Data access, discovery and interoperability in the European context

Françoise Genova, CDS







# Astronomy has been a pioneer of scientific data sharing

- Basic elements
  - A common data format since the 70s (FITS)
  - Strong tradition of international collaboration
  - Open data (in general after a proprietary period)
  - Driven by community needs (on-line observation archives, on-line services)
- Rapid development of services and links (also with journals) at the advent of the internet, in Europe and elsewhere

## Early networking, still in use

	Eichier Édition Affichage Historique Marque-pages Qutils ?	Eichier Edition Affichage His	orique Marque-pages Qutils ?					
Eichier Édition Affichage Historique Marque-pages Qutils 2	SAO/NASA ADS Data Links 🗴 🕂	WizieR	× (+					
The ISO-SWS post-helium × +	C dsads.u-strasbg.fr/cgi-bin/np      C Q ads harvard      A	vizier.u-strasbg.fr/viz-b	n/VizieR-4	⊽ ଫ	🥄 ads harvard -	☆ 自 ↓ 余 9		
Clsads.u-strasbg.fr/cgi-bin/nph-bib_query?2002A%26A390.1033V&db_key=AST8	noseti Sign on	Portal Sin	ibad VizieR Aladin X-Match	i Other∞ Help		1		
EAO/NAEA ADE Astronomy Abstract Sources	SAO/NASA ADS Abstract Service	<b>? B * *</b>	Show the target form     Show constraint information	VizieR Result Page	•	Send	i to VO tools	
	Links for 2002A&A390.1033V	Search Criteria Save in CDSportal Keywords Back	The 5 columns in <i>color</i> are comp <u>J/A+A/390/1033/table1</u> The IS Post mutation	uted by VizieR, and are not part of SO-SWS post-helium atlas (Vande	f the original data.	ReadMe+ftp	=	
<u>Find Similar Abstracts</u> (with <u>default settings below</u> ) <u>Electronic Refereed Journal Article (HTML)</u>	CDS, Strasbourg, France	Tables Add	rows)	es in the ISO-SWS post-neitum at	as (293 Spectrum 20			
• <u>Full Refereed Journal Article (PDF/Postscript)</u> • <u>On-line Data</u>	Data Table/Catalog	J/A+A/390/1033 ^ table1	Eull R4.12000 DE.12000	Source Simbad Alias	SnTyne TD	RA DE		
• <u>References in the article</u> • <u>Citations to the Article (20)</u> (Citation History)	Resource at <u>ida esac.esa.int:8080</u> <u>Infrared Space Observatory Obsno 04800954, AOT S06</u>		"hm:s"         "d:m:s"           1         20 03 29.4         +36 01 31 HD	190429 Simbad	O4 I f 89300	deg deg 101 300.8725 +36.0251		
<u>Refereed Citations to the Article</u> <u>SIMBAD Objects (295)</u> terreited Article	Resource at <u>ida.esac.esa.int.8080</u> Infrared Space Observatory Obsno 07100250, AOT S01	Constraints Modify Query	2 20 56 34.8 +44 55 29 HD 3 21 38 57.6 +57 29 21 HD	199579 Simbad HR 8023 206267 Simbad HR 8281	O6 V ((f)) 89300 O6.5 V ((f)) 90001	01 314.1449 +44.9247 01 324.7401 +57.4890		
Associated Articles     Also-Read Articles (Reads History)	Resource at <u>ida.esac.esa.int:8080</u>	Preferences max: 50	<u>4</u> 19 52 21.8 +18 40 19 HD <u>5</u> 22 02 04.6 +58 00 01 HD <u>6</u> 20 18 07.0 +40 43 55 HD	Simbad         QZ SGE           209481         Simbad         LZ Cep           193322         Simbad HR 7767	O7.5 la f 90000 O9 V : 90001 O9 V :(n)) 88201	01 298.0907 +18.6719 01 330.5191 +58.0004 01 304 5291 +40 7321		
