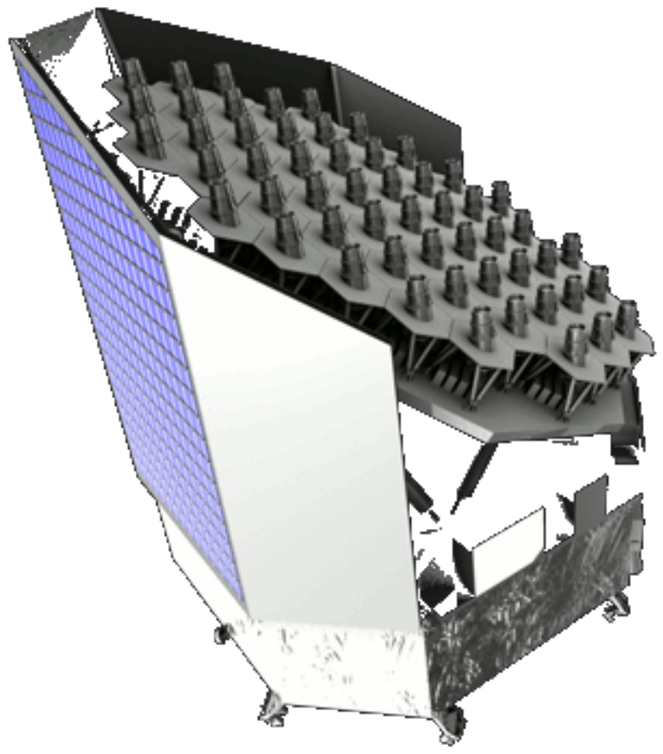
On the analysis of large data sets in the PLATO Data Center

M. Ammler-von Eiff, R. Burston, R. Heller, J. Knoche, H. Schunker, L. Gizon
(MPI for Solar System Research, Göttingen)

and the PLATO Mission Consortium

AG meeting 2016, Splinter: E-Science & Virtual Observatory, 14/09/2016



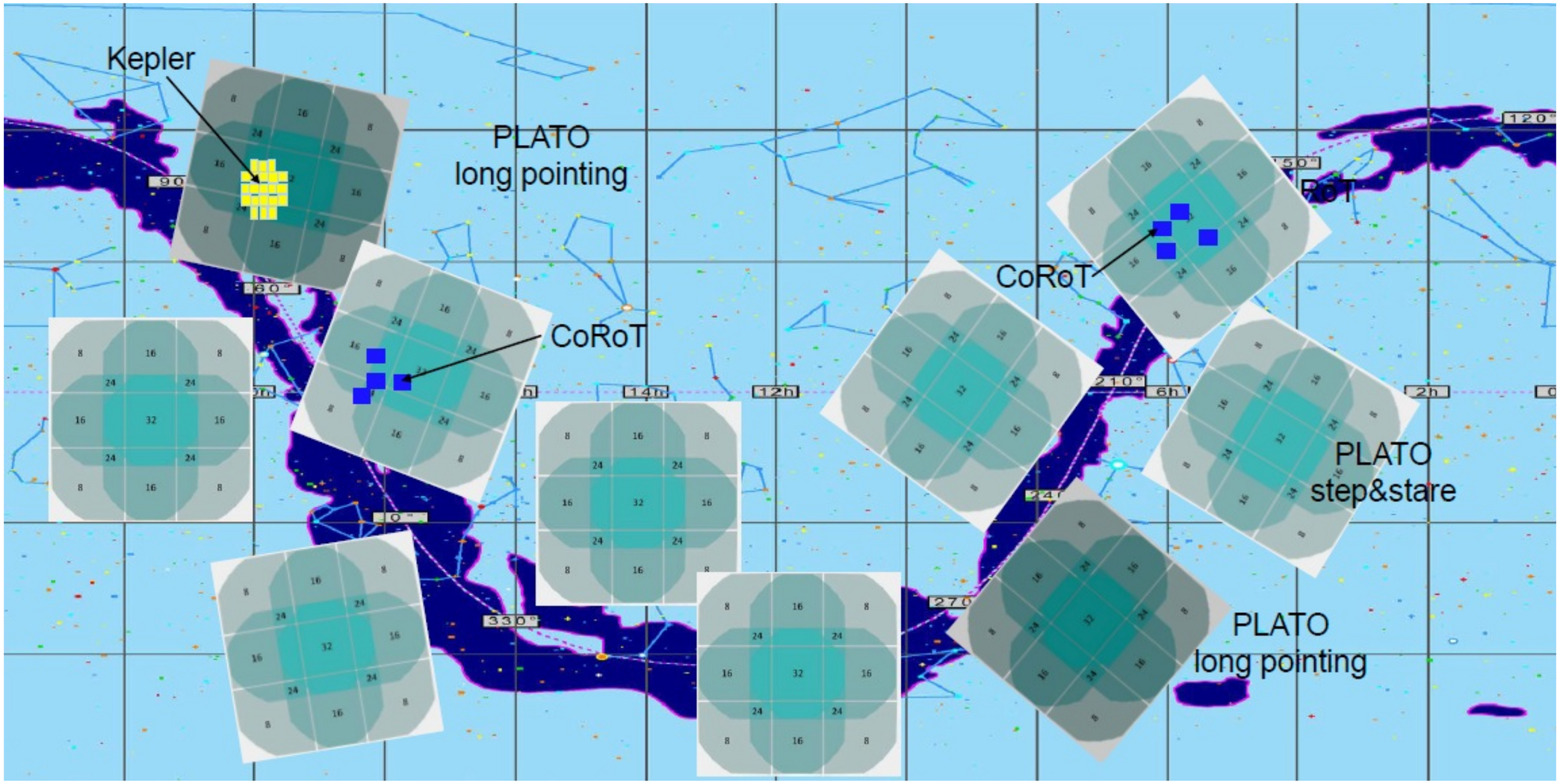
PLAnetary Transits and Oscillations of stars

