Global Scientific Data Sharing and Big Data Knowledge discovery

* * * * * *

Czech Virtual Observatory and COST action Big Sky Earth

Petr Škoda

Astronomical Institute of the Czech Academy of Sciences

Supported by grant COST LD-15113 of the Czech Ministry of Education Youth and Sports And COST Action TD1403 Big Sky Earth

RDA Meets Czech Researchers Prague, Czech Republic, 27th October 2017

Credits

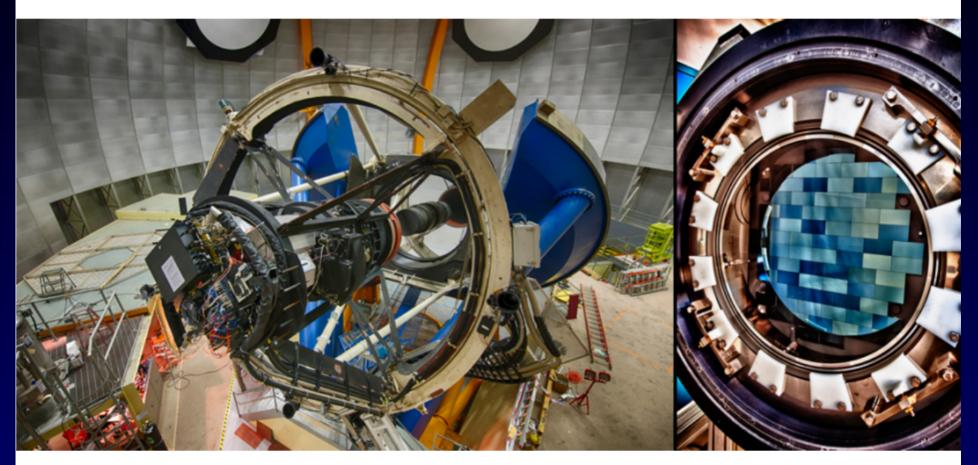
- The presentation is based on many different sources – mainly the on-line published slides from IVOA meetings, slides from Astroinformatics and COST meetings or pictures found on Internet.
- We acknowledge namely materials of B.Hanish, G. Djorgovski, G. Longo, T. Hey and M. Breddels, D Vinkovic, P. Baumann, A Nina and presentations from AI2016 in Sorrento

Outline of the Talk

- Data Avalanche in astronomy
- Virtual Observatory
- Astroinformatics
- Big Sky Earth
- Transfer of technology
- CZVO
 - Be stars discovery in LAMOST surveys
 - Ondrejov Southern Photometry Survey
- Future

Dark Energy Survey Camera

Dark Energy Camera (DECam)



~0.4 PB/yr

74 chips – 570 Mpixels - 4m Cerro Tololo

Large Synoptic Survey Telescope



201 CCD 4kx4k, 3.2 Gpix every 20 sec 3.5 deg FOV (64cm) 20 TB/day=6 PB/yr RAW 1.5 PB catalogue !!! detection of changes 60s!

38 billion objects x 1000 32 tril. meas. -5 PB table Cerro Pachón – Future site of the LSST





Cherenkov Telescope Array

Cherenkov Astronomy and CTA



- Two arrays of 100 (South) et 20 (North) telescopes
- July 2015: sites selection, Chile (ESO) and La Palma
- 2016: pre-production phase
- 2018-2013: production phase
- Observatory open to the community



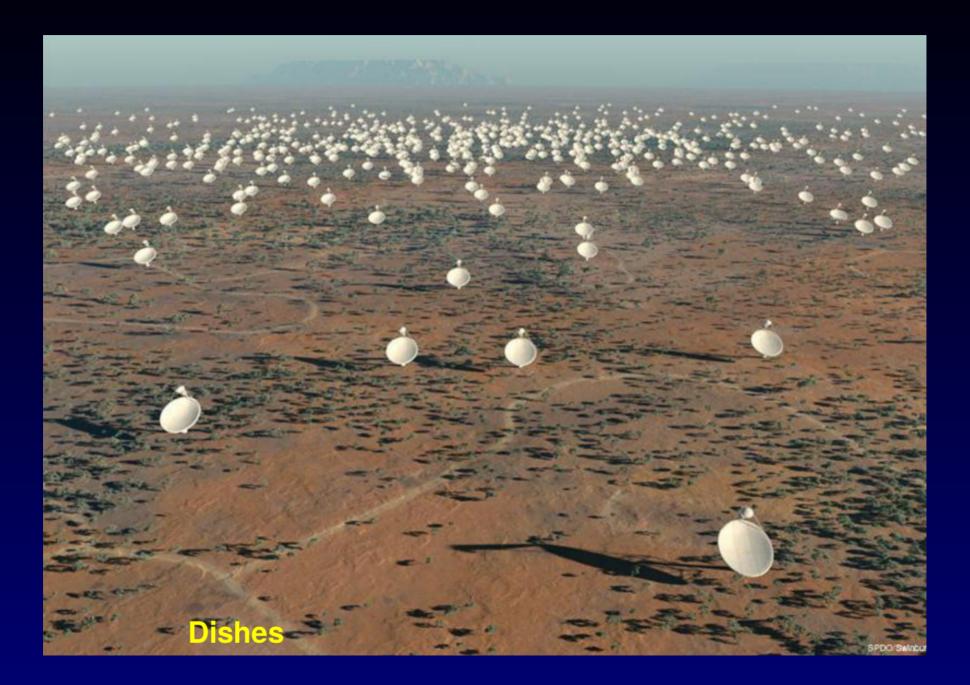
Mathieu Servillat

Cherenkov Telescope Array

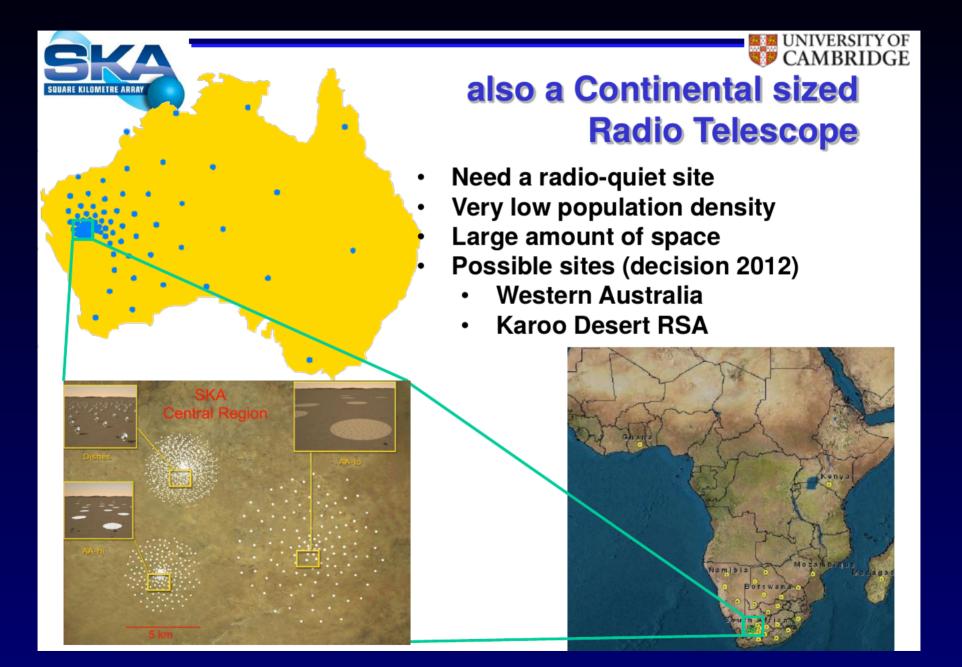
5

27 PB/year in archive





SKA

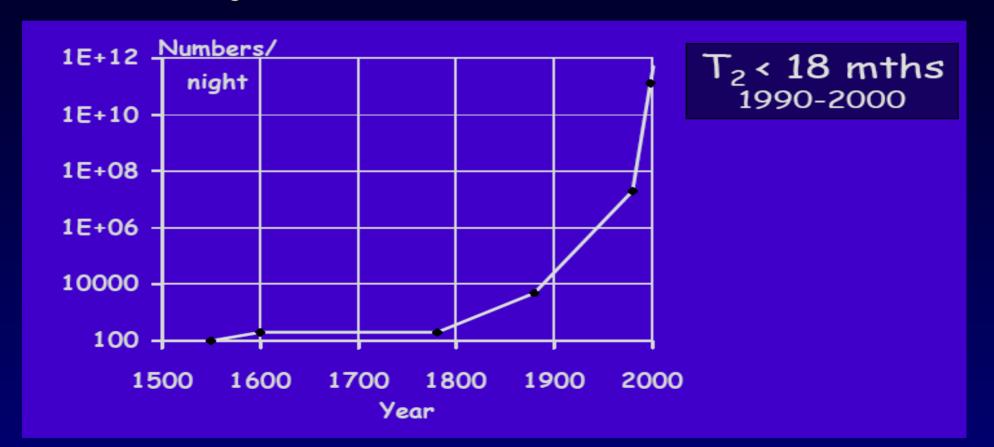


LOFAR network



Data Avalanche

Moore law for chips —doubling 1.5 year Data in astronomy — doubling < 1 yr ! (1000/10 yr) ====> Processing to data



Virtual Observatory : Key Definitions

- "The Virtual Observatory will be a system that allows astronomers to interrogate multiple data centers in a seamless and transparent way, which provides new powerful analysis and visualization tools within that system, and which gives data centers a standard framework for publishing and delivering services using their data".
- Standardization of data and metadata, and of data exchange methods.
- Registry, listing available services and what can be done with them.

