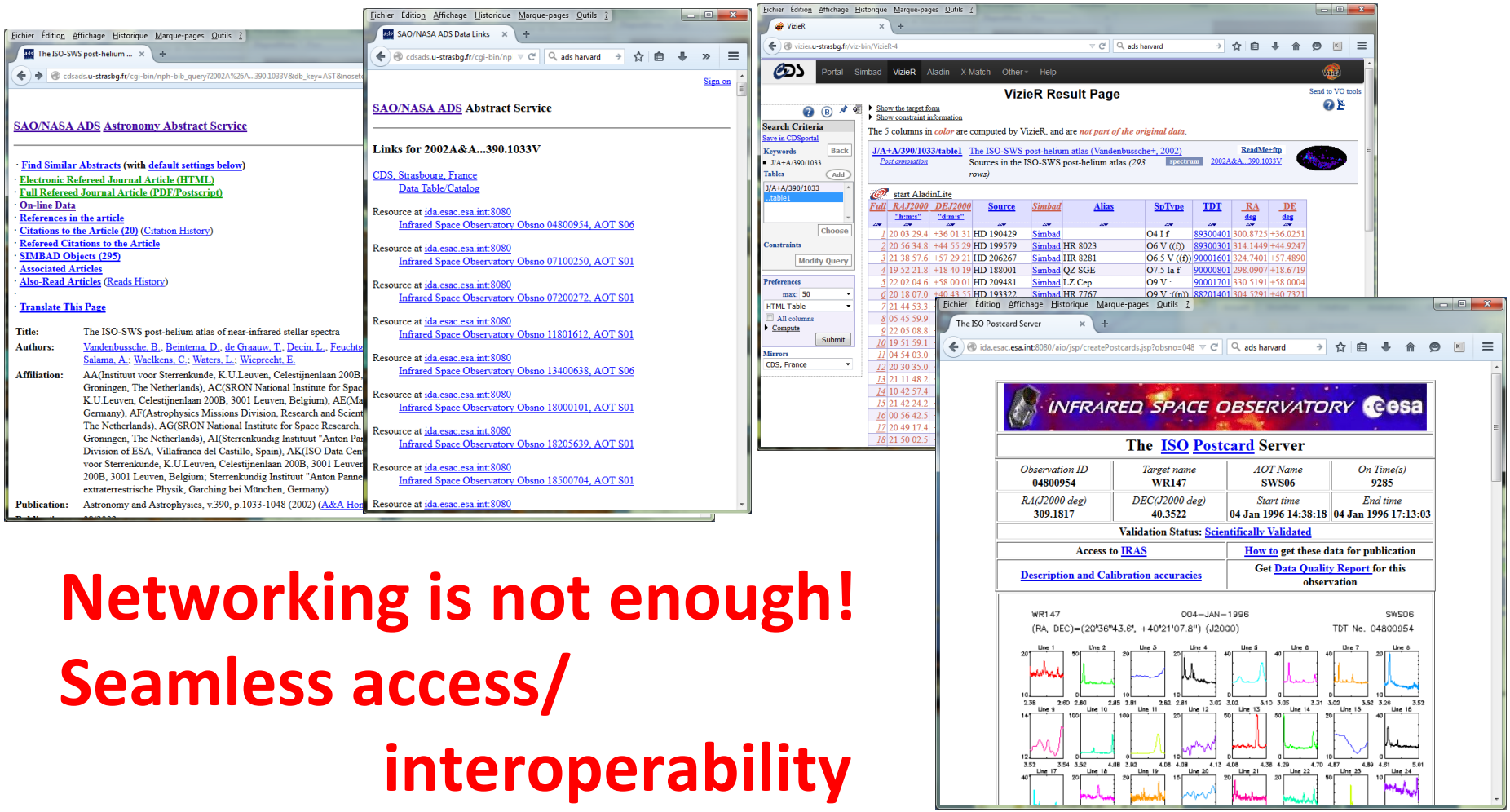
Data access, discovery and interoperability in the European context

