Archivy 2m dalekohledu

Od Reticonu k umělé inteligenci

Virtuální observatoř a Astroinformatika

Petr Škoda

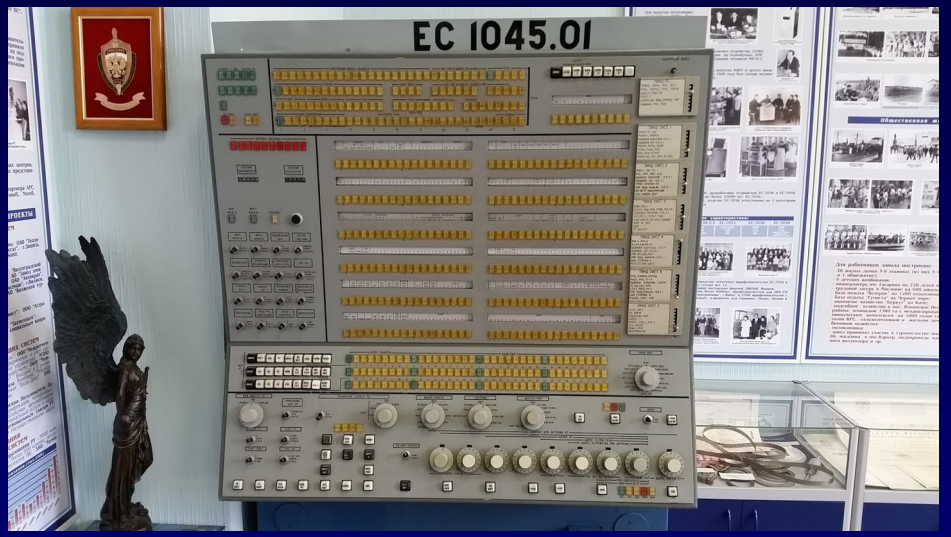
Astronomický ústav AVČR
Ondřejov ČR

Posezení k 50. výročí Perkova dalekohledu
Sídlo AVČR, Národní třída, Praha

Sálové počítače

- Teorie + zpracování měření EC1040 (<1MB)
- EC 1045 – 4 MB – terminál sítě – TSO
- Děrné štítky , pásky, mg pásky

  FORTRAN66 (EXEC FORTGCLG, FORT SYSIN DD *)



Přechod k PC

- 1988 rekonstrukce TCS 2m – Villati - 1. PC !
  - PC AT 6/10MHZ , 1MB RAM, EGA 640x350, 16c
- RS232 – komunikace s VILATI (TMS9900)
- Disky 8“ 360KB
- Totéž řízení 2m (ASCOL)

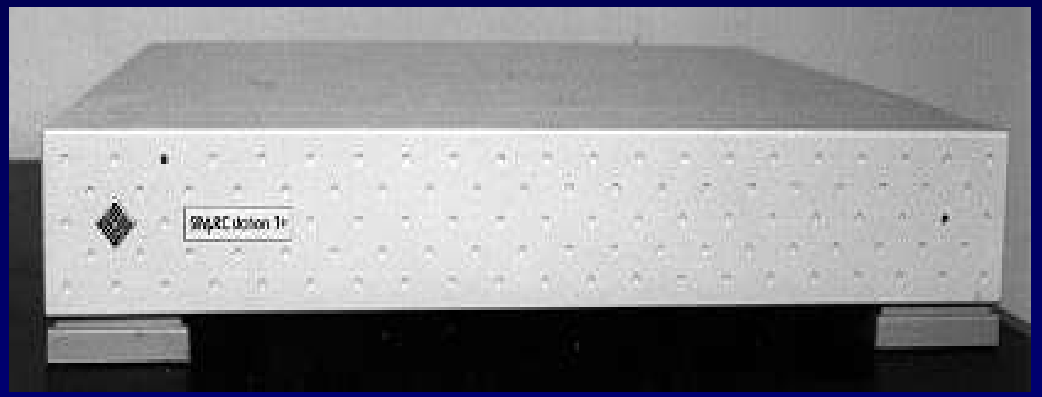
Rozvoj malé výpočetní techniky na AÚ

- 1989 (AÚ – Několik AT, XT, PP06), MS XENIX
- 1990 Několik AT + 386/33 MHz, 4M,
  - 80MB disk
  - Fireball (Ceplecha), PCMOS, Windows 2.0
  - LaTeX (ChiWriter, T602)

- Stelární - první LAN – 10Net (2linka 1Mb)
  - Mezi 3 PC AT

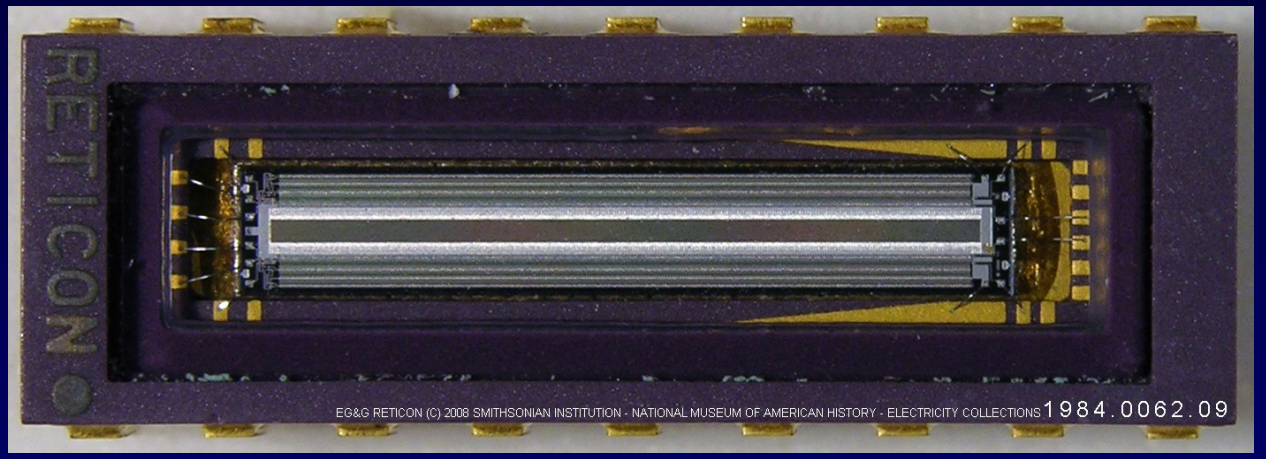
Rozvoj IT 2m - UNIX

- 1991 1 UNIX WS = Sun SPARC 1 CompuAdd
  - SPARC 20MHz, 16MB RAM, 210 MB SCSI
  - Ethernet – WD8003, 10Mb, koax , T 50 Ohm
  - PCTCP (DOS) + Syntax Lan Manager (SMB)
  - TCPIP LAN
  - Několik 486/66 + modem 2400 baud
  - Mail server – e-MAIL na IBM VM/SP Dejvice
  - EARN



Začátek digitální éry 2m

- 1992 Reticon 1872 AF (Lickova obs.,S Vogt)



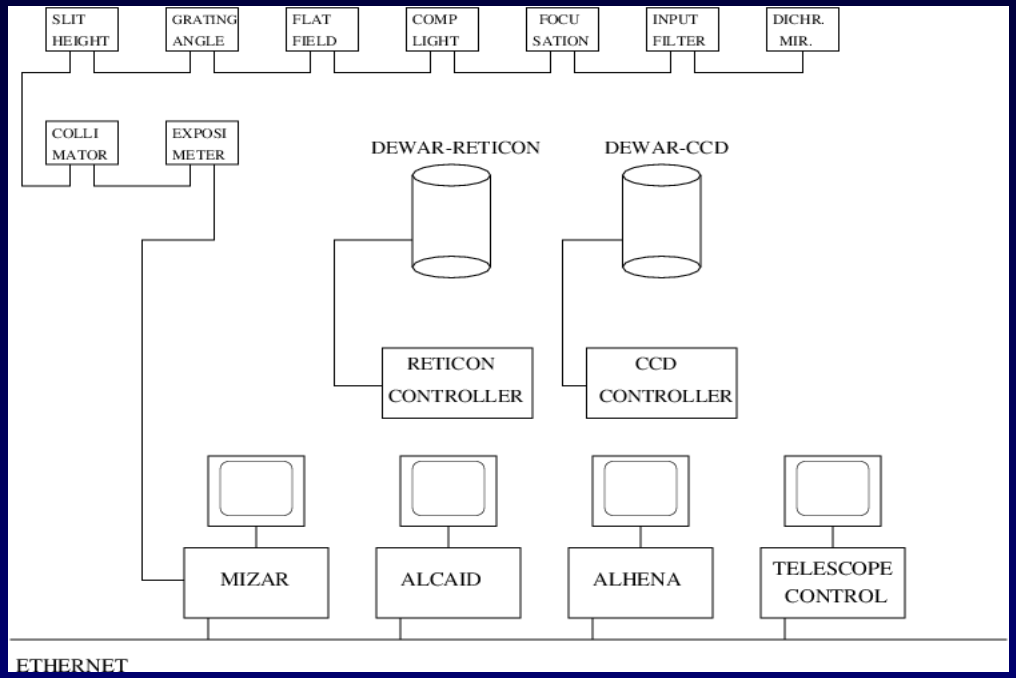
Řídící SW Reticonu

- Modifikace SPEFO (1988)
- Horn - RETICON.EXE, prosinec 1994 umírá
- ERETICON.EXE

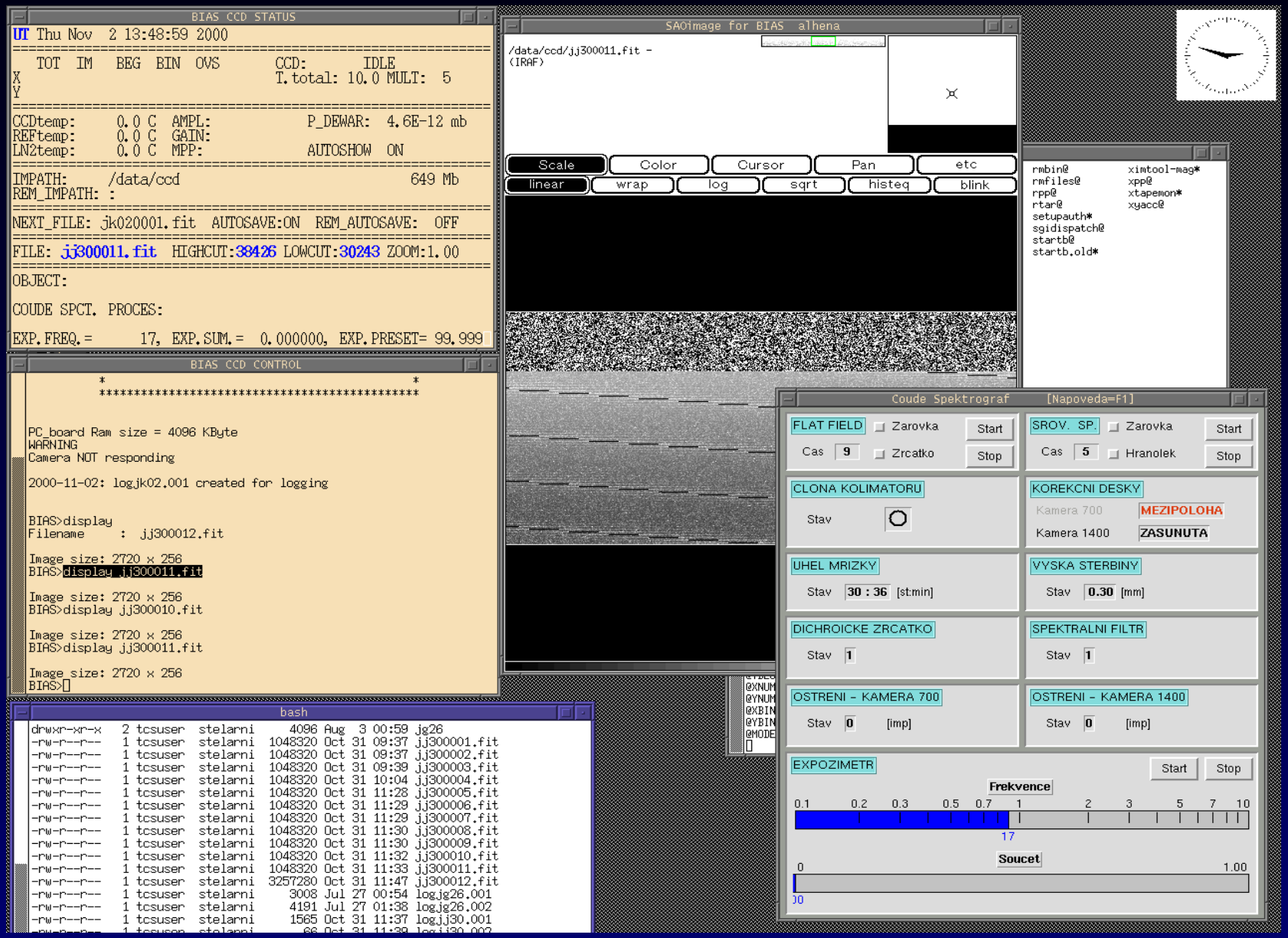
Robotizace spektrografu 1996

- CRETICON.EXE
  - řízení coudé krabiček – Ing. Wudia
  - váhovaní seeingu , zaostřeni cross-korelací

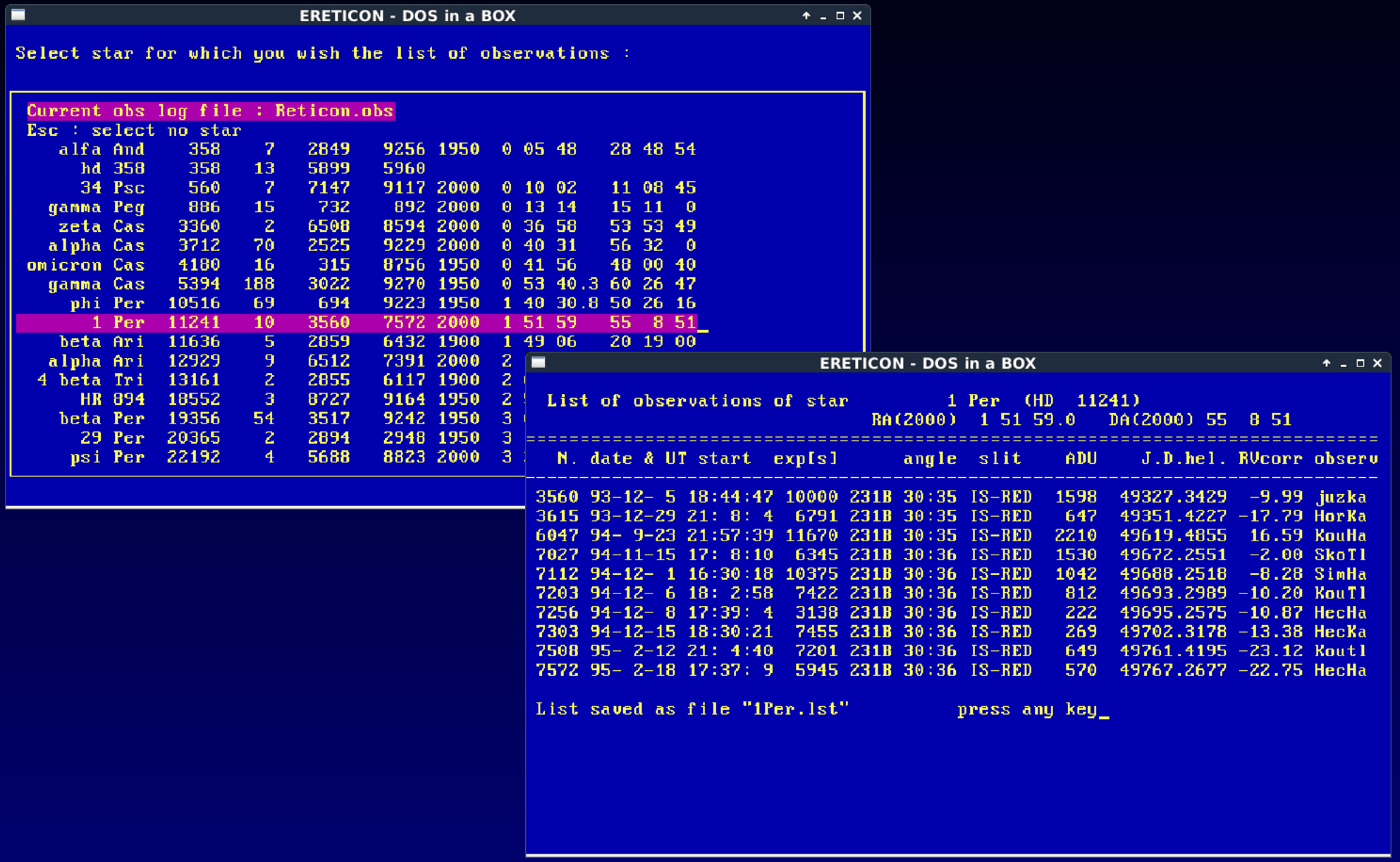
komunikace DOS-UNIX , centrální archív



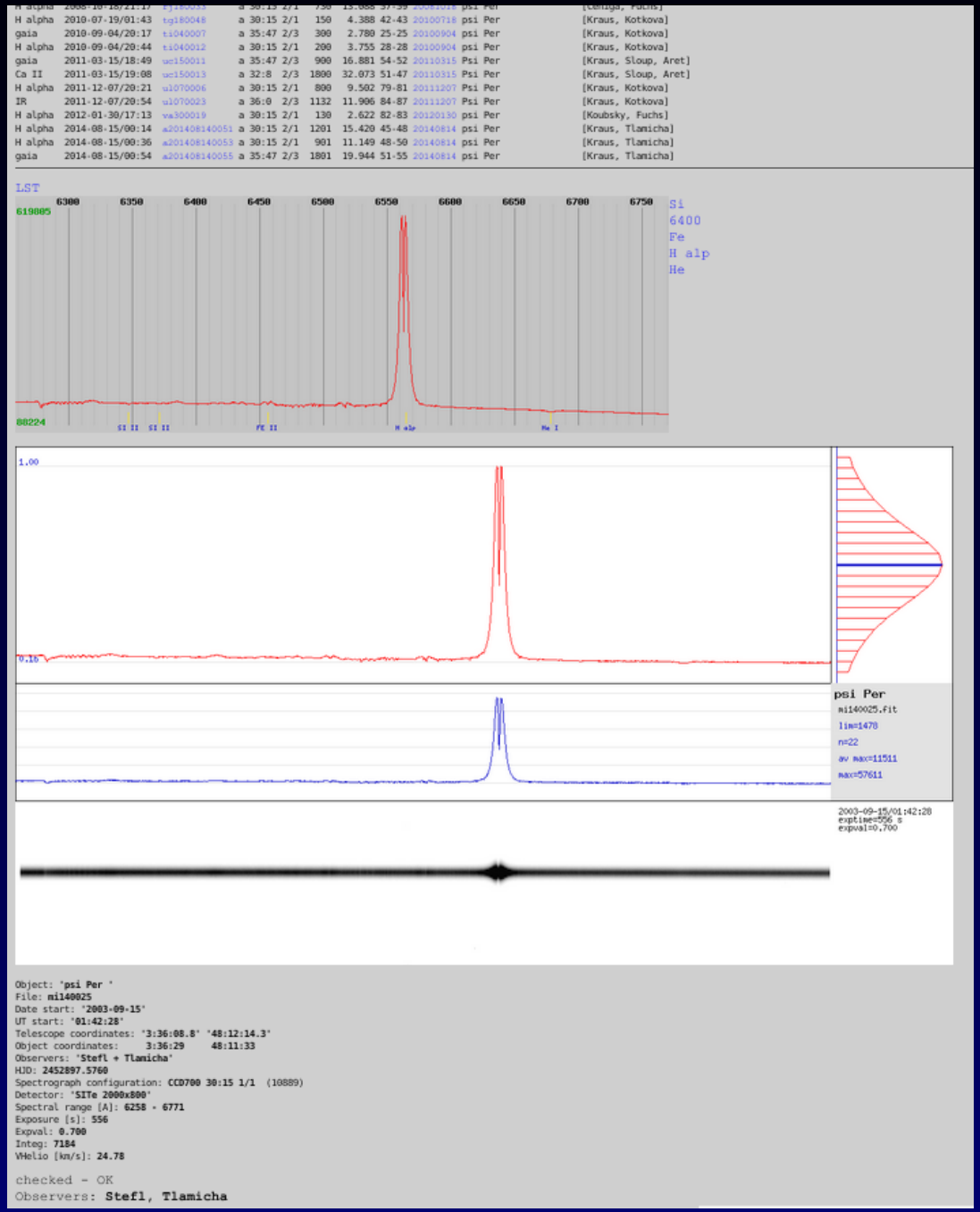
BIAS -CCD LORAL 2720x256 – od 1996
SiTe 2000x800 -od 2000



