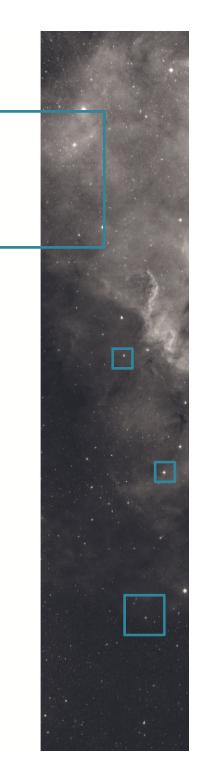
How astronomy shares and reuses scientific data

Françoise Genova





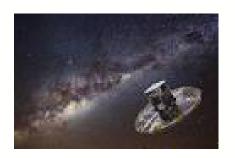
Astronomy research infrastructures

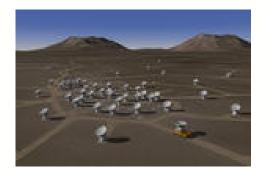












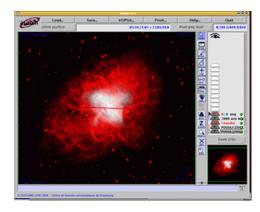
And data!

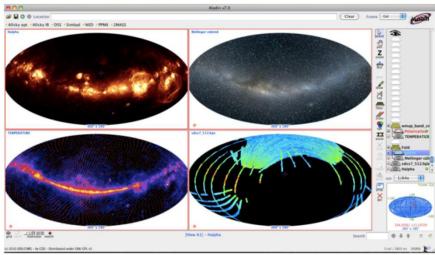
Fourth Symposium Challenge of Big Data in Science, Karlsruhe

Astronomical data

- Observations from ground- and space based telescopes (in general competitive calls for proposals)
- Sky surveys (homogeneous data set with up to billions of objects, measurements, images, spectra)
- Modelling results
- Data from publications
- Value-added data bases, which gather homogenized information in particular from publications
 e.g. SIMBAD, names and papers where the object is cited:
 8 000 000 objects, 22 300 000 object names, 380 000 references, begun ~1970

Why sharing data?





At the core of astronomy scientific needs !

- Multi-wavelengths, multi-technique astronomy
- Time variability
- Comparison of theoretical models with observations
- Etc.

Optimize the scientific return of the large infrastructures

Data is available and used

- Change of paradigm done: astronomers use remote distributed data in their everyday work
- More papers from data retrieved from archives than from original observations (HST, ...)
- More than 800.000 queries/day on the CDS services in 2014 – only one of the components of the astronomical data network

How? 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)

Networking and interoperability

• Networking of on-line resources from 1993-4 (added-value services, journals, archives)