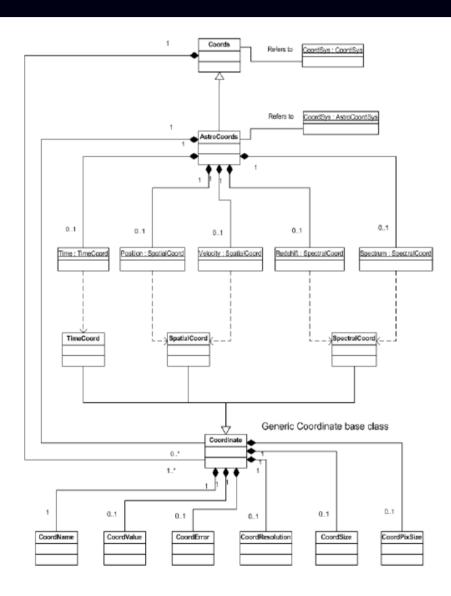
R.J.Hanisch, P.J.Quinn, in "IVOA – Guidelines for participation"

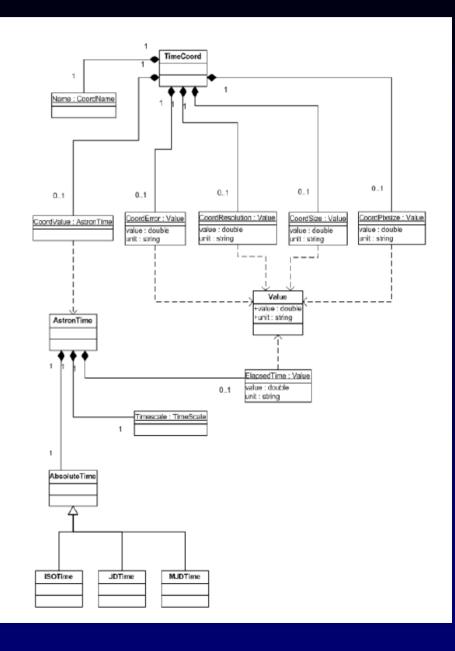
Start of VO – early 2000





Space-Time-Coordinate Data Model





Technology of VO

Unified data format– VOTable, UCD (semantics) Web services (WS) VOregistry – Google for data+WS protocols

ConeSearch (searching in circle on sky) SIAP (Simple Image Access Protocol) SSAP(Simple Spectral Access Protocol) SLAP(Simple Line Access Protocol) - VAMDC TAP (Table Access Protocol) – query e.g. whole SDSS VOEVENT (transients, robotic telescopes,Sun) datacubes, DATALINK on-the-fly data processing

Technology of VO

ADQL (Astronomical Data Query Language) XMATCH, REGION (2 catalogues – shifted)

Application interoperability – SAMP

Allows develop applications as bricks sending VOTABLES (catalogue-spectra-images)

Big Data handling

- VO Space Moving big tables across (load only results)
- SSO Authentication, authorization, groups and consortia
- UWS Universal worker service (job synch, asynch)
- PDL Parameter Description Language
- SIM-DB Simulations, theory data

Simple Spectra Access Protocol Spectral Data Model

Simple Spectral Access Protocol V1.04



International Virtual

Observatory

Alliance

Simple Spectral Access Protocol

Version 1.04 IVOA Recommendation Feb 01, 2008

This version: http://www.ivoa.net/Documents/REC/DAL/SSA-20080201.html Latest version: http://www.ivoa.net/Documents/latest/SSA.html Previous version(s): Version 1.03, December 2007 Version 1.02, September 2007 Version 1.01, June 2007 Version 1.00, May 2007 Version 0.97, November 2006 Version 0.96, September 2006 Version 0.95 May 2006 Version 0.91 October 2005 Version 0.90 May 2005 Editors: D.Tody, M. Dolensky Authors:

D.Tody, M. Dolensky, J. McDowell, F. Bonnarel, T.Budavari, I.Busko, A. Micol, P.Osuna, J.Salgado, P.Skoda, R.Thompson, F.Valdes, and the data access layer working group.



International Virtual Observatory

Alliance

IVOA Spectral Data Model Version 1.03 IVOA Recommendation 2007-10-29

This version (Recommendation Rev 1)

http://www.ivoa.net/Documents/REC/DM/SpectrumDM-20071029.pdf Latest version: http://www.ivoa.net/Documents/latest/SpectrumDM.html Previous versions:

http://www.ivoa.net/Documents/PR/DM/SpectrumDM-20070913.html

Editors:

Jonathan McDowell, Doug Tody Contributors:

Jonathan McDowell, Doug Tody, Tamas Budavari, Markus Dolensky, Inga Kamp, Kelly McCusker, Pavlos Protopapas, Arnold Rots, Randy Thompson, Frank Valdes, Petr Skoda, and the IVOA Data Access Layer and Data Model Working Groups.

SSAP Parameters

4.1.1 Mandatory Query Parameters

The following parameters must be implemented by a compliant service:

Parameter	Sample value	Physical unit	Datatype
POS	52,-27.8	degrees; defaults to ICRS	string
SIZE	0.05	degrees	double
BAND	2.7E-7/0.13	meters	string
TIME	1998-05-21/1999	ISO 8601 UTC	string
FORMAT	votable	-	string

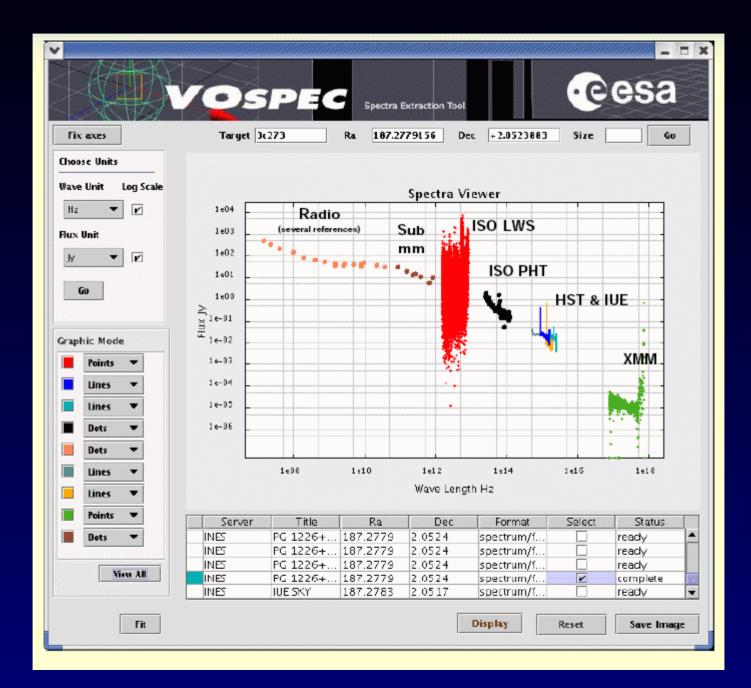
4.1.2 Recommended and Optional Query Parameters

Parameter	Sample value	Unit	Req	Datatype	
APERTURE	0.00028 (=1*)	degrees	OPT	double	
SPECRP	2000	$\lambda/d\lambda$	REC	double	
SPATRES	0.05	degrees	REC	double	
TIMERES	31536000 (=1yr)	seconds	OPT	double	
SNR	5.0	dimensionless	OPT	double	
REDSHIFT	1.3/3.0	dimensionless	OPT	string	
VARAMPL	0.77	dimensionless	OPT	string	
TARGETNAME	mars		OPT	string	
TARGETCLASS	star		OPT	string	
FLUXCALIB	relative		OPT	string	
WAVECALIB	absolute		OPT	string	
PUBDID	ADS/col#R5983		REC	string	
CREATORDID	ivo://auth/col\$R1234		REC	string	
COLLECTION	SDSS-DR5		REC	string	
TOP	20	dimensionless	REC	int	
MAXREC	5000		REC	string	
MTIME	2005-01-01/2006-01-01	ISO 8601	REC	string	
COMPRESS	true		REC	boolean	
RUNID			REC	string	

SPLAT-VO (Starlink, Heidelberg, Ondrejov)

	k SPLAT-VO: <pl< th=""><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th>_ = ×</th></pl<>										_ = ×
File Analy	sis Edit Optio	ns Graphics	Help								
	JPEG ⇔ Į		xX	≁ [‡] ,‡ 4	<u>}</u>		2				
Displaying:	: D:\SPE	EFO\LA280060.	RUI				▼	Y limits (%):	automatic 💌		🗌 :V-hair
LAMBDA :	6528.643	🗌 :log			D:\SPE	FO\LA2800	060.RUI:	0.9882333		🗌 :log	🗌 :Track free
X scale:	1.0	• +	-				Y scale:	0.5	V	+	-
(u.w.				2-d com	pound coor	dinate sys	stem				
1 9.9 10 10 10 10 10 10 10 10 10 10 10 10 10	6300	6350	6400	6450	6500 LAMBD	6550 A	660	0 665	0 6700		
4											

VOspec (ESAC)

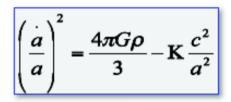


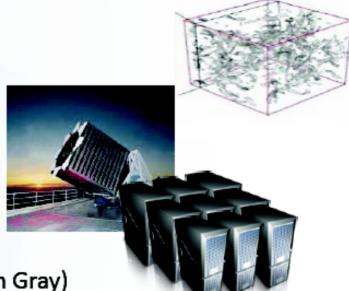
Emergence of a Fourth Research Paradigm

- 1. Thousand years ago Experimental Science
 - Description of natural phenomena
- 2. Last few hundred years Theoretical Science
 - Newton's Laws, Maxwell's Equations...
- 3. Last few decades Computational Science
 - Simulation of complex phenomena
- 4. Today Data-Intensive Science
 - Scientists overwhelmed with data sets
 - from many different sources
 - Data captured by instruments
 - Data generated by simulations
 - Data generated by sensor networks
 - eScience is the set of tools and technologies
 - to support data federation and collaboration
 - For analysis and data mining
 - For data visualization and exploration
 - For scholarly communication and dissemination

(With thanks to Jim Gray)







From T. Hey, AI2010

T. S. Kuhn – Structure of Scientific Revolutions

X-informatics



FOURTH PARADIGM

DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLS

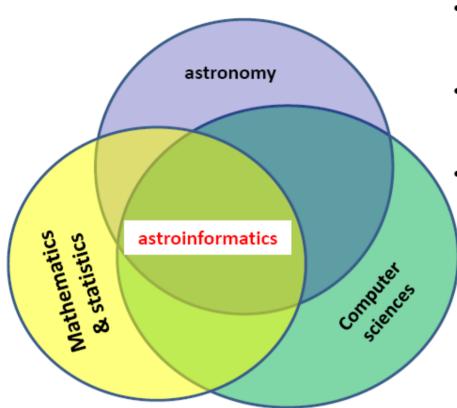
Downloadable at Microsoft Research site

Changing methodology of the Science

Synergy between different worlds

Sociological aspects (net-based research communities)

Experimental astronomy has become a three players game



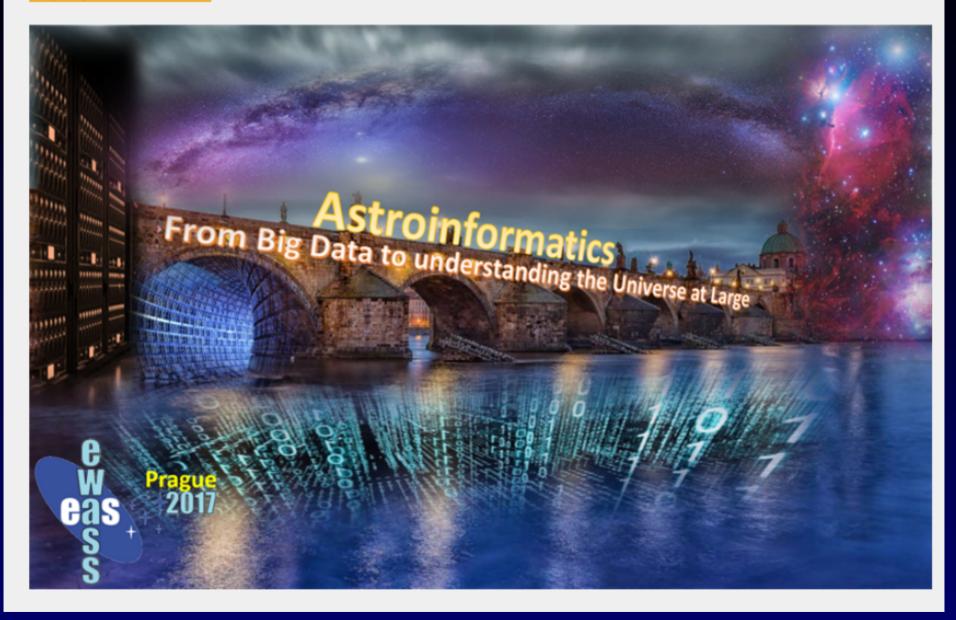
- astronomy: problems, data, understanding of the data structure and biases
- mathematics: evaluation of the data, falsification/validation of theories/models, etc
- computer science: implementation of infrastructures, databases, middleware, scalable tools, etc

- Astroinformatics: AAS n. 215, Washington, December 2009, chairperson: K. Borne
- Astroinformatics 2010: Caltech (USA) June 16-19 2010; co-chairpersons: S.G. Djorgovski, G. Longo
- Astroinformatics 2011: UNINA Sorrento, co-chairpersons: S.G. Djorgovski, G. Longo

Longo 2010

Symposium S14

29 - 30 June 2017



http://eas.unige.ch/EWASS2017/session.jsp?id=S14

Need for a new science: Astroinformatics Knowledge Discovery in Databases

Data Gathering (e.g., from sensor networks, telescopes...)