Archiv součást RETICONu



Spectra archive – raw + reduced



List of files – main metadata

Reduced spectrum (if exists)

Raw spectrum – maximum in COP

Raw spectrum – averaged all COP

Raw spectrum – inverse image

All metadata about frame

Lenka Šarounová-Kotková PHP

Internet v astronomii

- 1992 únor – Internet v ČSFR (EARN-> FERNET,CESNET)
- Na ASU – cca 1993 duben (brána EARN)
- Gopher, WAIS, Veronica, začátek WWW, Telnet
- Duben 1995 CDS Strasbourg

  Weaving the Astronomy Web
  - Použití WWW pro astronomii
  - Databáze ADS
  - Astrophysical Journal – on-line
  - Proroctví o konci tištěné verze za 20 let
  - NCSA Mosaic -> Netscape 1.0

WWW Archiv Reticonu

- ADASS – říjen 1995 – Tucson
  - komunita na celý život -SW, detektory, algoritmy
  - Školení IRAF, CCD laboratoř , prohlídka KPNO
- 1998 GRANT MŠMT INFRA2 LB98251
  - Budování archivu Reticonu
  - SQL Adabas, PostgresQL, WDB (ESO 1995)
  - 2000 - Archiv 2M byl jedním ze 6 světových archivů pozemních dalekohledů na WWW
    - ESO, CFHT, AAO Okyama, LaPalma INT
    - +logy Asiago, OHP ELODIE
- 2005 Šárka Zehnalová UJEP - do Java Struts

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"

IVOA

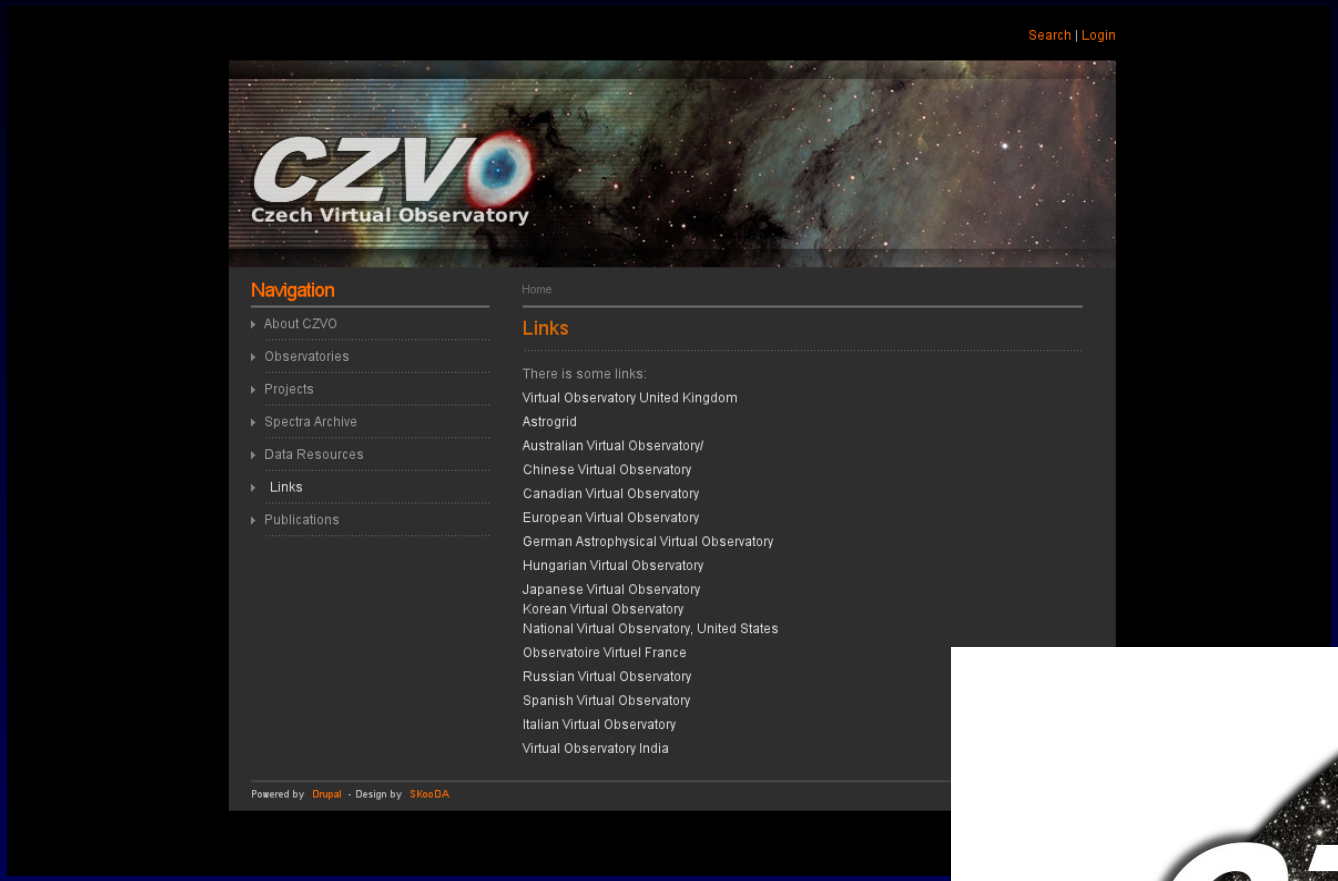

Astro Grid
ArVO
American Virtual Observatory
Aus-VO
BRAVO
China-VO
CVO
esa
France
ESO
IVOA
GERMAN ASTROPHYSICAL
GAVO
VIRTUAL OBSERVATORY
EURO VO
HVO
Hungarian Virtual Observatory
JVO
NOVA
RVO
SVO
vo-i
VAO
VIRTUAL ASTRONOMICAL OBSERVATORY
VObs.it

Czech VO - CZVO

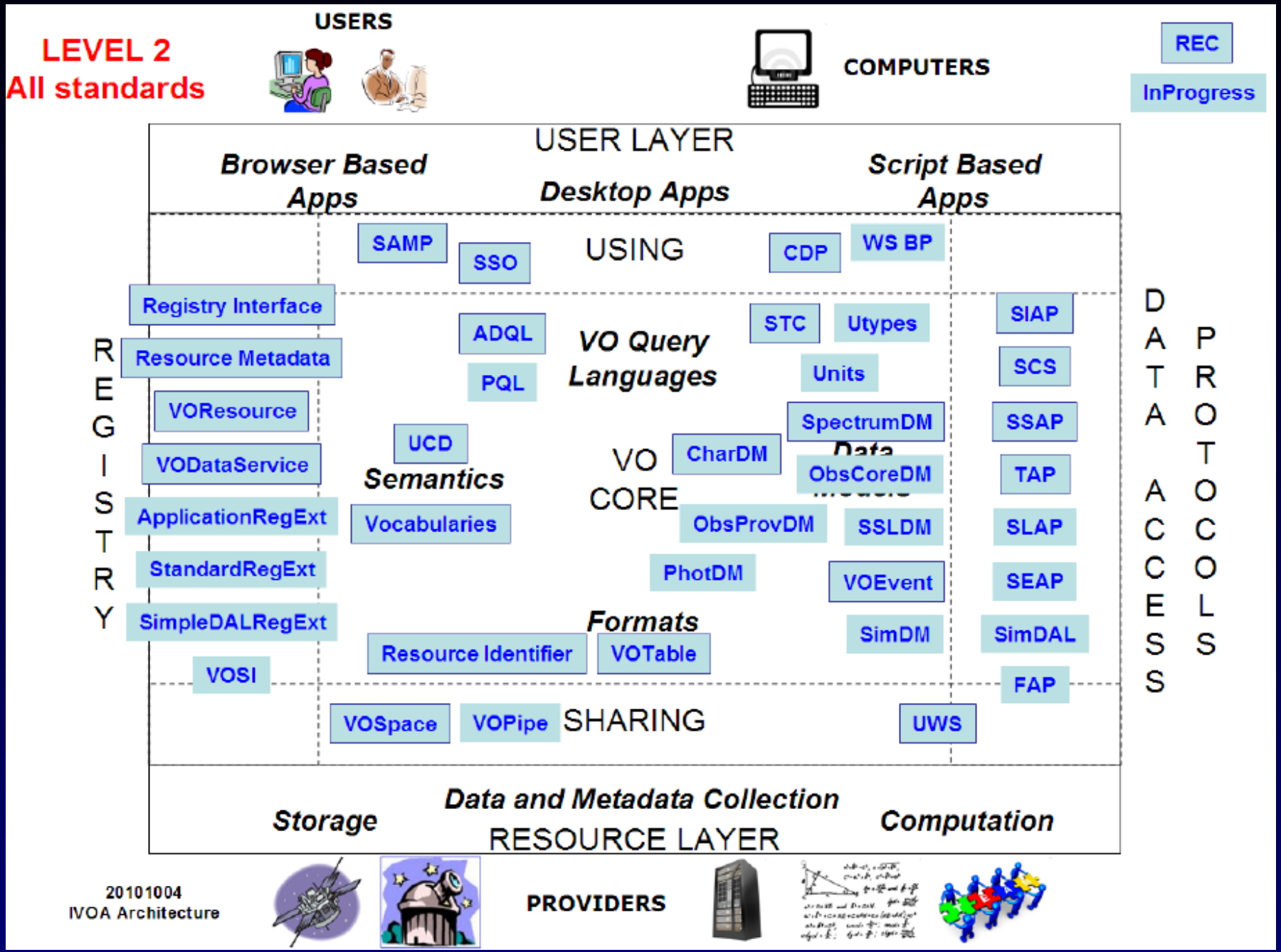
Od 2006 po IAU

Podpora EURO-VO

2008 HEROS archive ve VO Prugniel OHP
Cutout server



Ecosystem of VO – level 2



Simple Spectra Access Protocol
Spectral Data Model

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

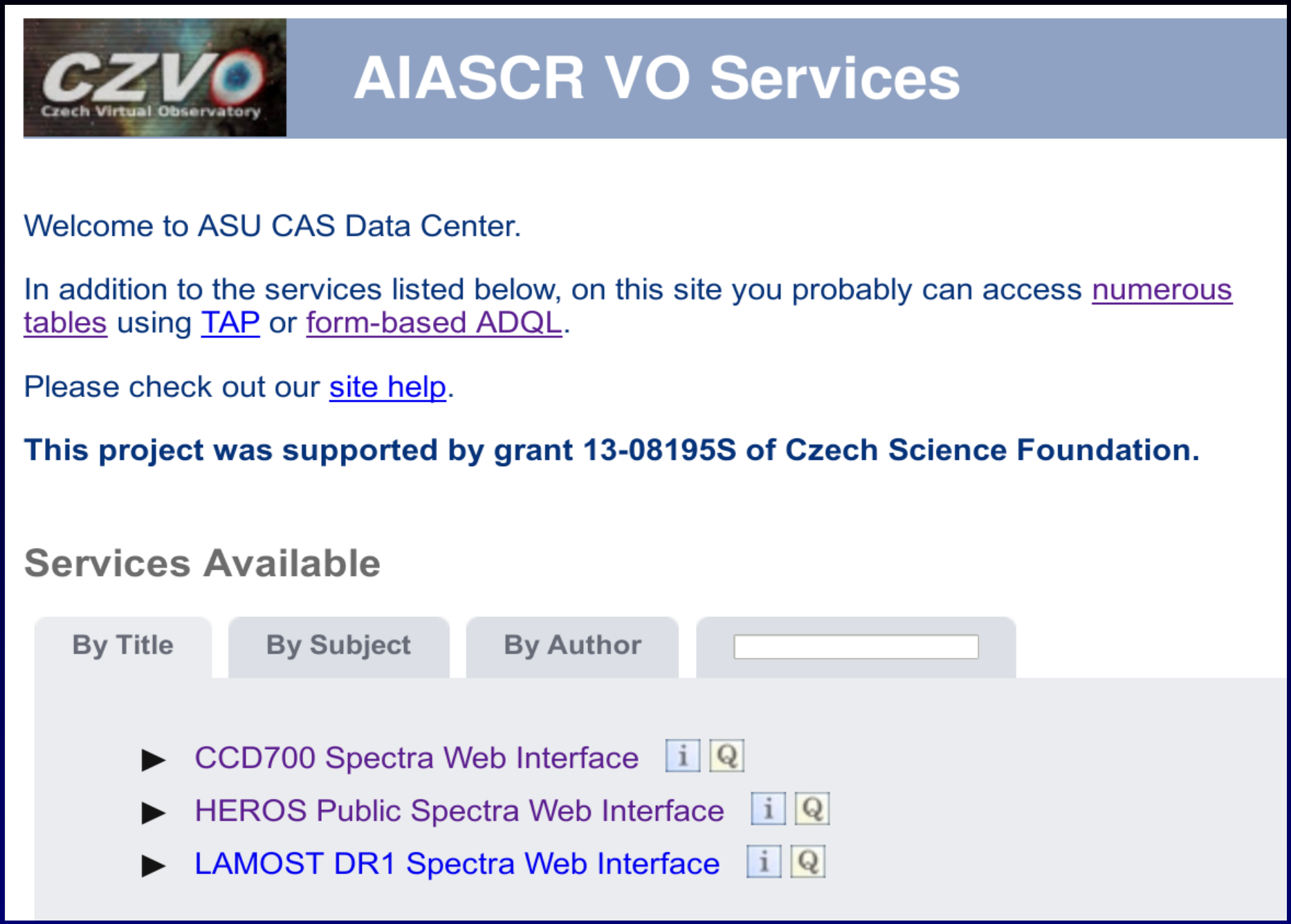
Version 1.03

IVOA Recommendation 2007-10-29

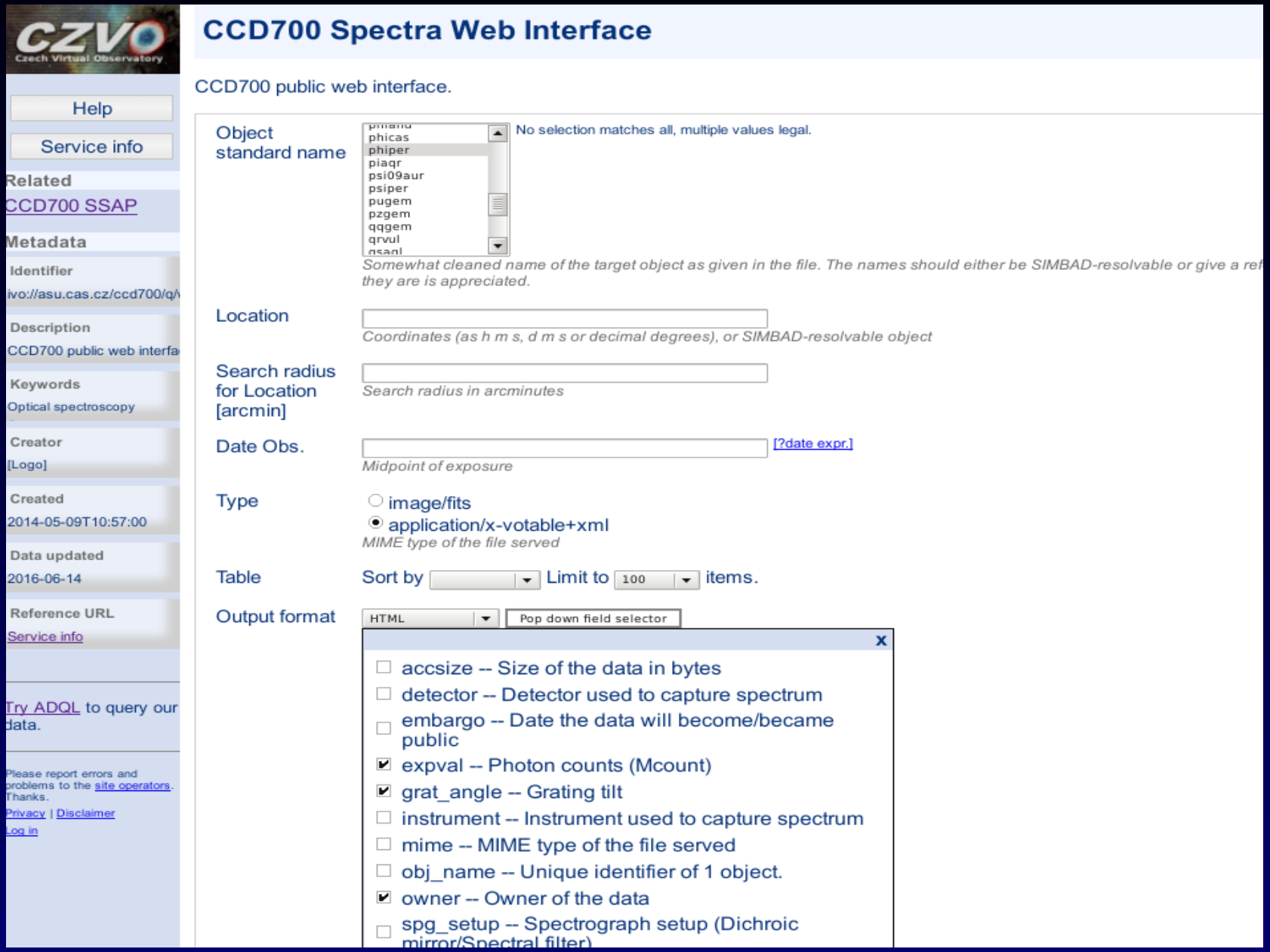
This version (Recommendation Rev 1)
http://www.ivoa.net/Documents/REC/DM/SpectrumDM-20071029.pdf
Latest version:
http://www.ivoa.net/Documents/latest/SpectrumDM.html
Previous versions:
http://www.ivoa.net/Documents/PR/DM/SpectrumDM-20070913.html

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.