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**

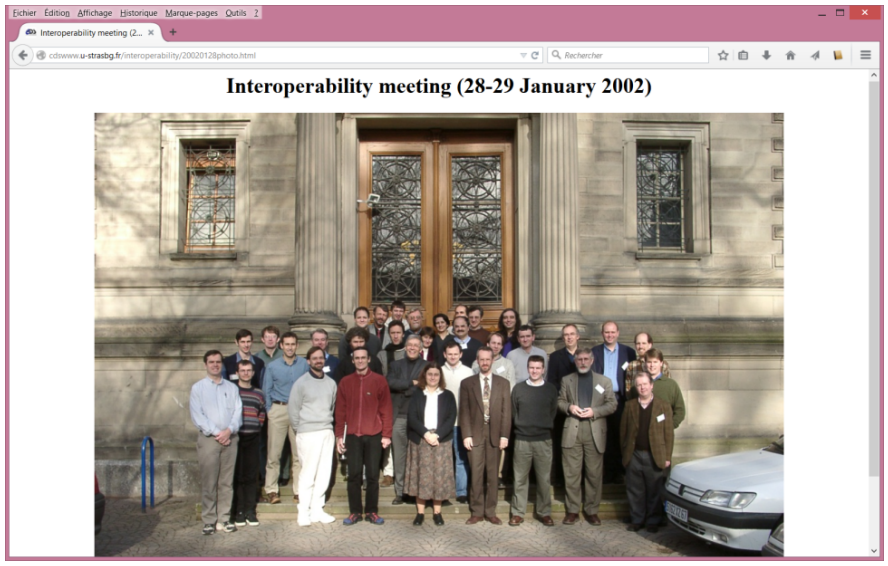
Early networking, still in use



Networking is not enough!

Seamless access/ interoperability

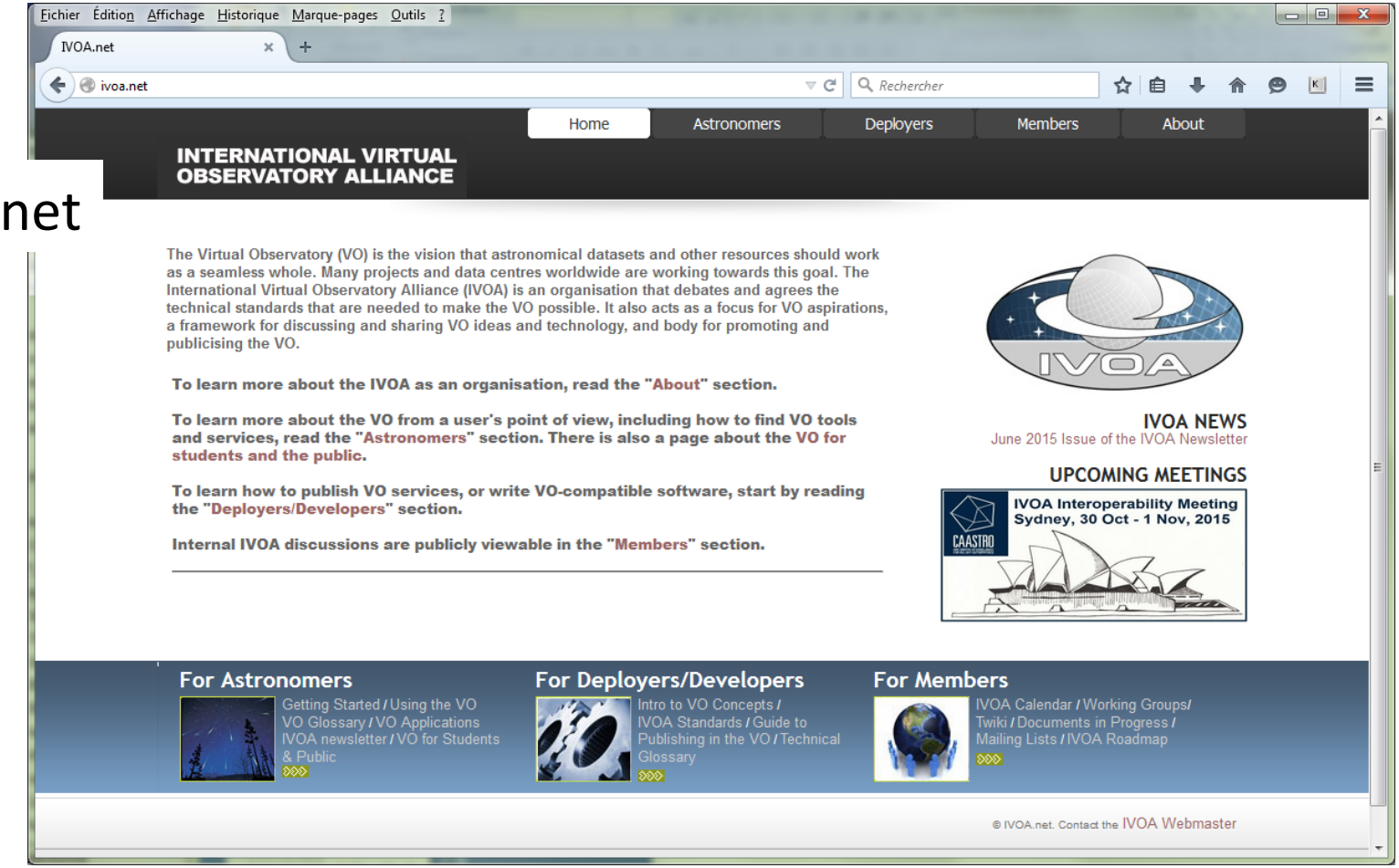
A global endeavour with strong European participation