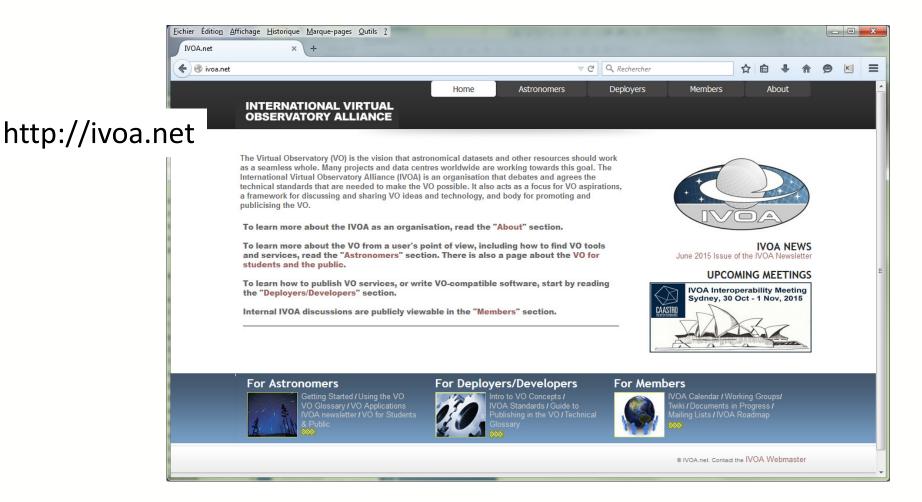
Early networking, still in use

	ſ	Fichier Édition Affichage Historique Margue-pages Outils ?	_ 0 _ X		fistorique Marque-pages Outils	2	
Fichier Édition	Affichage <u>Historique M</u> arque-pages <u>O</u> utils <u>?</u>	SAO/NASA ADS Data Links × +		vizier.u-strasbq.fr/viz	his OficiaR 4	v c Q ads harvard →	
The ISO-SV	VS post-helium × +	SAO/NASA ADS Data Links × +			_	<u>n</u>	
🔶 🔶 🥶 cds	ads. u-strasbg.fr /cgi-bin/nph-bib_query?2002A%26A390.1033	(←)	∔ » ≡	Portal S	Simbad VizieR Aladin)	<-Match Other∼ Help	
			C	*		VizieR Result Page	Send to VO tools
			Sign on		Show the target form Show constraint information		92
SAO/NASA	ADS Astronomy Abstract Service	SAO/NASA ADS Abstract Service		Search Criteria Save in CDSportal	The 5 columns in <i>color</i> ar	e computed by VizieR, and are not part of the original data.	
I		SHOTHER HOST ADSTRUCT SHARE		Keywords Back	J/A+A/390/1033/table1 Post annotation	The ISO-SWS post-helium atlas (Vandenbussche+, 2002)	ReadMe+ftp 02A&A_390.1033Y
	r <u>Abstracts</u> (with <u>default settings below</u>) efereed Journal Article (HTML)			J/A+A/390/1033 Tables Add	Post annotation	Sources in the ISO-SWS post-helium atlas (293 spectrum 20 rows)	JARA. 390.1033V
	d Journal Article (PDF/Postscript)	Links for 2002A&A390.1033V		J/A+A/390/1033 ^	Ø start AladinLite		
On-line Data References in					Full RAJ2000 DEJ200		
	n the article the Article (20) (Citation History)	CDS, Strasbourg, France		Choose	1 20 03 29.4 +36 01	Δ Ψ Δ Ψ Δ Ψ Δ Ψ	deg deg 401 300.8725 +36.0251
	ations to the Article	Data Table/Catalog		Constraints	2 20 56 34.8 +44 55 1	29 HD 199579 Simbad HR 8023 O6 V ((f)) 893003	301 314.1449 +44.9247
SIMBAD OF Associated A		Resource at ida.esac.esa.int:8080		Modify Query	<u>3</u> 21 38 57.6 +57 29 2 4 19 52 21.8 +18 40		601 324.7401 +57.4890 801 298.0907 +18.6719
	rticles (Reads History)	Infrared Space Observatory Obsno 04800954, AOT S06		Preferences	5 22 02 04.6 +58 00	01 HD 209481 Simbad LZ Cep O9 V : 90001	701 330.5191 +58.0004
· · Translate Th	is Page	Resource at ida.esac.esa.int:8080		max: 50 • HTML Table •	<u>6</u> 20 18 07.0 +40.43 7 21 44 53.3 Eichie	r Édition Affichage Historique Marque-pages Outils ?	401 304 5291 +40 7321
		Infrared Space Observatory Obsno 07100250, AOT S01		All columns Compute	8 05 45 59.9 9 22 05 08.8	ne ISO Postcard Server × +	
Title: Authors:	The ISO-SWS post-helium atlas of near-infrared Vandenbussche, B.; Beintema, D.; de Graauw, T.;			Submit	10 10 51 50 1	ida.esac. esa.int :8080/aio/isp/createPostcards.isp?obsno=048 ♥ C	🔍 ads harvard 🔿 🛧 🖨 🛔 🏠 😁 🕷
	Salama, A.; Waelkens, C.; Waters, L.; Wieprecht,	Resource at ida.esac.esa.int:8080		Mirrors CDS, France	<u>11</u> 04 54 03.0 + <u>12</u> 20 30 35.0 +		
Affiliation:	AA(Instituut voor Sterrenkunde, K.U.Leuven, Ce	Infrared Space Observatory Obsno 07200272, AOT S01			13 21 11 48.2		
	Groningen, The Netherlands), AC(SRON Nation K.U.Leuven, Celestijnenlaan 200B, 3001 Leuven	Resource at ida.esac.esa.int:8080			<u>14</u> 10 42 57.4 <u>15</u> 21 42 24.2		ARCERIATORY CORR
	Germany), AF(Astrophysics Missions Division, I	Infrared Space Observatory Obsno 11801612, AOT S01			16 00 56 42.5	INFRAKED SPACE	OBSERVATORY Cesa
	The Netherlands), AG(SRON National Institute f Groningen, The Netherlands), AI(Sterrenkundig 1	D			<u>17</u> 20 49 17.4 - 18 21 50 02.5 -		
	Division of ESA, Villafranca del Castillo, Spain),	Resource at ida.esac.esa.int:8080 Infrared Space Observatory Obsno 13400638, AOT S06				The <u>ISO</u> Pos	icard Server
	voor Sterrenkunde, K.U.Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium; Sterrenkundig Inst	Infrared Space Observatory Obsito 15400036, AOT 300				Observation ID Target name	AOT Name On Time(s)
	extraterrestrische Physik, Garching bei München,	Resource at ida.esac.esa.int:8080				04800954 WR147	SWS06 9285
Publication:	Astronomy and Astrophysics, v.390, p.1033-1048	Infrared Space Observatory Obsno 18000101, AOT S01				RA(J2000 deg) DEC(J2000 deg) 309.1817 40.3522	Start time End time 04 Jan 1996 14:38:18 04 Jan 1996 17:13:03
•		Resource at ida.esac.esa.int:8080				Validation Status: Sci	entifically Validated
		Infrared Space Observatory Obsno 18205639, AOT S01				Access to IRAS	How to get these data for publication
						Description and Calibration accuracies	Get Data Quality Report for this
		Resource at ida.esac.esa.int:8080 Infrared Space Observatory Obsno 18500704, AOT S01				Description and Cambration accuracies	observation
		Infrared Space Observatory Obsho 18500704, AOT SUT				WR147 004-JAN	I-1996 SWS06
		Resource at ida.esac.esa.int:8080		-		(RA, DEC)=(20*36*43.6*, +40*21'07.8") (J2	
						Une 1 Une 2 Une 3 Une 4	Use 5 Use 6 Use 7 Use 6
							+0 +0 20
						have been been been and been been been been been been been be	hand blacked have
						10 0 10 10 10 10 10 10 10 10 10 10 10 10	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
							50 50 10 20 10 40 10
						A LAN	hull have been hered
							0 10 0 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
						40 20 20 20 15 15 15	20 20 50 10 23 Une 24
						I mlu which m	Henry marked

