PLATO Data Analysis System: Status and Structure

Laurent Gizon

DAS coordinator, Max Planck Institute for Solar System Research (MPS, Germany)

Documentation

- PLATO Data Center Assessment Study (July 2009, appendix submitted to ESA)
- DAS section of PLATO Yellow Book (December 2009)

→ Based on 3 DAS meetings over past 2 years
→ assessment study by L. Gizon (MPS), N. Walton (Cambridge), T. Appourchaux (IAS), with input from DLR-PF, LAM, LESIA, Leicester, Leuven, MPIA

In short

The Data Analysis System (DAS, on the ground) is in charge of the validation, calibration, and analysis of the PLATO observations.

It delivers the final science PLATO Data Products.

PLATO Rationale

- high-precision photometry of a very large sample of stars to search for transits
- observation of bright stars, amenable to high-precision follow-up observations (spectroscopy, Gaia photometry)
- High-cadence long-duration observations for the asteroseismology of planet-host stars (Rs, Ms, age_s → Rp, Mp, age_p)

Three Data Levels

- Level 0: delivered by individual telescopes: light curves, centroid curves, and selected imagettes. Analysis of imagettes to **validate** performance of on-board treatment and **optimize** it.

- Level 1: On-ground instrumental **corrections**, such as those related to temperature sensitivity, some specific CCD corrections, and most importantly jitter a posteriori correction. Also, calculation of suitable averages of individual light curves and centroid curves for each star.

- Level 2: includes transit detection and measurements, stellar oscillation mode parameters, as well as star and planet characteristics. High scientific added value, and uses L1 and information from an ancillary data base.

