



EuroVO-AIDA

Euro-VO Astronomical Infrastructure for Data Access

D3.2

-

Registry curation tools

Final Version

Grant agreement no: 212104

Combination of Collaborative Projects & Coordination and Support Actions



DOCUMENT INFORMATION

Project

Project acronym: EuroVO-AIDA
Project full title: Euro-VO Astronomical Infrastructure for Data Access
Grant agreement no.: 212104
Funding scheme: Combination of Collaborative Projects & Coordination and Support Actions
Project start date: 01/02/2008
Project duration: 30 months
Call topic: INFRA-2007-1.2.1 Scientific Digital Repositories
Project web sites: <http://www.euro-vo.org/pub/general/intro.html>
<http://cds.u-strasbg.fr/twikiAIDA/bin/view/EuroVOAIDA/WebHome>

Document

Deliverable number: D3.2
Deliverable title: Registry curation tools
Due date of deliverable: November 2009 (as indicated earlier on in 2009)
Actual submission date: 02 December 2009
Authors: ESA
Work Package no.: WP3-SA1
Work Package title: Service Activities in support of deployment of IVOA protocols and standards
Work Package leader: ESA
Lead beneficiary: ESA
Dissemination level: PU
Nature: Other
No of pages (incl. cover): 13

TABLE OF CONTENT

1. GENERAL WORKPLAN.....	4
2. CURATION TOOL FOR THE REGISTRY RESOURCES	6
3. CURATION OF THE SERVICE INTERFACE : DAL VALIDATERS.....	8
ACRONYM LIST	13

1. GENERAL WORKPLAN

Earlier on in 2009, taking into account the discussion within the IVOA about Registry Curation tools, the work plan for this deliverable had been re-organized as follow:

TASK	DATE	COMMENTS
Inputs gathering	EuroVO-AIDA Technology Forum 16 - 19 March 2009	Gather possible inputs in relation to Curation tools from the European community
Discussion and crosscheck with IVOA	IVOA Interoperability meeting 25 - 29 May 2009	Fill the IVOA community in with inputs. Discuss applicability of needs within the IVOA community and gather requirements
Basic design of tool capabilities	June 2009	Make a top level design of possible curation tool with the possibilities available at the time
Tool presentation	IVOA Interoperability meeting 9 - 13 November 2009	Show the tool implementation at the interoperability meeting

As planned, specific report and discussion about Registry Curation tools took place at the EuroVO-AIDA Technology Forum (16-18 March 2009), as reflected in: http://cds.u-strasbg.fr/twikiAIDA/pub/EuroVOAIDA/ThirdTechnologyForum/WP3_-_Registry-Curation.pdf

This has been the occasion to discuss within EuroVO-AIDA about the plans for Registry Curation tools, and to prepare a dedicated session on this subject at the IVOA Interoperability Meeting (28 May 2009, Strasbourg): http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/InterOpMay2009ResReg#Reg_3_Curation_Practices_and_Too

The subject was introduced by Christophe Arviset, recalling the work that had been done in the past within the IVOA context and the specific plans within EuroVO-AIDA WP3.

There were presentations from various IVOA partners (US-NVO, Canadian-CVO, EuroVO), which showed some existing curation tools effort and which led to discuss collaboration and determine altogether how to move forward on this matter.

The conclusion from this discussion was to focus the effort on two areas:

1. Curation of the Registry Resource
2. Curation of the Service Interface

We commonly agreed not to address the Curation of the astronomical data itself, as this is the responsibility of the Data Centres themselves who know better the data and therefore are more able to determine their quality and if they should be published through the VObs.

There was quite some discussion about the validity of the Curation Tools, which then led to the conclusion that we needed to find a process to ensure that these Curation Tools can be “validated” by the IVOA community so they are trustful for the data centres.

On one hand, within EuroVO-AIDA WP3, we developed specific curation tools for the Registry Resource and we deployed them on the Euro-VO Registry (<http://registry.euro-vo.org/index.jsp>) on 06 November 2009.