- transit detection and characterisation of thousands of exoplanets, including Earth-like planets in the habitable zone
- M3 mission selected in ESA's Cosmic Vision 2015-2025 Programme
- launch: end of 2025

The PLATO Concept

- ratio of stellar and planetary radius from transits (1,000,000 stars surveyed)
- stellar radius from Gaia
- mass ratio from ground-based radial velocities (for V<11)
- stellar mass and age from asteroseismology
- radius, mass and age of exoplanets
- The PLATO photometric precision (34 ppm) is such that it enables the measurements of modes of oscillations (80,000 cool dwarfs).

PLATO Observation Strategy



- monitor several fields to cover close to 50% of the sky
- about 1,000,000 stars brighter than V=13 in total
- monitor fields for up to 2-3 years at a cadence of 600s and shorter

PLATO Data Volume Estimates

data volume	data level	per day [Gbytes]	per year [TBytes]
raw data (uncompressed telemetry)		201	72
light curves and centroid curves with corrections applied on-board, imagettes, images of sky regions	L0	214	76
calibrated data with all instrumental corrections applied (L1 data level)	L1	245	87
final science data products, including catalogue of confirmed planetary systems and their characteristics	L2/L3	>>51	>>19
total		**711**	**254**

➔

>1.5 PBytes for 6 years of observations; w/o intermediate data, reprocessing, and ground-based follow-up (data level Lg)

PLATO Science Ground Segment

TC

TM

GS/MOC (ESOC)

SOC (ESAC)

L0/L1 generation

generation of on-board catalogue

PLATO Archive

L3

L2

L0/L1

Lg

L1 calibration software

Input Catalogue

PMC

PMC Calibration/ Operation Team

PLATO Data Center

- data processing algorithms
- development support

Input Catalogue generation

- exoplanet analysis system
- stellar analysis systems
- data analysis support tools