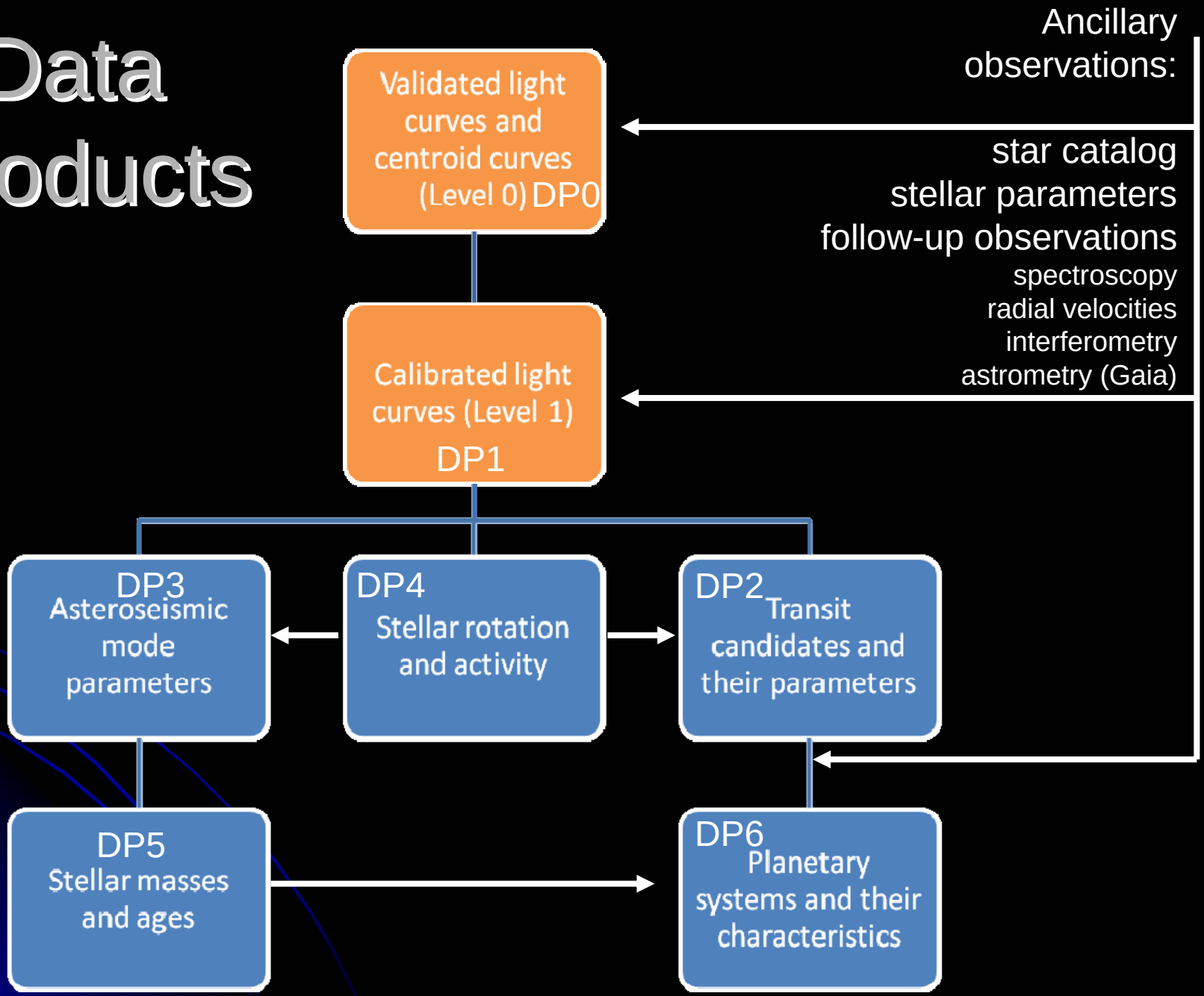
Telemetry

Baseline: 109 Gb/day uncompressed (8.7 Mb/s compressed during 3.5 hr each day)

Light curves downloaded for all stars and all 32+2 telescopes, and many thousands of imagettes at high cadence

If additional NASA ground station, then telemetry rate may double → Many more imagettes can be downloaded.

7 Data Products



PLATO Data Products: Exoplanet DPs

- **Transit candidates and their parameters (DP2)**
  - List of transit candidates,
  - List includes candidates from centroid curves (astrometry)
  - Ranking of candidates according to planetary likelihood,
  - Basic characteristics of the transits (depth, duration, period, and ephemerids).

- **Planetary systems and their characteristics (DP6)**
  - The most important PLATO deliverable
  - List of confirmed planets, using follow-up observations
  - Assessment of false alarm probability
  - Potentially several hundreds of planetary systems for which the seismology of the central stars is possible.
  - Determination of the planet parameters: orbital parameters, planet size, mass, density (average composition), age (from central stars)
  - Any additional characterization of planet properties from follow-up observations and light curves analysis, e.g. planetary atmospheres etc.

PLATO Data Products
Stellar DPs

- **Asteroseismic mode parameters (DP3)**
  - For most stars
  - Frequencies, amplitudes, and lifetimes of the modes of oscillation.
  - From fits to spectra of stellar oscillations.

- **Stellar rotation and stellar activity (DP4)**
  - Rotation periods from activity-induced periodicities.
  - Whenever possible, characterization of stellar activity: activity level from low-frequency power spectrum, star spot models.

- **Stellar masses and ages (DP5)**
  - For cool stars with magnitude less than 11.
  - Stellar parameters are obtained from stellar model fits to the frequencies of oscillation
  - Also, chemical composition etc. (seismo + spectro)

Ancillary Observations

- Essential information for the success of the mission.
- Support for on-board processing, on-ground calibration, and scientific data analysis
- Is a formidable, joint effort.