' <u>Translate This Page</u>	Resource at ida.esac.esa.int:8080	HTML Table	7 21 44 53.3 Eichier Éditi 8 05 45 59.9 The ISO P	ion Affichage <u>H</u> istorique <u>M</u> arq Postcard Server × +	ue-pages <u>Q</u> utils <u>?</u>			
Title:         The ISO-SWS post-helium atlas of near-infrared stellar spectra           Authors:         Vandenbussche, B.; Beintema, D.; de Graauw, T.; Decin, L.; Fe	Infrared Space Observatory Obsno 11801612, AOT S01 Jchtg	Submit	9 22 05 08.8 10 19 51 59.1 12 04 54 02 0 ida	a.esac. <b>esa.int</b> :8080/aio/jsp/createPos	tcards.jsp?obsno=048 ⊽ C	Q ads harvard →	☆自◆合	⊜ ⊾ ≡
Salama, A.; Waelkens, C.; Waters, L.; Wieprecht, E.           Affiliation:         AA(Instituut voor Sterrenkunde, K.U.Leuven, Celestijnenlaan 1	Resource at <u>ida.esac.esa.mt:8080</u> 100B, <u>Infrared Space Observatory Obsno 13400638, AOT S06</u>	CDS, France	<u>12</u> 20 30 35.0 13 21 11 48.2					
Groningen, The Netherlands), AC(SRON National Institute for K.U.Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium), A Germany), AF(Astrophysics Missions Division, Research and 4	Spac Spac Resource at <u>ida.esac.esa.int.8080</u> Signa <u>Infrared Space Observatory Obsno 18000101, AOT S01</u> cient		<u>14</u> 10 42 57.4 <u>15</u> 21 42 24.2 <u>16</u> 00 56 42.5			OBSERVATO	DRY Cesa	
The Netherlands), AG(SRON National Institute for Space Ress Groningen, The Netherlands), AI(Sterrenkundig Institutut "Antt Division of ESA, Villafranca del Castillo, Spain), AK(ISO Dat	rch, Resource at <u>ida.esac.esa.int.8080</u> n Par Infrared Space Observatory Obsno 18205639, AOT S01 Cen		<u>17</u> 20 49 17.4 <u>18</u> 21 50 02.5		The <u>ISO</u> Pos	tcard Server		<u> </u>
voor Sterrenkunde, K.U.Leuven, Celestijnenlaan 200B, 3001 L 200B, 3001 Leuven, Belgium, Sterrenkundig Instituut "Anton I avtraterastrische Physik Garbing hai Minchen Garban	<sup>suver</sup> Resource at <u>ida esac esa int:8080</u> anne <u>Infrared Space Observatory Obsno 18500704, AOT S01</u>			Observation ID 04800954	Target name WR147	AOT Name SWS06	On Time(s) 9285	
Publication: Astronowy and Astrophysics, v.390, p.1033-1048 (2002) ( <u>A&amp;r</u>	Hon Resource at ida.esac.esa.int:8080			RA(J2000 deg) 309.1817	DEC(J2000 deg) 40.3522	Start time 04 Jan 1996 14:38:1	End time 8 04 Jan 1996 17:13:0	)3
					Validation Status: <u>Sci</u>	entifically Validated		
				Access to	IRAS	How to get these	data for publication	
Notworki	ng is not once	ughl		Description and Cali	bration accuracies	Get <u>Data Qual</u> obse	<u>ity Report f</u> or this ervation	
Networki	ng is not enot	ugn:		WR147 (RA, DEC)=(20*36**4	004-JAN 3.6°, +40°21'07.8'') (J2	-1996 000)	SWSD6 TDT No. 04800954	
Seamless	access/			20 Line 1 50 Line 2	20 Une 3 20 Une 4	40 Line 8	40 Line 7 20 Line 8	

interoperability

## A global endeavour with strong European participation





- The first

   Interoperability »
   Working Group was
   under OPTICON
- WG with international participation from the beginning
- VOTable, March 2002
- Succeeded by the IVOA in June 2002

