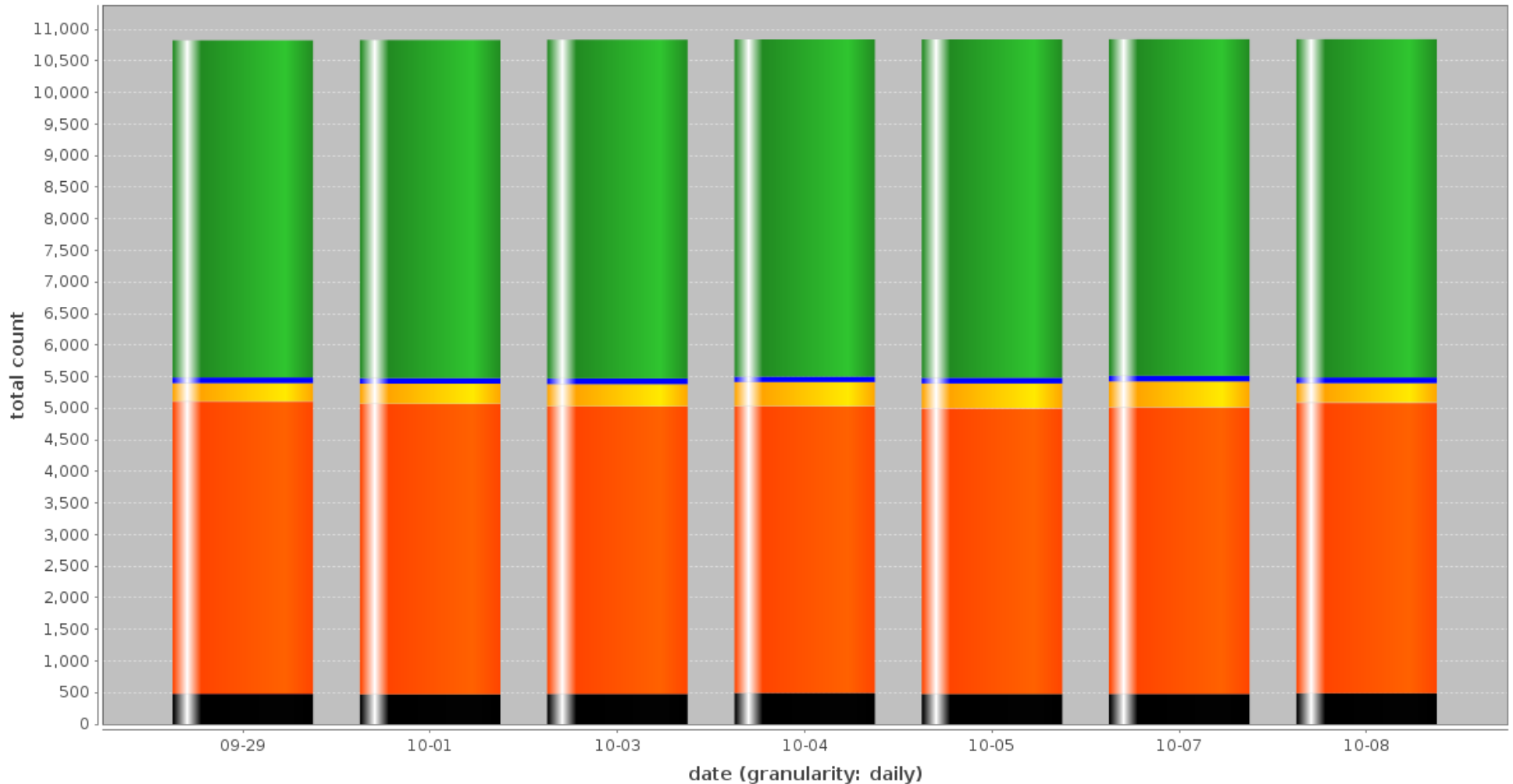
```xml
<VOTABLE xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.ivoa.net/xml/VOTable/v1.1" version="1.1"
 xsi:schemaLocation="http://www.ivoa.net/xml/VOTable/v1.1 http://www.ivoa.net/xml/VOTable/v1.1">
<!--

    ! VOTable written by STIL version 2.8 (uk.ac.starlink.votable.VOTableWriter)
    ! at 2014-05-19T11:37:40
    !
 -->
<RESOURCE>
<TABLE>
<FIELD datatype="double" name="cx" ucd="pos.distance;pos.cartesian.x">
  <DESCRIPTION>unit vector of spherical co-ordinate</DESCRIPTION>
  </FIELD>
<FIELD datatype="double" name="cy" ucd="pos.distance;pos.cartesian.y">
  <DESCRIPTION>unit vector of spherical co-ordinate</DESCRIPTION>
  </FIELD>
<FIELD datatype="double" name="cz" ucd="pos.distance;pos.cartesian.z">
  <DESCRIPTION>unit vector of spherical co-ordinate</DESCRIPTION>
  </FIELD>
<FIELD datatype="long" name="htmID" ucd="pos.eq">
  <DESCRIPTION>HTM index, 20 digits, for co-ordinate</DESCRIPTION>
  </FIELD>
<FIELD datatype="double" name="ra" ucd="POS_EQ_RA_MAIN" unit="deg">
  <DESCRIPTION>J2000 right ascension with respect to the ICRS</DESCRIPTION>
  </FIELD>