On the other hand, within EuroVO-AIDA WP3 and in collaboration with NVO, we developed Curation Tools for the Service Interface called *DALValidators* and we deployed them on the Euro-VO Registry on 06 November 2009.

As planned both efforts have been presented at the IVOA Interoperability Meeting in Garching, Germany (9-12 November 2009) in the Registry Session at: <http://www.ivoa.net/internal/IVOA/InterOpNov2009Reg/20091110RWG-Curation.pdf>

This has been very well received by the IVOA community, who showed a clear interest on their existence and appreciated the efforts done to build these as IVOA common tools, including a path for their “validation” and future collaborative improvements, which will lead to their wider acceptance in the IVOA community.

More details of both curation tools are given in the following sections.

This deliverable is then considered **complete**.

2. CURATION TOOL FOR THE REGISTRY RESOURCES

To ensure the quality of the Registry Resources entered into the Euro-VO Registry, some checks need to be performed when service provider insert or update new resources, mainly the compliancy with the corresponding resource XML schema and specific checks on the various inputs parameters.

This has been implemented through the Euro-VO Registry HTML Forms (<http://registry.euro-vo.org/insert.jsp>), and it includes:

- checking the compulsory and optional metadata
 - Title, Identifier, Publisher, Description, etc.
- checking the correct XML schema data types
 - Identifier, Date, Reference URL, etc.
- Regular expression restrictions
 - Email, Identifier, all ivo-ids, etc.
- Other specific restrictions
 - Short Name, etc ...

Such checks can be seen in the following figure:

The screenshot displays a web form titled "Resource Type : Resource". The form is divided into several sections: "Resource Curation" and "Dates".

Resource Curation Section:

- Title:** A text input field with a red error message: "Title is a required 'String' value."
- Identifier:** A text input field with a red error message: "Identifier is a required 'Identifier URI' value." Above the field, the text "not/an/IVOA identifier" is displayed.
- Short Name:** A text input field.
- Publisher:** A text input field with a red error message: "Publisher is a required 'String' value."
- Publisher ivo-id:** A text input field.

Dates Section:

- Buttons: "Insert after selected" and "Delete selected".
- Date:** A text input field with a red error message: "Date is a required 'Date' value." Above the field, the text "not a date" is displayed.
- Date Role:** A text input field.
- Email:** A text input field with a red error message: "Email is an optional 'Email' value." Above the field, the text "not an email" is displayed.

Once the HTML Form has been filled in, some final validation steps are performed prior to the final insertion of the resource in the Euro-VO Registry, in particular:

- Correct Identifier for Authority Resources
 - Format and not already existent in IVOA
- Correct Identifier for all Other Resources
 - Format and locally controlled AuthorityID
- XML Schema validation against XSDs
- Final database metadata validation

3. CURATION OF THE SERVICE INTERFACE : DAL VALIDATORS

The second area of Registry Curation Tool relates to the curation of the Service Interface, i.e. to determine if the Service is compliant with the IVOA service specifications.

This is mainly concentrated on IVOA most popular and widespread DAL (Data Access Layer) services such as:

- SCS : Simple Cone Search
- SIAP : Simple Image Access Protocol
- SSAP : Simple Spectra Access Protocol
- SLAP : Simple Line Access Protocol

Based on the existing efforts in that context within IVOA, we initiated a specific development collaboration with NVO to extend their DALValidators tool with the following work plan:

1. Share knowledge and expertise about the DALValidators software so it can be extended to more IVOA DAL services specifications
2. Make the DALValidators web pages modular so it can be deployed at various Registries in an easy way with project associated logo and with clear NVO, Euro-VO and IVOA reference
3. Fix minor problems and improve the SCS and SIAP DALValidators
4. Develop the SSAP and SLAP DALValidators
5. Write associated documentation on SCS, SIAP to SSAP and SLAP DALValidators on the IVOA Registry wiki pages
6. Determine an IVOA wide "review/validation" process for the DALValidators to be accepted and improved by the IVOA community, being open to future collaboration

Steps 1 to 4 have been performed with the resulting software being available in a common project and code repository available at: <http://trac.us-vo.org/project/nvo/wiki/DALValidator>

As such the software is open for further collaboration and improvements from more IVOA partners. In fact, it is interesting to note that Canadian-VO project already indicated that they would be ready to participate by adding the IVOA TAP (Table Access Protocol) to the DALValidators.