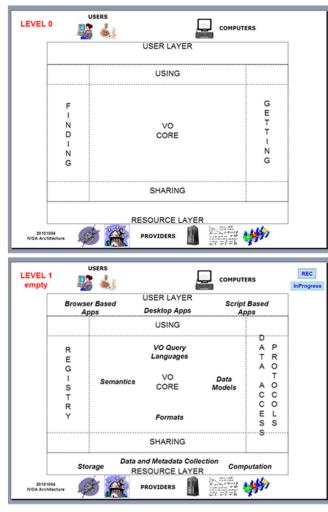
Networking and interoperability

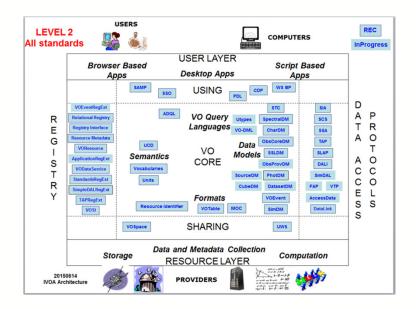
- Networking of on-line resources from 1993-4 (added-value services, journals, archives)
- Seamless access to on-line data (~2000) The astronomical Virtual Observatory
- The VO framework : standards and data access tools – discover, access, use data
- Standards defined by the International Virtual Observatory Alliance (IVOA)
 - Procedure inspired from W3C
 - When possible generic elements (OAI-PMH, SKOS/RDF)

The Virtual Observatory



The IVOA standard framework

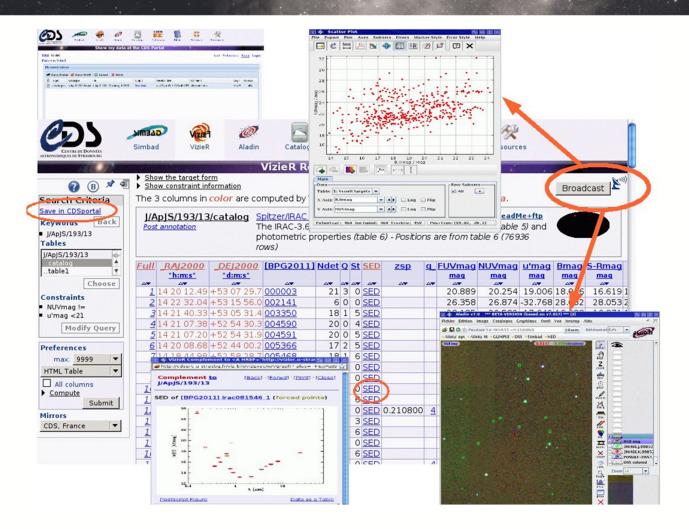




An inclusive and open framework

- 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)
- The VO is invisible but used because people use the services and the tools!
- Data providers also imbed VO building blocks in their archives and services