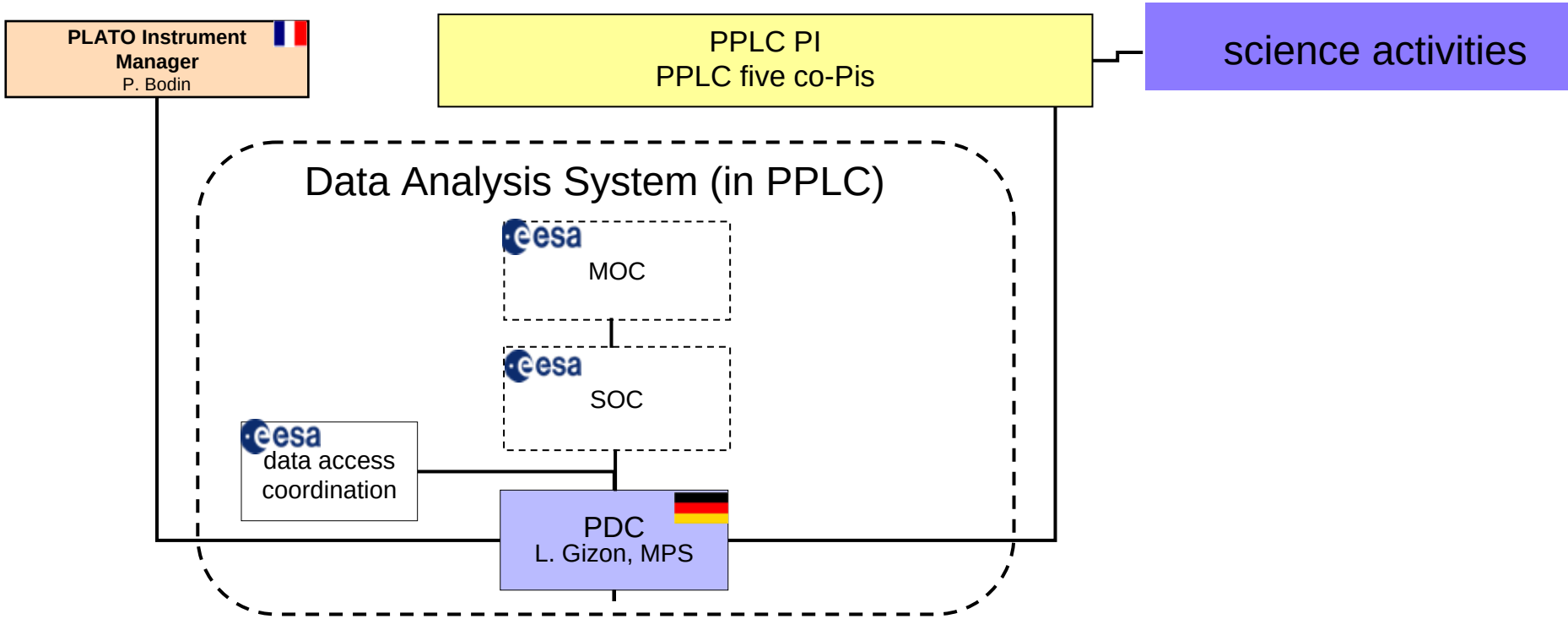
- Star catalogues
- basic stellar properties: effective temperature, absolute luminosity, radius [Gaia], chemical abundances, v sin i, activity, properties specific to multiple stars.
- Follow-up observations (RV) to confirm planets (at several wavelengths when possible)
- Other relevant complementary observations: spectra, astrometry, imaging, spectro-polarimetry, etc..

DAS Structure

The DAS, a component of the PPLC, includes:

- a Mission Operations Center (MOC, flight-critical)
- a Science Operations Center (SOC, mission-critical)
- a PLATO Data Center (PDC, science-critical)
- the DAS includes several Data Processing Centers (DPC), including one at the SOC