The DALValidators software can be deployed at various registries through:

- a web service with an HTML result page
- a direct HTML check page

Both pages have a common look and feel for the common checking / resulting parts, but the header can be adapted to the specific project registry (ie NVO, Euro-VO) and the footer makes clear reference to the common NVO, Euro-VO effort in the context of the IVOA.

Examples on how the DALValidators have been deployed on the Euro-VO Registry are given in the figures below.

Validating your resources before you insert them in the Euro-VO Registry

<http://registry.euro-vo.org/validate.jsp>

last updated: 04-Nov-2009

co-funded project



Developed by ESAVO in the context of the EURO-VO AIDA project using the DAL Validator software source code from NVO, which was developed with the support of the National Science Foundation under Cooperative Agreement AST0112449 with The Johns Hopkins University.



Result of a SCS Validation

<http://registry.euro-vo.org:8080/dalvalidate/ConeSearchValidator?endpoint=http%3A%2F%2Fheasarc.gsfc.nasa.gov%2Fcgi-bin%2Fvo%2Fcone%2FconeGet.pl%3Ftable%3Da2led%26&RA=180.0&DEC=60.0&SR=1.0&format=html&show=fail&show=warn&show=rec&op=Validate>



Cone Search Verification Results

Base URL of Service: <http://heasarc.gsfc.nasa.gov/cgi-bin/vo/cone/coneGet.pl?table=a2led&>

The following 3 test queries were sent to the service:

- user: RA=180.0&DEC=60.0&SR=1.0
- user input error: RA=as3f&DEC=13.0&SR=0.25
- metadata: RA=270&DEC=13.0&SR=0

Test Query Name: user

Description: user-provided cone

Type: normal query

Arguments: RA=180.0&DEC=60.0&SR=1.0

URL: <http://heasarc.gsfc.nasa.gov/cgi-bin/vo/cone/coneGet.pl?table=a2led&RA=180.0&DEC=60.0&SR=1.0>

VOTable version returned: 1.0 (dtd)

Number of records returned: 0

Compliance Errors

None found.

Warnings

None found.

Recommendations

Below are a list of suggested changes that should be made to the service response to make it more consistent with the standard and common usage and, thus, easier for clients to use. These changes, however, are not required for compliance.

- 2b Recommend specifying datatype, ucd, and DESCRIPTION for FIELD with name='unique_id'
- 2b Recommend specifying datatype, ucd, and DESCRIPTION for FIELD with name='othername'
- 2b Recommend specifying datatype, ucd, and DESCRIPTION for FIELD with name='Search_Offset'
- 2.3a Recommend including DESCRIPTION under VOTABLE indicating catalog name and that it is a Cone Search response.

Test Query Name: user input error

Description: test handling of an erroneous query.

Type: error handling

Arguments: RA=as3f&DEC=13.0&SR=0.25

URL: <http://heasarc.gsfc.nasa.gov/cgi-bin/vo/cone/coneGet.pl?table=a2led&RA=as3f&DEC=13.0&SR=0.25>

VOTable version returned: (dtd)

Number of records returned: 0

Compliance Errors

None found.

Warnings

None found.

Recommendations

No additional recommendations.

Test Query Name: metadata

Description: test response to a metadata query where SR=0.

Type: metadata query (SR=0)

Arguments: RA=270&DEC=13.0&SR=0

URL: <http://heasarc.gsfc.nasa.gov/cgi-bin/vo/cone/coneGet.pl?table=a2led&RA=270&DEC=13.0&SR=0>

VOTable version returned: 1.0 (dtd)

Number of records returned: 0

Compliance Errors

None found.