<FIELD datatype="double" name="dec" ucd="POS_EQ_DEC_MAIN" unit="deg">
  <DESCRIPTION>J2000 declination with respect to the ICRS</DESCRIPTION>
  </FIELD>
<FIELD datatype="float" name="err_maj" ucd="phys.angSize;src" unit="arcsec">
  <DESCRIPTION>
    Semi-major axis length of the one sigma position uncertainty ellipse
  </DESCRIPTION>
  </FIELD>
<FIELD datatype="float" name="err_min" ucd="phys.angSize;src" unit="arcsec">
  <DESCRIPTION>
    Semi-minor axis length of the one sigma position uncertainty ellipse
  </DESCRIPTION>
  </FIELD>
<FIELD datatype="short" name="err_ang" ucd="pos.posAng" unit="deg">
  <DESCRIPTION>
    Position angle on the sky of the semi-major axis of the position uncertainty ellipse (East of North)
  </DESCRIPTION>
  </FIELD>
<FIELD arraysize="*" datatype="char" name="designation" ucd="??">
  <DESCRIPTION>
    Sexagesimal, equatorial position-based source name in the form: hhmmssss+ddmmsss[ABC...].
  </DESCRIPTION>
  </FIELD>
<FIELD datatype="float" name="j_m" ucd="phot.flux" unit="mag">
  <DESCRIPTION>Default J-band magnitude</DESCRIPTION>
  </FIELD>
```
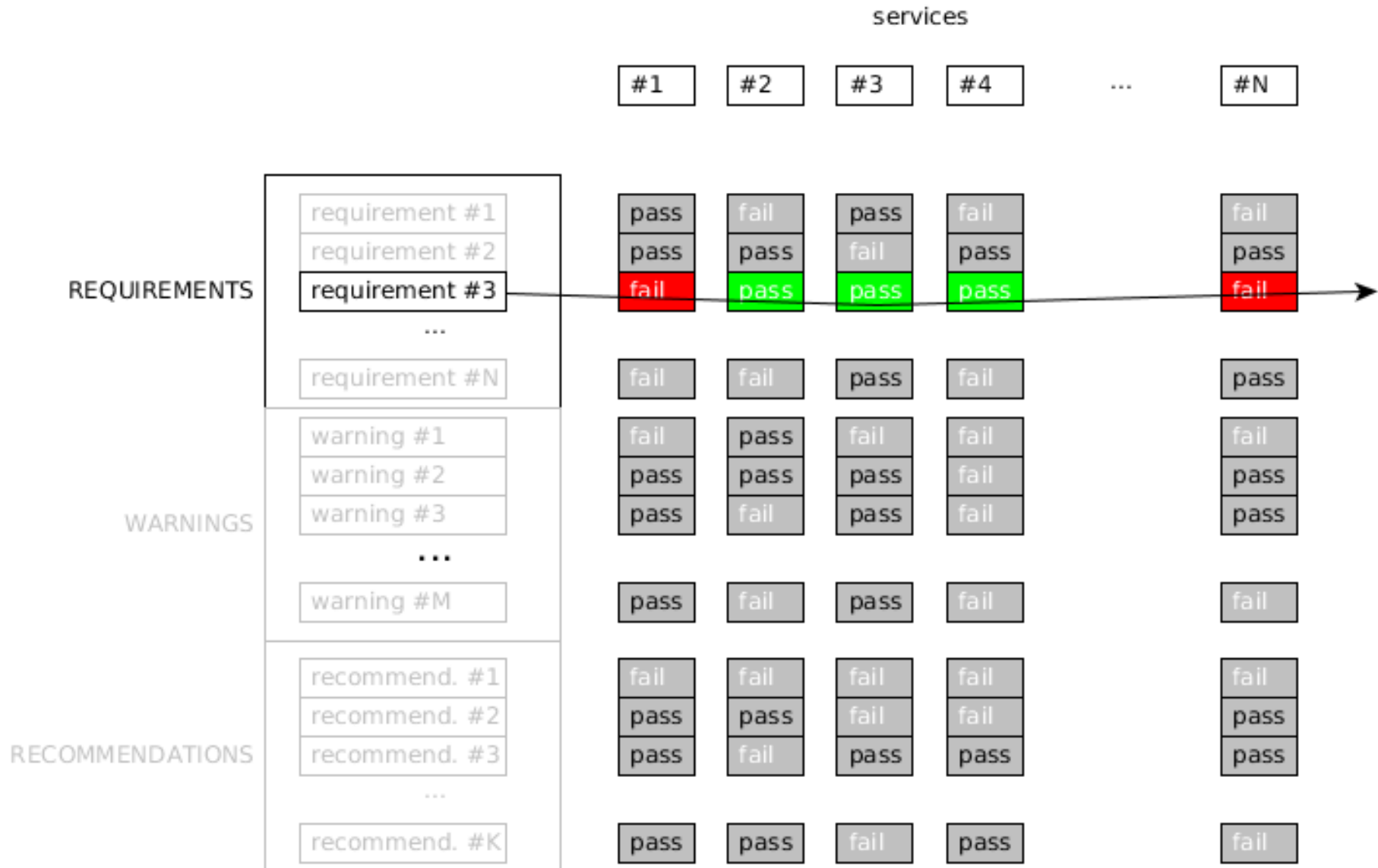
from a criteria vector to service compliance classes