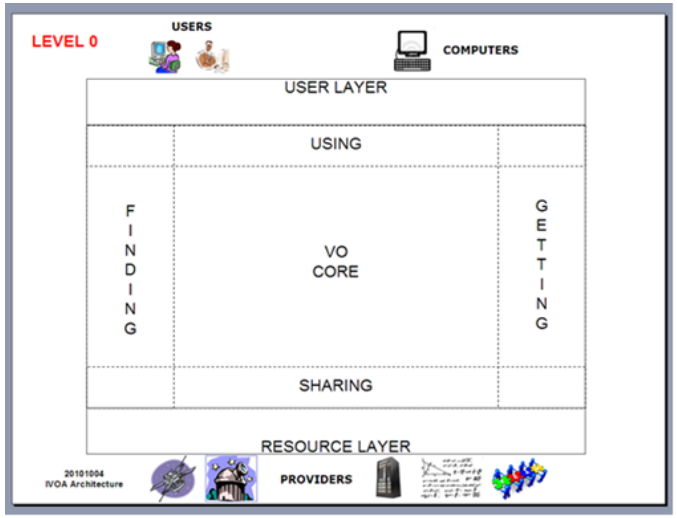
- **The first « Interoperability » Working Group was under OPTICON**
- **WG with international participation from the beginning**
- **VOTable, March 2002**
- **Succeeded by the IVOA in June 2002**