VO access to spectra - Browser



Spectra query - web browser



Spectra - query output and previews

CZVO Czech Virtual Observatory

Help

Service info

Related

CCD700 SSAP

Metadata

Identifier

ivo://asu.cas.cz/ccd700/q/

Description

CCD700 public web interfa

Keywords

Optical spectroscopy

Creator

[Logo]

Created

2014-05-09T10:57:00

Data updated

2016-06-14

Reference URL

Service info

Try ADQL to query our data.

Please report errors and problems to the site operators. Thanks.

Privacy | Disclaimer

Log in

CCD700 Spectra Web Interface

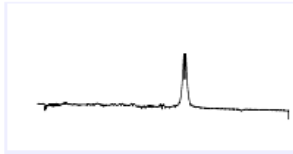
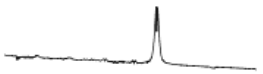
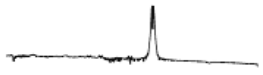
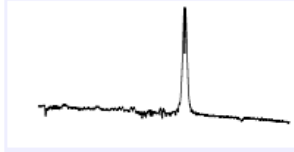
Parameters

- Object standard name: ['psiper']

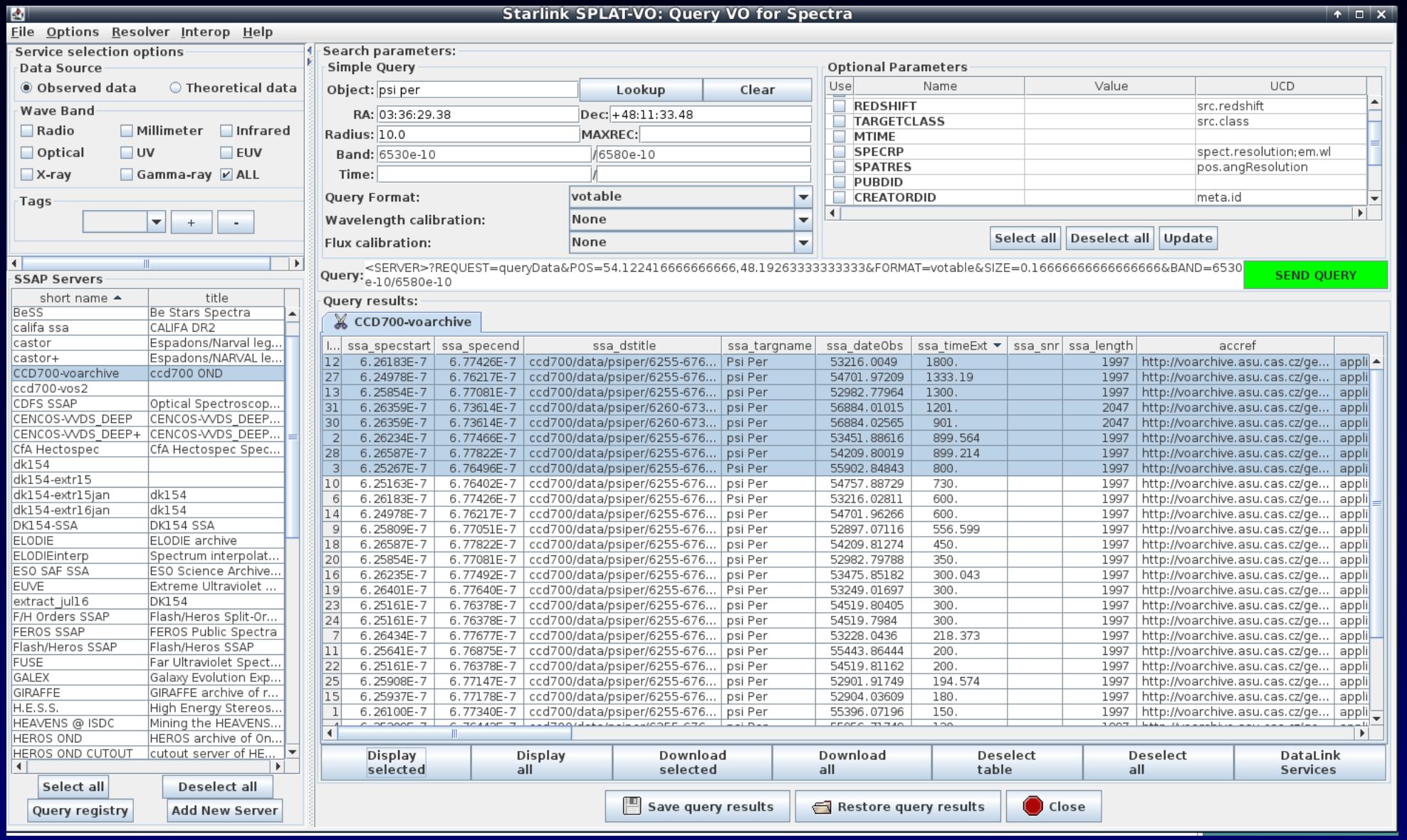
Result

Matched: 44

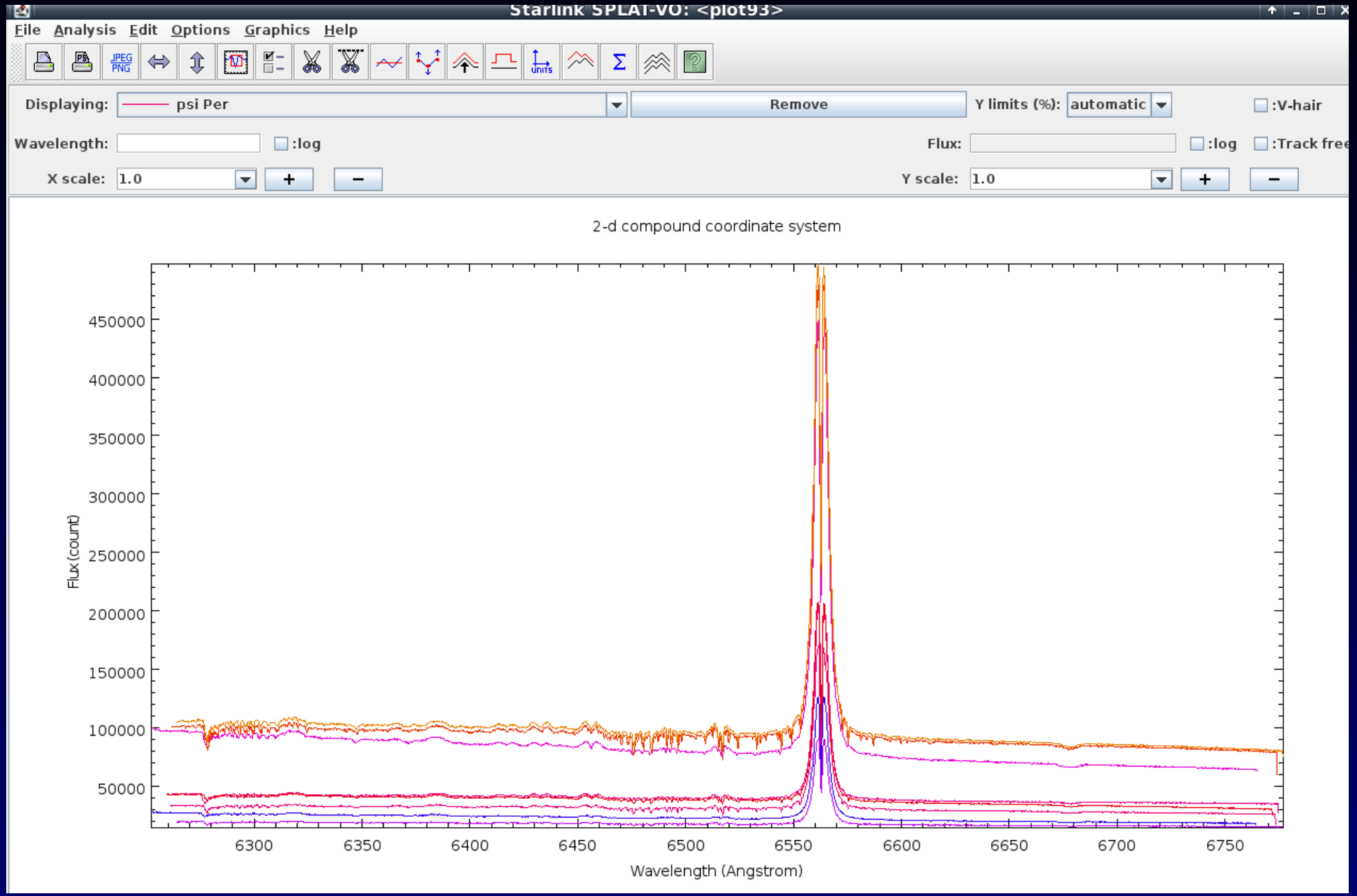
Send via SAMP | Quick Plot

Product key	Object	Raj2000	Dej2000	Band start [Angstrom]	Band end [Angstrom]	Date Obs.	Observer	Exp. Time [s]	MHJD
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
 oc220022.fit	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
 ul070006.fit	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
va300019.fit	psi Per	03:36:29.380	+48:11:33.40	6252.09	6764.42	2012-01-30T17:13:10Z		130.0	55956.7202446
mi180074.fit	psi Per	3:37:03.2	48:10:44.2	6259.63	6772.03	2003-09-19T02:34:39Z		60.0	52901.1096669
ng290040.fit	Psi Per	3:36:46.8	48:11:39.4	6261.83	6774.26	2004-07-30T00:40:28Z	Kubat, Kalas	600.0	53216.0292701
 nh100015.fit	psi Per	3:36:42.4	48:13:02.0	6264.34	6776.77	2004-08-11T01:02:47Z	Stefl + Rezna	218.373	53228.0435205
 nh310030.fit	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