MPS

PDC main database

follow up and preparatory database management

Input catalogue specifications

L2/L3 processing specifications

L2/L3 scientific validation

PLATO Science Management

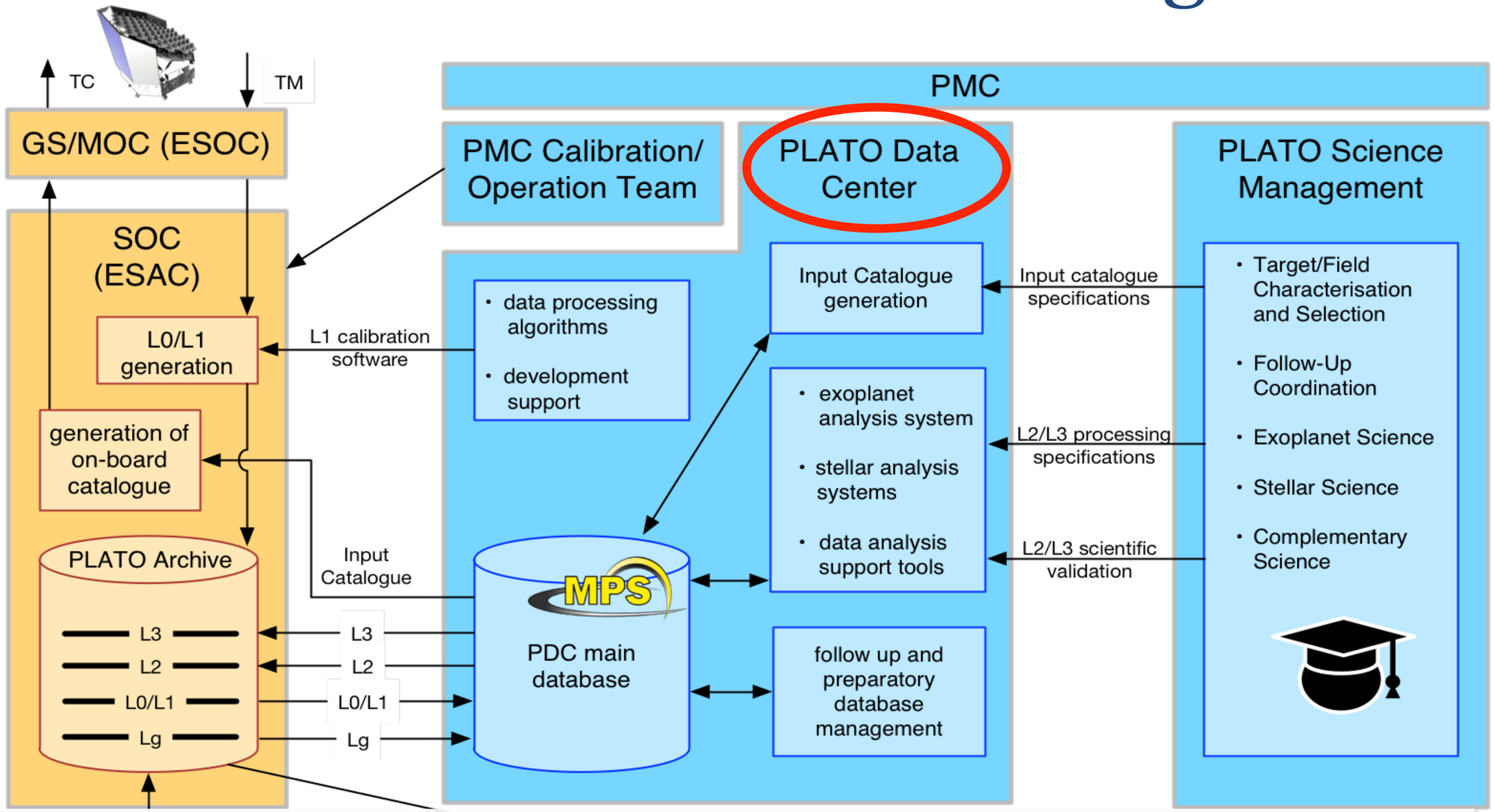
- Target/Field Characterisation and Selection
- Follow-Up Coordination
- Exoplanet Science
- Stellar Science
- Complementary Science