Warnings

None found.

Recommendations

Below are a list of suggested changes that should be made to the service response to make it more consistent with the standard and common usage and, thus, easier for clients to use. These changes, however, are not required for compliance.

- 2b Recommend specifying datatype, ucd, and DESCRIPTION for FIELD with name='unique_id'
- 2b Recommend specifying datatype, ucd, and DESCRIPTION for FIELD with name='othername'
- 2b Recommend specifying datatype, ucd, and DESCRIPTION for FIELD with name='Search_Offset'
- 2.3a Recommend including DESCRIPTION under VOTABLE indicating catalog name and that it is a Cone Search response.

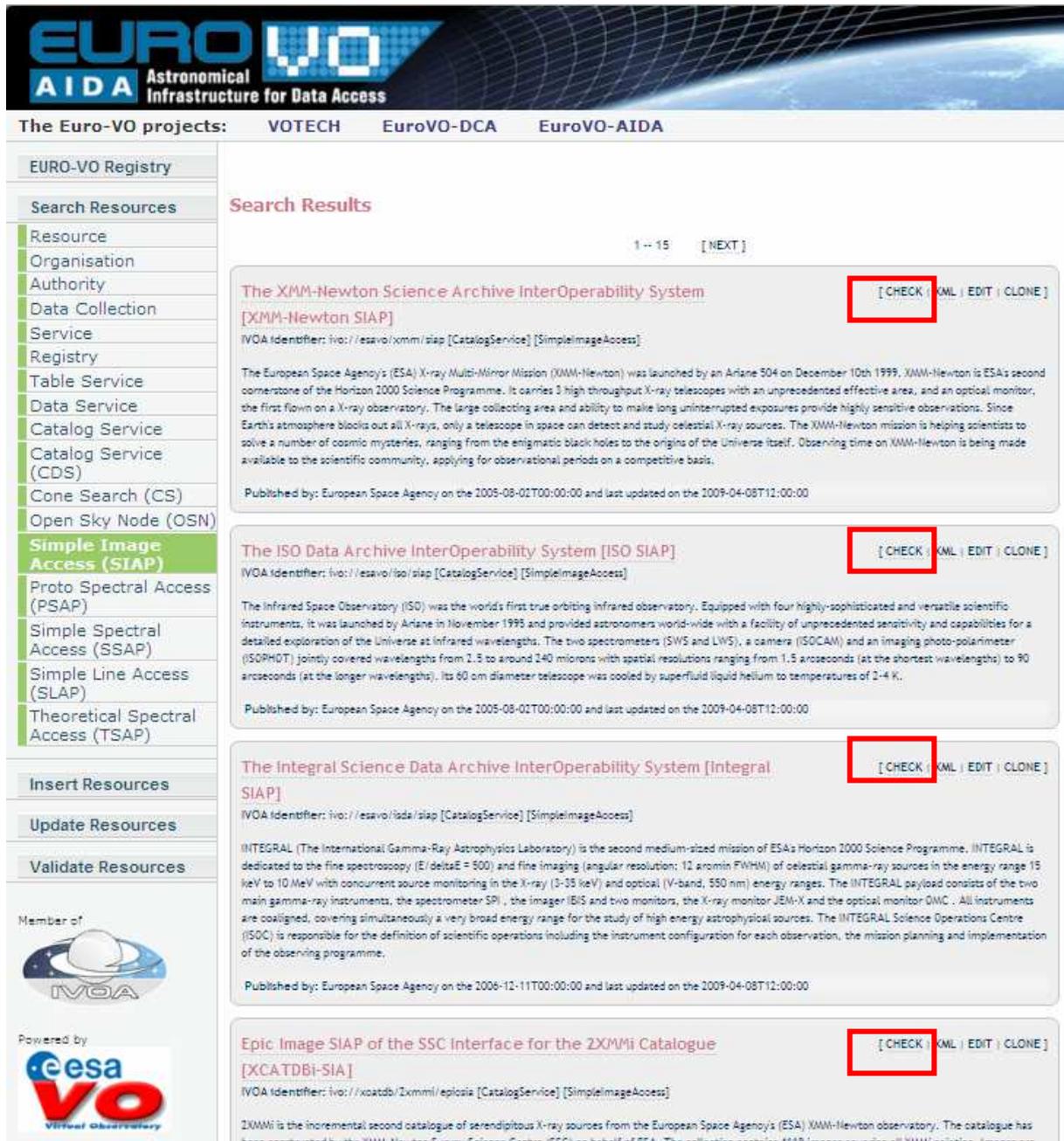


Developed by ESAVO in the context of the EURO-VO AIDA project using the DALValidator software source code from NVO, which was developed with the support of the National Science Foundation under Cooperative Agreement AST0122449 with The Johns Hopkins University.



Furthermore, for each resource already present in the Euro-VO Registry, a check button (calling the DALValidaters) has been added as shown below:

<http://registry.euro-vo.org/result.jsp?searchMethod=XPathQLSearch&select=%23ServiceType%23%3D%27SimpleImageAccess%27>



The screenshot shows the Euro-VO Registry search results page. The header includes the Euro-VO logo and navigation links for VOTECH, EuroVO-DCA, and EuroVO-AIDA. A sidebar on the left lists various search and access services, with 'Simple Image Access (SIAP)' highlighted. The main content area displays search results for SIAP services, each with a title, IVOA identifier, description, and a red-bordered button labeled '[CHECK]'. The results include:

- The XMM-Newton Science Archive InterOperability System [XMM-Newton SIAP]**: IVOA Identifier: `ivo://esa/vo/xmm/siap`. Description: The European Space Agency's (ESA) X-ray Multi-Mirror Mission (XMM-Newton) was launched by an Ariane 504 on December 10th 1999. XMM-Newton is ESA's second cornerstone of the Horizon 2000 Science Programme. It carries 3 high throughput X-ray telescopes with an unprecedented effective area, and an optical monitor, the first flown on a X-ray observatory. The large collecting area and ability to make long uninterrupted exposures provide highly sensitive observations. Since Earth's atmosphere blocks out all X-rays, only a telescope in space can detect and study celestial X-ray sources. The XMM-Newton mission is helping scientists to solve a number of cosmic mysteries, ranging from the enigmatic black holes to the origins of the Universe itself. Observing time on XMM-Newton is being made available to the scientific community, applying for observational periods on a competitive basis. Published by: European Space Agency on the 2005-08-02T00:00:00 and last updated on the 2009-04-08T12:00:00.