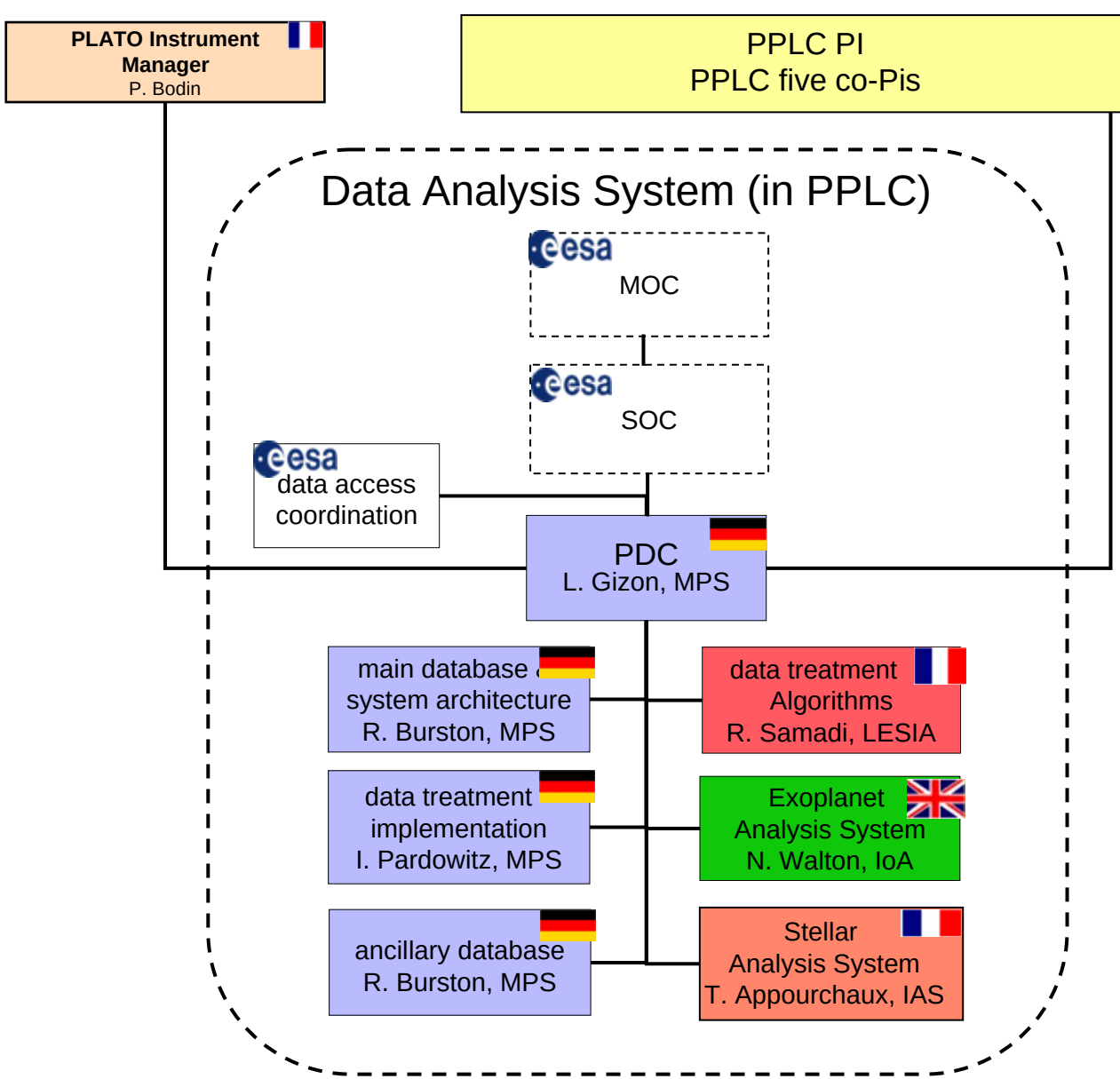
There are a number of high-level Science Activities in support of the DAS, but not included in the Payload Consortium.