Ground-based Observation Programme Team

Scientific community

access to public data and proprietary data

PLATO Science Ground Segment



PDC:

- in charge of the calibration and processing of the PLATO observations
- delivers final PLATO science data products to the SOC

PLATO Science Ground Segment

TC

TM

GS/MOC (ESOC)

SOC (ESAC)

L0/L1 generation

generation of on-board catalogue

PLATO Archive

L1 calibration software

Input Catalogue

PMC

PMC Calibration/ Operation Team

PLATO Data Center

- data processing algorithms
- development support

Input Catalogue generation

- exoplanet analysis system
- stellar analysis systems
- data analysis support tools

MPS

Input catalogue specifications

L2/L3 processing specifications

L2/L3 scientific validation

PLATO Science Management

- Target/Field Characterisation and Selection
- Follow-Up Coordination
- Exoplanet Science
- Stellar Science
- Complementary Science

Ground-based Observation Programme Team

Scientific community

access to public data and proprietary data

PSM:

- provides scientific specifications of algorithms that run at PDC
- consortium scientists scientifically validate final science data products using PDC tools

PLATO Science Ground Segment

TC

TM

PMC

PMC Calibration/ Operation Team

PLATO Data Center

PLATO Science Management

GS/MOC (ESOC)

catalogue

PLATO Archive

L3

L2

L0/L1

Lg

Input Catalogue

MPS

PDC main database

data analysis support tools

follow up and preparatory database management

L2/L3 scientific validation

Stellar Science

Complementary Science

Ground-based Observation Programme Team

Scientific community

access to public data

and proprietary data

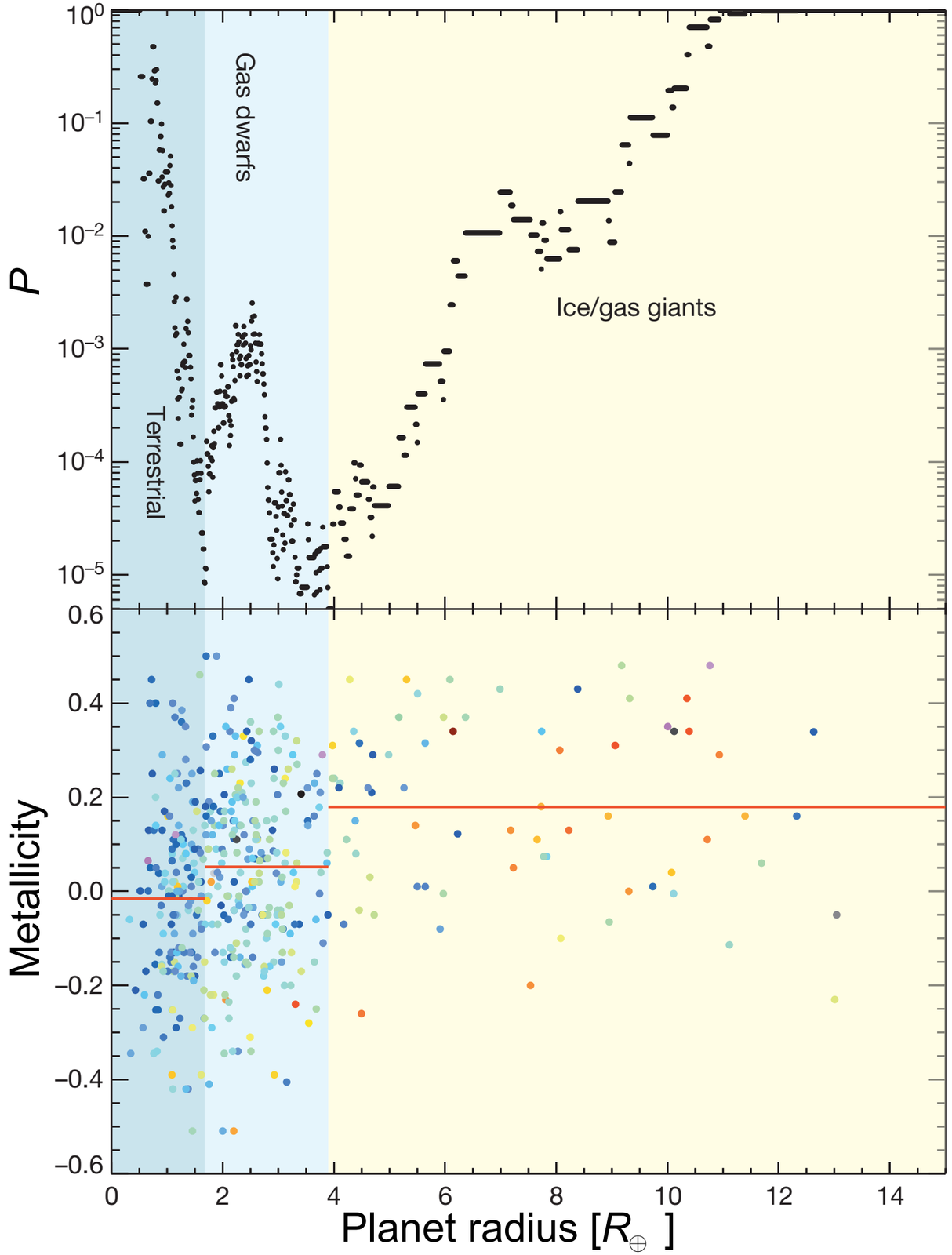
Data Analysis Support Tools:

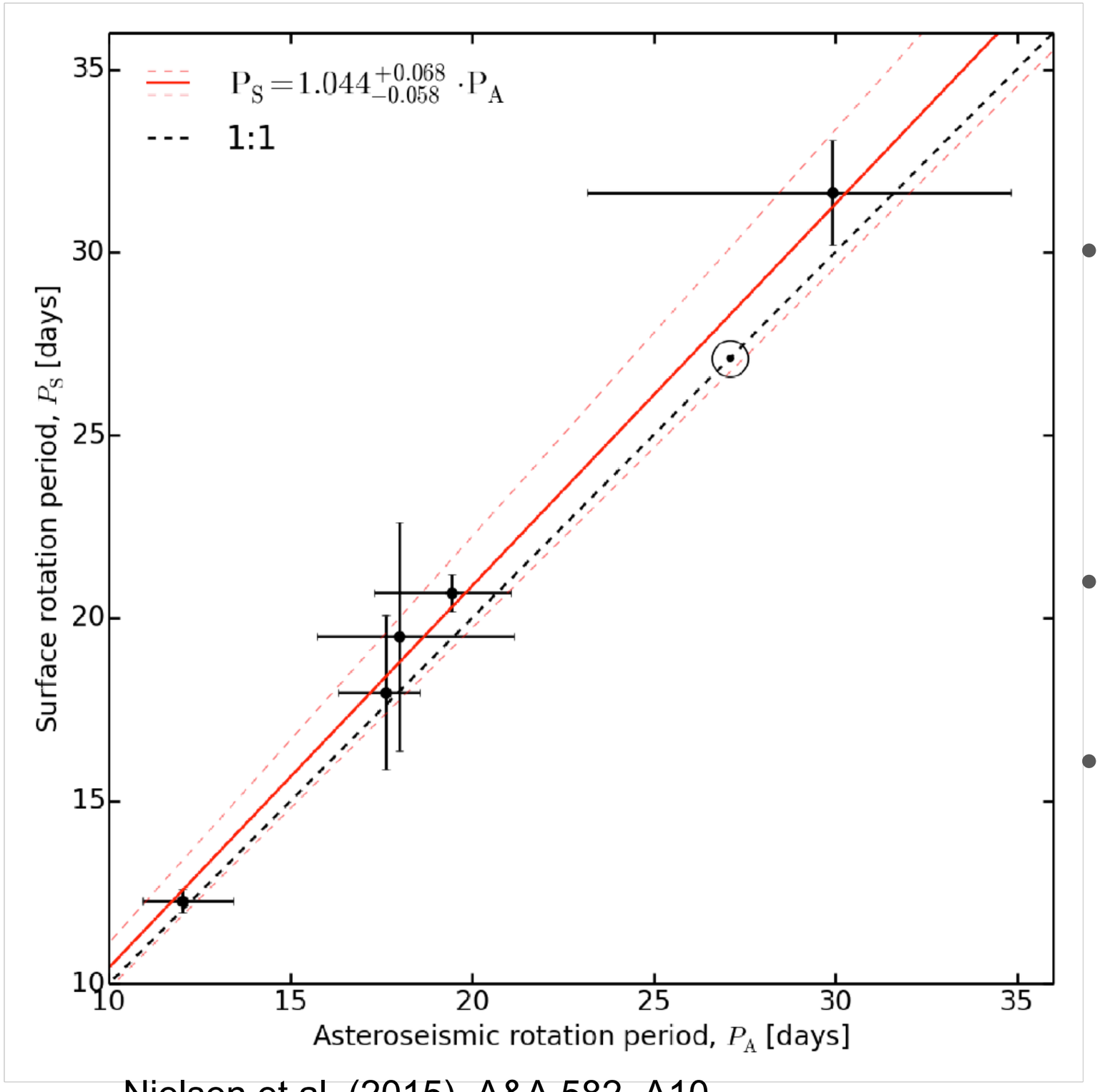
- interactive tools to assist consortium scientists to inspect and scientifically validate PLATO Data Products on a case-by-case basis and update ranking of planetary systems
- also include tools for statistical analysis and data mining of large samples
- access and inspect according to VO standards