└→ Data Farming:

Storage/Archiving Indexing, Searchability Data Fusion, Interoperability, ontologies, etc.

→ Data Mining (or Knowledge Discovery in Databases):

Pattern or correlation search Clustering analysis, automated classification Outlier / anomaly searches Hyperdimensional visualization

→ Data understanding

Computer aided understanding KDD Etc. Ongoing research

Key mathematical

Database

issues

technologies

→ New Knowledge



Data Driven Science

What is Fundamentally New Here?

- The *information volumes and rates* grow exponentially
- Most data will never be seen by humans

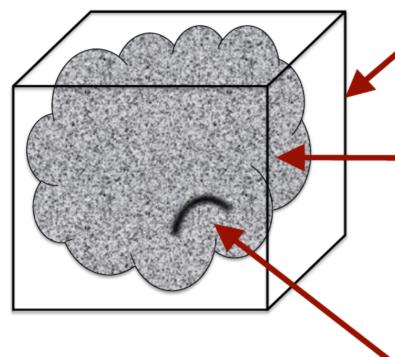


- A great increase in the data *information content*
- → Data driven vs. hypothesis driven science
- A great increase in the information complexity
- There are patterns in the data that cannot be comprehended by humans directly



Hidden Patterns in Data

Pattern or structure (Correlations, Clustering, Outliers, etc.) Discovery in High-Dimensional Parameter Spaces



D >> 3 parameter space hypercube

> High-D data cloud: mostly noise, of an arbitrary distribution

But in some corner of some sub-D projection of this data space, there is *something ≠ noise*

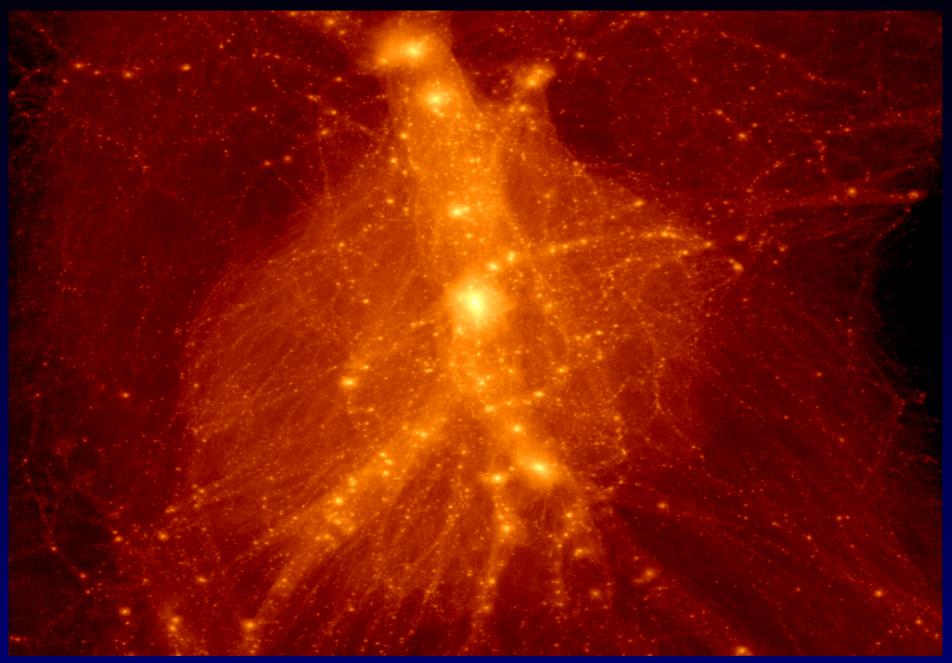
Visualization in Machine Learning

A Key Challenge: Visualisating Multidimensional Data Spaces

- Hyperdimensional structures (clusters, correlations, etc.) may be present in many complex data sets, whose dimensionality may be D ~ 10² – 10⁴, or higher
- It is a matter of *data understanding*, choosing the right data mining algorithms, and interpreting the results
- We are biologically limited to perceiving up to ~ 3 - 12(?) dimensions

What good are the data if we cannot effectively extract knowledge from them?

Visualization of Big Data



Breddels 2016

Big Data Era in Sky and Earth Observation – TD 1403 COST action

P. Škoda Czech Representative in MC



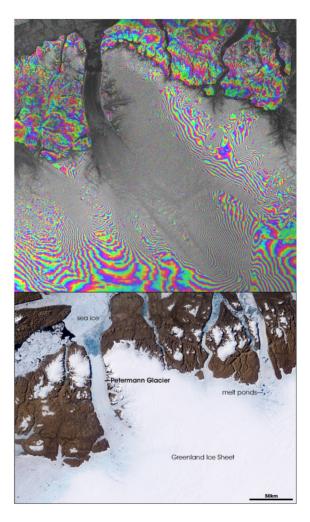
sentinel-1A

Launch date: 3. April 2014 up to 2.4 TB/day of imaging radar data for 7 years (fully open and free data access policy)

Applications: Oceans and ice, Changing lands, Emergency response

The era of Big Data has arrived!

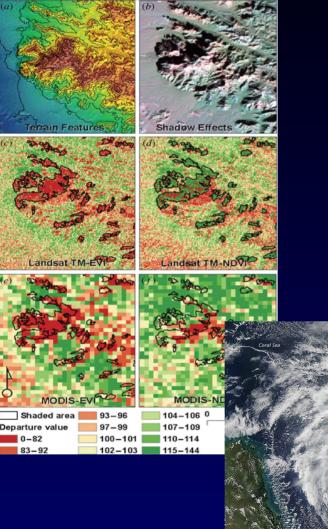
Example: images + time = surface movements



Part of the **European Earth Observation Programme Copernicus:** the most ambitious Earth observation programme to date: 30 satellites: peta-bytes now: zetta-bytes in a decade



Remote sensing - Big Data Machine Learning





Precise farming

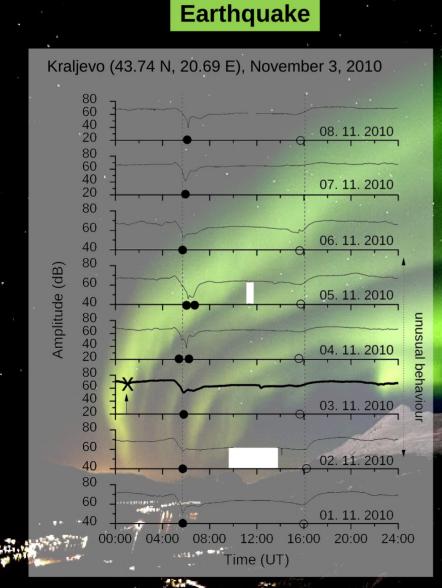
Forestry

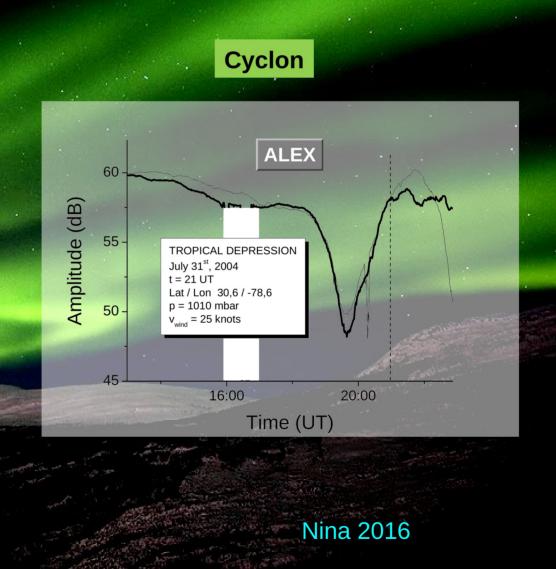
Ore mining

Water resources monitoring

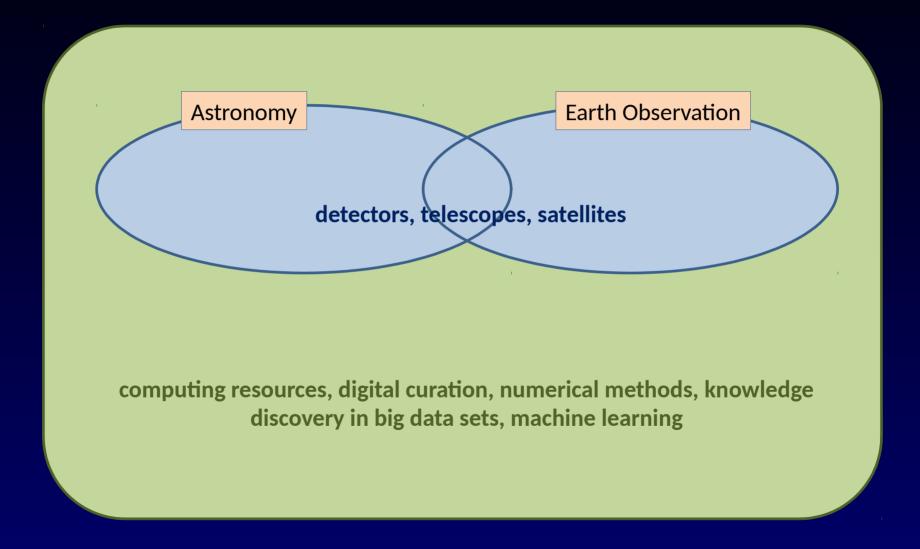
Automatic classification of terrain Resistence of buildings (Aquilla) Mobile network structure – disasters Customized product for every user

Ionospheric variations and natural disasters predictions





Big Sky Earth Commons



Vinkovic 2016

Impact

Big Data is not just bigger, it is different!

Success in research will depend on the ability to mine knowledge from that data.

And some of the most interesting science probably hasn't even been imagined!