SOC: includes a DPC for validation and calibration of data

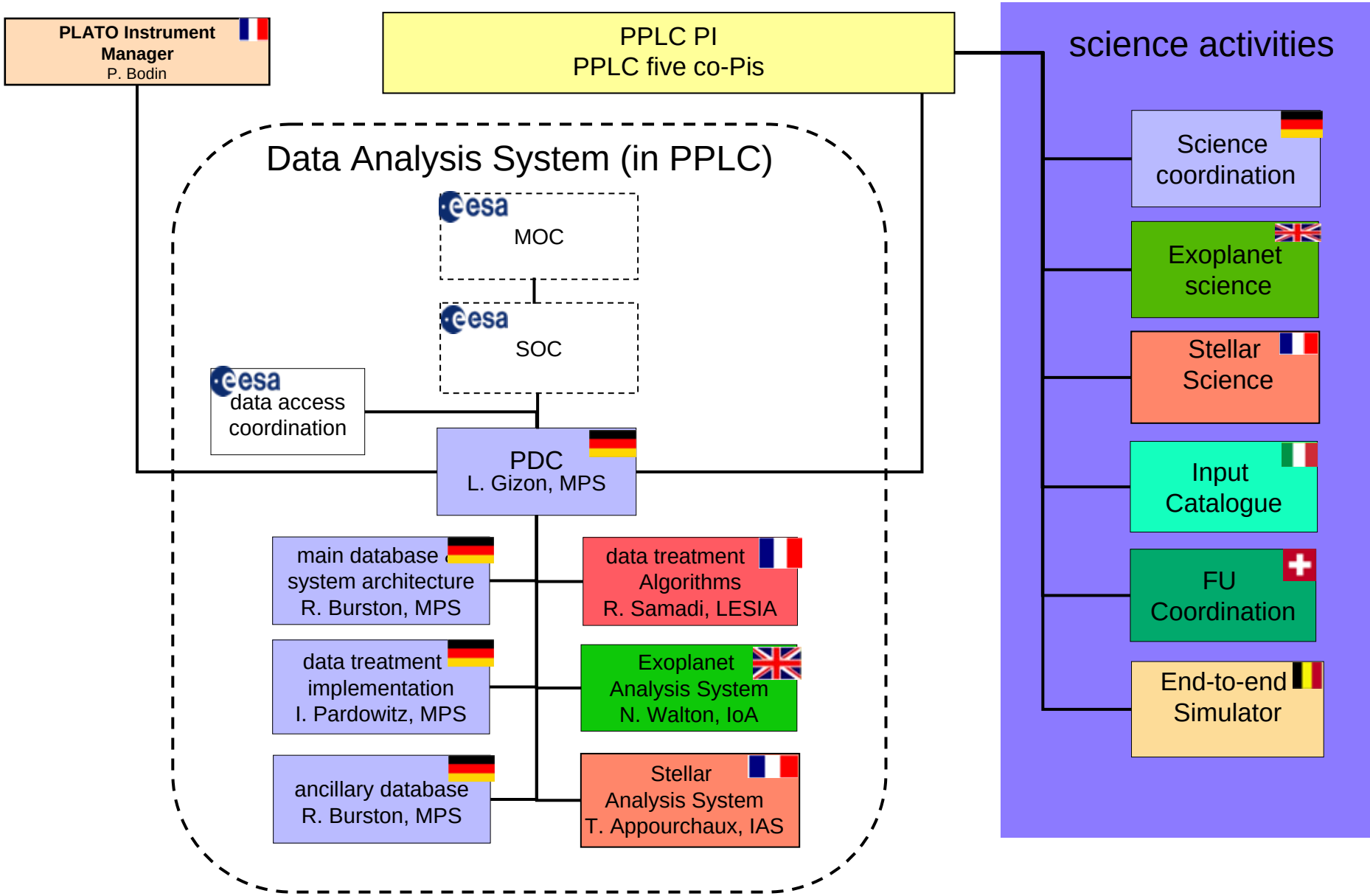
ESA overall coordination (oversight) of science data releases, data access and distribution

Note: PDC designs and implements software to be run at the SOC



PDC:

6 top-level WPs

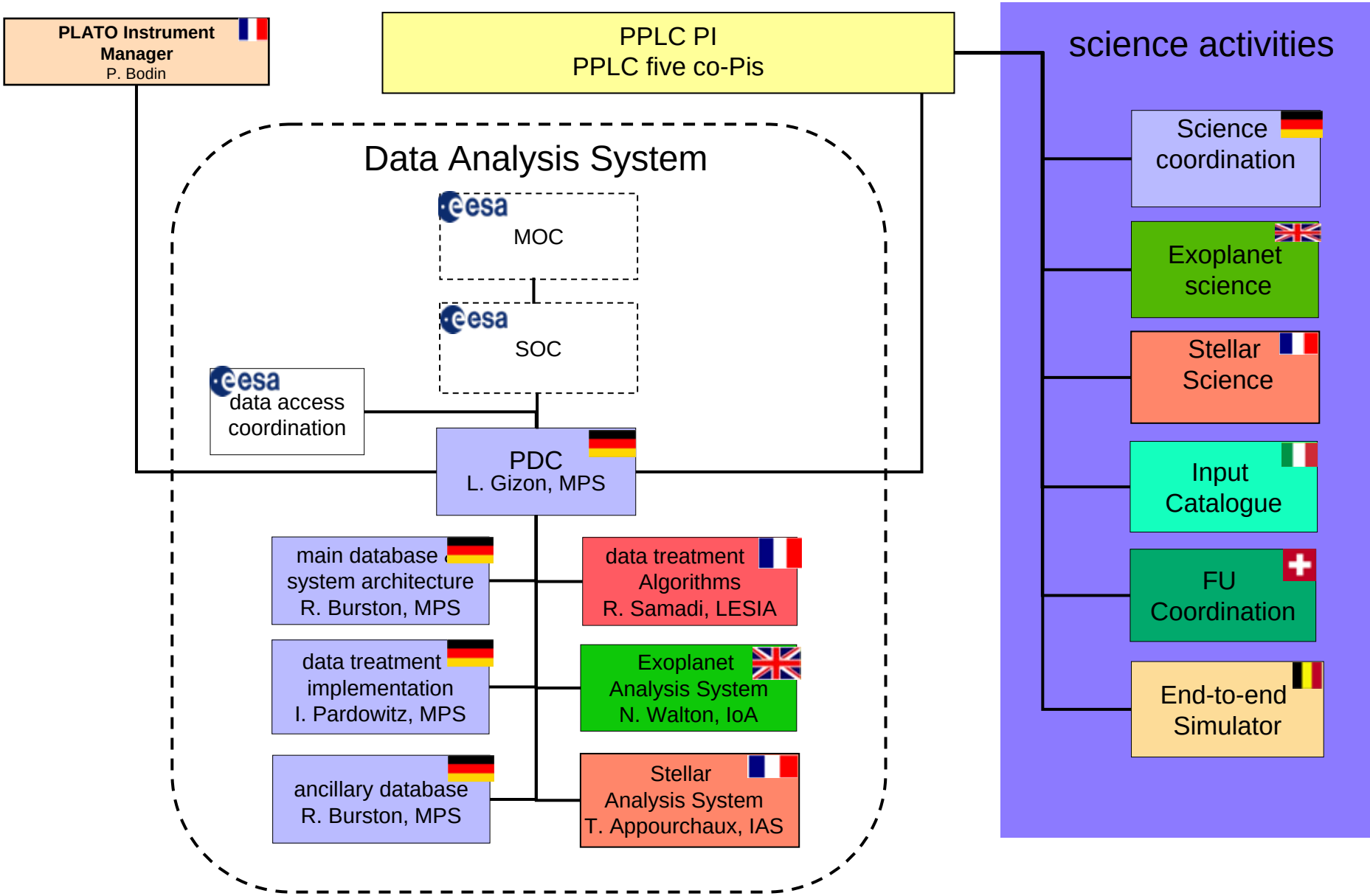

PLATO Instrument Manager
P. Bodin
PPLC PI
PPLC five co-Pis
science activities
Data Analysis System (in PPLC)
MOC
SOC
data access coordination
PDC
L. Gizon, MPS
main database system architecture
R. Burston, MPS
data treatment implementation
I. Pardowitz, MPS
ancillary database
R. Burston, MPS
data treatment Algorithms
R. Samadi, LESIA
Exoplanet Analysis System
N. Walton, IoA
Stellar Analysis System
T. Appourchaux, IAS
Science coordination
Exoplanet science
Stellar Science
Input Catalogue
FU Coordination
End-to-end Simulator