Sample validation statistics report

all repos, SCS compliance category distribution over time
month 10-2014 (Σ#=75857) (loops:1→7 | 2014-09-29→2014-10-08)
total @ earliest timepoint with actual data: 10824, @ latest: 10843

- ■ E:no criterion succeeds ■ D:all requirements fail but some other criterion (of lesser importance) succeeds
- ■ C--:at least three (3) requirements fails, but at least one (1) requirement equally succeeds ■ C-:at least two (2) requirements fails, but at least one (1) requirement equally succeeds
- ■ C:at least one (1) requirement fails, but at least one (1) requirement equally succeeds ■ B:no requirement fails, but at least one warning fails
- ■ A:no requirement or warning fails, but at least one recommendation fails ■ A+:no requirement, warning or recommendation fails ■ A++:no criterion fails

Criteria-focused reports

Registries anomalies report

Consistently Failing Harvests

Harvest operations that have not succeeded in the last 30 days.
Report run time: 2014-10-10T00-20-04.

IVOID	Metadata	Operation
ivo://archive.stsci.edu/nvoregistry	oai_dc	FH
ivo://archive.stsci.edu/nvoregistry	oai_dc	IH
ivo://astronet.ru/cas/registry	oai_dc	FH
ivo://astronet.ru/cas/registry	oai_dc	IH
ivo://cadc.nrc.ca/registry	ivo_vor	FH
ivo://cadc.nrc.ca/registry	ivo_vor	IH
ivo://cadc.nrc.ca/registry	oai_dc	FH
ivo://cadc.nrc.ca/registry	oai_dc	IH
ivo://CDS.VizieR/registry	oai_dc	FH
ivo://CDS.VizieR/registry	oai_dc	IH
ivo://jvo/publishingregistry	oai_dc	FH
ivo://jvo/publishingregistry	oai_dc	IH
ivo://nasa.heasarc/registry	oai_dc	FH
ivo://nasa.heasarc/registry	oai_dc	IH
ivo://wfau.roe.ac.uk/org.astrogrid.registry.RegistryService	ivo_vor	FH
ivo://wfau.roe.ac.uk/org.astrogrid.registry.RegistryService	ivo_vor	IH
ivo://wfau.roe.ac.uk/org.astrogrid.registry.RegistryService	oai_dc	FH
ivo://wfau.roe.ac.uk/org.astrogrid.registry.RegistryService	oai_dc	IH

example reason for failed harvest

Cell Form (Editable) - "stacktrace"