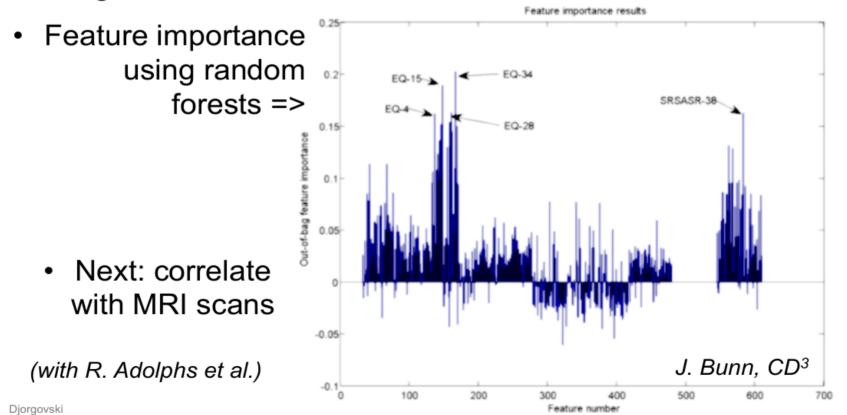
The FOURTH PARADIGM DATA-INTENSIVE SCIENTIFIC DISCOVERY

DITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

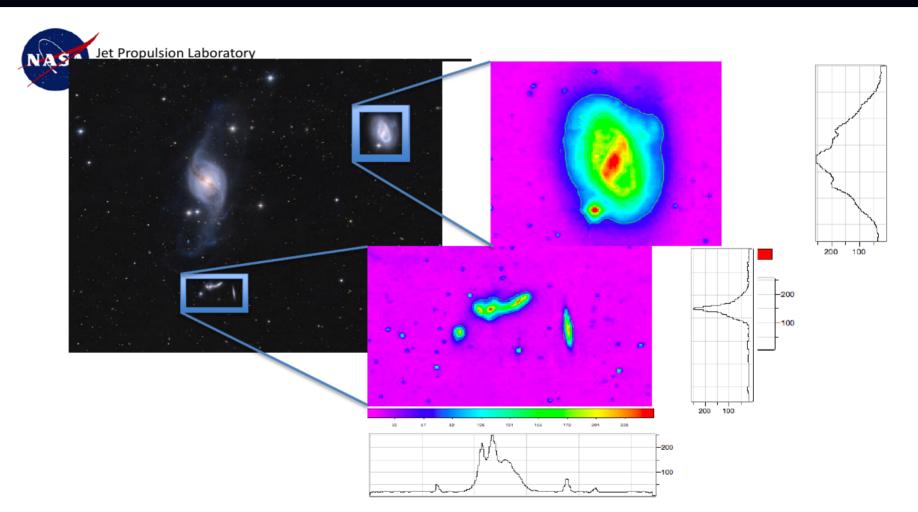
Astro-Neurology

From Sky Surveys to Neurobiology

 Using the data analytics tools based on ML, developed for the analysis of sky surveys, to design a better diagnostics for autism



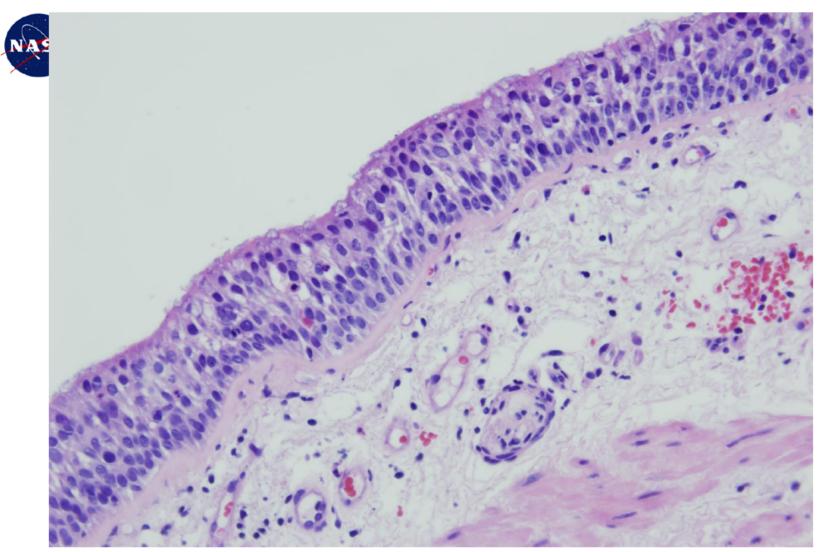
Finding Galaxies by Shape NASA



Description: Detecting objects from astronomical measurements by evaluating light measurements in pixels using intelligent software algorithms.

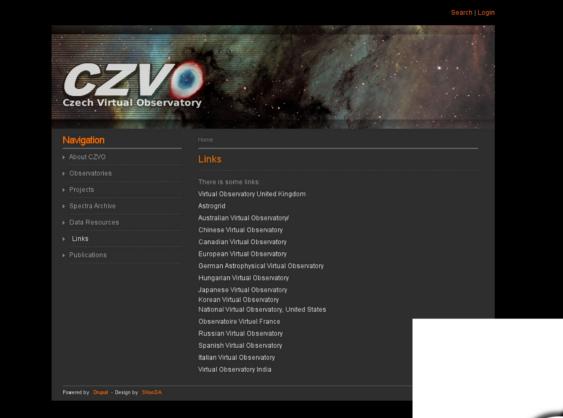
Image Credit: Catalina Sky Survey (CSS), of the Lunar and Planetary Laboratory, University of Arizona, and Catalina Realtime Transient Survey (CRTS), Center for Data-Driven Discovery, Caltech.

Finding Cancer Signatures NASA



Description: Detecting objects from oncology images using intelligent software algorithms transferred to and from space science. Image Credit: EDRN Lung Specimen Pathology image example, University of Colorado

Czech VO - CZVO (not in IVOA)





BT and MT at Faculties of IT astroinformatics and VO

•FIT VUT Brno 2011 1 BT (Random Forests in Astronomy) 1 PhD – Wavelets Dimensionality Reduction (pending)

•VŠB-TU Ostrava 2013+2015 1 BT + 1 MT - SPLAT-VO

•FIT ČVUT (8 BT+5 MT)

- 2012 2 BT (VO-Korel+SSA proxy)
- 2013 2 BT (OSPS Image + Catalogue Server)
- 2014 2 BT (Random Forests + SOM)
- 2015 1 BT (VO-Cloud)
 - 2 MT (Clustering OSPS + Deep Learning)
- 2016 2 MT (Semisupervised learning + Outlier finding)
- 2017 1 MT (VO Cloud) + 1 BT (deep learning)+ 1 PhD (VO light curve)

VO Services of CZVO



Welcome to ASU CAS Data Center.

In addition to the services listed below, on this site you probably can access <u>numerous</u> tables using <u>TAP</u> or <u>form-based ADQL</u>.

Please check out our site help.

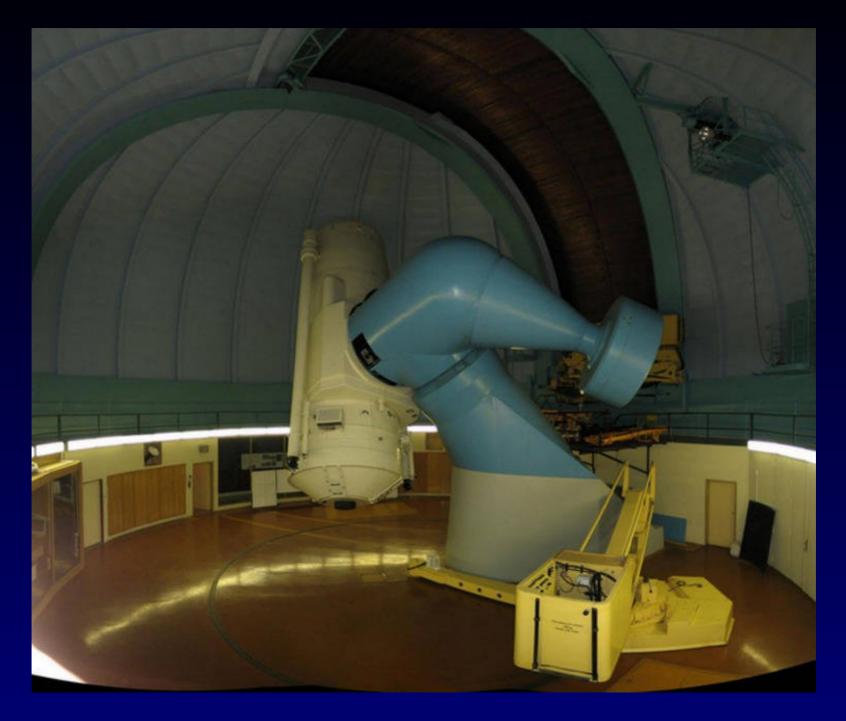
The early stages of development of this archive in years 2013-2015 was supported by grant 13-08195S of Czech Science Foundation.

Its current extended version has been funded by grants COST LD-15113 (spectra and light curves) and INGO LG-15010 (images and photometry) of Czech Ministry of Education Youth and Sport.

Services Available

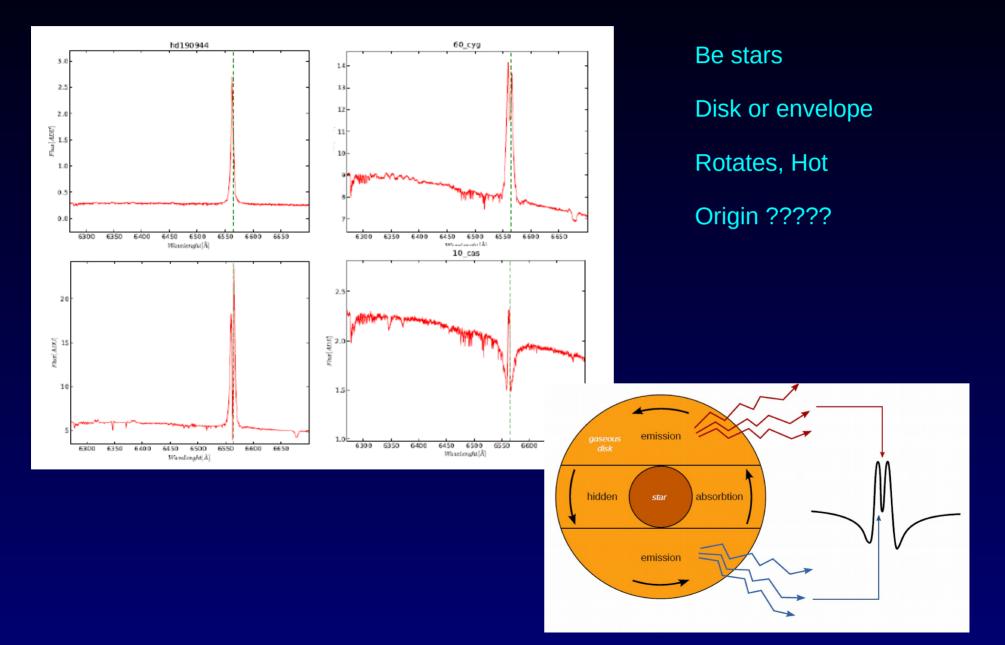
By Title	By Subject	By Author		
►	CCD700 Spectra W	eb Interface	iQ	
►	DK154 Lightcurves	Web Interface	i Q	
	DK154 Ondrejov RA	W observation	ns SIAP 🚺 Q	
	DK154 Ondrejov RE	EDUCED obse	rvations SIAP 🚺 Q	
	DK154 SCS for obs	erved objects	iQ	
►	HEROS Public Spe	ctra Web Interf	face i Q	
	LAMOST DR1 Spec	tra Web Interf	ace i Q	
►	LAMOST DR3 Spec	tra Web Interf	ace i Q	
	LAMOST PILOT Sp	ectra Web Inte	erface i Q	