Notable changes since last meeting

- PDAAS → DAS
- Exoplanet/Stellar Analysis Systems now explicitly part of DAS
- L1 data calibration under ESA responsibility, but PDC still in charge of developing and implementing data treatment algorithms
- Simplified and more efficient PDC structure with 6 top-level WPs (instead of 15) -- including essentially the same tasks, though.
- Additional people have expressed interest in participating in the DAS Phase A activities

Top-level DAS Work Packages

PLATO Payload Consortium: PLATO Data Center (PDC) and Science Activities



WP System architecture and main database (Lead: Burston)

- System architecture, archives, database, system management
- Pipeline, workflow management system
- Data flow management, export system, network
- Simulation of data stream (uses simulations of telemetry as input)

WP Data Treatment Algorithms (Lead: Samadi)

Data validation

- Validate onboard software:
  - Check onboard processing using ground copy of onboard software and the imagettes of ~1600 stars
  - Validate distortion matrix model, 2D sky background model, PSF model fits
  - Validate computation of masks and windows
- Validate onboard setup:
  - Fine tuning of onboard software algorithm. For example choose number of parameters needed to describe PSF. Especially during configuration mode.
- Monitor health of each telescope and assess quality of the data

Data corrections

- Correction for jitter. Performed independently for each telescope; requires PSF knowledge, stellar catalog, and distortion matrix.
- Integration time correction, sampling time correction
- Statistical analysis over the 40 telescopes to identify cosmic ray hits, hot pixels, and possibly deficient telescopes
- Average light curves and centroid curves over all telescopes (weighted average).
- Compute error based on scatter
- The ~1600 stars for which imagettes are available receive a more sophisticated treatment. PSF fits to improve photometry (contamination from neighboring sources taken into acount). →Imagettes are downloaded for all stars for which a serious planetary candidate has been identified.
- Long term detrending probably moved to PDC

WP Data Treatment Implementation (Lead: Pardowitz)

- Write and implement core-processing software that will run at the SOC
- Requires a good understanding of system interfaces, esp. with ESA

- The procedure to run the data treatment algorithms at the SOC is still TBD
- For phase A, we should try to implement jitter correction in PDC demonstrator (Burston talk) to prove feasibility

WP Exoplanet Analysis System (Lead: Walton)

WP Stellar Analysis System (Lead: Appourchaux)

The focus of this meeting
Two splinter sessions

WP Ancillary Data Base (Lead: R. Burston)

The coordination of the preparatory and follow-up observations is part of the science activities that do not belong to the PDC.

Within the PDC:

- Interface PDC ←→ ancillary data from observatories
- Tools to reduce and analyze FU data (e.g. extract Teff from spectra)
- Tools to visualize the ancillary data (e.g. star field)
- The database itself (technical aspects)

The ancillary data are in support of the processing activities and are accessed by the PDC via the main data base.

Data volumes

- Telemetry rate: 109 Gb/day uncompressed
- Over a 6 yr mission: 30 TB uncompressed
- The volume of archived L0, L1 and HK data is expected to be 10-50 times this amount (reformatting and calibration history), i.e. 300-1500 TB
- The volume of the science data products is likely to be negligible in comparison (although the complexity of the data may be high).
- Ancillary data base: basic stellar observations and parameters, spectra, Gaia specific obs, etc. How big?

→The overall data volume should not exceed a few PB, which is not problematic.

DAS Time schedule
→ see Nic's talk

- **Now-June 2011: Definition Phase A**
- **June 2011- December 2011: Definition Phase B1 (detailed definition phase)**
- January 2012. Setup of project office
- June 2012. PDAS System Requirements Review
- June 2013. PDAS Preliminary Design Review
- June 2016. PDAS Critical Design Review
- June 2017. PDAS Flight acceptance Review
- December 2017: Launch of PLATO
- 3+2+1 years in space
- Several releases of science data products during and after space mission
- After end of mission in space: Several years of follow up observations to confirm a planet with e.g. T=3 yr. During this time the PDC must remain operational: 15 years

Cost

In Assessment Study (July 2009) we had estimated

- ~25 FTEs over 15 years (2012-2023+3)
- Hardware (hundreds of cores, PB storage), network, software: ~1 MEUR
- Cost around 30 MEUR

EXCLUDING:

- MOC, SOC operational tasks under ESA responsibility
- Science studies, e.g. stellar model grid computations etc. etc.
- **The Stellar and Exoplanet Analysis Systems**

How much more this time? Twice more? It must be figured out during Phase A studies.