## The Virtual Observatory



## The IVOA standard framework





# Standards AND interoperable tools

### An inclusive and open framework

- No central point, a multi-polar world, a global endeavour
- "Open" and inclusive model
  - A thin interoperability layer on top of the data holdings
  - Anyone can register a data service or build a tool (more than 100 "authorities" with a registered service)
- Data providers also imbed VO building blocks in their archives and services, "minimizing the developments in certain areas"
- The VO is invisible from the end users but it is used because data providers use it (as well shown in this meeting) and people use the services and the tools!

# Europe in the international VO landscape

- IVOA: an Alliance of national & agency initiatives
- In Europe like elsewhere, the core of the work is performed by the national & agency « initiatives », which are very diverse
- Euro-VO: Coordination through projects funded by the European Commission
- Euro-VO is a member of the IVOA

## Why Euro-VO?

- Structuring role of European coordination (ESA/ ESO, EC funding)
- A role to play beyond the « VO countries »
  - Desirable to set up coordination of European VO activities
- A light structure interfaced with the national initiatives and the agencies
  - Evolving partnership
  - Reduced organising overhead
- Through a series of European projects since 2001
- Formalized by a MoU since 2014 (CNRS, INAF, INTA, UEDIN, UHEI)

## A series of European projects



### Reference: Genova et al., Astronomy & Computing 11, 181 (2015)

## ... with evolving partnership

Project	FP	ID	Start-End date	Partners
AVO	FP5	HPRI-CT-2001-50030	2001-2004	ESO ESA (ST-ECF), AstroGrid, Univ. Louis Pasteur/CNRS (CDS) Jodrell Bank, Terapix
VO-TECH	FP6	011892	2005-2009	Univ. Edinburgh Univ. Cambridge, Univ. Leicester, CNRS/ULP (CDS), ESO, INAF
EuroVO-DCA	FP6	RI031675	2006-2008	CNRS (CDS) ESA, ESO, INAF (VObs.it), INTA (SVO), Univ. Leicester (AstroGrid),
EuroVO-AIDA	FP7	212104	2008-2010	Univ. Groningen, MPG (GAVO) CNRS (CDS) ESA, ESO, INAF, INTA, Univ. Edinburgh (AstroGrid), Univ. Groningen, Univ. Heidelberg (GAVO)
EuroVO-ICE	FP7	261541	2010-2012	CNRS (CDS) ESO_INAE_Univ_Edinburgh (AstroCrid)
CoSADIE	FP7	312559	2012-2015	CNRS/U. Strasbourg (CDS) INAF, INTA, Univ. Edinburgh (AstroGrid), Univ. Heidelberg (GAVO)

### FP5 ... FP6 ... FP7 Horizon 2020 : ASTERICS

## Proof of concept: AVO (2001-2004)

- Astrophysical Virtual Observatory
   Definition of Euro-VO approach to the VO
   development
  - User-centered approach
    - Science Working Group, Science Reference
       mission
    - Demos defined by the SWG
  - Interoperable tools
    - Not a single portal
    - ALADIN : portal for images



## AVO products



A&A 424, 545-559 (2004) DOI: 10.1051/0004-6361:20041153 © ESO 2004 Astronomy Astrophysics

#### Discovery of optically faint obscured quasars with Virtual Observatory tools

P. Padovani<sup>1</sup>, M. G. Allen<sup>2</sup>, P. Rosati<sup>3</sup>, and N. A. Walton<sup>4</sup>

<sup>1</sup> ST-ECF, European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany e-mail: Paolo.Padovani@eso.org

- <sup>2</sup> Centre de Données astronomiques de Strasbourg (UMR 7550), 11 rue de l'Université, 67000 Strasbourg, France e-mail: allen@astro.u-strasbg.fr
- <sup>3</sup> European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
- e-mail: Piero.Rosati@eso.org
- <sup>4</sup> Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK e-mail: naw@ast.cam.ac.uk