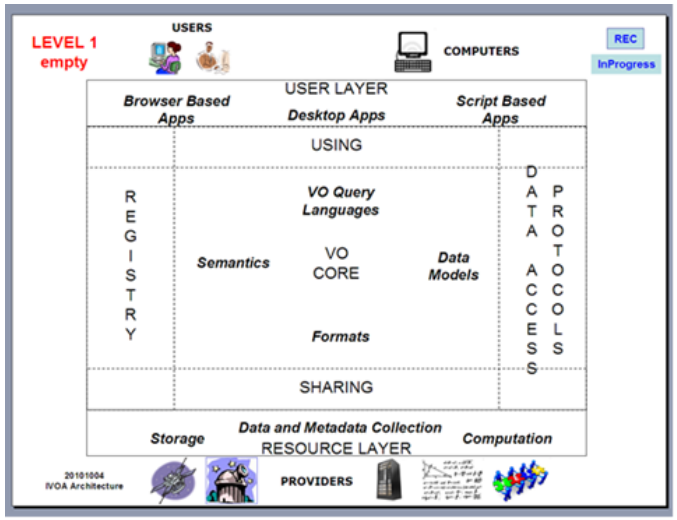
The Virtual Observatory

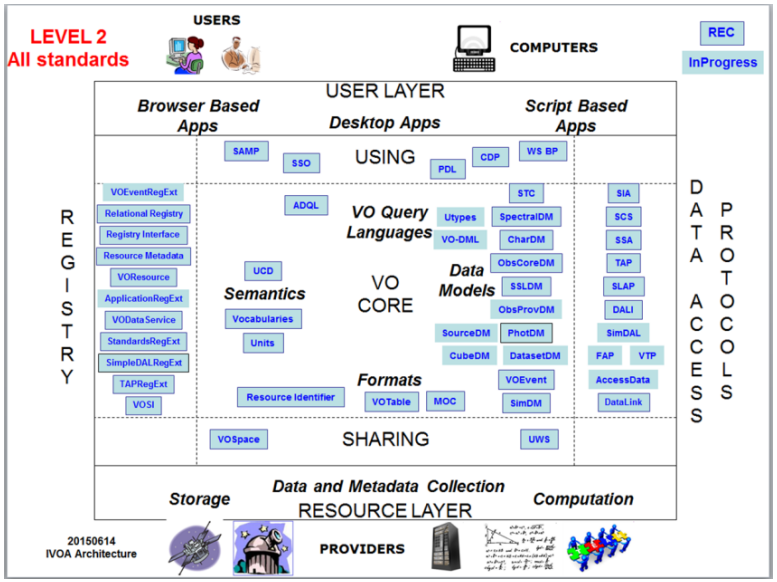
http://ivoa.net



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!**

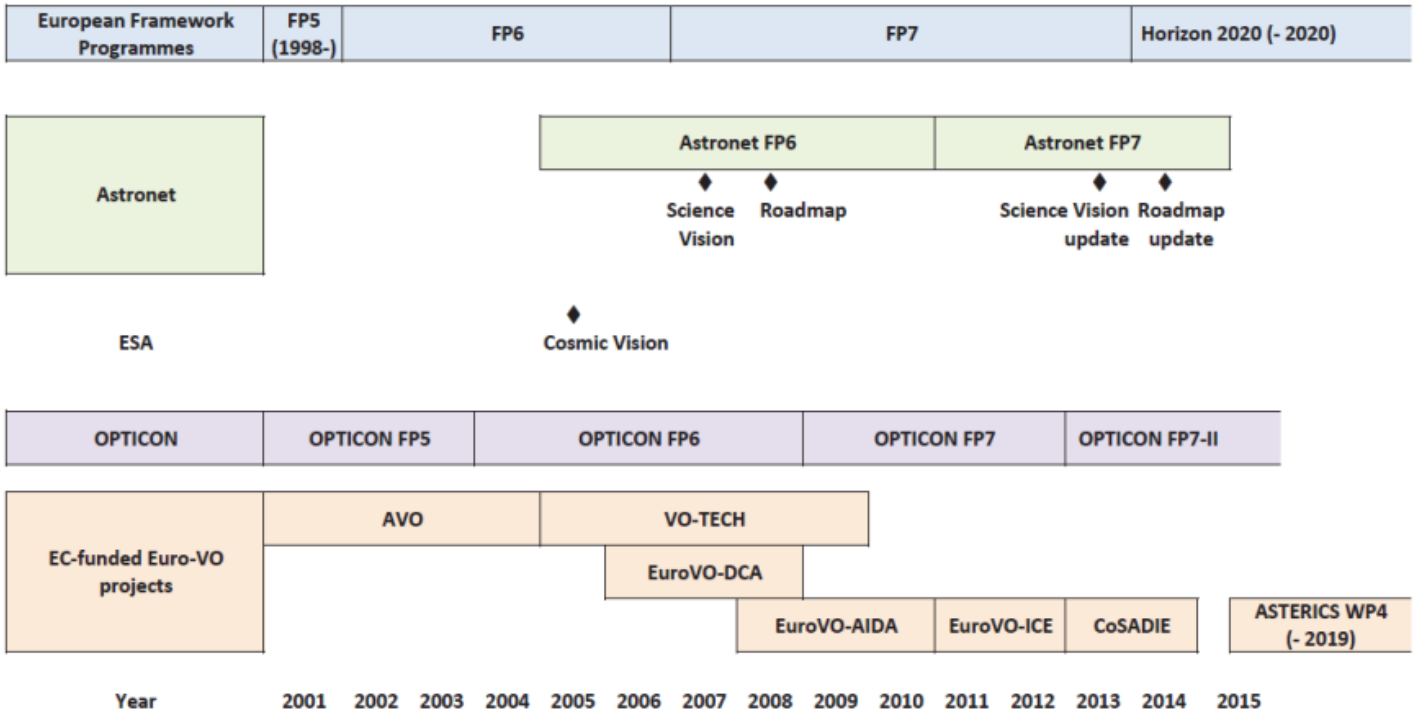
Europe in the international VO landscape