Interoperable tools and data services



Fourth Symposium Challenge of Big Data in Science, Karlsruhe

Keys for success

- Key for success (science users): seamless access to data AND interoperable tools
- Key for success (data providers)
 - More visibility for their data
 - No need to reinvent the wheel, people already worked

Big and smaller data

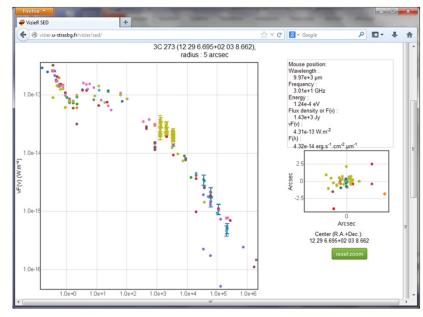
astronomical data

- Observatory archives + disciplinary data centres
- Also data from publications
 - Agreement between CDS and the journals (started in 1993 with *Astronomy & Astrophysics*)
 - tabular data from publications (also images, spectra, time series)
 - together with catalogues from sky surveys, space missions (up to 2 billion rows)
 - 14 000 "catalogues", i.e. data sets
 - Homogeneous metadata describing the very heterogenous content

« Long Tail » data in VizieR

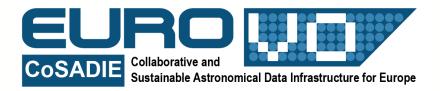


"Photometry viewer": Spectral points extracted from the collection Data validated by a publication Fully discoverable and usable Together with the very large surveys



The European Virtual Observatory

- Coordination of VO activities in Europe
- Three pillars
 - Support for the data providers to provide their data in the VO framework
 - Support for the astronomers in their scientific usage of the VO
 - Technological activities to update the VO framework of standards and tools
- Several EC-funded projects



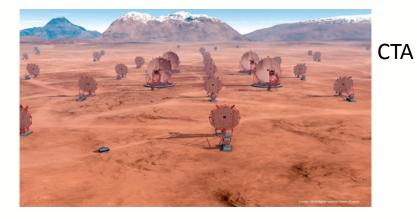


Next step: ASTERICS WP4 – large projects

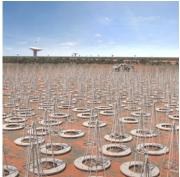
- Data Access, Discovery and interoperability
- Make the ESFRI and pathfinder project data available for discovery and usage by the whole astronomical community, interoperable in a homogeneous international framework, and accessible with a set of common tools.
- Fully aligned with the current
 IVOA priorities

Cluster

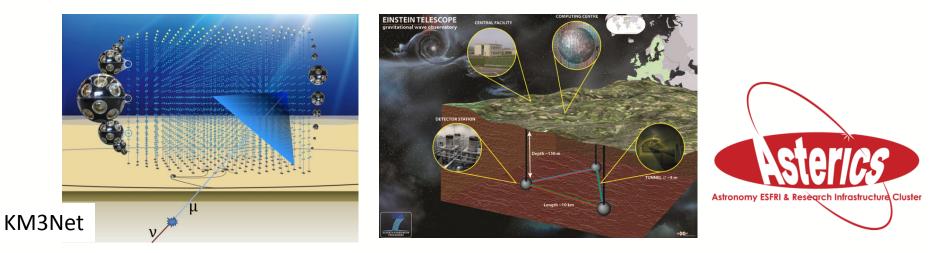
□ Who is involved







SKA



01/10/2015

Fourth Symposium Challenge of Big Data in Science, Karlsruhe

19

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 Euro-VO



20

Conclusions

- The way astronomers do science changed a lot with availability of open data on-line and interoperable data and tools
- Data in astronomy includes « Big » and smaller data
- Smaller data = research results, also discoverable, usable and used
- « Long Tail » = huge diversity, also characteristic of Big Data

Conclusions

- User-centered approach
 - Technology is a tool, not an aim
 - Data sharing in astronomy is science-driven, not technology-driven
- Evaluation of relevant new technologies to identify whether they fit our needs
- Beware of the buzz around seducing but volatile technologies – implementation is resource consuming and technologies must be « sustainable enough »!

Conclusions

- Elements of the VO framework are customized and reused by « nearby » disciplines – planetary studies, the Virtual Atomic and Molecular Data Centre
- The generic elements (registry of resources, vocabulary concepts) allow astronomy data infrastructure to interface with the generic data framework

Sharing scientific data – Open Science

- Strong Open data statement (among others) of G8 Ministers of Research in June 2013
- More and more demand from funding agencies (at least for a data management plan)
- Astronomy has been a pionneer and shares lessons learnt in the Research Data Alliance

The Research Data Alliance

- Founded in March 2013 by EC, NSF and Australia
- 3000 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!