2m Perek telescope (1967)



Machine Learning of Spectra

Use case: ML of spectra profile of Halpha line (Be stars)



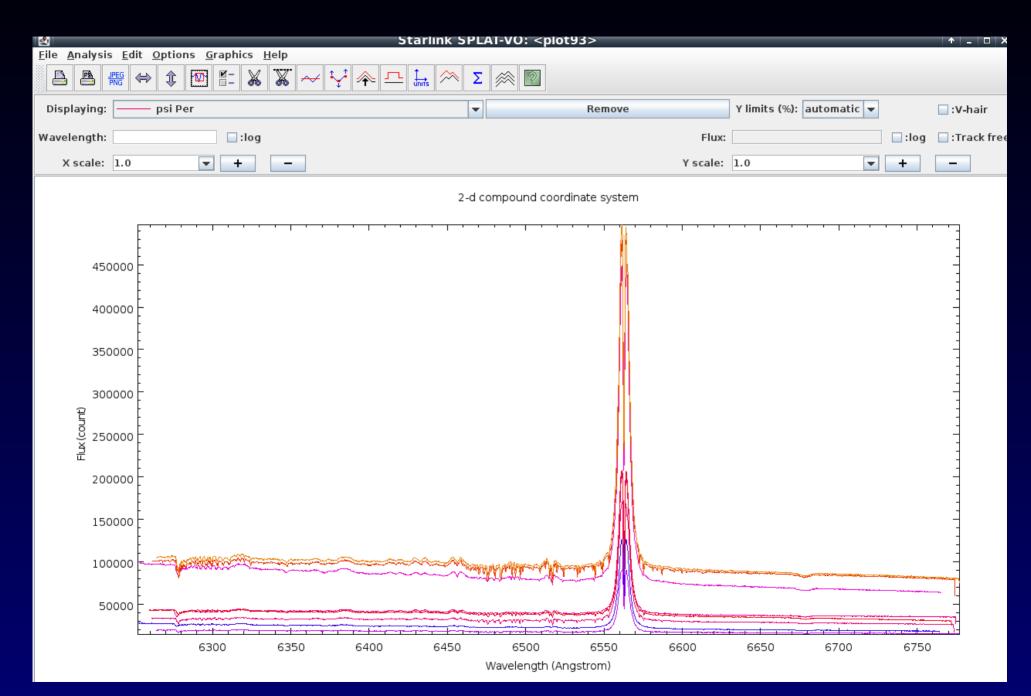
Spectra - query output and previews

CZVO Crach Vintasi Diservatory	CCD700 Spectra V	Veb Inte	rface	•							
Help Service info	Parameters Object standard name: [' 	nsiner'l									
Related CCD700 SSAP Metadata	Result Matched: 44	bolber 1									
ivo://asu.cas.cz/ccd700/q/ Description CCD700 public web interfa	Send via SAMP Quick Plot Product key		Object	Raj2000	Dej2000	Band start [Angstrom]		Date Obs.	Observer	Exp. Time [s]	MHJD
Keywords Optical spectroscopy	tg180048.fit		psi Per	03:36:29.380	+48:11:33.40	6261.00	6773.40	2010-07-19T01:43:36Z		150.0	55396.0696765
Creator [Logo] Created 2014-05-09T10:57:00		<u>oc220022.fit</u>	psi Per	3:37:01.1	48:12:17.1	6262.34	6774.66	2005-03-22T21:16:04Z	Kubat, Sarounova	899.564	53451.8889388
Data updated 2016-06-14 Reference URL Service info		<u>ul070006.fit</u>	Psi Per	03:36:29.380	+48:11:33.40	6252.67	6764.96	2011-12-07T20:21:43Z	Kraus, Kotkova	800.0	55902.8579726
	<u>va300019.fit</u>		psi Per	03:36:29.380	+48:11:33.40	6252.09	6764.42	2012-01-30T17:13:10Z		130.0	55956.7202446
Try ADQL to query our	<u>mi180074.fit</u> ng290040.fit			3:37:03.2 3:36:46.8	48:10:44.2 48:11:39.4	6259.63 6261.83	6772.03 6774.26	2003-09-19T02:34:39Z 2004-07-30T00:40:28Z	Kubat, Kalas	60.0 600.0	52901.1096669 53216.0292701
data. Please report errors and problems to the <u>site operators</u> . Thanks. <u>Privacy Disclaimer</u> Log in		<u>nh100015.fit</u>		3:36:42.4	48:13:02.0	6264.34	6776.77	2004-08-11T01:02:47Z		218.373	53228.0435205
		<u>nh310030.fit</u>	Psi Per	3:36:46.3	48:13:02.4	6264.01	6776.40	2004-09-01T00:22:31Z	Libich, Sarounova	60.0	53249.016454

Spectra in SPLAT-VO - query

2		_	_	Sta	arlink SPL	AT-VO: Query	y VO for S	pect	tra	_	_	_	_	_	↑ □ X
<u>F</u> ile <u>O</u> ptions <u>R</u> eso	lver <u>I</u> nterop <u>H</u> elp														
Service selection of	otions	1 - S	earch parameters	s:											
Data Source		K)	Simple Query						Optional Par	ameters					
Observed data	O Theoretical data	3	Object: psi per			Lookup	Clear		Use	Name		Value		UCD	
	- medicilical da					· ·	Clear		REDSHIF				STC TE	edshift	A
Wave Band			RA: 03:36:29.38	8	Dec:	+48:11:33.48			TARGETC				src.cl		
🗌 Radio 👘 🗌 M	lillimeter 🛛 🗌 Infrare	I R	tadius: 10.0		MAX	REC:			MTIME	2,100			0.0101	000	
🗌 Optical 👘 🗌 U	V 📃 EUV		Band: 6530e-10		/65	580e-10			SPECRP				spect	resolution; em.v	vl
X-ray G	amma-ray 🔽 ALL	100	Time:		/			= [SPATRES				pos.a	angResolution	
	annia-ray 🗹 ALL						1	-111-	PUBDID						
Tags		Q	Query Format:		votab	le		-	CREATOR	IDID			meta	.id	_
	▼ + -	v	Vavelength calibr	ation:	None			↓ Ľ	•						•
		F	lux calibration:		None			-		Se	elect all	Deselect	all Update		
4		TRANSPORT		OUEST auto		1 1 2 2 4 1 6 6 6 6 6 6 6	666 40 10262	22222	222226 50044		-01660		SEEELDAND CE	20	
SSAP Servers		Q	uery: <pre></pre>	=y∪Es1=quer 0	yData&P05=5	4.122416666666	000,48.19263	33333	333330FURMA	ai=votable&SIZE	=0.1006	000000000000	0000%BAND=02	30 SEND Q	UERY
short name 🔺	title		uery results:	~											
BeSS	Be Stars Spectra														
califa ssa	CALIFA DR2		🔏 CCD700-voarcl	hive											
castor	Espadons/Narval leg		ssa specstart s	sa specend	SS	a dstitle	ssa targnar	ne s	ssa dateObs	ssa timeExt 🔻	ssa snr	ssa length	а	ccref	
castor+	Espadons/NARVAL le								53216.0049	1800.		1997		e.asu.cas.cz/ge	. appli 🔺
CCD700-voarchive	ccd700 OND	2		6.76217E-7		psiper/6255-676			54701.97209	1333.19		1997		e.asu.cas.cz/ge	
ccd700-vos2		1	3 6.25854E-7	6.77081E-7	ccd700/data/	psiper/6255-676	psi Per		52982.77964	1300.		1997		e.asu.cas.cz/ge	
CDFS SSAP	Optical Spectroscop	3	1 6.26359E-7	6.73614E-7	ccd700/data/	psiper/6260-673	psi Per		56884.01015	1201.		2047	http://voarchiv	e.asu.cas.cz/ge	. appli
CENCOS-W/DS_DEEP	CENCOS-WDS_DEEP	3	0 6.26359E-7	6.73614E-7	ccd700/data/	psiper/6260-673	psi Per		56884.02565	901.		2047	http://voarchiv	e.asu.cas.cz/ge	. appli
CENCOS-W/DS_DEEP+	CENCOS-WDS_DEEP		2 6.26234E-7	6.77466E-7	ccd700/data/	psiper/6255-676	. psi Per		53451.88616	899.564		1997	http://voarchiv	e.asu.cas.cz/ge	. appli
CfA Hectospec	CfA Hectospec Spec			6.77822E-7		psiper/6255-676			54209.80019	899.214		1997		e.asu.cas.cz/ge	
dk154				6.76496E-7		psiper/6255-676			55902.84843	800.		1997		e.asu.cas.cz/ge	
dk154-extr15	dk154			6.76402E-7		psiper/6255-676			54757.88729	730.		1997		e.asu.cas.cz/ge	
dk154-extr15jan dk154-extr16jan	dk154			6.77426E-7		psiper/6255-676			53216.02811	600.		1997		e.asu.cas.cz/ge	
DK154-SSA	DK154 SSA					psiper/6255-676			54701.96266	600.		1997		e.asu.cas.cz/ge	
ELODIE	ELODIE archive			6.77051E-7		psiper/6255-676			52897.07116	556.599		1997		e.asu.cas.cz/ge	
ELODIEinterp	Spectrum interpolat			6.77822E-7		psiper/6255-676			54209.81274	450.		1997		e.asu.cas.cz/ge	
ESO SAF SSA	ESO Science Archive					psiper/6255-676			52982.79788	350.		1997		e.asu.cas.cz/ge	
EUVE	Extreme Ultraviolet			6.77492E-7 6.77640E-7		psiper/6255-676 psiper/6255-676			53475.85182 53249.01697	300.043 300.		1997 1997		e.asu.cas.cz/ge	
extract jul16	DK154					psiper/6255-676 psiper/6255-676			54519.80405	300.		1997		e.asu.cas.cz/ge e.asu.cas.cz/ge	
F/H Orders SSAP	Flash/Heros Split-Or	2				psiper/6255-676 psiper/6255-676			54519.80405	300.		1997		e.asu.cas.cz/ge e.asu.cas.cz/ge	
FEROS SSAP	FEROS Public Spectra	4		6.77677E-7		psiper/6255-676			53228.0436	218.373		1997		e.asu.cas.cz/ge e.asu.cas.cz/ge	
Flash/Heros SSAP	Flash/Heros SSAP	1		6.76875E-7		psiper/6255-676			55443.86444	200.		1997		e.asu.cas.cz/ge	
FUSE	Far Ultraviolet Spect					psiper/6255-676			54519.81162	200.		1997		e.asu.cas.cz/ge	
GALEX	Galaxy Evolution Exp					psiper/6255-676			52901.91749	194.574		1997		e.asu.cas.cz/ge	
GIRAFFE	GIRAFFE archive of r			6.77178E-7		psiper/6255-676			52904.03609	180.		1997		e.asu.cas.cz/ge	
H.E.S.S.	High Energy Stereos			6.77340E-7		psiper/6255-676			55396.07196	150.		1997		e.asu.cas.cz/ge	
HEAVENS @ ISDC	Mining the HEAVENS		4 0 050005 7	C 70440E 7			Den Den		FF0F0 71740	100		1007	hate the ended.		
HEROS OND	HEROS archive of On														▶
	cutout server of HE	- 8	Display		isplay	Downloa			ownload	Dese			eselect	DataLi	
	•		selected	a	Ш	selected		al	11	table	e	a	Ш	Service	es
Select all	Deselect all				Г	(111)									
Query registry	Add New Server	and a second				Save quer	y results		Restore qu	ery results	Clo	ose			