- The ISO Data Archive InterOperability System [ISO SIAP]**: IVOA Identifier: `ivo://esa/vo/iso/siap`. Description: The Infrared Space Observatory (ISO) was the world's first true orbiting infrared observatory. Equipped with four highly-sophisticated and versatile scientific instruments, it was launched by Ariane in November 1995 and provided astronomers world-wide with a facility of unprecedented sensitivity and capabilities for a detailed exploration of the Universe at infrared wavelengths. The two spectrometers (SWS and LWS), a camera (ISOCAM) and an imaging photo-polarimeter (ISOPHOT) jointly covered wavelengths from 2.5 to around 240 microns with spatial resolutions ranging from 1.5 arcseconds (at the shortest wavelengths) to 90 arcseconds (at the longer wavelengths). Its 60 cm diameter telescope was cooled by superfluid liquid helium to temperatures of 2-4 K. Published by: European Space Agency on the 2005-08-02T00:00:00 and last updated on the 2009-04-08T12:00:00.
- The Integral Science Data Archive InterOperability System [Integral SIAP]**: IVOA Identifier: `ivo://esa/vo/ida/siap`. Description: INTEGRAL (The International Gamma-Ray Astrophysics Laboratory) is the second medium-sized mission of ESA's Horizon 2000 Science Programme. INTEGRAL is dedicated to the fine spectroscopy ($E/\delta E = 500$) and fine imaging (angular resolution: 12 arcmin FWHM) of celestial gamma-ray sources in the energy range 15 keV to 10 MeV with concurrent source monitoring in the X-ray (3-35 keV) and optical (V-band, 550 nm) energy ranges. The INTEGRAL payload consists of the two main gamma-ray instruments, the spectrometer SPI, the imager IBIS and two monitors, the X-ray monitor JEM-X and the optical monitor OMC. All instruments are co-aligned, covering simultaneously a very broad energy range for the study of high energy astrophysical sources. The INTEGRAL Science Operations Centre (ISOC) is responsible for the definition of scientific operations including the instrument configuration for each observation, the mission planning and implementation of the observing programme. Published by: European Space Agency on the 2006-12-11T00:00:00 and last updated on the 2009-04-08T12:00:00.
- Epic Image SIAP of the SSC Interface for the 2XMMi Catalogue [XCATDBI-SIA]**: IVOA Identifier: `ivo://xcatdb/xmmi/epic/siap`. Description: 2XMMi is the incremental second catalogue of serendipitous X-ray sources from the European Space Agency's (ESA) XMM-Newton observatory. The catalogue has been constructed by the XMM-Newton Survey Science Centre (SSC) on behalf of ESA. The collection contains 4119 images covering all XMMi pointing. Images are