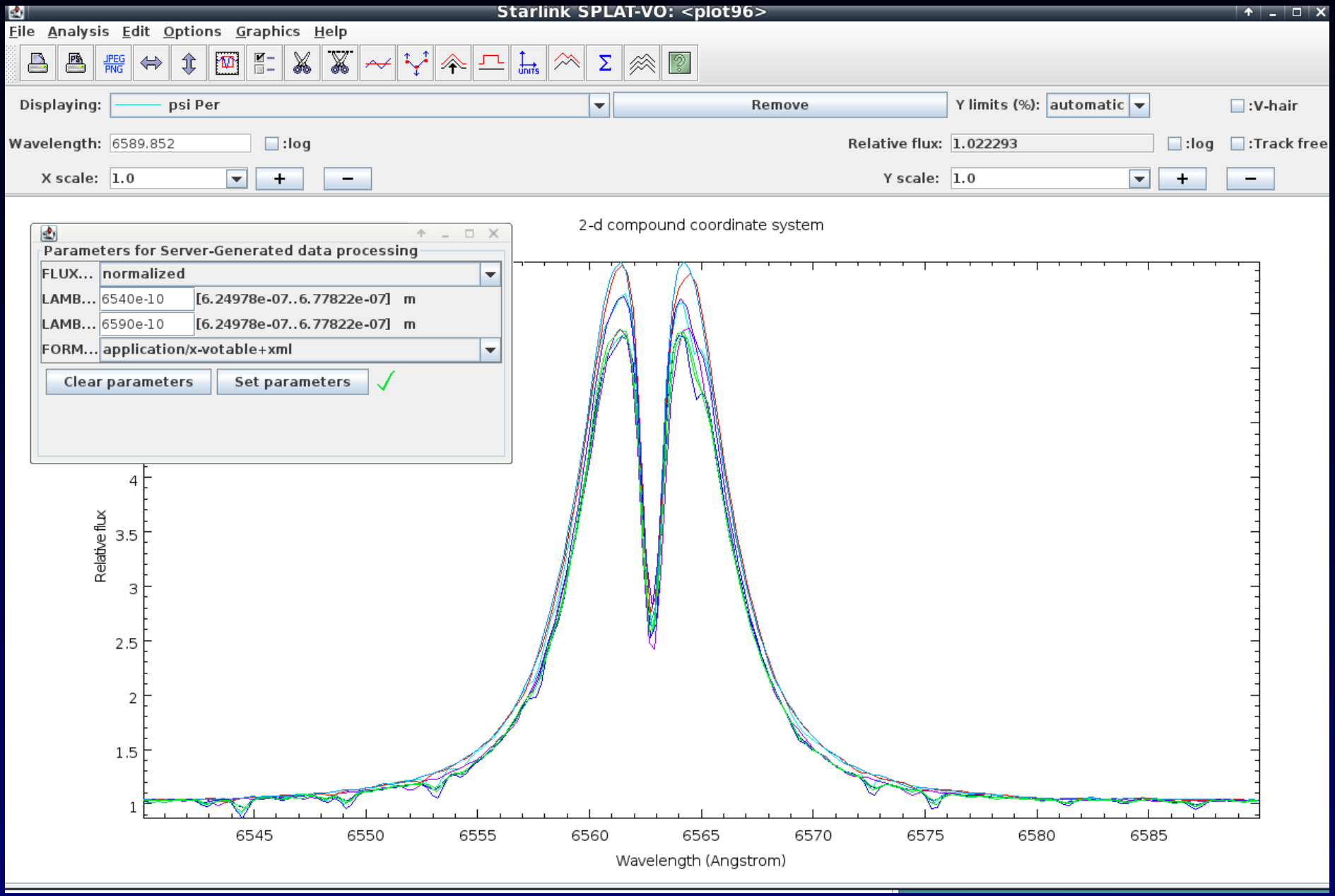
Spectra in SPLAT-VO - query



Spectra in VO – direct access plot



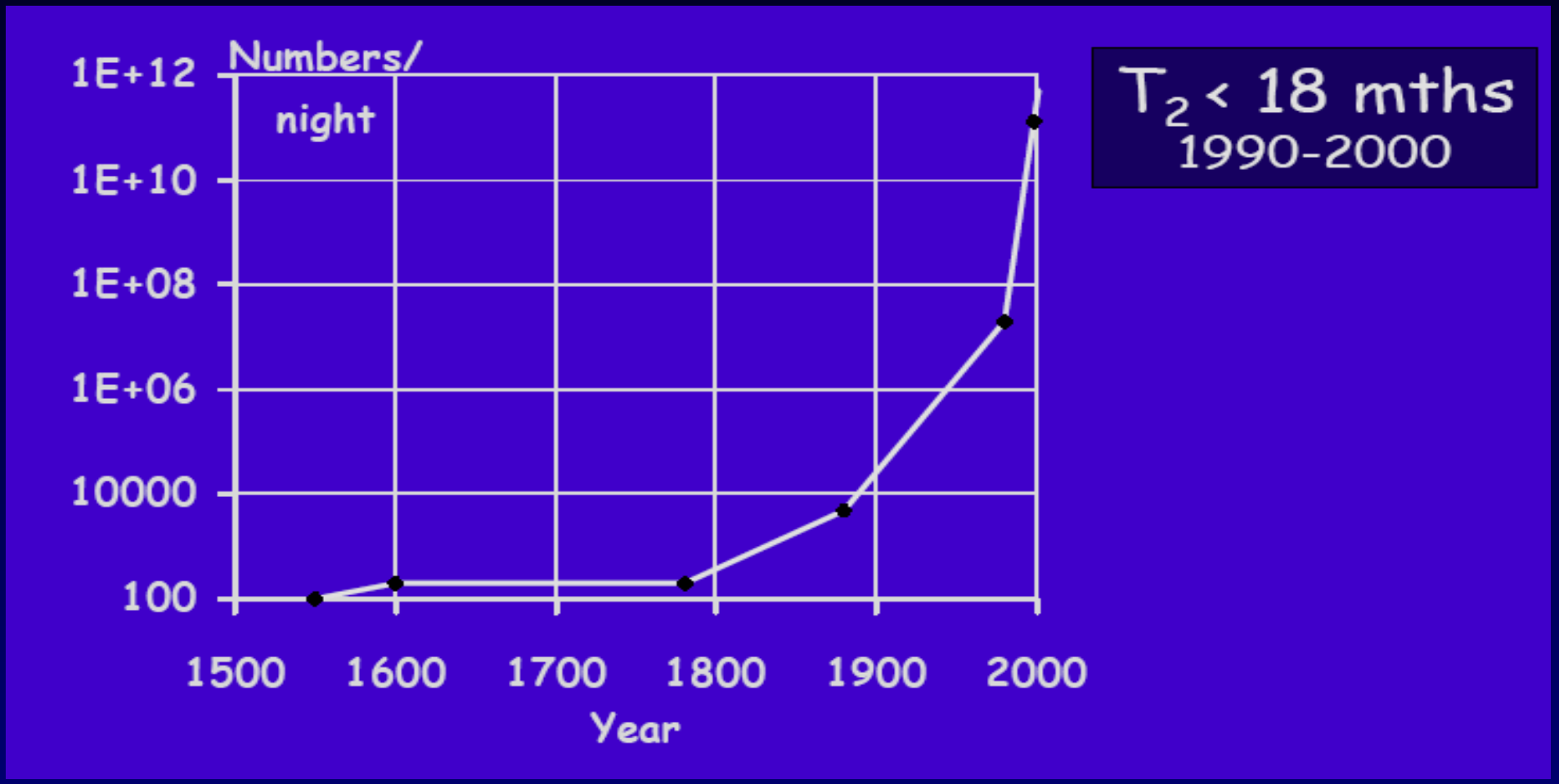
Spectra in SPLAT-VO - DataLink



Lavina dat

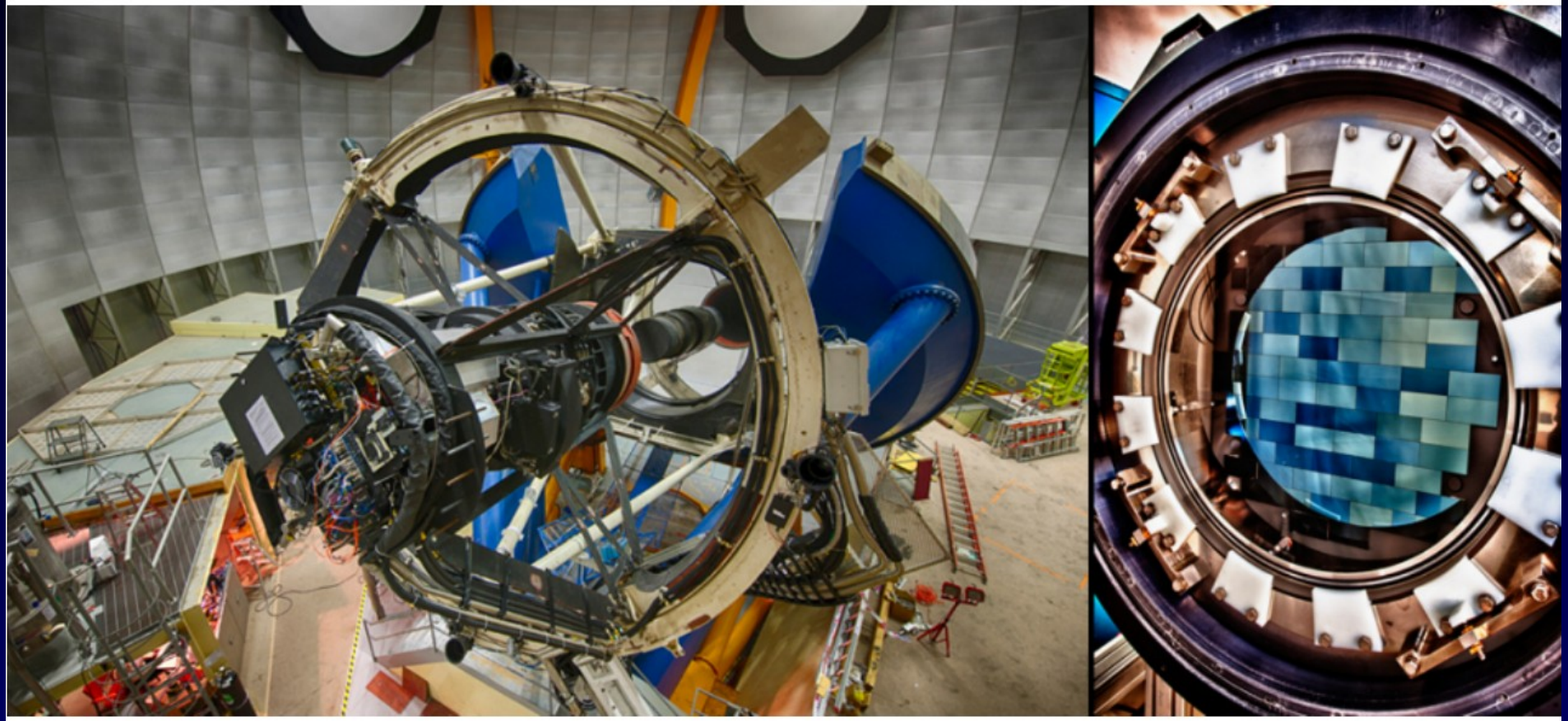
Moorův zákon chipy – zdvojení 1.5 roku (1000/10 let)

Data v astronomii – zdvojení < 1 rok !



Dark Energy Survey Camera

Dark Energy Camera (DECam)

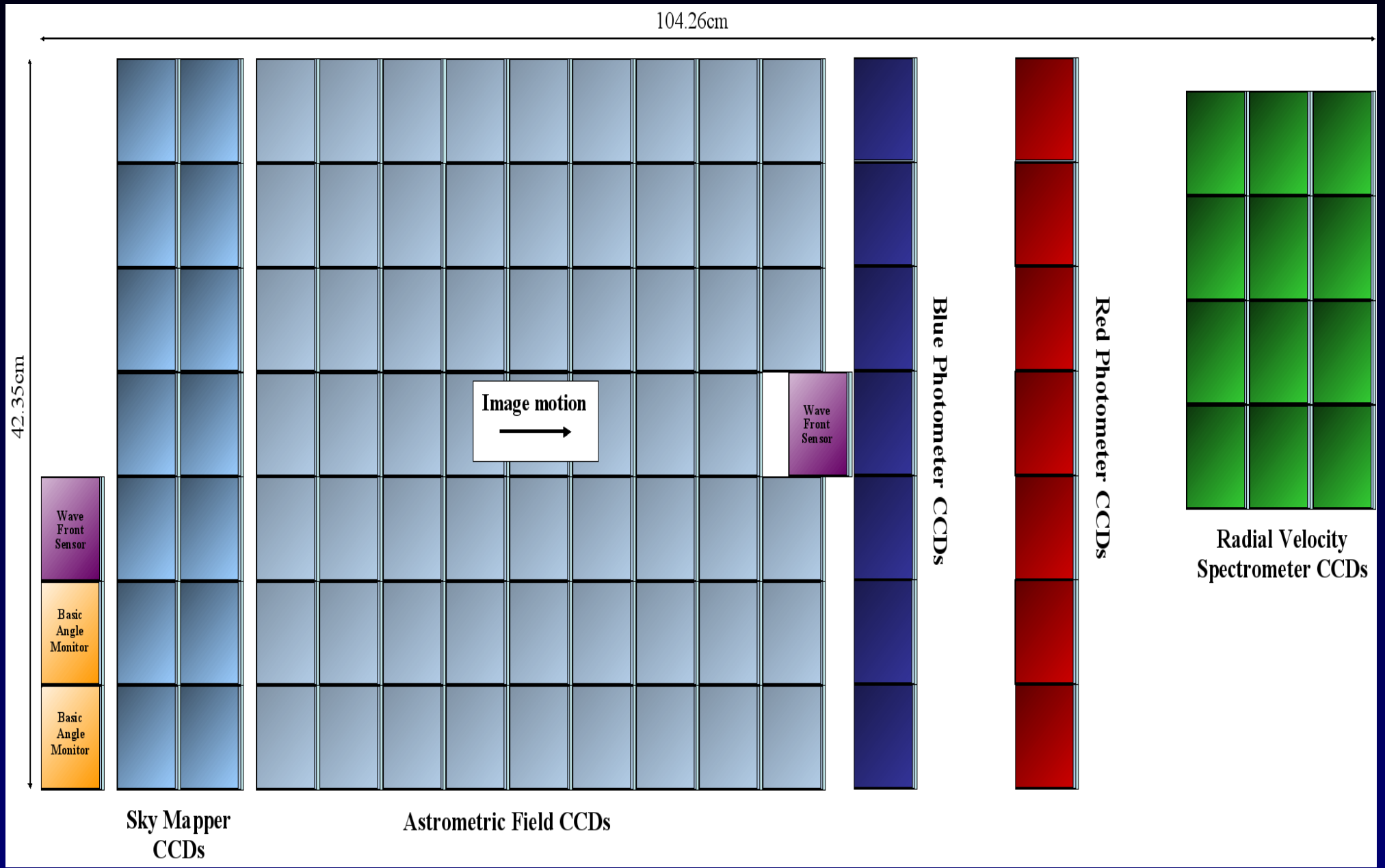


~0.4 PB/yr

74 chips – 570 Mpixels - 4m Cerro Tololo

GAIA CCDs

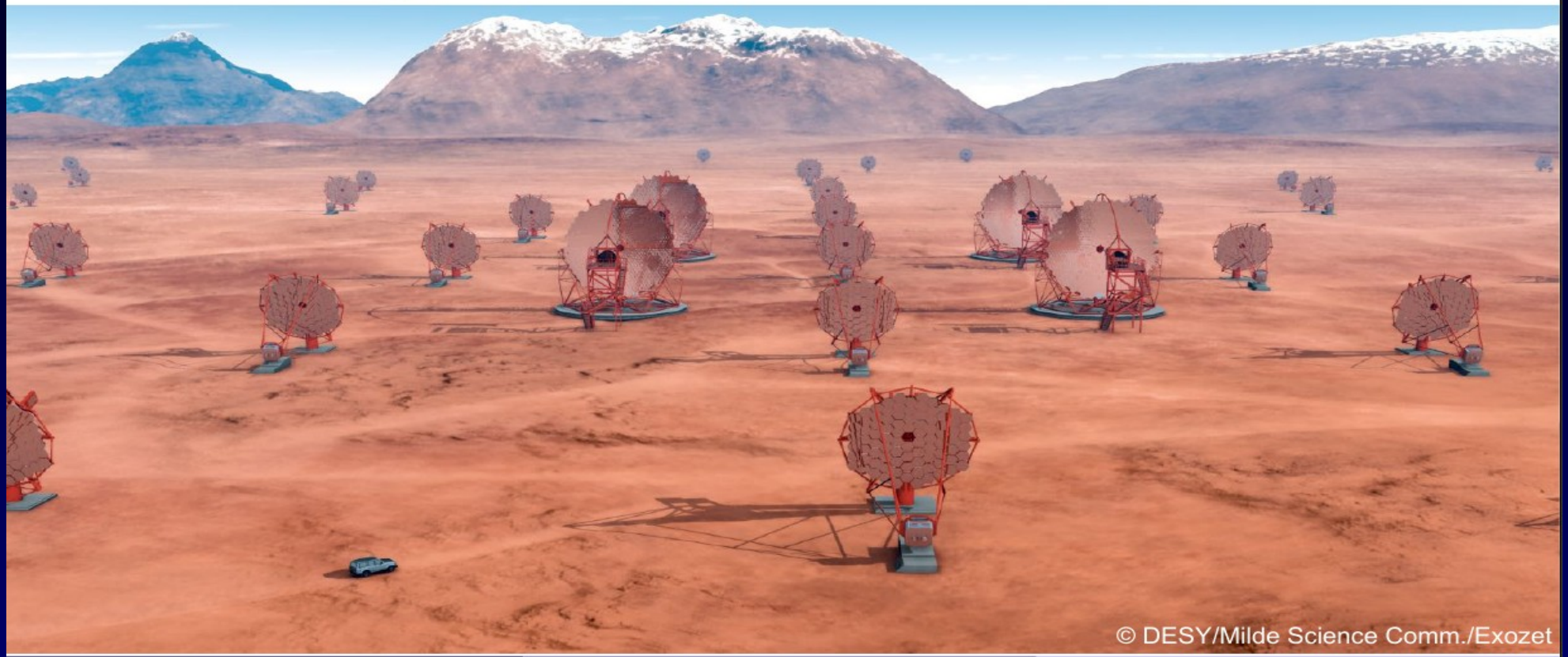
106 CCDs
938 Mpix
2800cm²



Cherenkov Telescope Array

cta
cherenkov telescope array
Observatory

- 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



27 PB/year in archive

LOFAR network

	LOFAR	SKA
Raw Telescope	112 PB/yr	60 EB/yr
Archive Rate	6 PB/yr	100 PB/yr

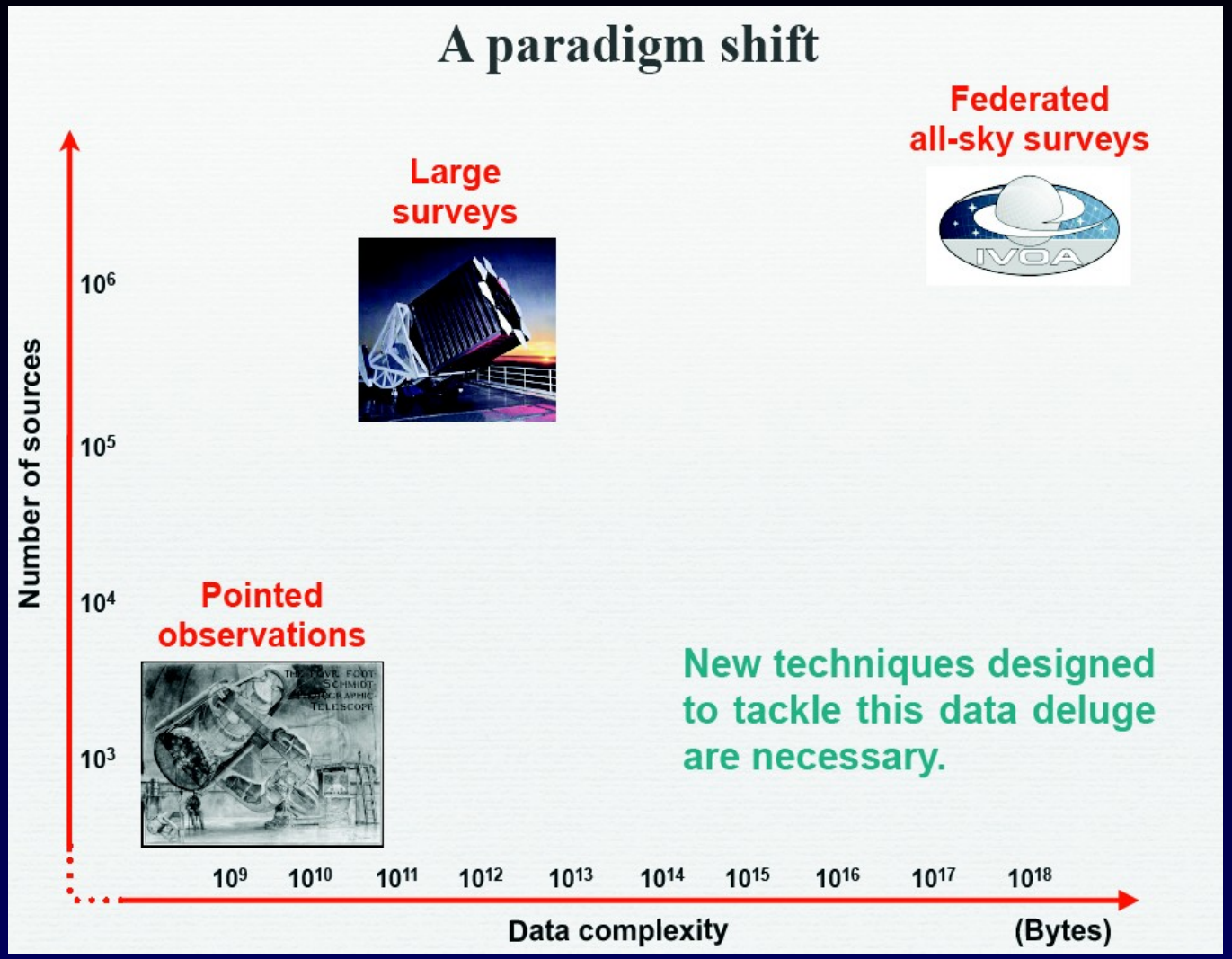
SKA

SKA Archive Volumes

- ~0.5 – 10 PB/day of image data
- Source count ~10^6 sources per square degree
- ~10^{10} sources in the accessible SKA sky, 10^4 numbers/record
- ~1 PB for the catalogued data

100 Pbytes – 3 EBytes / year of fully processed data

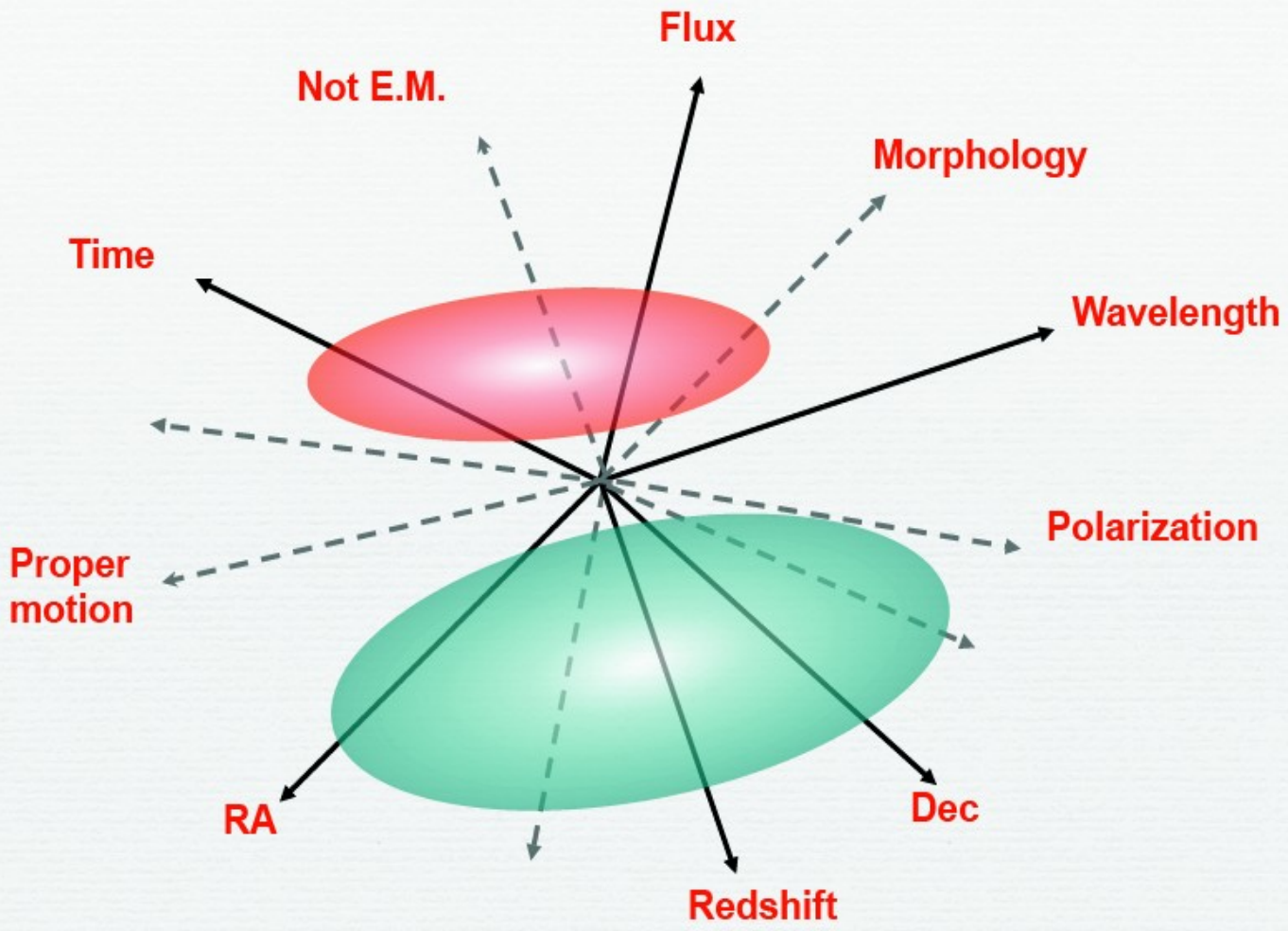
Alexander 2016



D'Abrusco 2010

Data nutno analyzovat na místě uložení
Move processing = not data !