Spectra in VO - direct access plot

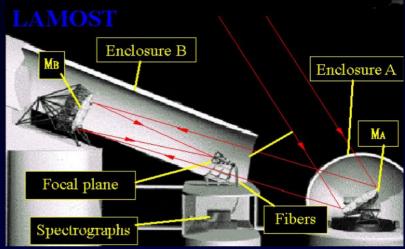


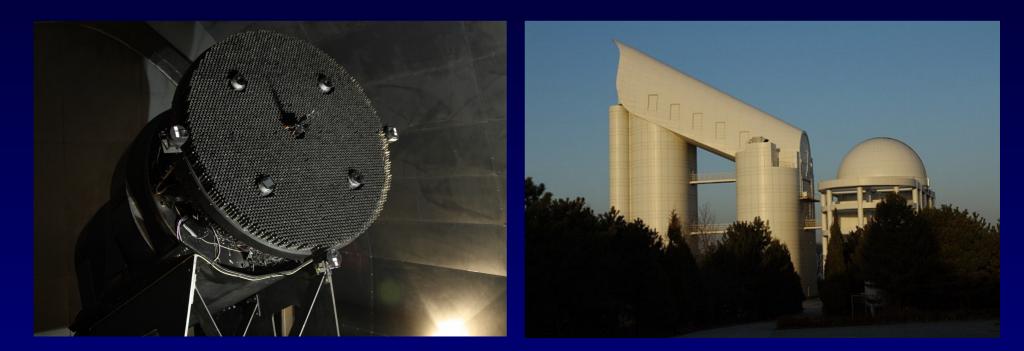
Spectra in SPLAT-VO - DataLink

🖞 Starlink S	5PL	AT-VO: <plot96></plot96>	_		_	↑ _ □ ×
<u>File Analysis Edit Options G</u> raphics <u>H</u> elp						
	Σ					
Displaying: psi Per	-	Remove		Y limits (%): automatic 💌		🗌 :V-hair
Wavelength: 6589.852			Relative flux:	1.022293	🔲 : log	🗌 :Track free
X scale: 1.0 💌 🛨 🗕			Y scale:	1.0	+	-
Parameters for Server-Generated data processing FLUX normalized LAMB 6540e-10 [6.24978e-076.77822e-07] m LAMB 6590e-10 [6.24978e-076.77822e-07] m FORM application/x-votable+xml Clear parameters Set parameters 4 4 4 2.5 2 1.5 2 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5	L .	ompound coordinate system				
		Wavelength (Angstrom)				

LAMOST (Guoshoujing)

Xinglong- China 4m mirror (30 deg meridian) 4000 fibers 10 mil spectra / 5 yr Automatic RV-z





LAMOST Spectral Surveys

DR1 (end 2013) 2 204 860 spectra 1 085 404 stars

DR3 (half 2015)5 755 126 spectraDR4 (Feb 2016)+ 741 522

Each Fiber – 2 motors double arm 33mm circle

Fibre collects light from 3.3 arcsec circle on sky

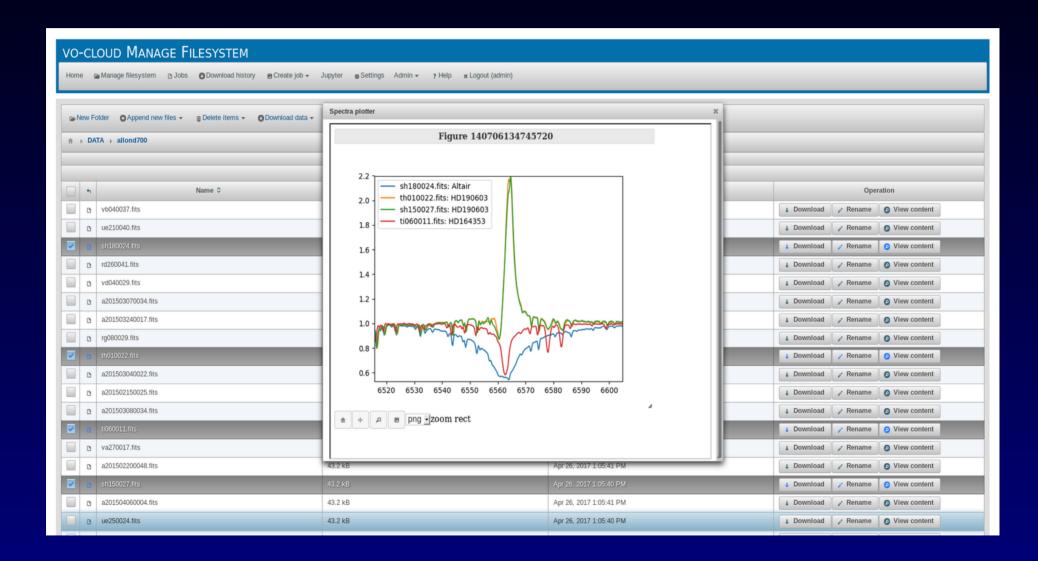


SOM Worker example

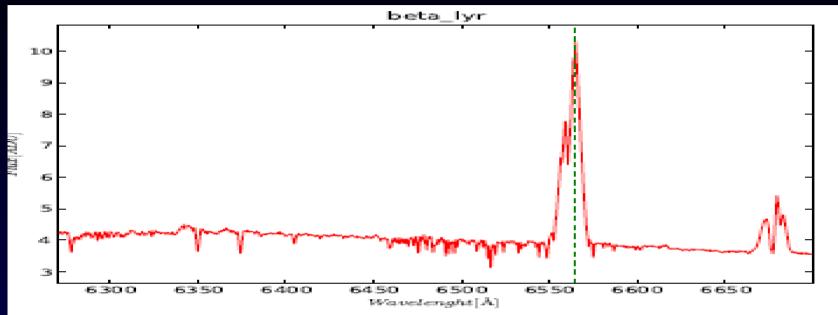
Type Id 1-92										
ın again 🗙 Delete	Phase COMPLETED	Worker local worker	Created 2/14/17 11:32:47 PM	Started 2/14/17 11:32:4	Finished 7 PM 2/14/17 11:32:59	Execution PM 11 sec	g time			
v x.html - Fullscreen						_				
0				_	VO-CLOUD DE	TAILS OF JOI	В			
					Home 🕞 Manage files	stem 🖻 Jobs 🔘	Download history	Create job 👻 Jupyter	ø Settings Admin + ?	Help x Logout (admin)
				•	AllSpecOndSOM Type	Id	Phase	Worker	Created	Started Finished
5 -		• •		• -		1-92	COMPLETED	local worker	2/14/17 11:32:47 PM	2/14/17 11:32:47 PM 2/14/17 11:32:59 PM
	•••				Preview	ele				
		• • • •	•••	•	index.html - Fullscre	n				
10 -		•••		• -	4.4					Neuron x: 18 y: 1
	• • • • •		•	:	4.2		1			All Associated Spectra
					3.8					Y+
15 -	•••••		•••••		3.6					X- HOME X+
15				•	3.4					Y-
	• • • • •	• • • •	• • • • •		3					Display reference vector
	• • • • •				2.8					Display all spectra
20 -			1		2.4					1. no name class: x
	0 5	10	15	20	2.2					2. no name class: x 3. no name class: x
					1.8					4. no name class: x
					1.6					5. no name class: x 6. no name class: x
					1.4					7. no name class: x 8. no name class: x
					1.2		barghorm		m	9. no name class: x
					100-201					
					0.8	M.J. M. M. M. K.				
					0.8	Mili, Mi Miliha				

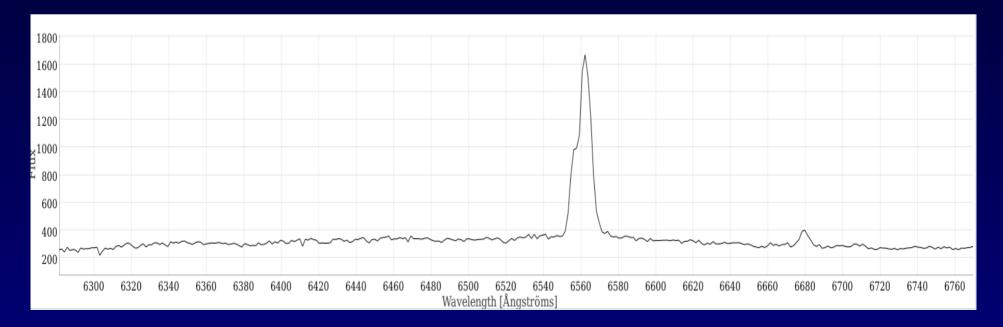
xecuting time

VO-CLOUD spectra visualisation

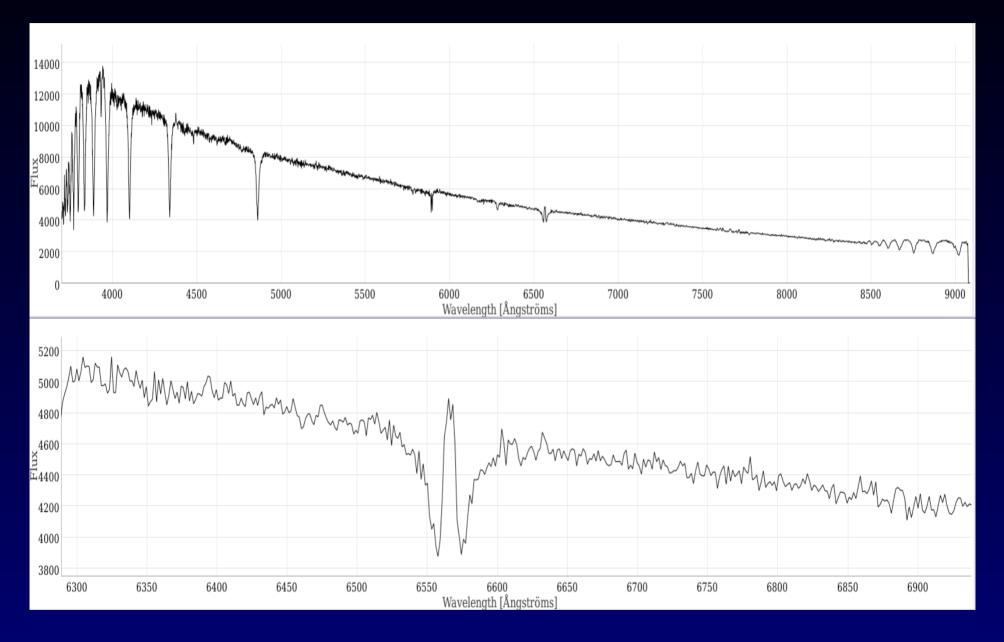


Be Candidates Found





Yet Unknown Be Star (UCAC ...)