At the bottom of the sidebar, there are logos for 'Member of IVOA' and 'Powered by eesa vo Virtual Observatory'.

Step 5 consisted in writing the appropriate documentation about the exact checks performed by the DALValidators so the IVOA community can understand better the results when validating their DAL services.

This has been done on the IVOA Registry wiki pages available at:

- <http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/ServiceValidation>
- <http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/SIAValidatorTests>
- <http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/SSAValidatorTests>
- <http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/SLAValidatorTests>

We encouraged the IVOA community to check these pages, to verify if the DALValidators actually check what they are supposed to check and if our interpretation of the corresponding IVOA DAL standards specifications is correct.

Furthermore, these pages will also be used to indicate potential future developments, indicating in particular corrections of the DALValidators for proper validation, extra checks and further improvements and potential collaboration from other partners.

Step 6 ("review/validation" process for the DALValidators) represents the "last but not least" area of work which is crucial for the confidence and wide acceptance of the DALValidators within the IVOA community.

This process has been presented and accepted at the IVOA Interoperability Meeting in Garching, Germany (9-12 November 2009) and people recognized its value for the overall improvements of the DALValidators and ultimately of IVOA DAL services as a whole.

The process consists of entering a "virtuous circle" as follow:

- Release the DAL Validators
- Checking the resources
- Checking back the DAL Validators
- Fixing the resources (if needed)
- Correcting the DAL Validators (if needed)
- Defining DAL Validators improvements (new / stronger checks)

We will start this circle for existing DAL services currently covered by DALValidators (ie SCS, SIAP, SSAP and SLAP) and this will be naturally extended for new IVOA DAL services to be covered in the future (e.g. TAP, SIAP v2, ...).

ACRONYM LIST

AIDA	Astronomical Infrastructure for Data Access
CVO	Canadian Virtual Observatory
D#	Deliverable number
DAL	Data Access Layer
ESA	European Space Agency
Euro-VO	European Virtual Observatory
EuroVO-AIDA	Euro-VO Astronomical Infrastructure for Data Access (EC funded, FP7 Call "Scientific Digital Repositories")
HTML	Hypertext Markup Language
IVOA	International Virtual Observatory Alliance
NVO	National Virtual Observatory (US)
PU	Public
SA	Service Activity
SCS	Simple Cone Search
SIAP	Simple Image Access Protocol
SLAP	Simple Line Access Protocol
SSAP	Simple Spectra Access Protocol
TAP	Table Access Protocol
URL	Uniform Resource Locator
US	Unites States
VO	Virtual Observatory
WP#	Work Package number
XML	Extensible Markup Language
XSD	XML Schema Definition