A growing parameter space



Most discoveries were made in small regions of subspaces or along some of these axes

D'Abrusco 2010

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.

Database technologies

→ Data Mining (or Knowledge Discovery in Databases):

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

Key mathematical issues

→ Data understanding

Computer aided understanding
KDD
Etc.

Ongoing research

→ 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***

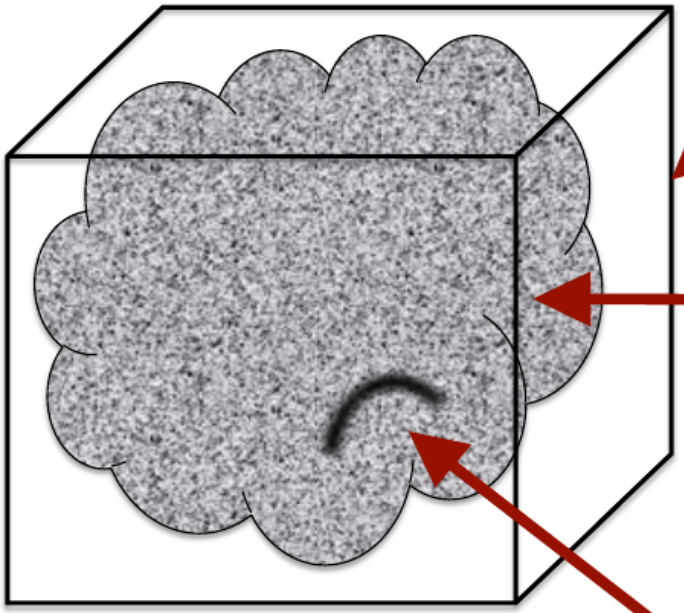
➡ ***Data driven vs. hypothesis driven science***

- 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

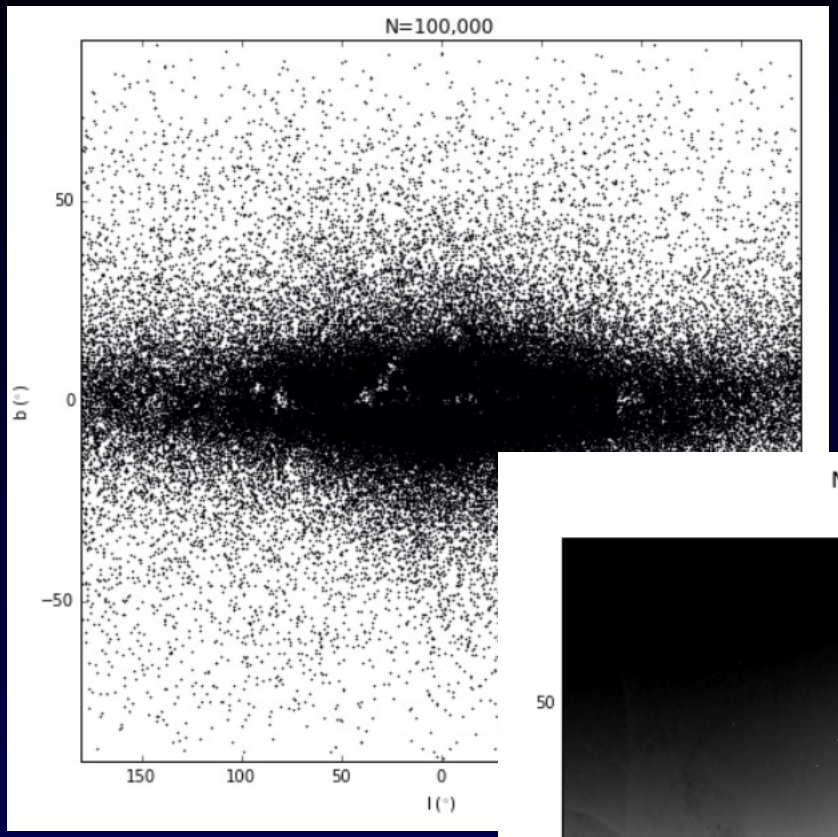
A Key Challenge: Visualisating Multidimensional Data Spaces

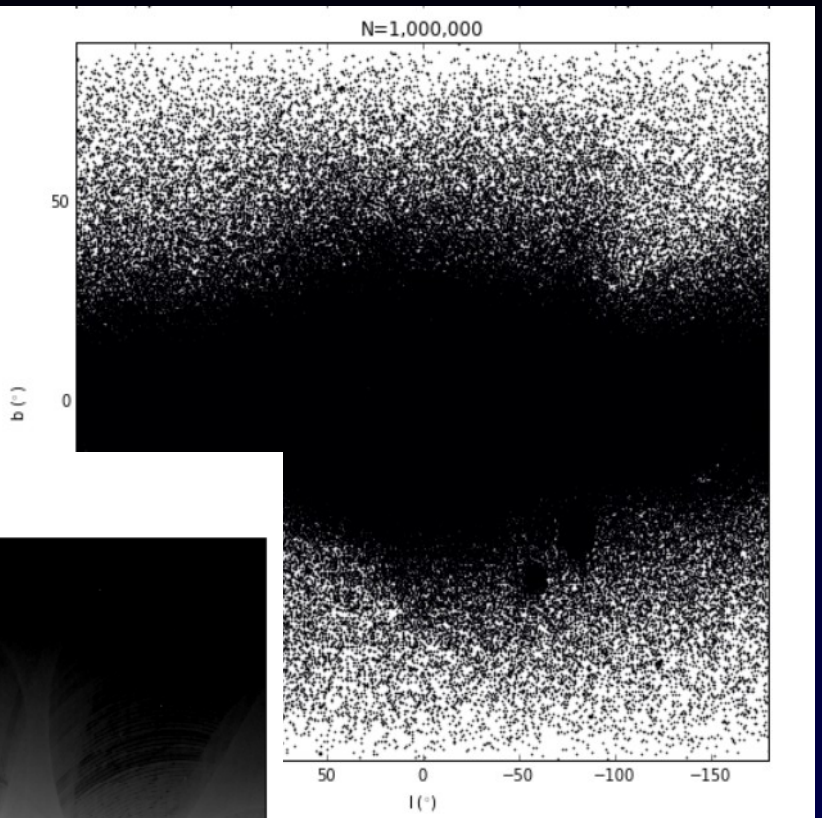
- Hyperdimensional structures (clusters, correlations, etc.) may be present in many complex data sets, whose dimensionality may be D ~ 10^2 – 10^4, or higher
- It is a matter of ***data understanding***, choosing the right data mining algorithms, and interpreting the results
- We are biologically limited to perceiving up to ~ 3 - 12(?) dimensions

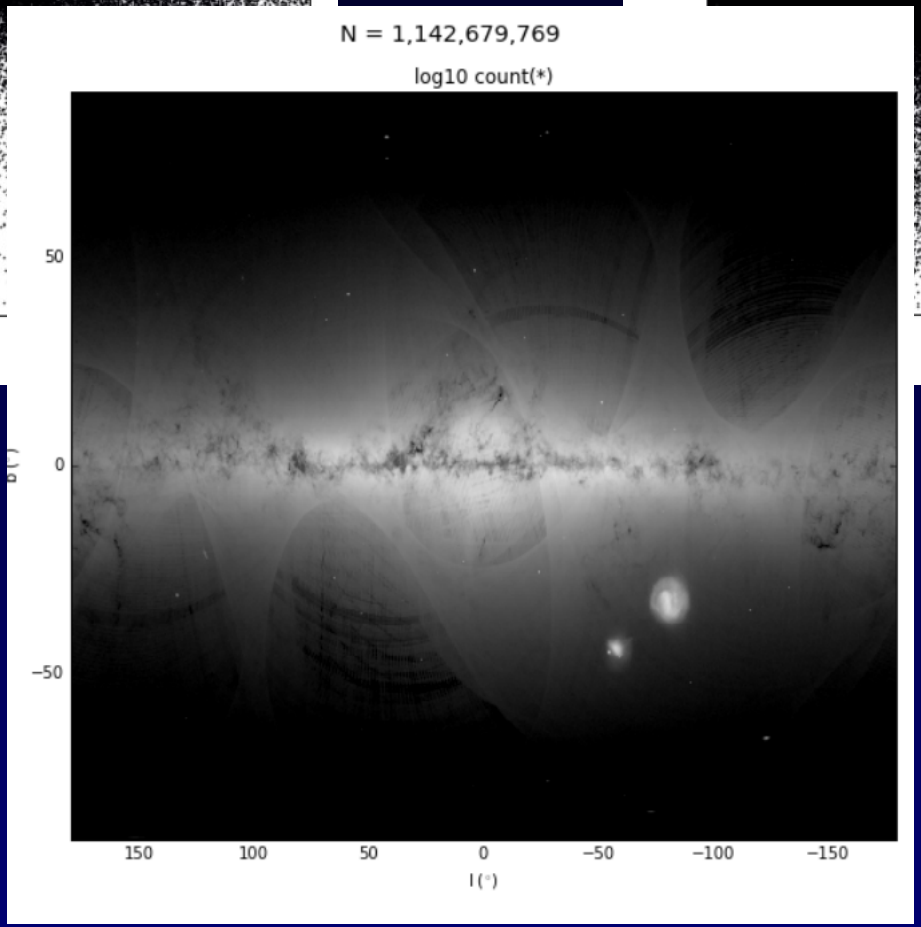
What good are the data if we cannot effectively extract knowledge from them?

Djorgovski

Visualization of 1 B points – Gaia DR1



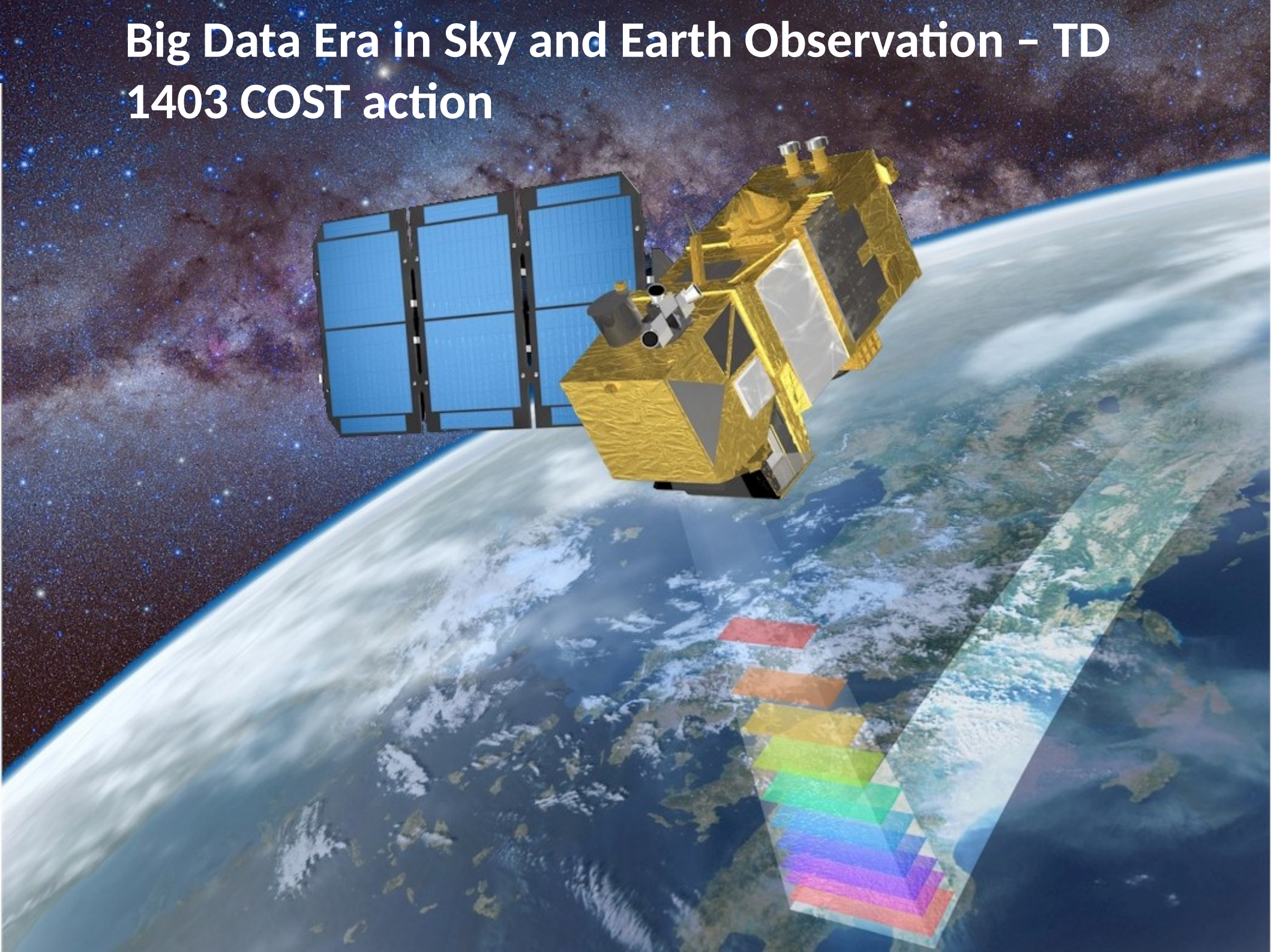




10^9 points In 1 sec !