#### Received 23 April 2004 / Accepted 2 June 2004

Abstract. We use Virtual Observatory (VO) tools to identify optically faint, observed (i.e., type 2) active galactic nuclei (AGN) in the two Great Observatories Origins Deep Survey (GOODS) fields. By employing publicly available X-ray and optical data and catalogues we discover 68 type 2 AGN candidates. The X-ray powers of these sources are estimated by using a previously known correlation between X-ray luminosity and X-ray-to-optical flux ratio. Thirty-one of our candidates have high estimated powers ( $L_x > 10^{44}$  erg/s) and therefore qualify as optically obscured quasars, the so-called "QSO 2". Based on the derived X-ray powers, our candidates are likely to be at relatively high redshifts,  $z \sim 3$ , with the QSO 2 at  $z \sim 4$ . By going  $\sim 3$  mag fainter than previously known type 2 AGN in the two GOODS fields we are sampling a region of redshift – power space which was previously unreachable with classical methods. Our method brings to 40 the number of QSO 2 at faint the GOODS fields, an improvement of a factor  $\sim 4$  when compared to the only 9 such sources previously known. We derive a QSO 2 surface density down to 10<sup>-15</sup> erg cm<sup>-2</sup> s<sup>-1</sup> in the 0.5–8 keV band of  $\gtrsim 320$  deg<sup>-2</sup>,  $\sim 30\%$  of which is made up of previously knows ources. This is larger than current estimates and soome predictions and suggests that the surface density QSO 2 at faint flux limits has been underestimated. This work demonstrates that VO tools are mature enough to produce cutting-edge science results by exploiting astronomical data beyond "classical" identification limits ( $R \leq 25$ ) with interoperable tools for statistical identification fination.

Key words. astronomical data bases: miscellaneous - methods: statistical - galaxies: quasars: general - X-rays: galaxies

#### 1. Introduction

The unified model for active galactic nuclei (AGN) is largely

AGN models, type 2 sources are expected to make a significant fraction of the X-ray background (see, e.g., Comastri et al. 2001) and are therefore also cosmologically very relevant. These sources are heavily reidened and therefore fall through The three pillars of the VO development

... were identified thanks to the AVO project

- Support for scientific users
- Support for data providers
- Technological activities

## (Padovani, 2006)

# ... structured the Euro-VO development since AVO

## FP6 projects: VO-TECH/Euro-VO

Two of the required strands of work

- VO-TECH: Design Study Technological development
  - Architecture/New user tools/Intelligent resource discovery/Data exploration
- EuroVO-DCA: Engagement with data centres
  - First European astronomical Data Centre Census
  - Liaison with the community beyond the partners (VO Days Czech Republic, Lithuania, Poland, ...)
  - Inclusion of modelling results



## FP7 projects: EuroVO-AIDA et al



- EuroVO-AIDA: the full set of networal Astronomical technological activities
  - **Significant legacy:**
  - Schools for scientists, Registry of resources, Standards, Tools, Semantics, ...
  - Close collaboration between the core Euro-VO partners
- **EuroVO-ICE, CoSADIE: (very) small coordination** actions to continue coordination and assess sustainability
  - **Representation of Europe in IVOA**
  - **Close collaboration with ASTRONET**
  - MoU