Statistical tools: example for exoplanets

- planet radius and host star metallicity of ~600 Kepler planets (bottom panel)
- Does metallicity distribution vary over planet radius?
- *P* value (top panel) of two-sample KS test suggests 2 transitions at ~1.7 and ~3.9 $R_\oplus$
- large number of planets (hundreds or thousands) is required to determine these transitions/ patterns precisely.

Buchhave et al. 2014, Nature 509, 593





Nielsen et al. (2015), A&A 582, A10

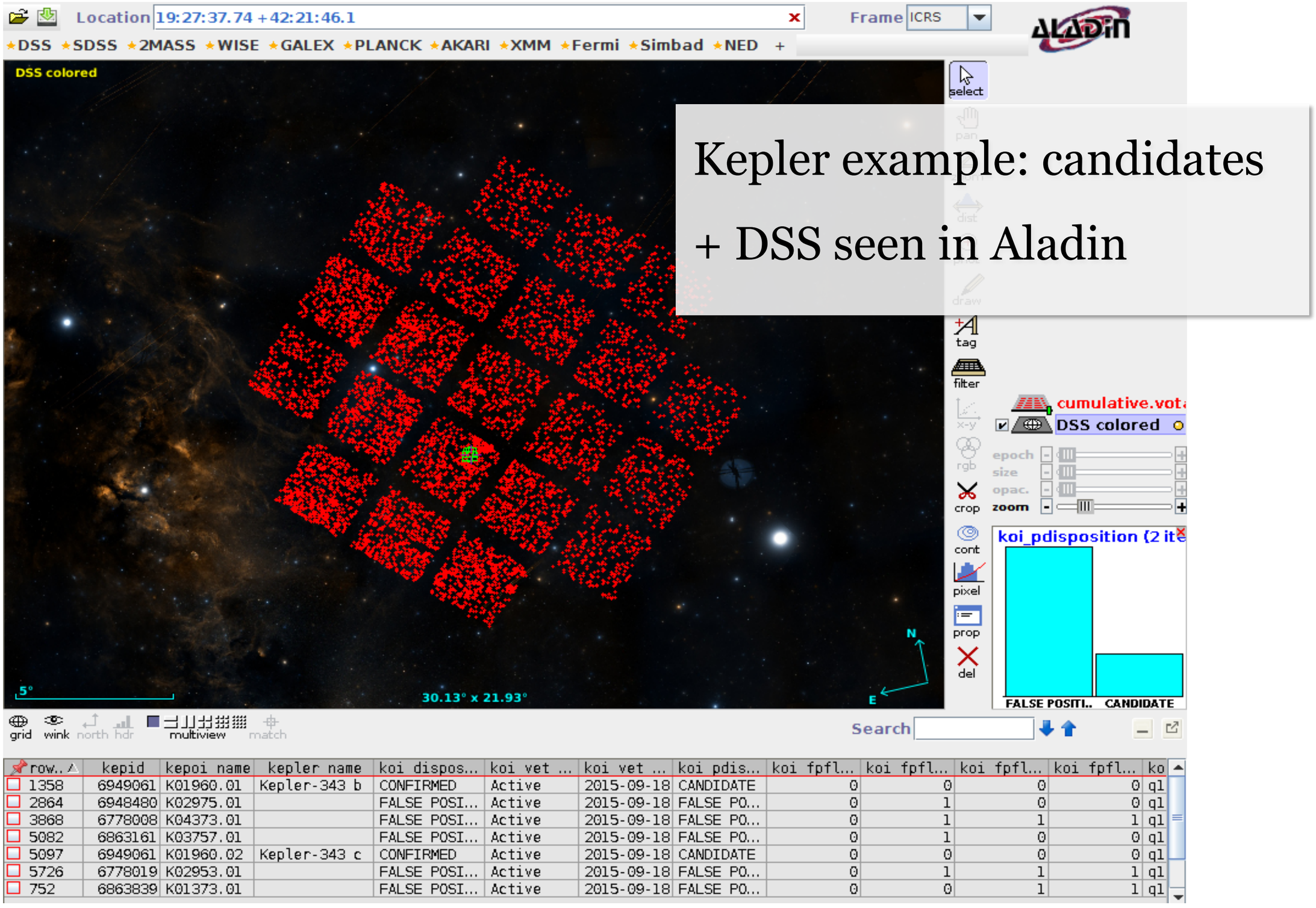
Statistical tools: example for stars

- rotation period derived with Kepler based on star spots and astero-seismology
- How well do the period measurements agree?
- PLATO: number of objects will be higher and error bars smaller

PLATO and the Virtual Observatory

- VO standards will be included in PLATO software development
- SAMP and VOTable support planned for dedicated tools at PDC main database as well as data analysis support tools
- other VO standards not yet considered
- Interaction with other software like Aladin, Topcat, …

PLATO and the VO: example



Kepler example: candidates

+ DSS seen in Aladin