Breddels 2016

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

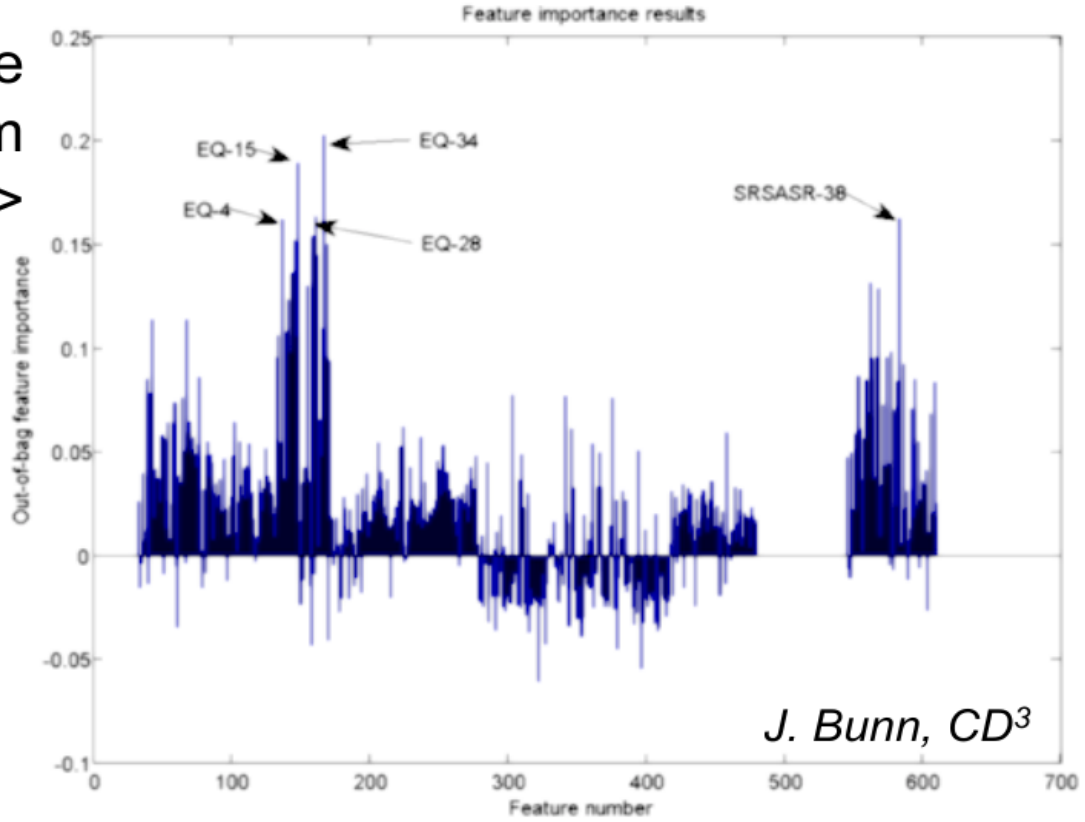


Astro-Neurology

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

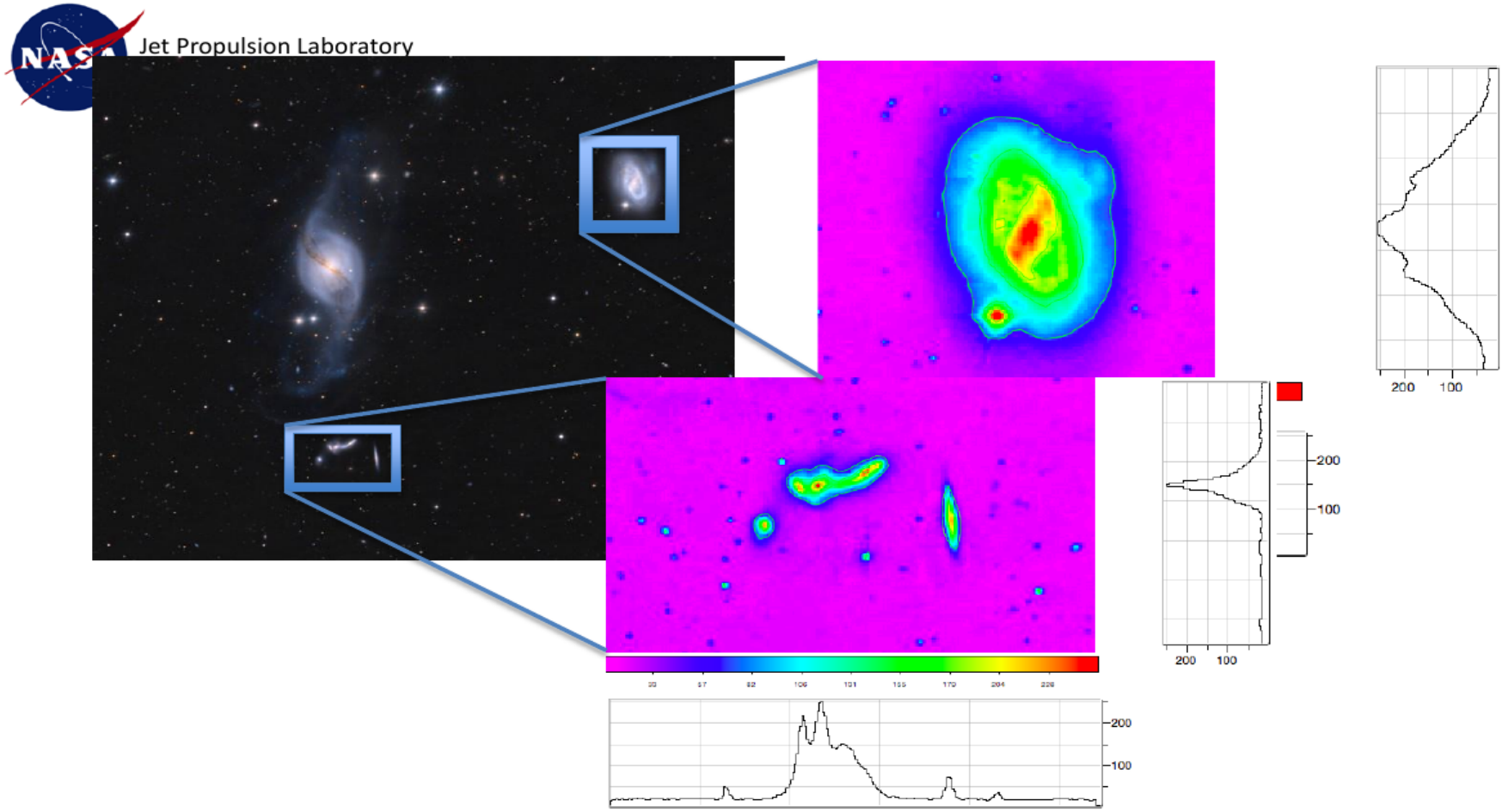
- Feature importance using random forests =>



- Next: correlate with MRI scans

(with R. Adolphs et al.)

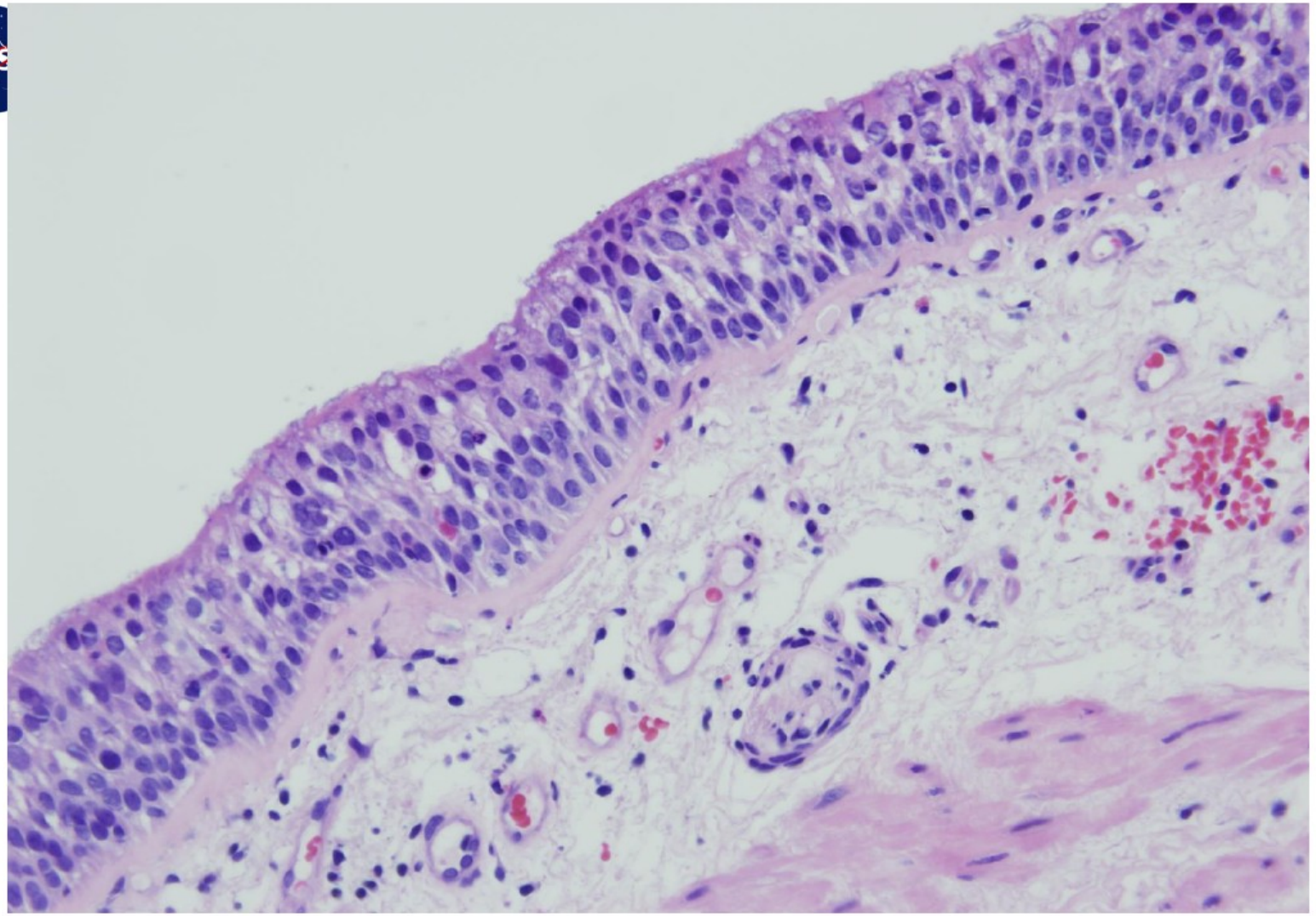
Finding Galaxies by Shape NASA



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.

Crichton 2016

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

Crichton 2016

BP a DP na FITs
z astroinformatiky a VO

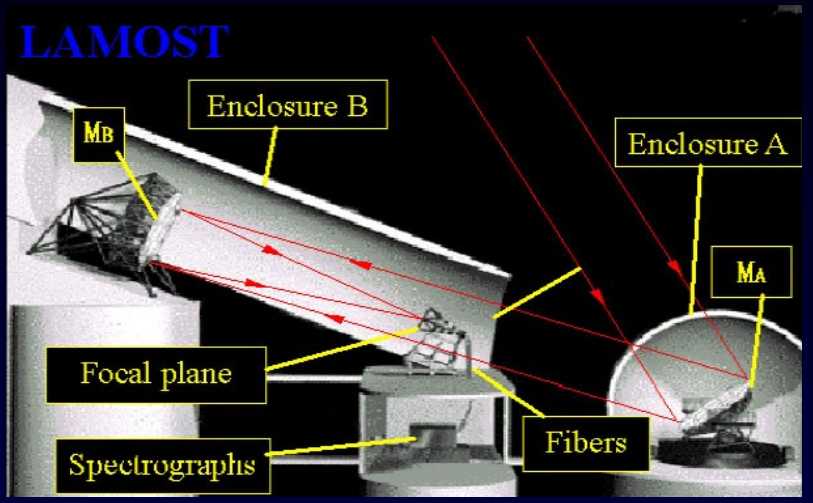
- FIT VUT Brno
  - 2011 1 BP (Random Forests in Astronomy)
  - 1 PhD – Wavelets Dimensionality Reduction (pending)

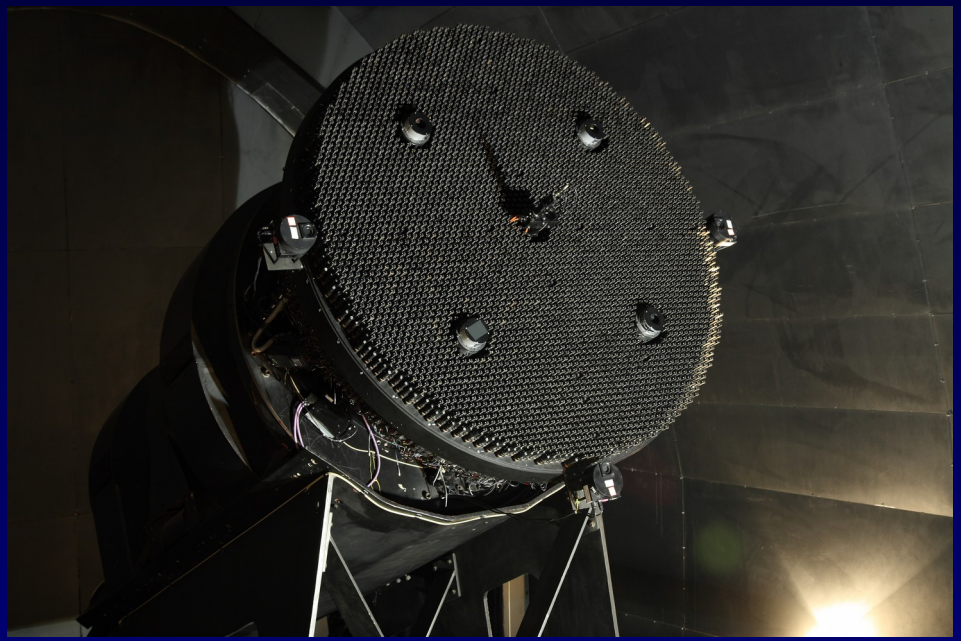
- VŠB-TU Ostrava
  - 2013+2015 1 BP + 1 DP - SPLAT-VO

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

LAMOST (Guoshoujing)

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





LAMOST Spectral Surveys

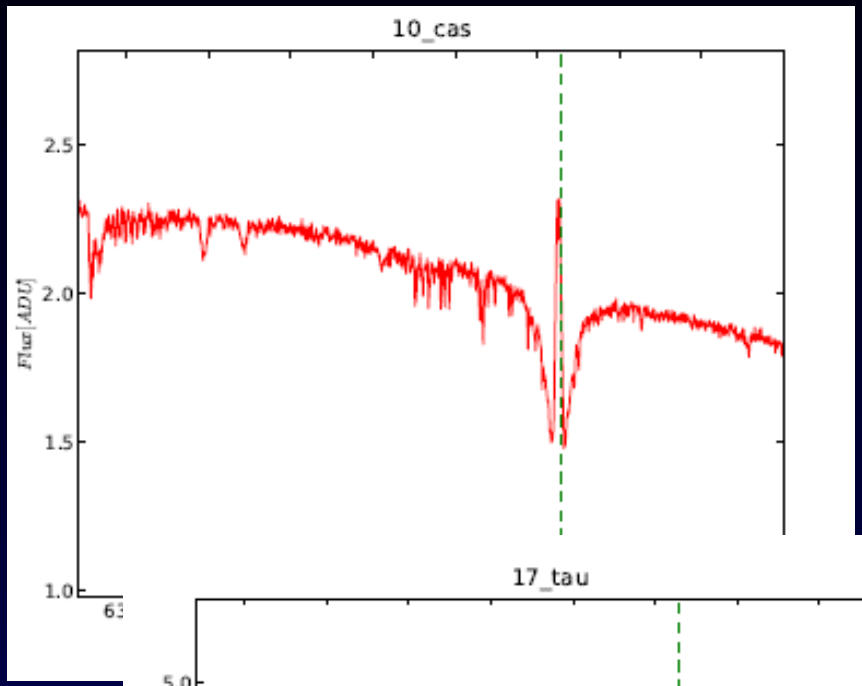
DR1 (end 2013) **2 204 860** spectra
1 085 404 stars

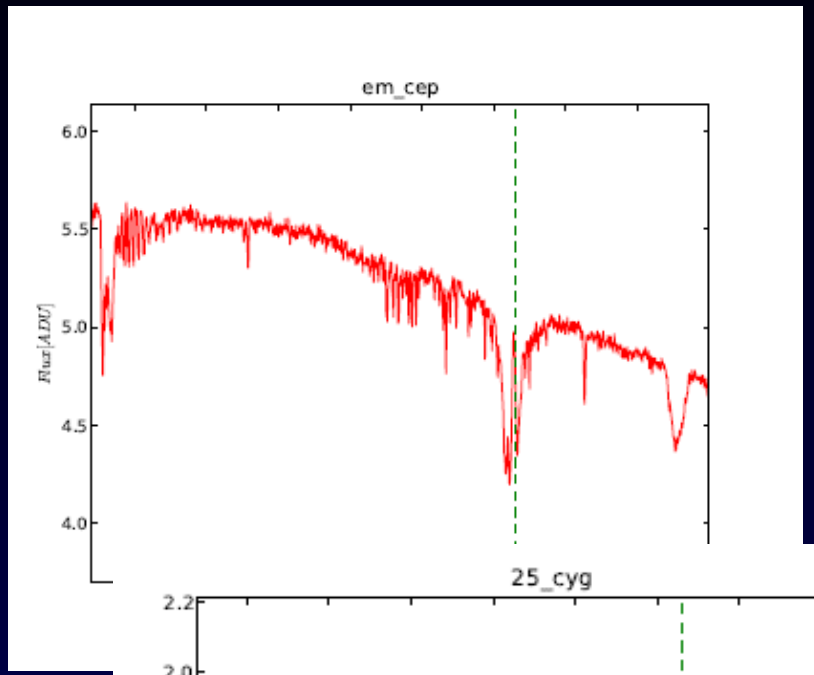
DR3 (half 2015) **5 755 126** spectra
DR4 (Feb 2016) **+ 741 522**

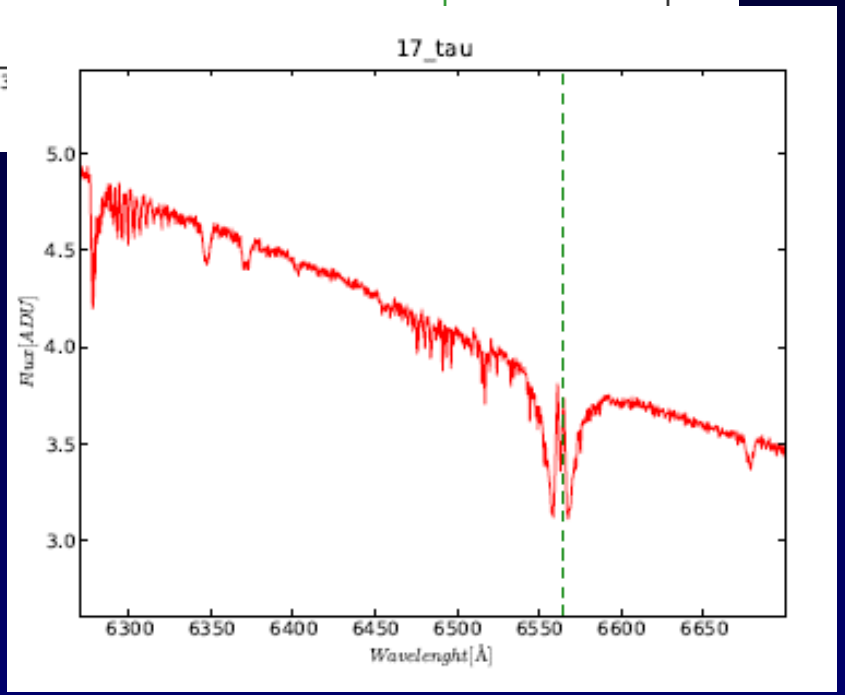
Each Fiber – 2 motors
double arm 33mm circle

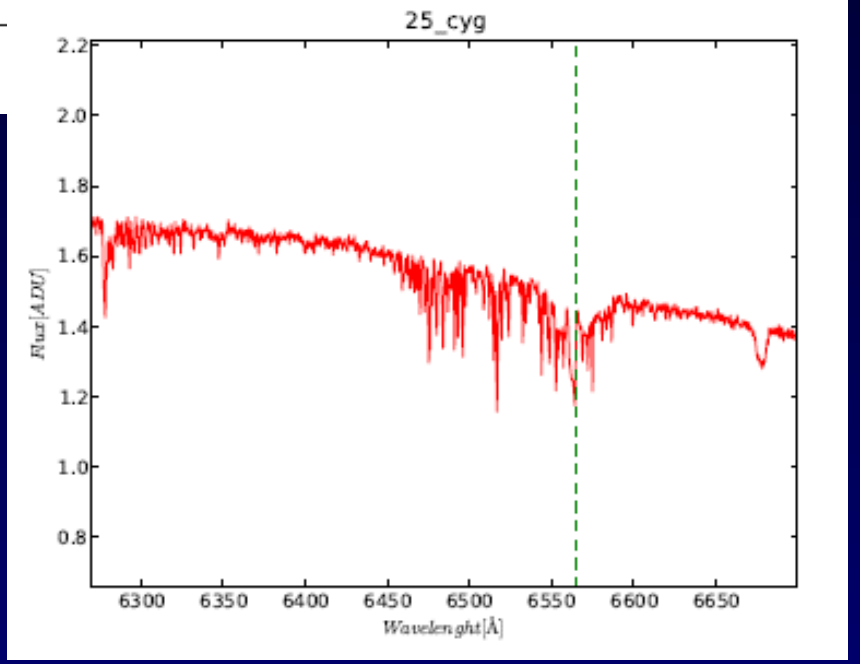
Fibre collects light from
3.3 arcsec circle on sky