DAS Phase A Study

Goal of DAS Phase A studies is to deliver a document by May 2011 that will

- say what steps are involved in producing the various DPs
- demonstrate that it is feasible, i.e. that there is no show stopper
- Study in detail (and simulate?) the most important parts of the data processing: e.g., jitter correction
- provide a cost estimate
- provide a schedule
- be in good shape for phase B1 (will go fast)

DAS Definition Phase Milestones

- This meeting. **Focus on Stellar and Exoplanet Analysis Systems. Confirm responsibilities.**
- Mid-june 2010: AO to create a nationally-funded payload consortium for activities not funded by ESA
- end-august 2010: answer to AO
- Oct 2010 → Kickoff meeting of consortium activities. PDC splinter during this meeting.
- End Feb 2011 → PDC Meeting in Lindau. Identify final problems. Invite Science Activities Team.
- WP Leaders deliver reports to LG by March 2011
- Integrated document delivered by LG in May 2011
- June 2011: Decision on Plato selection
- After selection! End June 2011, Meeting for Phase B1. Phase B1 will last only 3 months.

Responsibilities during DAS Phase A studies

- 3 levels of responsibilities
  - LG → delivers DAS study to P Baudin (TBC)
  - 6 top-level WP leaders → deliver WP studies to LG
  - WP contributors, to be identified at this meeting and confirmed before the summer vacations
- Top-level WPs: Deliverables, Interfaces and dependencies
  - Must be defined urgently and agreed to by WP leaders.
  - Between and within WPs
  - Between WPs and science activities
  - A communication channel is needed between DAS and science activities coordinators

important schedule dependencies must be discussed at this meeting

Just one example:

In parallel (due e.g. Dec 2010)

- Jitter correction algorithm/code
- Simulation of telemetry (end-to-end simu)
- DAS demonstrator

Then (due e.g. February 2011)

- Implementation and execution of jitter correction in DAS demonstrator to produce feasibility study

“Missing” people

- Spain, through J Miguel Mas-Hesse: i) pre-treatment of light-curves for seismology and ii) Assessment of validity of seismic models
- A Rainers, S Dreizler (Goettingen): stellar activity and ancillary data base
- M Guedel (Vienna): stars
- Belgium: asteroseismology
- T Appourchaux replaced by P Boumier

PLATO Data Analysis System: Schedule, Next Steps

Laurent Gizon

DAS coordinator, Max Planck Institute for Solar System Research (MPS, Germany)

DAS Schedule

- Oct 2010 → Kickoff meeting of consortium activities. PDC splinter during this meeting.
- End Feb 2011 → PDC Meeting in Lindau. Identify final problems. Invite Science Activities Team.
- WP Leaders deliver reports to LG by **March 2011**
- Integrated document delivered by LG in May 2011
- June 2011: Decision on Plato selection
- End June 2011, Phase B1 meeting.
- December 2011: End phase B1 (few months of work)

- Leaders of WPs defined.
- One WP has not be discussed in much detail: ancillary db.

- WP leaders identify tasks
- WP leaders identify and select contributors, and seek approval from LG and Claude, before the summer vacations

- Feasibility: specific points that require attention
  - Calibration: jitter correction
  - Stellar analysis: automatic processing
  - Stellar analysis: quality checks on mode frequency fits
  - Exoplanet analysis: complexity of processing workflow?

- LG: Talk to ESA wrt SOC interfaces
- Work needed to specify interfaces between WPs, and between WPs and science activities
- Need to improve communication
  PDC ←→ “science activities”
- Example potential inputs from “science activities”:
  - Stellar models: Do they exist? If no, what needs to be done? Is it feasible?
  - Statistical study of stellar DPs to calibrate e.g. a rotation-age relationship