- **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), Univ. Groningen, MPG (GAVO)
EuroVO-AIDA	FP7	212104	2008–2010	CNRS (CDS) ESA, ESO, INAF, INTA, Univ. Edinburgh (AstroGrid), Univ. Groningen, Univ. Heidelberg (GAVO)
EuroVO-ICE	FP7	261541	2010–2012	CNRS (CDS) ESO, INAF, Univ. Edinburgh (AstroGrid)
CoSADIE	FP7	312559	2012–2015	CNRS/U. Strasbourg (CDS) INAF, INTA, Univ. Edinburgh (AstroGrid), Univ. Heidelberg (GAVO)

FP5 … FP6 … FP7
Horizon 2020 : ASTERICS

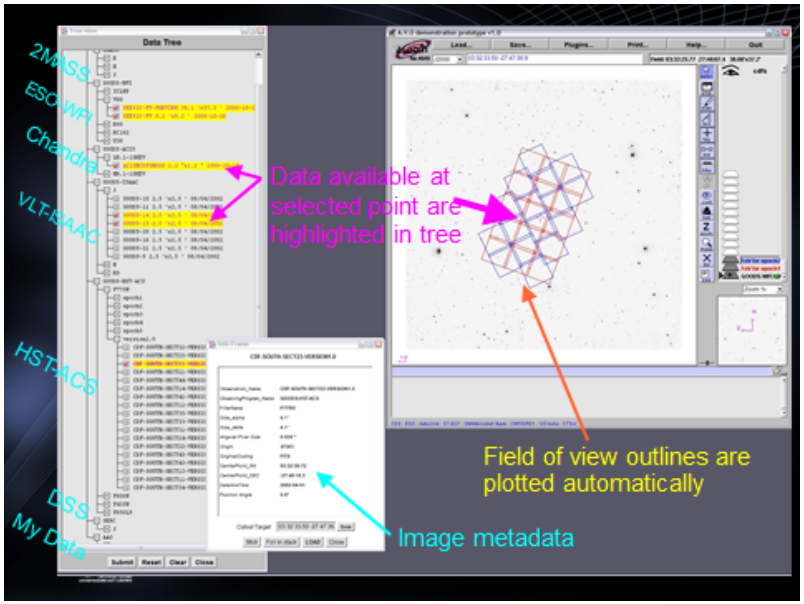
Proof of concept: AVO (2001-2004)

- **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

Astronomy & Astrophysics

Discovery of optically faint obscured quasars with Virtual Observatory tools

P. Padovani[1], M. G. Allen[2], P. Rosati[3], and N. A. Walton[4]

[1] ST-ECF, European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
e-mail: Paolo.Padovani@eso.org
[2] Centre de Données astronomiques de Strasbourg (UMR 7550), 11 rue de l'Université, 67000 Strasbourg, France
e-mail: allen@astro.u-strasbg.fr
[3] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
e-mail: Piero.Rosati@eso.org
[4] 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, obscured (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 ~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 in the GOODS fields, an improvement of a factor ~ 4 when compared to the only 9 such sources previously known. We derive a QSO 2 surface density down to 10^{-15} erg cm^{-2} s^{-1} in the 0.5–8 keV band of $\gtrsim$330 deg^{-2}, ~30% of which is made up of previously known sources. This is larger than current estimates and some predictions and suggests that the surface density of QSO 2 at faint flux limits has been underestimated. This work demonstrates that VO tools are mature enough to produce cutting-edge science results by exploiting astronomical data beyond "classical" identification limits ($R \lesssim 25$) with interoperable tools for statistical identification of sources using multiwavelength information.

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 reddened 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 networking and 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**