Be Stars : Emission in absorption



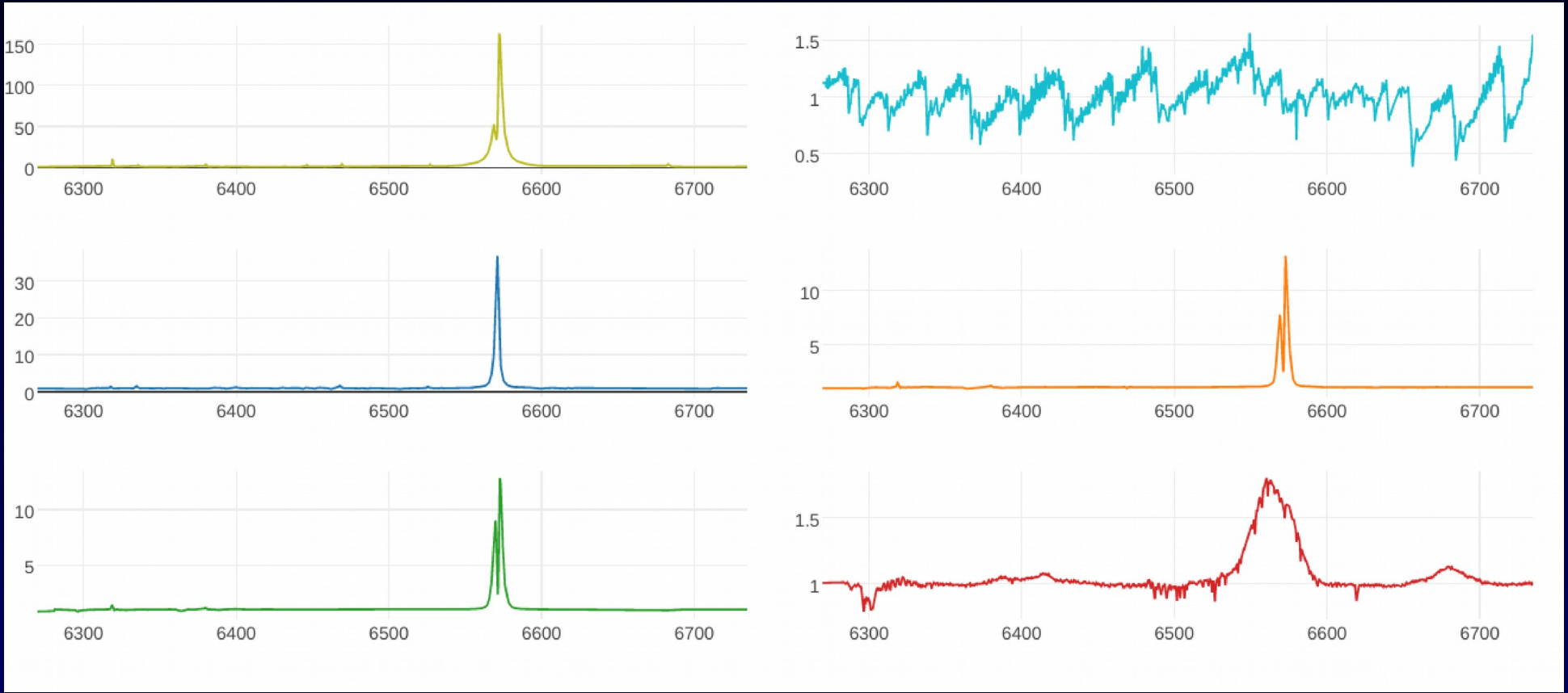






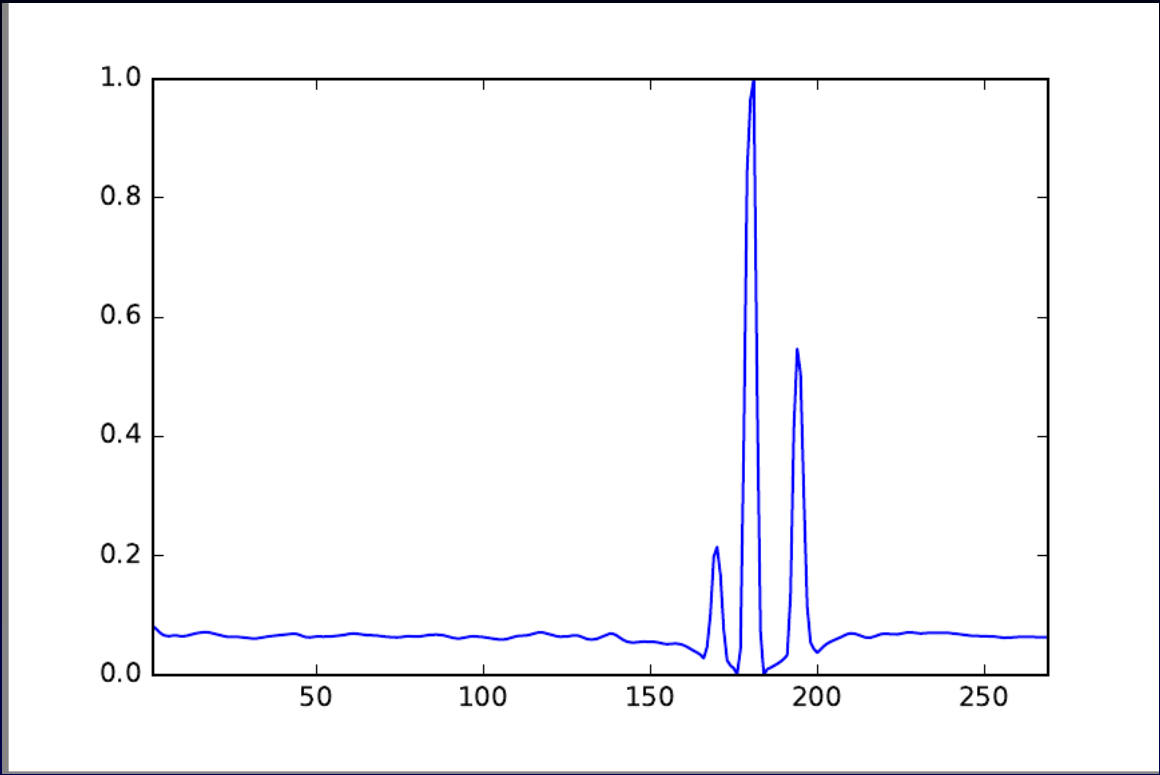
CCD700 Outliers

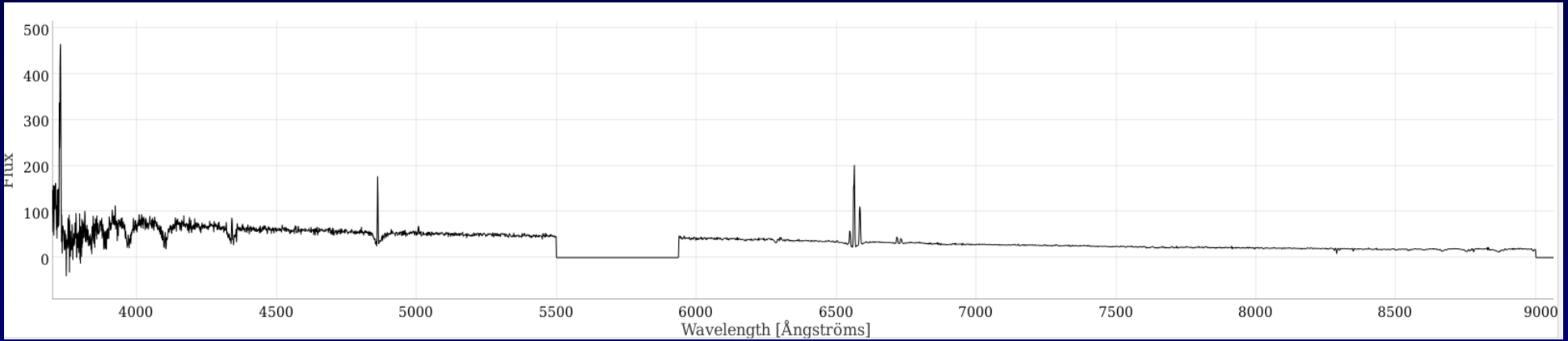
Unsupervised learning – Local Outlier Factor - LOF