## Elements of Euro-VO sustainability

ΑCTIVITY	Description - Current Support - Sustainability
Core Service Support	<ul> <li>Registry, Validation, Monitoring</li> <li>Euro-VO Web Portal</li> </ul>
	• Help-desk
<i>VO Infrastructure Support</i>	<ul> <li>Development and maintenance of IVOA standards</li> <li>Publication Tools and Software Libraries</li> <li>Coordination of European VO technology activities</li> <li>Coordination with large astronomy infrastructures</li> <li>Participation in global e-Infrastructures</li> <li>Liaison with neighbouring disciplines</li> </ul>
VO Publishing Support	<ul> <li>Coordination of Data Centre Community</li> <li>Data Centre Training Workshops</li> <li>Direct Assistance to Data Providers</li> <li>Maintenance of reference implementations</li> </ul>
Astronomy Community Support	<ul> <li>Engagement to define science priorities</li> <li>Development and maintenance of science tools</li> <li>Training events</li> <li>Training materials</li> <li>Support of Education and Outreach activities</li> </ul>



### Allen et al. 2014

## Data Centres in the VO context

#### 🕹 DataCentresinVO < EuroVODCA < TWiki - Mozilla Firefox Fichier Édition Affichage Historique Marque-pages Outils ? S 🗘 🔹 💽 • Google C 1 1 http://cds.u-strasbg.fr/twikiDCA/bin/view/EuroVODCA/DataCentresinVO X Data Centres in the Virtual Observatory context EuroVODCA Web Create New Topic Index Data Centres are an essential component of the Virtual Observatory, publishing data, Q Search metadata and services, and providing the physical storage and computational fabrics. The VO 📭 Changes development is a strong incentive to share data and knowledge, and many teams are willing to Notifications provide data and services in their domains of expertise, 'Classical' data centres, such as Statistics ground-and space-based observatory archives, and generalist data centres, are key Preferences providers of added-value services and tools. More and more teams are willing to join with 1 Help value-added services and tools in specific domains, and VO 'data centres' work in very different contexts - national or international Agencies, scientific laboratories - , and are highly Webs EuroVODCA diverse in size and objectives, from small and specific to large and general. Common Main keywords are the willingness to provide a service to the community, provision of Sandbox added-value built on expertise, some kind of sustainability, and concern for quality. TWiki Lessons learnt from the long term history of astronomical data centres show that when beginning these activities, critical parameters are in particular having a critical mass adapted Links to the goals, and ensuring medium-term sustainability, which requires at least a strong Euro-VO support from the local authorities. An important factor to win community support, which is IVOA indispensable to secure funding, is to find a national and/or international niche VOTech Communication Network Many types of contribution are possible: data archives, with a particular emphasis put on Development in FP6 'science ready' data; added-value data bases, services; tools, software suites and algorithms, for instance for data visualisation, data analysis and data mining; thematic services to help solving a well-defined science question; full data analysis or reasearch environments. New types of services are emerging, with in particular theoretical services, providing modelling results, or matching models with observations.

## Data Centre Censuses

- Inclusive definition data centres and services
- Several updates between 2008 and 2014
- A lively snapshot of the European data centre landscape
- Large variety of approaches and sizes, from large data centres maintained by European or National Agencies to small teams maintaining a specific service, including reference data centres such as the CDS
- All disciplines represented, as well as theory

## Next step: ASTERICS WP4 – large projects

- Data Access, Discovery and interoperability
- Make the ESFRI and pathfinder project data available for discovery and usage by the whole astronomical community, interoperable in a homogeneous international framework, and accessible with a set of common tools.
- Fully aligned with the current IVOA priorities

Cluster

Astronomy ESFRI & Research Infrastructure



DADI (WP4) - Francoise Genova (CNRS-OAS):
 4.5 M€

ASTERICS WP4: DADI (Data Access, Discovery and Interoperability)

Figure 6: The ESFRI projects integrated in the VO Framework offers users uniform access.