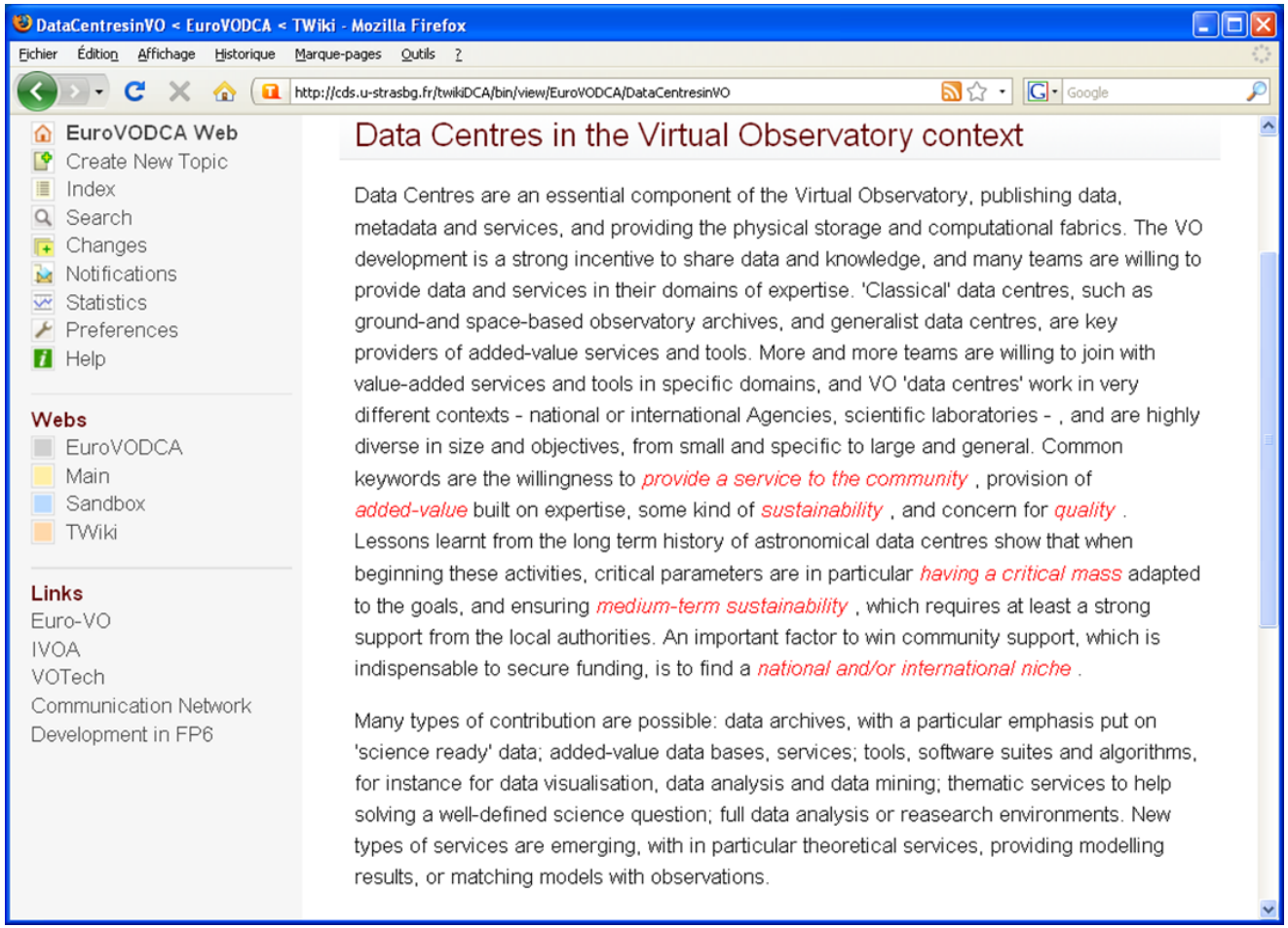
Elements of Euro-VO sustainability

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

EURO VO CoSADIE Collaborative and Sustainable Astronomical Data Infrastructure for Europe

Allen et al. 2014

Data Centres in the VO context



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**

- DADI (WP4) - Francoise Genova (CNRS-OAS):