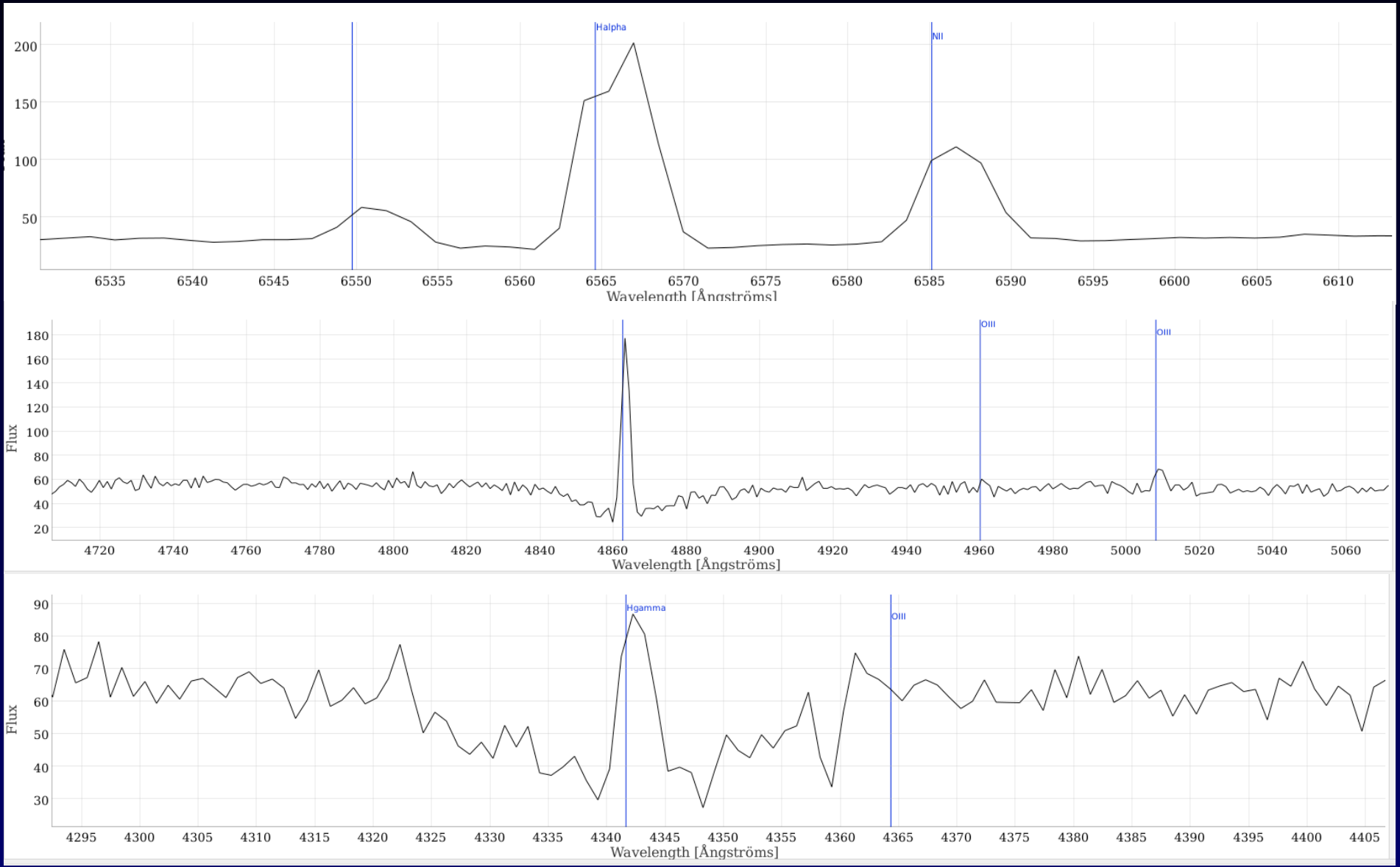
Shakurova 2016

Be Candidates Foud

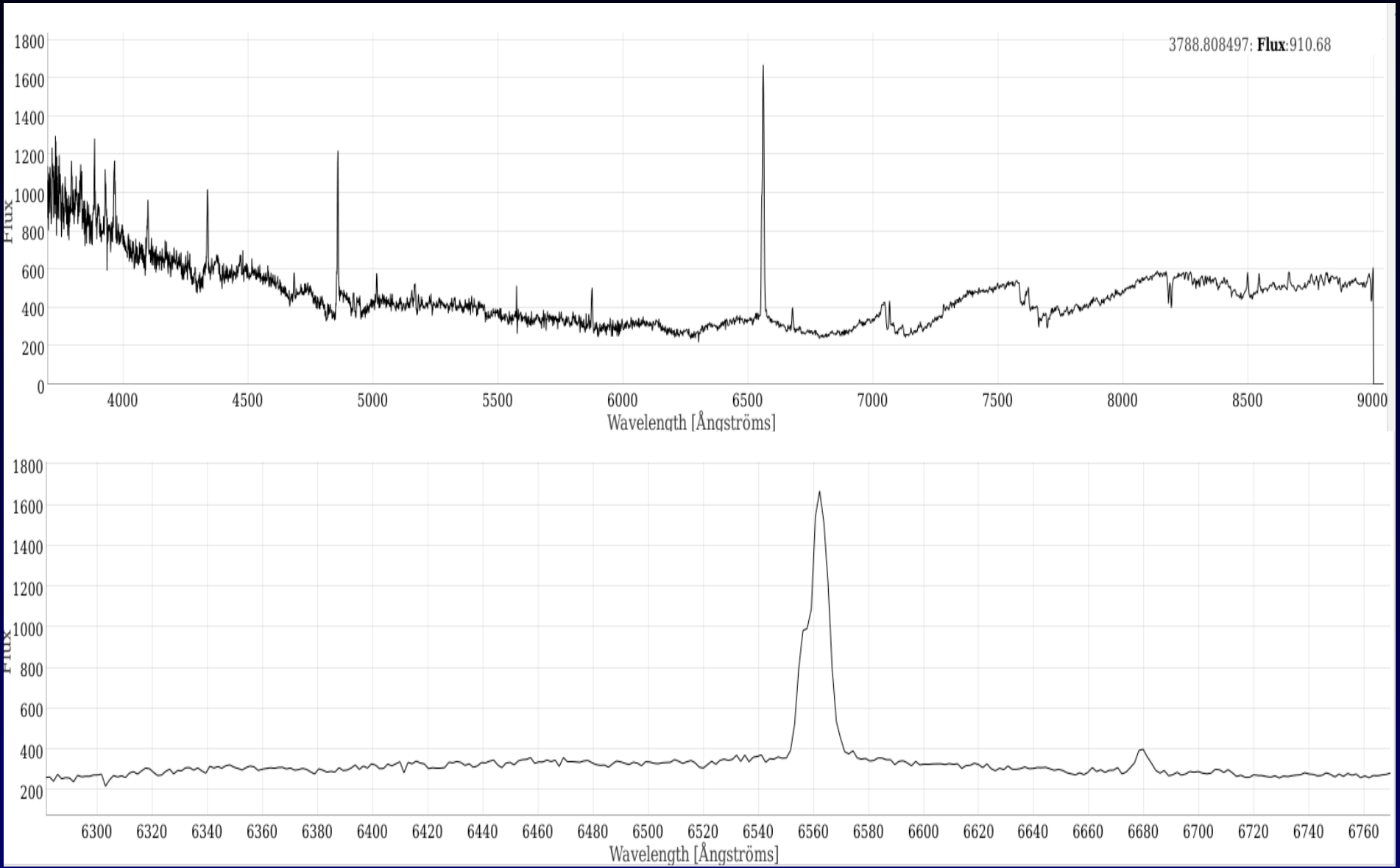




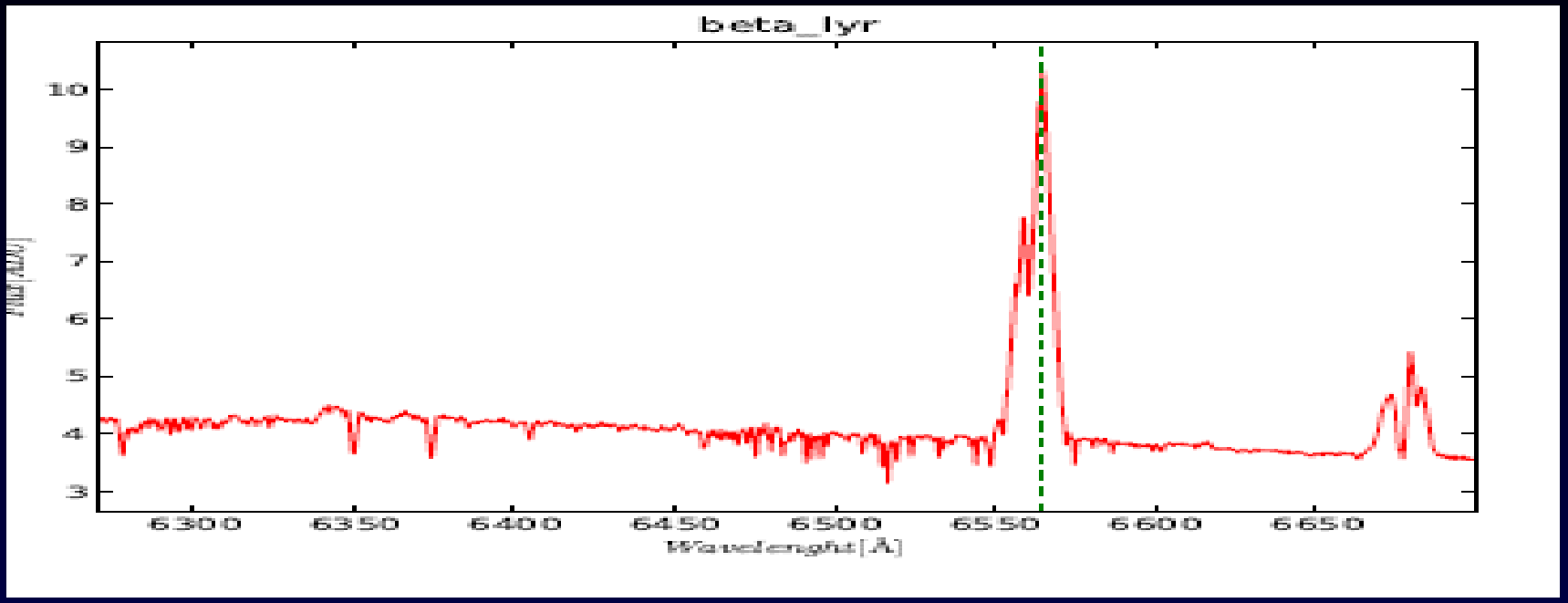
Be Candidates Found

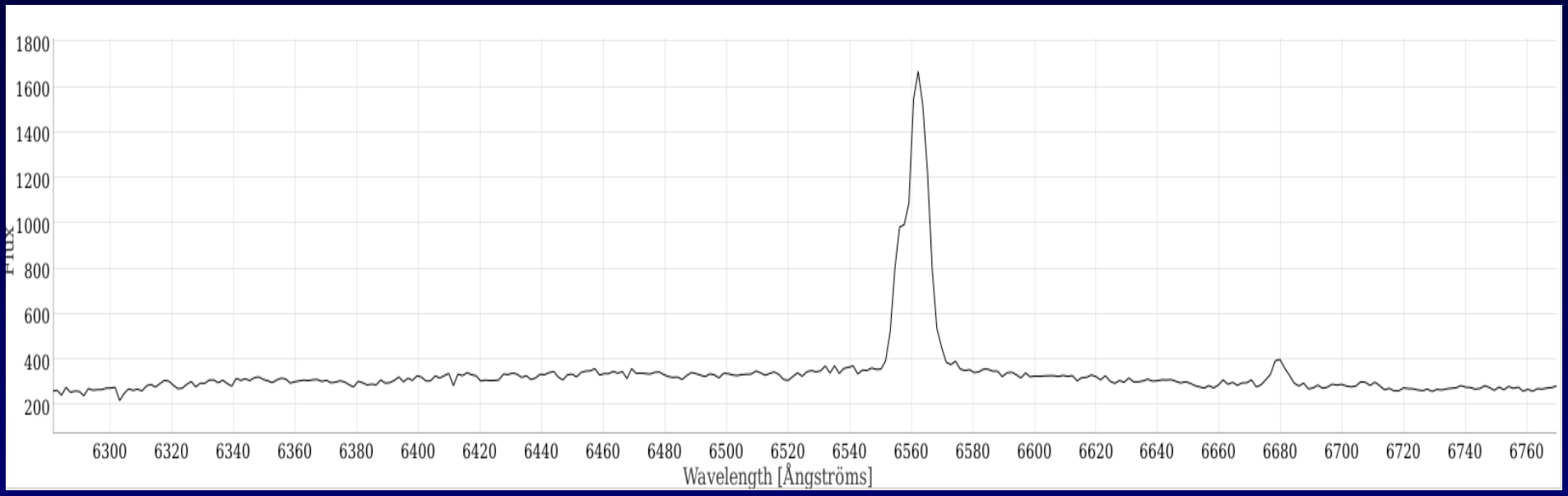


Be Candidates Found

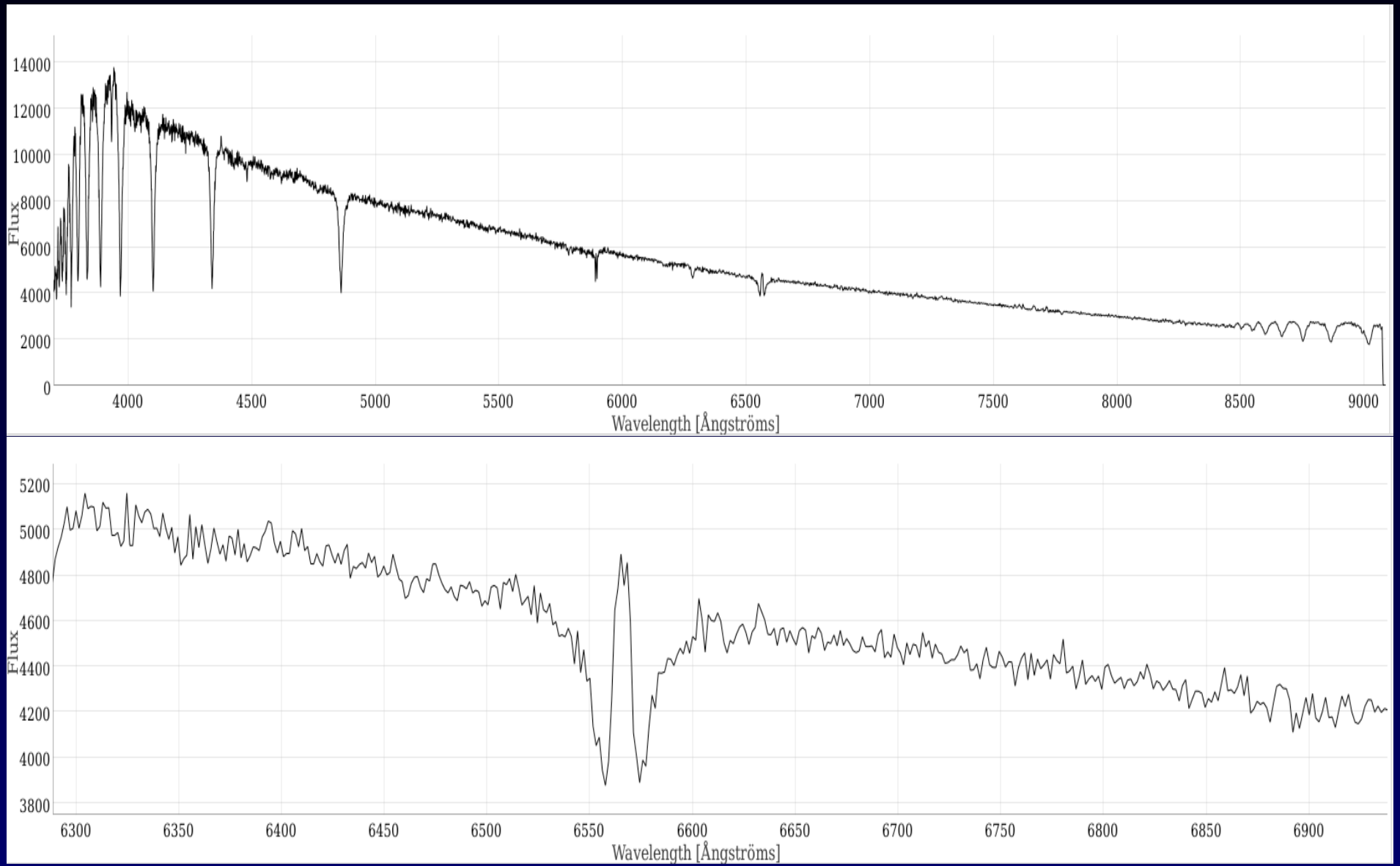


Be Candidates Found

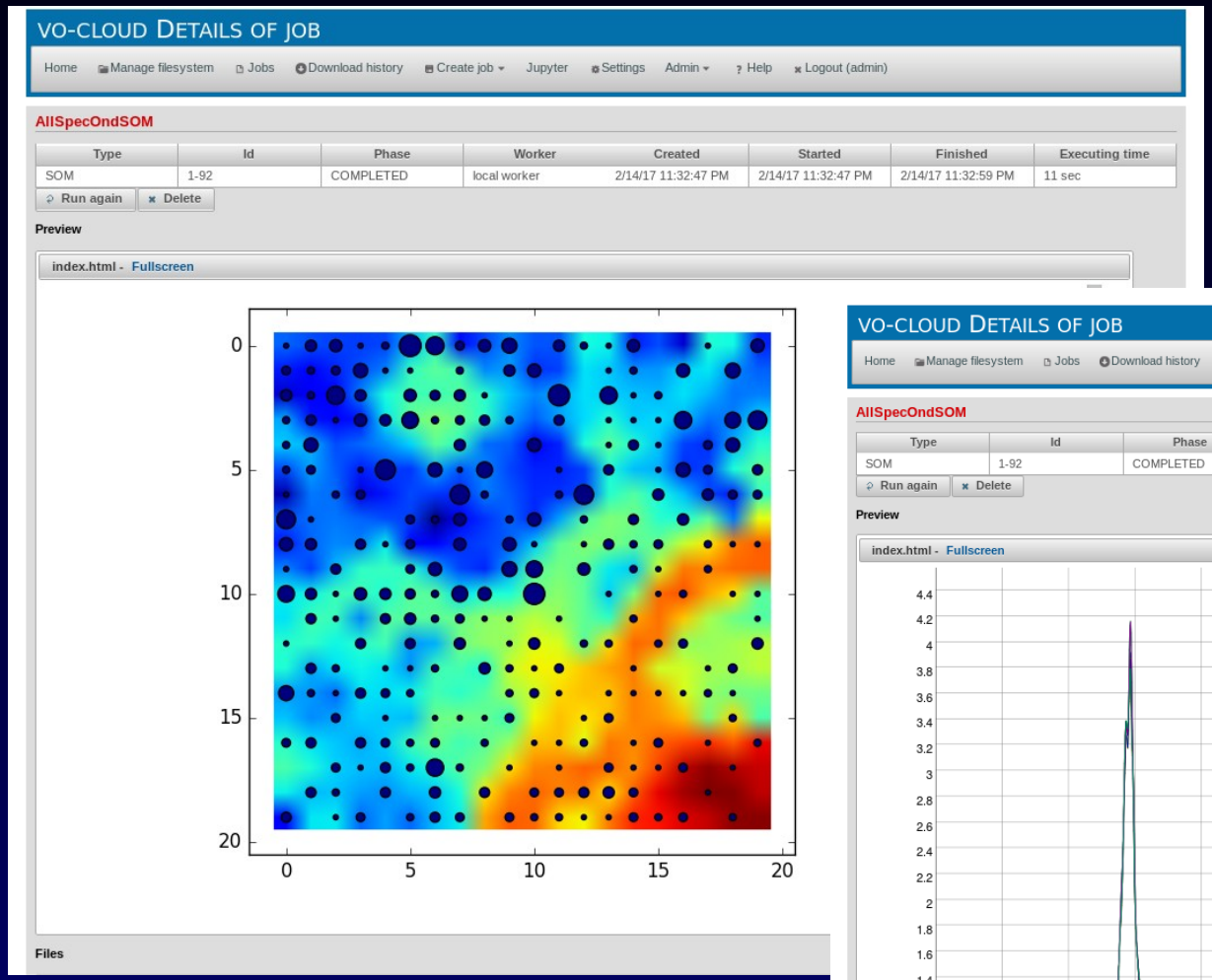


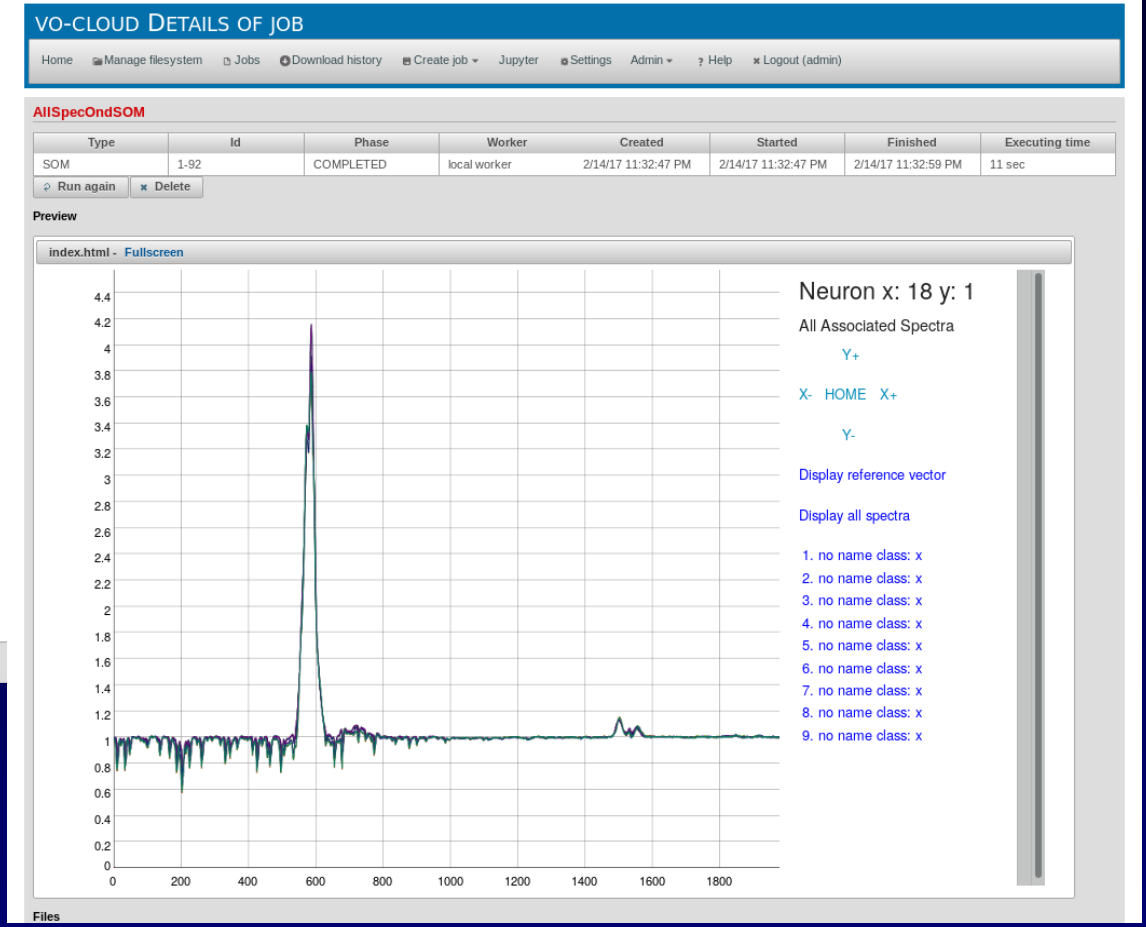


Yet Unknown Be Star

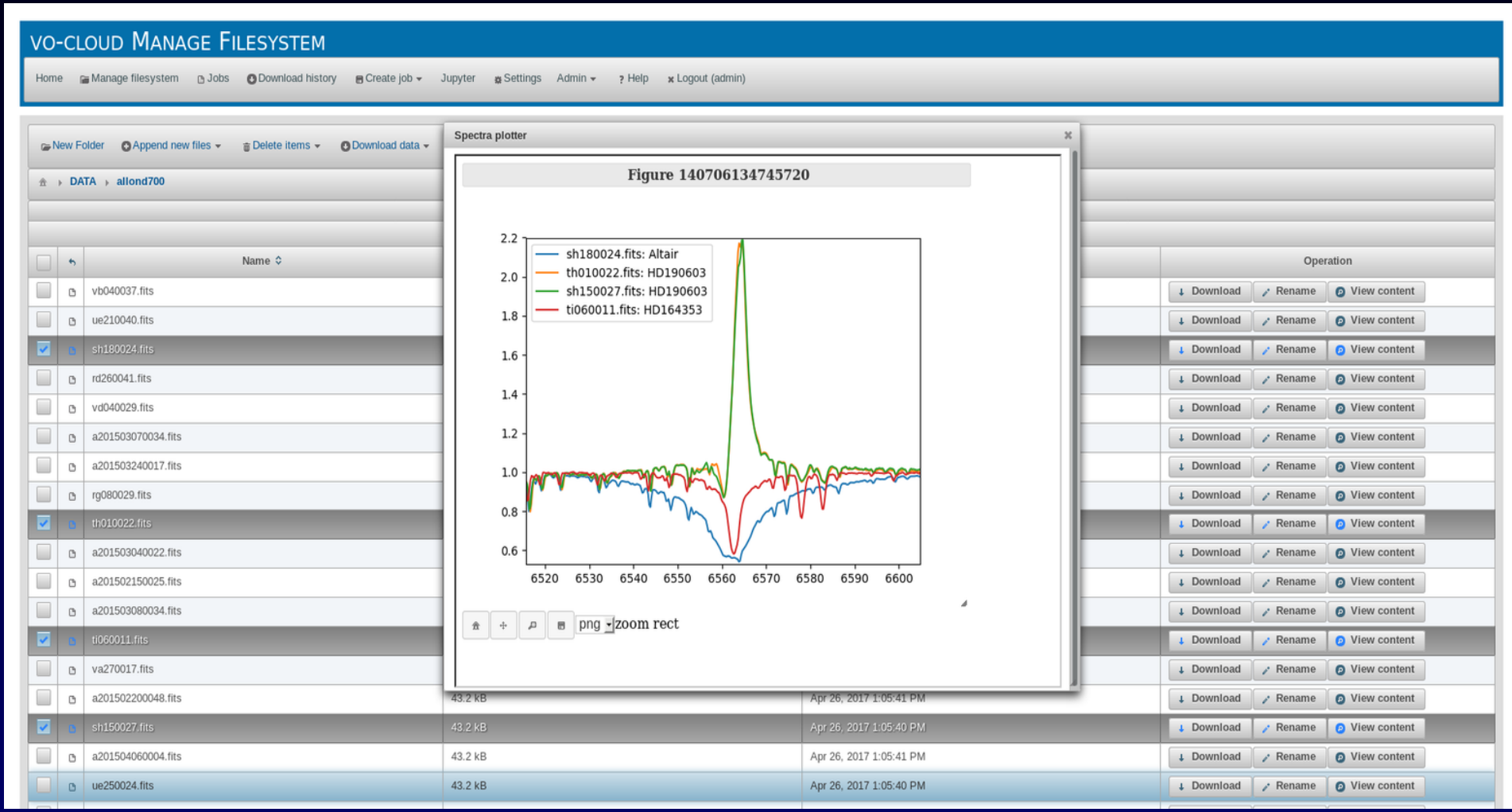


Samoorganizační mapa (Kohonen)





VO-CLOUD vizualizace spekter



Budoucnost 2m

Určitě velmi moderní

V souladu s dobovými trendy

Moderní technologie

Mnoho produkce z archivu – followup, hledání zpět