Text | Hex Viewer

Search for: 421R

Editor (automatic word wrap)

```
1738 <oaidc:dc xmlns="http://purl.org/dc/elements/1.1/" xmlns:oaidc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001
1739     <title>Wavelengths and transition probabilities in FeIII (Ekberg 1993)</title>
1740     <creator>EKBERG J. O.</creator>
1741     <publisher>CDS</publisher>
1742     <date>1997-12-09T18:10:21</date>
1743     <identifier>http://cdsarc.u-strasbg.fr/cgi-bin/Cat?J/A+AS/101/1</identifier>
1744     <language>en-us</language>
1745     <relation>http://cdsweb.u-strasbg.fr</relation>
1746     <rights>unrestricted</rights>
1747 </oaidc:dc>
1748 </metadata>
1749 </record>
1750 <record>
1751 <header>
1752 <identifier></identifier>
1753 <datestamp>Z</datestamp>
1754 <setSpec>J_A_AS</setSpec>
1755 <setSpec>ivo_managed</setSpec>
1756 </header>
1757 <metadata>
1758 <oaidc:dc xmlns="http://purl.org/dc/elements/1.1/" xmlns:oaidc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001
1759     <title>uvby-beta photometry of GP And (Rodriguez+ 1993)</title>
1760     <creator>RODRIGUEZ E., ROLLAND A., LOPEZ DE COCA P. Astron. Astrophys. Suppl. Ser. 101, 421 (1993) =1993A&AS..101..421R</creator>
1761     <publisher>CDS</publisher>
1762     <date>1997-12-09T18:10:41</date>
1763     <identifier>http://cdsarc.u-strasbg.fr/cgi-bin/Cat?J/A+AS/101/421</identifier>
1764     <language>en-us</language>
1765     <relation>http://cdsweb.u-strasbg.fr</relation>
1766     <rights>unrestricted</rights>
```

stacktrace varchar (2147483647)
Not NULL
Text, 89,011 Bytes, text/plain

Close

wrong-set records

- harvesting for the `ivo_managed`
- records whose IVOIDs do not belong to a declared `managedAuthority`

▓▓▓▓▓ :: ivo_vor :: FH

Harvested at: **2014-10-09** *, total number of wrong-set records:* **492**

Wrong set records found when harvesting registry with IVOID **ivo://a▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓** for the **ivo_vor** metadata format, in the **FH** operation # **12**.

At the OAI-PMH interface:
http://▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓?verb=Identify
... the registry declared the following 38 managed authorities: sdss.fnal, nrao, fr.oamp.lam, nrao.archive, archive.noao.edu, voevent.noao.edu, svo.iaa, lca, stecf.euro-vo, obs-lyon.fr, opo.stsci, obs-hp.fr, giraffe, cvo.naoc, iap.fr, archive.stsci.edu, arecibo.cornell, rvo.ru, nasa.gsfc.nssdc, gcp, badc.naoc.cn, ia2.inaf, nvo.noao.edu, msc.koa, anusf.anu.au, svo, gov.nasa.gsfc.nssdc, mast.stsci, ads.harvard.edu, mast.stsci.edu, nao.ua, sdss.jhu, obelix.lambrate.inaf.it, nvo.jhu, astro-wise, KeckObs, svo.laeff, basebe.obspm.fr.
The harvest contained 492 records (out of a total of 792) that did not belong to any of the above. Below are 5 such records:

- ivo://uk.ac.le.star.tmpledas/ledas/ledas/swiftbalog
- ivo://uk.ac.le.star.tmpledas/ledas/ledas/rasscns3
- ivo://uk.ac.le.star.tmpledas/ledas/ledas/xrbcat
- ivo://uk.ac.le.star.tmpledas/ledas/ledas/wgacat95
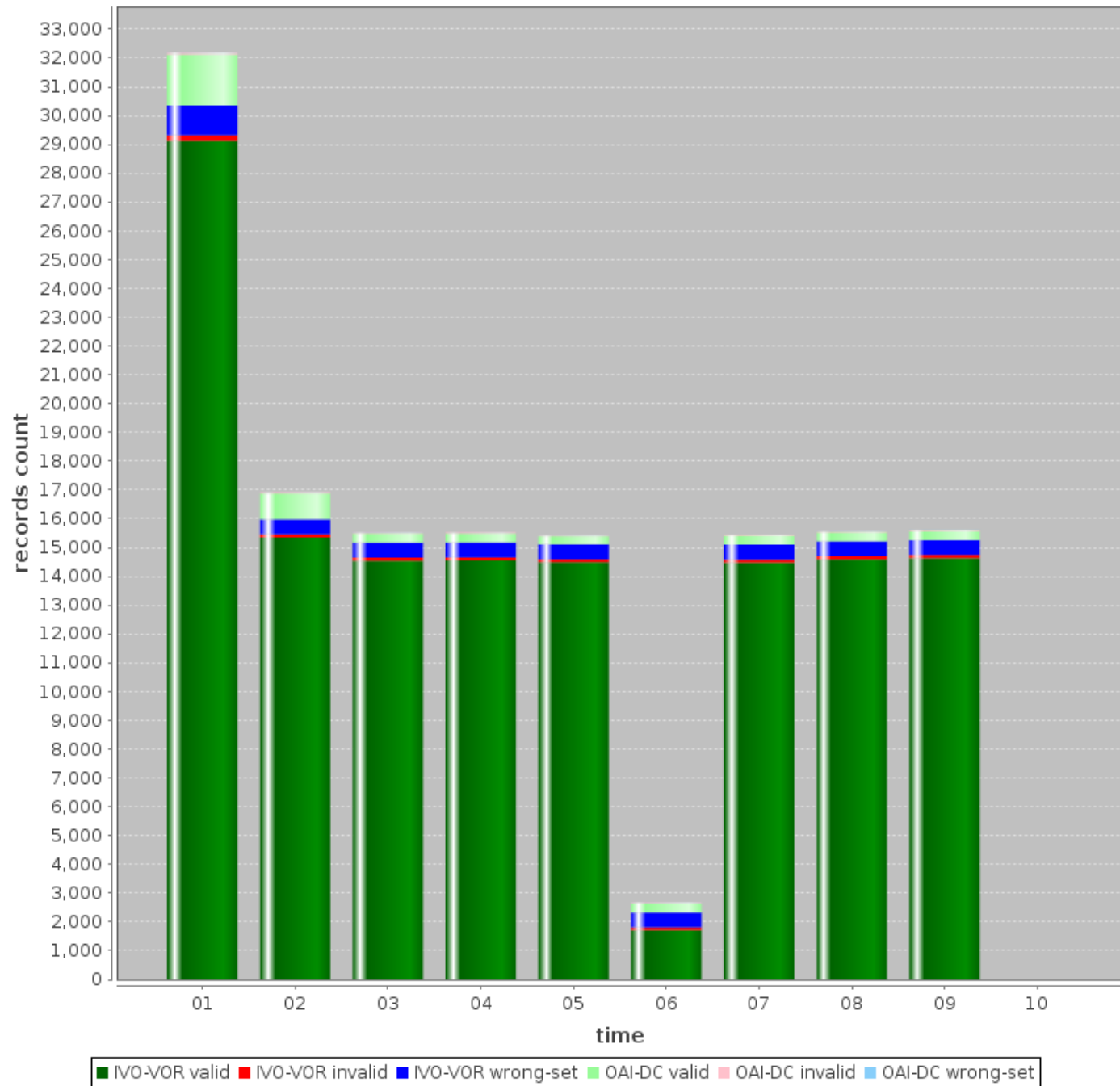- ivo://uk.ac.le.star.tmpledas/ledas/ledas/ipc

The below 1 contact detail elements are declared for said registry (in the registry resource):

- **name:** ▓▓▓▓▓▓▓▓▓▓
- **address:**
- **email:** ▓▓▓▓▓▓▓▓▓▓
- **phone:**

XSD-invalid records

Harvested records counts by validity (month 10-2014)
Granularity: day

IVO-VOR valid ■ IVO-VOR invalid ■ IVO-VOR wrong-set ■ OAI-DC valid ■ OAI-DC invalid ■ OAI-DC wrong-set