## Who is involved

- Euro-VO partners, i.e. VO initiatives from France, Germany, Italy, Spain, UK
- Representatives of ESFRI and pathfinders
- Astronomy & Astroparticle physics, including new messengers
- ESO is associated to the project
- ESA (ESAC) is working in close collaboration with





Three Tasks in support to three complementary targets, all partners involved in all activities

- Support to astronomy ESFRI facilities, their pathfinders and other infrastructures of pan-European interest for implementation of their data in the VO framework (INAF, UHEI)
- Support to the astronomical community (INTA, CDS)
- Updates of the VO framework from feedback
   and requirements (UEDIN, CDS)

Astronomy ESERI & Res

## Among DADI Activities



- EFRI Forums, Data Provider Forums, to discuss data provider requirements
- VO Schools for scientists, feedback from scientists
- Support to implementation in the ESFRIs
- Tutorials
- Technology Forums, for technological collaboration
- Participation in IVOA
- Participation in RDA Plenary meetings
- Make the large projects actors of the VO development

## **Euro-VO** impact

- Core Euro-VO partners, a European VO community
- Strong collaboration on technical aspects

   e.g., seen from CDS, VizieR<>TOPCAT, Aladin/Aladin Lite
   usage; close collaboration within IVOA; etc
- Coordination of activities with an European-wide impact, increasing effectiveness and visibility
- ASTRONET Roadmap since 2008
- Strategy definition, in collaboration with the funding agencies (ASTRONET)
- 40 IVOA standards, 22 with at least one European editor, only 2 with no European author

## Conclusion

- Astronomy has been a pioneer of scientific data sharing
- We have built a global data infrastructure which is used by the community in its daily work
- Scientific data sharing is currently a very hot topic, including the highest political level (Research Ministers of the G8)

## The political dimension knows the right keywords!

Eichier Édition Affichage H	Historique Marque-pages Qutils ?	3 Science Ministers Supp.,	ninisters state 🗶 🌆	G8 Science Ministers Sta	ate × +			×
A https://www.gov.uk/g	/government/publications/g8-science-ministers-statement-	ondon-12-june-2013	▼ C Q, G8 minister	rs scientific data June 201	10→ ☆ @	+ ń -	4 🖌	≡
G	G8 science ministers written st need global cooperation.	atement from their UK me	eting on inter	national issu	ies that			^
D	Document	G8 science UK, 12 June PDF, 58.6KB, 5 page	e ministers st e 2013 s	tatement: L	<u>ondon</u> .			
D	Detail	G8 science ministers met in Lor respective national science aca discussed how their nations cor coordination of global scientific maximise the social and econor The statement proposes new at to consider These include:	ndon c demie uld im resea nic be reas fo	i. T se	o the cient	e gr ific erns	ea res in	tes sea re

lobal challenge

- whilst acknowledging the legitimate concerns of private partners. global research infrastructure open scientific research data ncreasing access to the peer-revi Open scientific research data should be easily discoverable, accessible. assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.
  - To maximise the value that can be realised from data, the mechanisms for delivering open scientific research data should be efficient and cost effective, and consistent with the potential benefits.
  - iv. To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including recognition of researchers fulfilling these principles, and appropriate digital infrastructure.

## We are not an isolated island!

- We used generic building blacks when possible, in particular for key building blocks, registry of resources (OAI-PMH) and vocabularies (SKOS, RDF)
- Strong interest in DOIs/PIDs to join the generic infrastructure which will allow data citation beyond our disciplinary solutions
- The IVOA Registry is included in EUDAT B2FIND



• Liaison with RDA

## The Research Data Alliance

- Founded in March 2013 by Australia, EC, and NSF and NIST
- 3200 members from more than 100 countries
- Bottom-up work to tackle all the aspects of scientific data sharing, technological as well as « sociological »
- Have a look at rd-alliance.org and join!







## ASTERICS

- Astronomy ESFRI Research Infrastructure Cluster
- Horizon2020 project Call September 2014 Topic : Implementation of cross-cutting solutions for clusters of ESFRI research infrastructures
- Focus: ESFRIs SKA, CTA, KM3Net, close links with E-ELT, plus EGO (new messengers)
- 22 partners, 15 M€, including 4.5 M€ for WP4, 4 years, began 1 May 2015
- Major collaboration Astronomy-Astrophysics/Astroparticle physics
- Multi-wavelength/multi-messenger