row..	kepid	kepoi name	kepler name	koi dispos...	koi vet ...	koi vet ...	koi pdis...	koi fpfl...	koi fpfl...	koi fpfl...	koi fpfl...	ko
1358	6949061	K01960.01	Kepler-343 b	CONFIRMED	Active	2015-09-18	CANDIDATE	0	0	0	0	ql
2864	6948480	K02975.01		FALSE POSI...	Active	2015-09-18	FALSE PO...	0	1	0	0	ql
3868	6778008	K04373.01		FALSE POSI...	Active	2015-09-18	FALSE PO...	0	1	1	1	ql
5082	6863161	K03757.01		FALSE POSI...	Active	2015-09-18	FALSE PO...	0	1	0	0	ql
5097	6949061	K01960.02	Kepler-343 c	CONFIRMED	Active	2015-09-18	CANDIDATE	0	0	0	0	ql
5726	6778019	K02953.01		FALSE POSI...	Active	2015-09-18	FALSE PO...	0	1	1	1	ql
752	6863839	K01373.01		FALSE POSI...	Active	2015-09-18	FALSE PO...	0	0	1	1	ql

Summary and Outlook

- The PLATO science data products will be generated at the PLATO Data Center which is led by MPS Göttingen.
- The PLATO Data Center will offer VO compliant tools and services for use by the mission consortium, including tools for statistical analysis of large data sets.
- A selection of the tools will be made available to the public by ESA.
- Open questions: the final list of tools for statistical analysis and data mining tools needs to be identified.
- Check PLATO M3 Yellow Book (2013). New M3 Red Book is being prepared.