EuroVO remaining plan

- 2014-Q4: Web (human) registry interface and testing

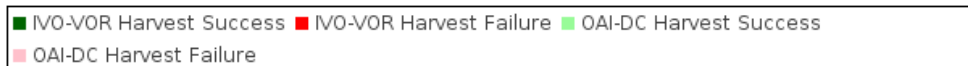- 2015-Q1: acceptance testing period and phase-out of existing Euro-VO registry
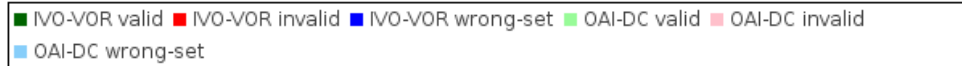
Back-up slides
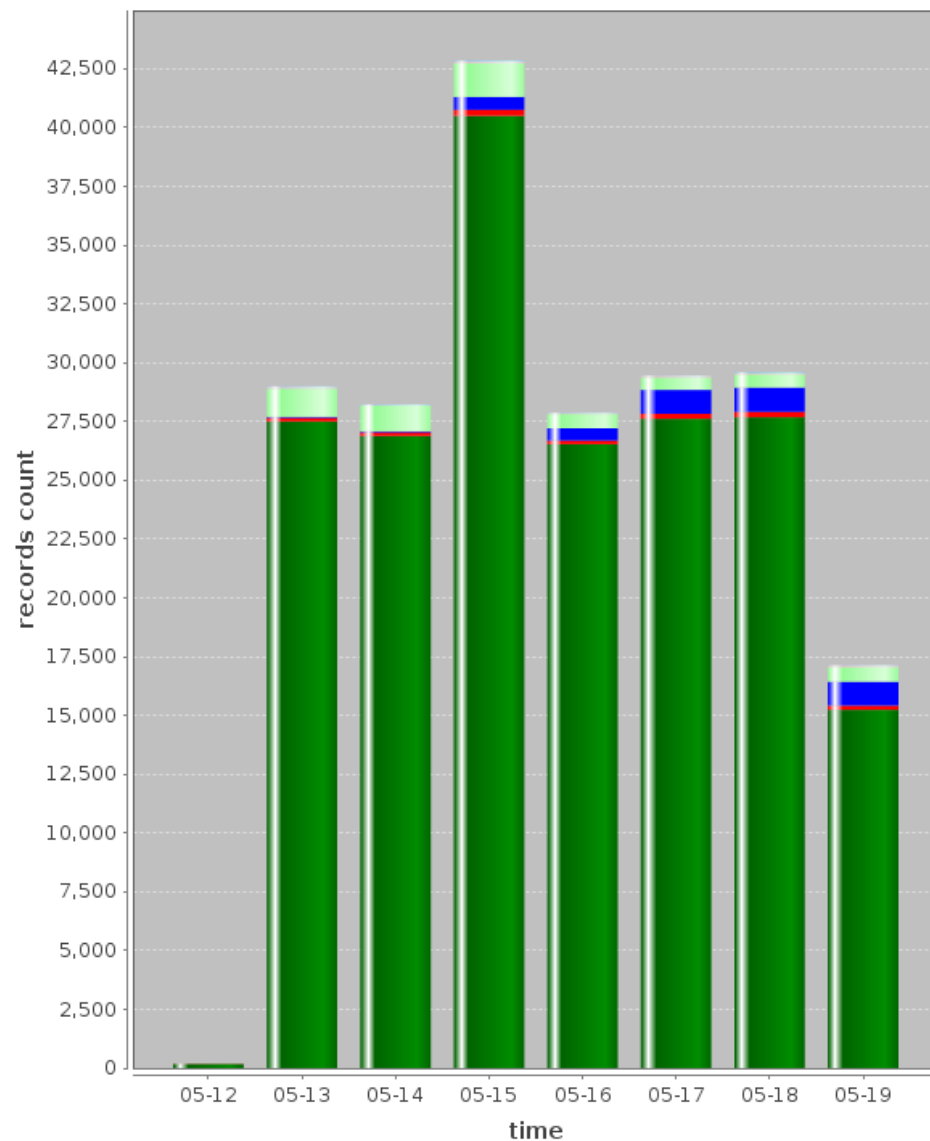
sample harvest operations reports

Harvest counts by prefix (month 5-2014)
Granularity: day

Harvested records counts by validity (month 5-2014)
Granularity: day

OAI-PMH handling

HTTP Response

HTTP Response Body Content

OAI-PMH envelope

OAI-PMH payload

IVOA Resource

... have separate libraries and
modules dealing with each envelope

LEGEND

yellow - governed by HTTP RFC

blue - governed by OAI-PMH specification

white - governed by IVOA specification