4.5 M€

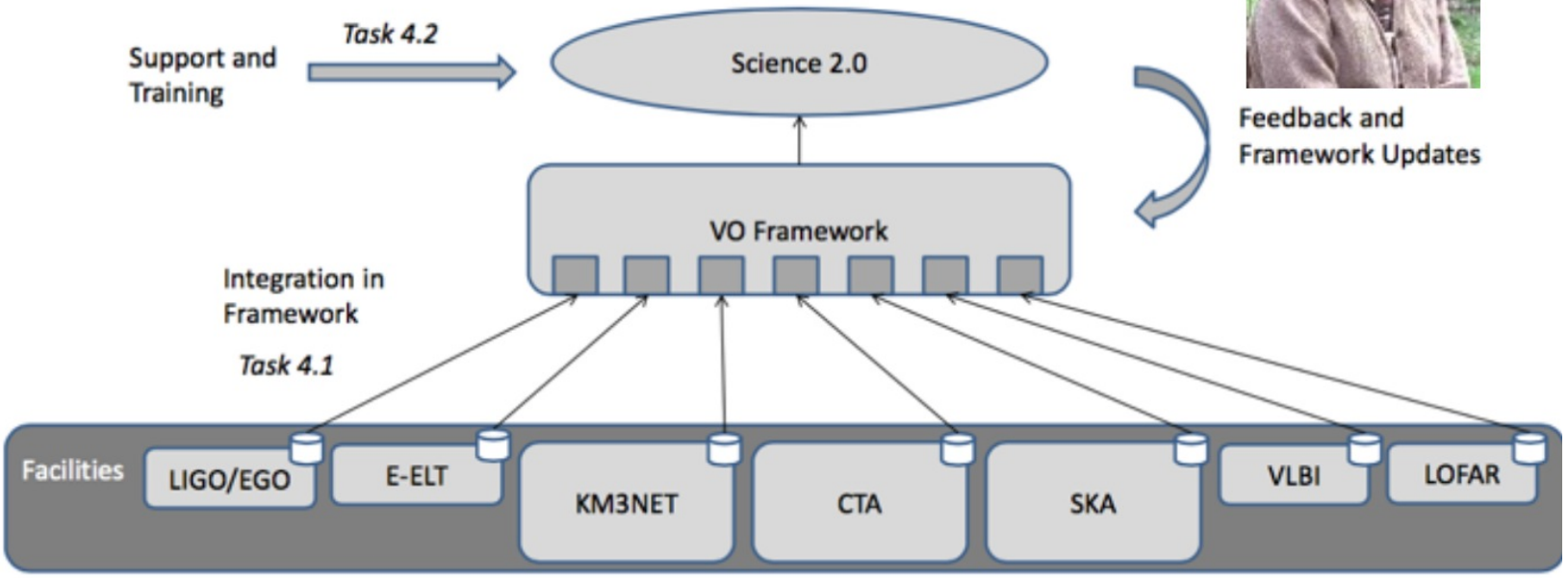


Figure 6: The ESFRI projects integrated in the VO Framework offers users uniform access.

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**

Astronomy ESFRI & Research Infrastructure Cluster

Targets

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)**

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

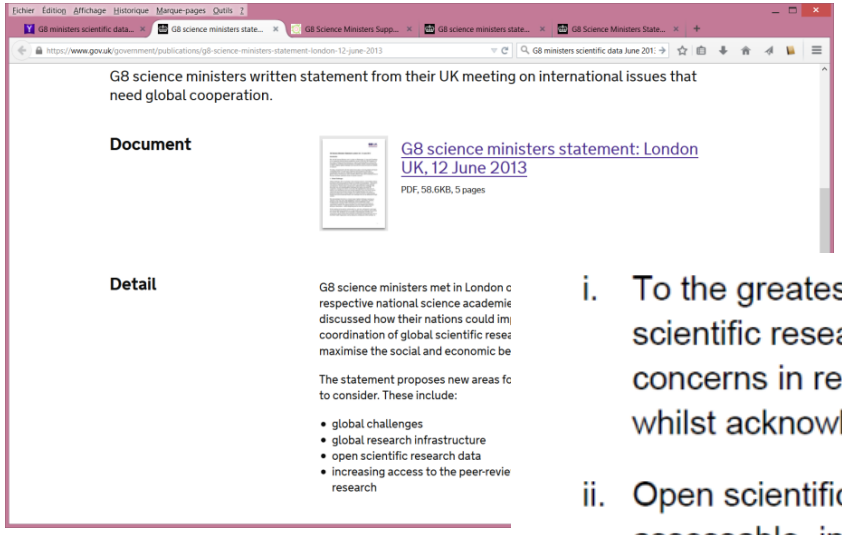
- **Core Euro-VO partners, a European VO community**
- **Strong collaboration on technical aspects**
  e.g., seen from CDS, VizieR<>TOPCAT, Aladin/Aladin Lite usage; close collaboration within IVOA; etc
- **Coordination of activities with an European-wide impact, increasing effectiveness and visibility**
- **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!



G8 science ministers written statement from their UK meeting on international issues that need global cooperation.

Document

G8 science ministers statement: London UK, 12 June 2013

PDF, 58.6KB, 5 pages

Detail

G8 science ministers met in London c
respective national science academie
discussed how their nations could im
coordination of global scientific resea
maximise the social and economic be

The statement proposes new areas fc
to consider. These include:

- global challenges
- global research infrastructure
- open scientific research data
- increasing access to the peer-revie research

i. To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.

ii. Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.

iii. 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