Virtual Observatory inside

- OND 2m archive on SSAP protocol (spectra access)
- LAMOST DR1 on SSAP (using DaCHS)
- Preprocessing (rectify, cutout) DataLink on server
- SAMP (send spectra to SPLAT-VO view details)
- Visualization on sky ALADIN, X-MATCH CDS
- VO-CLOUD cloud engine based on UWS REST jobs
- Cross-matching (ADQL, TAP, TOPCAT, TAPhandle, pyVO, Vizier)
- Very similar methodology in e.g. Mass spectrography
- DNA analysis search for patterns

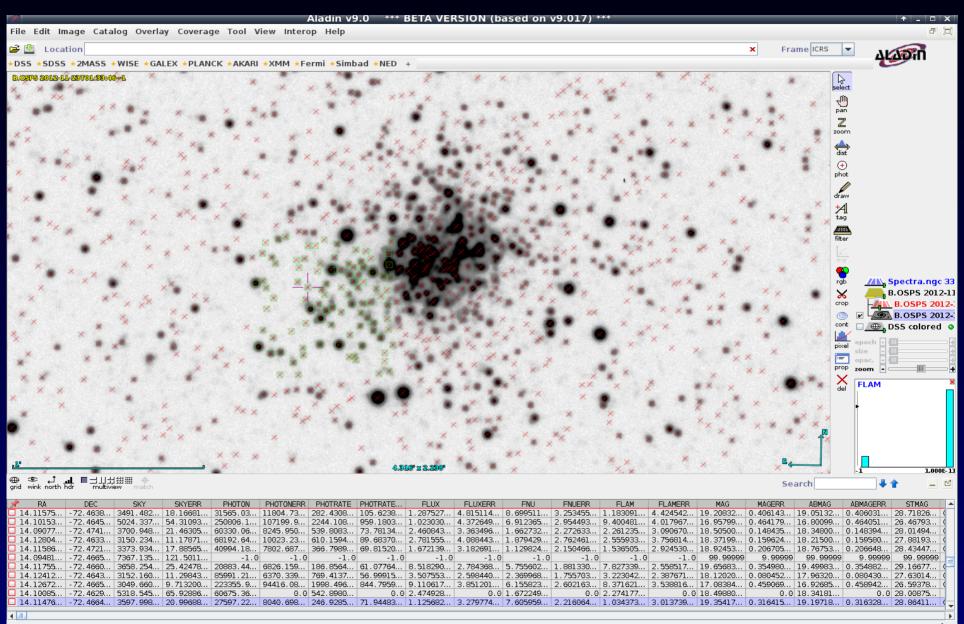
Danish 1.54m at La Silla robotized in Summer 2012



Danish 1.54m Telescope



Reduced OSPS image + bintable photometry in 2nd extension



P: Load files quickly => Drag&Drog from your desktop

SIAP - Raw images query

CZV0 Lech Virtual Diservatory	DK154 Ond	drejov RAW observations SIAP
Help	Observations cap	tured by ASU CAS facility by DK154 telescope
Service info	Position [deg]	ngc 330 ICRS Position, RA,DEC, or Simbad object (e.g., 234.234,-32.45)
Vetadata Identifier ivo://asu.cas.cz/dk154_rav	Field size [deg]	0.5 Size in decimal degrees (e.g., 0.2 or 1,0.1)
Description Observations captured by . Keywords DK154	Intersection type	 Image overlaps Rol Image covers Rol Rol covers image The given position is shown on image Relation of image and specified Region of Interest.
Creator [Logo]	File format	 ANY ○ image/png ● image/fits Requested format of the image data
Created 2012-04-27T00:00:00 Data updated	Img_type	ALL Type of observation (SCIENCE, FLAT, or BIAS)
2016-03-12	Band [m]	ALL Vavelength (range) of interest (or symbolic bandpass names)
Reference URL Service info	Minimum Date	Minimum date (If empty, returns everything until Maximum date)
<u>Try ADQL</u> to query our data.	Maximum Date Table	Image:
Please report errors and problems to the <u>site operators</u> . Fhanks. <u>Privacy Disclaimer</u>	Output format	Limit to 100 V items.
<u>_og in</u>		60

Raw images results

DK154 Ondrejov RAW observations SIAP

Parameters

Band: ALL
 Field size: 0.05
 File format: image/png

- Img_type: ALL
 Position: ngc 330

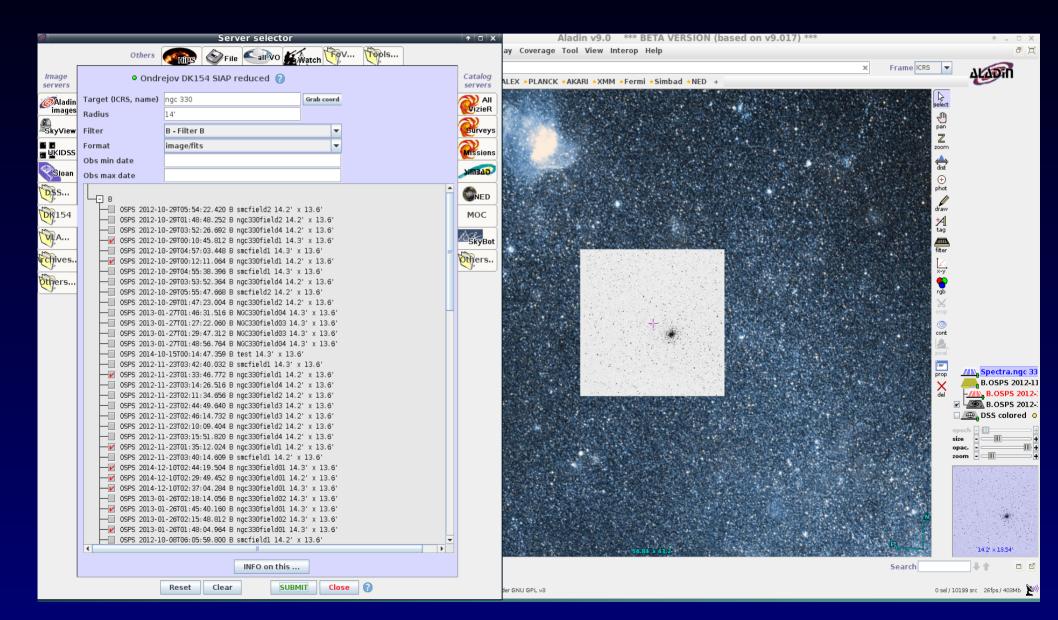
Result

Matched: 100

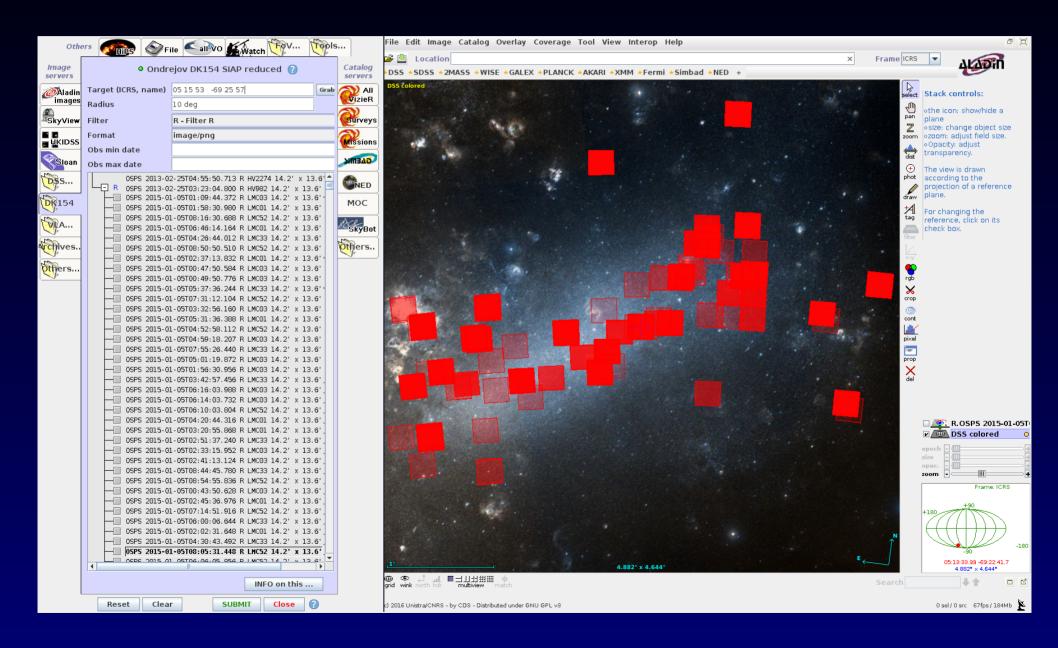
Send via SMIP Quick Plot

The query limit was reached. Increase it to retrieve more matches. Note that unsorted truncated queries are not reproducible (i.e., might return a different result set at a later time).																													
Accref	Owner	Embargo ends	Туре		ze F	Ctr. Ct RA Do deg][de	ec	Title	Instrument	Obs. date #		Axes Lengths [pix]	Scales [deg/pix]		Equino [yr]	ox Proj	. Ref. pixel [pix]	Ref. values [deg]	CD matrix [deg/pix]	Bandpass Ba	ndpass unit	Band Ref. [m]	Band upper [m]	Band Iower F [m]		[deg] 1	Exp. To time [s]	elescope	Img_type
nge330/ieid1_000003.png	beusers	2010-12-31 00:00:00	image/png	g 247.6	3kiB 14	ł.19 -72	2.45 20	DSPS 2012-10-08T03:32:47.212 / ngc330field1	DFOSC_FASU	2012-10-08T03:32:46Z 2		[2148, 2048]	[0.000109996] 0.000109996]	ICRS	2000.0	TAN	[1074.0, 1024.0]	[14.193875, -72.453333333333333	[-0.000109996 -9.82956e-07, -9.82956e-07, 0.000109996]	V m		5.4e-07	6.7e-07	4.85e-07 N	14 -72 14 1/A -72 13 -72 13	olygon ICRS 1.5844654144 2.5665328579 1.5862737491 2.3413642854 12.3392662543 3.7962514963 2.564408589	20.0 D)K-1.54	LIGHT
ngc330field1_000002.png	beusers	2010-12-31 00:00:00	image/png	234.7	7kiB 14	4.19 -72	2.45 20	DSPS 2012-10-08T03:29:15.728 3 ngc330field1	DFOSC_FASU	2012-10-08T03:29:15Z 2		[2148, 2048]	[0.000109996, 0.000109996]	ICRS	2000.0	TAN	[1074.0, 1024.0]	[14.193875, -72.4533333333333333	[-0.000109996 -9.82956e-07,] -9.82956e-07, 0.000109996]	Bm		4.4e-07	5.6e-07	3.6e-07 N	14 -72 14 1/A -72 13 -72 13	olygon ICRS 4.5844654144 2.5665328579 4.5862737491 2.3413642854 3.807794488 2.3392662543 3.7962514963 2.5664408589	80.0 D	IK-1.54	LIGHT
ngc330field1_000005.png	beusers	2010-12-31 00:00:00	image/png	266.2	žkiB 14	4.19 -72	2.45 20	DSPS 2012-10-08T03:37:43.924 R ngc330field1	DFOSC_FASU	2012-10-08T03:37:42Z 2	2	[2148, 2048]	[0.000109996, 0.000109996]	ICRS	2000.0	TAN	[1074.0, 1024.0]	[14.193875, -72.4533333333333333	[-0.000109996 -9.82956e-07,] -9.82956e-07, 0.000109996]	R m		6.2e-07	8.5e-07	5.8e-07 N	14 -72 14 1/A -72 13 -72 13	olygon ICRS 4.5844654144 2.5665328579 4.5862737491 2.3413642854 3.807794488 2.3392662543 3.7962514963 2.564408589	20.0 D	IK-1.54	LIGHT
ngc330field1_000004.png	beusers	2010-12-31 00:00:00	image/png	246.5	5kiB 14	4.19 -72	2.45 20	DSPS 2012-10-08T03:35:12.456 / ngc330field1	DFOSC_FASU	2012-10-08T03:35:12Z 2	2	[2148, 2048]	[0.000109996, 0.000109996]	ICRS	2000.0	TAN	[1074.0, 1024.0]	[14.193875, -72.4533333333333333	[-0.000109996 -9.82956e-07,] -9.82956e-07, 0.000109996]			5.4e-07	6.7e-07	4.85e-07 N	14 -72 14 13 -72 13	olygon ICRS 4.5844654144 (2.5665328579 4.5862737491 (2.3413642854 12.3413642854 (2.3392662543 3.7962514963 (2.564408589	20.0 D	IK-1.54	LIGHT
ngc330field1 000006.png	beusers	2010-12-31 00:00:00	image/png	g 260.8	3kiB 14	4.19 -72	2.45 20	DSPS 012-10-08T03:40:09.216 R ngc330field1	DFOSC_FASU	2012-10-08T03:40:09Z 2	2	[2148, 2048]	[0.000109996] 0.000109996]	' ICRS	2000.0	TAN	[1074.0] 1024.0]	[14. 193875, -72.4533333333333333	[-0.000109996 -9.82956e-07,]-9.82956e-07, 0.000109996]	R m		6.2e-07	8.5e-07	5.8e-07 N	14 -72 14 1/A -72 13 -72 13	olygon ICRS 1.5844654144 2.5665328579 1.5862737491 2.3413642854 2.3439642854 2.3392662543 3.7962514963 2.564408589	20.0 D	IK-1.54	LIGHT
ngc330field1_000001.png	beusers	2010-12-31 00:00:00	image/png	235.3	3kiB 14	4.19 -72	2.45 20	DSPS 2012-10-08T03:25:50.480 3 ngc330field1	DFOSC_FASU	2012-10-08T03:25:50Z 2	2	[2148, 2048]	[0.000109996, 0.000109996]	ICRS	2000.0	TAN	[1074.0, 1024.0]	[14.193875, -72.453333333333333	[-0.000109996 -9.82956e-07, 9.82956e-07, 0.000109996]	B m		4.4e-07	5.6e-07	3.6e-07 N	14 -72 14 I/A -72 13 -72	olygon ICRS 1.5844654144 2.5665328579 1.5862737491 2.3413642854 18.807794488 2.3392662543 3.7962514963 0.5514963	80.0 D)K-1.54	LIGHT

OSPS SIAP in Aladin (DSS in back)



OSPS Image coverage (footprints)



Data Cubes = Scientific Data Structure



International Virtual Observatory Alliance

IVOA N-Dimensional Cube Model Version 1.0 IVOA Working Draft 20150320

This version: WD-CubeDM-1.0-20140930 Previous version(s):

Editor(s): Mark Cresitello-Dittmar

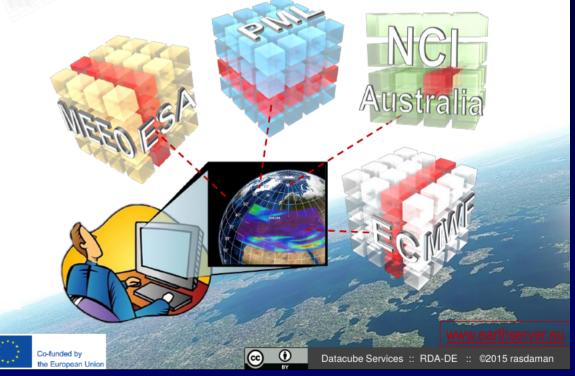
Authors:

Doug Tody, Francois Bonnarel, Omar Laurino, Mireille Louys, Arnold Rots, Jose Enrique Ruiz, Jesus Salgado, and the IVOA Data Model Working Group.

Datacubes as New Paradigm for Agile Analytics

rasdaman

IACOBS



OSPS Light curves - plot (customized)



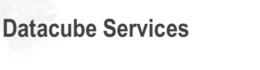
Time Series Cube Data Model

Version 1.1

IVOA Note 2017-02-05

Working group Time domain interest group This version http://www.ivoa.net/documents/cubeDM/20170205 Latest version http://www.ivoa.net/documents/cubeDM Previous versions Author(s)

Jiří Nádvorník, Petr Škoda, Dave Morris, Pavel Tvrdík Editor(s) Jiří Nádvorník

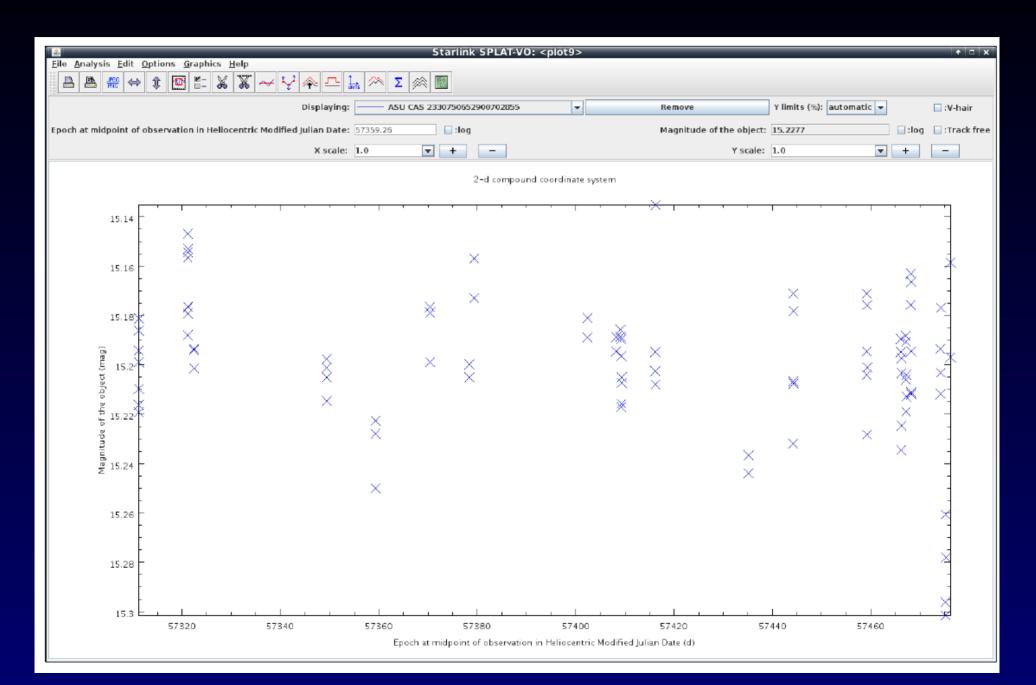




OSPS Light curves in SPLAT-VO using Sparse Cube DM

Starlink SPLAT-VO: Ouerv VO for Spectra 1 D X File Options Resolver Interop Help Search parameters: Service selection options Simple Ouerv Data Source Optional Parameters Use Name Value UCD Observed data Theoretical data Object: ogle Imc-dpv-056 Lookup Clear REDSHIFT src.redshift Wave Band RA: 05:15:53.21 Dec: -69:25:57.9 TARGETCLASS src.class Radio Millimeter Infrared MAXREC: Radius: 0.05 MTIME UV SPECRP spect.resolution:em.wl Optical EUV Band: SPATRES pos.angResolution 🔲 Gamma-ray 🔽 ALL X-ray Time: PUBDID votable Query Format: . CREATORDID meta.id Tags 4 Wavelength calibration: None . T + Deselect all Update Select all Flux calibration: None -Ouery: <SERVER>?REQUEST=queryData&POS=78.97170833333332,-69.43275&FORMAT=votable&SIZE=8.333333333333334E-4 SEND QUERY SSAP Servers short name 🔺 title Ouerv results: 6dF DR3 Simple Spe... 6dF Spectra 👗 dk154-extr16jan BEFS Berkeley Extreme an. max date ssa targname accref ssa bandpass ssa length ssa producttype location arr BeSS Be Stars Spectra mime (78.971725, -69.43276 1 ASU CAS 2329093547214535910 56954.38908 http://vos2.asu.cas.cz/getpro... B 9 timeseries application/x-votable+xml califa ssa CALIFA DR2 2 ASU CAS 2329093547214535910 56953.98628 http://vos2.asu.cas.cz/getpro... V 8 timeseries application/x-votable+xml (78.971725, -69.43276 Espadons/Narval leg. castor (78.971725, -69.43276 3 ASU CAS 2329093547214535910 56953.98739 http://vos2.asu.cas.cz/getpro... | I 8 timeseries application/x-votable+xml Espadons/NARVAL le. castor+ 4 ASU CAS 2329093547214535910 56953.99348 http://vos2.asu.cas.cz/getpro... U 8 timeseries application/x-votable+xml (78.971725, -69.43276 CCD700-voarchive ccd700 OND 5 ASU CAS 2329093547214535910 56953.98521 http://vos2.asu.cas.cz/getpro... R 9 timeseries application/x-votable+xml (78,971725, -69,43276 ccd700-vos2 CDFS SSAP Optical Spectroscop. CENCOS-WDS DEEP CENCOS-W/DS DEEP. CENCOS-W/DS DEEP+ CENCOS-WDS DEEP. CfA Hectospec Spec. CfA Hectospec dk154 dk154-extr15 dk154-extr15jan dk154 dk154-extr16ian dk154 DK154-SSA DK154 SSA ELODIE ELODIE archive ELODIEinterp Spectrum interpolat. ESO SAF SSA ESO Science Archive. Ъ EUVE Extreme Ultraviolet .. Download Download DV1 E 4 Display Display Deselect Deselect DataLink autroat julie **₽** selected all selected all table all Services Select all Deselect all Save query results Restore query results Close Query registry Add New Server

OSPS Light curves - plot (customized)



Conclusions

- Machine learning on Big Data archives may identify new interesting objects yet unknown
- Global Data Federation from Multiple fields
- Crucial is interactive visualization
- Future science will be multidisciplinary
- Wide collaboration of experts and informaticians
- Education of new expert Data Scientist
- Transfer of technology commercial interest
- VO-like technology helps in every step