CILIA DAMIANI

MPSSR GÖTTINGEN

PIC TO PDC-DB AND BACK TO PSM

PIC Meeting - 25/09/2019 - Padova

WE ARE HIRING! (JÖRG KNOCHE IS NOW WORKING AT HAMBURG OBSERVATORY)

Database Software Developer

The position is available for an initial period of 3 years, with the possibility of an extension that is contingent upon performance and project funding. The start date is negotiable.

Tasks

- Develop the PDC Database system
- Supervision of software developers
- Contribute to the production of PDC technical documentation
- Monitor and maintain the PDC schedule

Profile

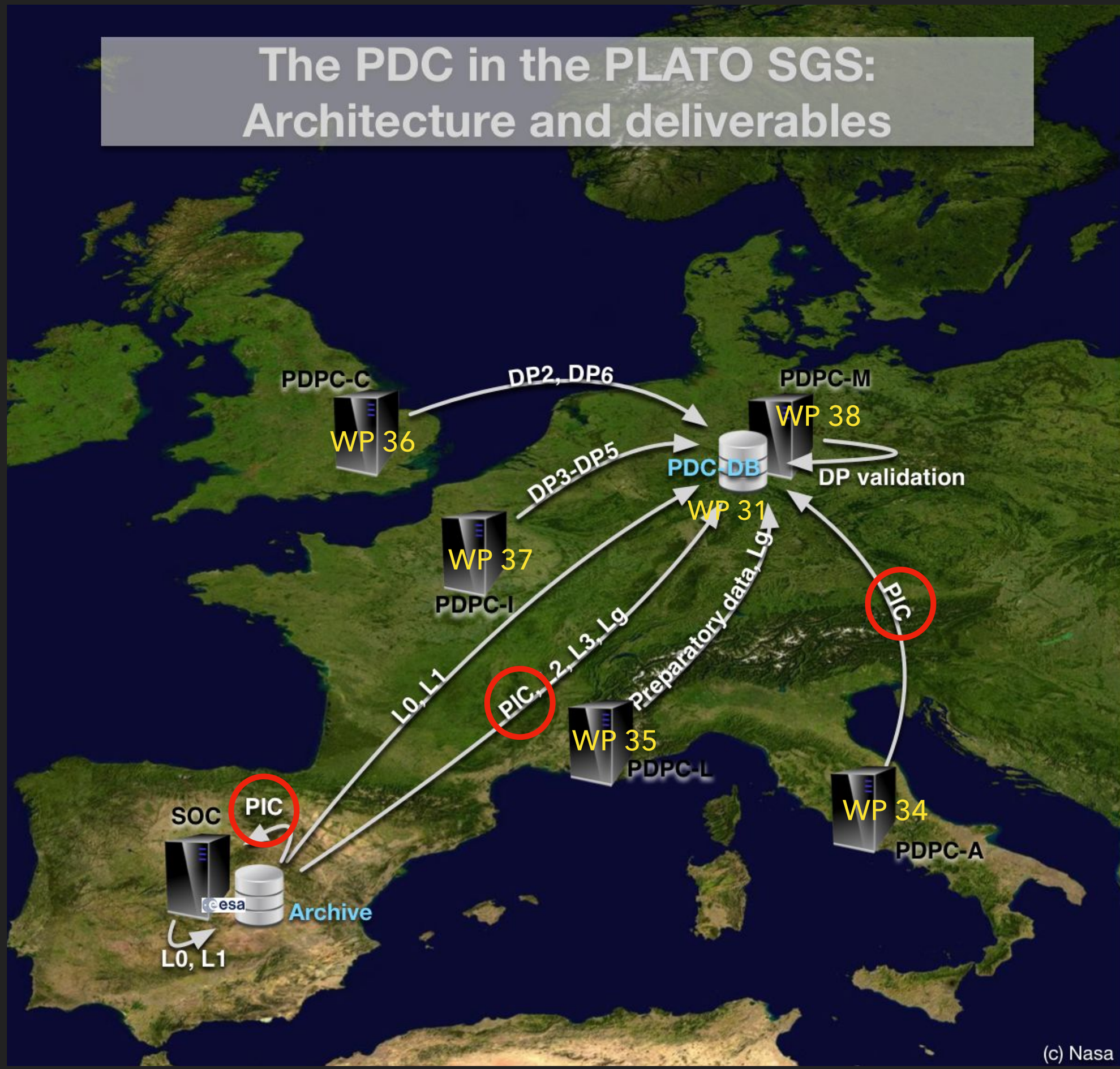
- Master of Science degree or relevant experience
- Expertise in databases, Java, Eclipse, Swing/Abstract Window Toolkit
- Experience with dev tools (e.g. Jenkins, JIRA, Bitbucket), Linux, filing systems
- Experience with space missions and scientific projects

CONTENTS

- Overall PDC organisation - Structure and data flow
- The PIC as an internal PDC interface:
  - What we think should be covered by requirements on the PIC
- … Back to PSM …

- PDC-DB is central data hub, located at MPSSR in Göttingen
- Data centres spread across Europe, will have local copies of the data
- Data to be moved back and forth between MPSSR and data centres
- ESAC (Spain) hosting SOC and public archive

- PDC-DB is central data hub, located at MPSSR in Göttingen
- Data centres spread across Europe, will have local copies of the data
- Data to be moved back and forth between MPSSR and data centres
- ESAC (Spain) hosting SOC and public archive



DELIVERY OF THE PIC : REQUIREMENTS ?

- The PIC is delivered by WP34 (Input Catalog) to the PDC-DB in Göttingen (represented by WP31, System Architecture and Management).
- In the context of the ECSS standard, this interface is considered a PDC-internal interface.
- The scientific requirements are the responsibility of the PSM (includes actual content, origin for the content)
- Need for a working interface between WP34 and WP31 early in the development phase.
- Topic will be covered later by the PMC SGS User Requirements Document (URD), Interface Requirement Document (IRD) and Interface Control Document (ICD).

WHAT WE THINK SHOULD BE COVERED BY REQUIREMENTS ON THE PIC (CHECKED BY P. MARESE)

- **PIC delivery** : The PDC-DB in Göttingen is planned to be the central data exchange hub for the PMC. Therefore the PIC will be delivered to other PDPCs via this system.
- **Completeness of PIC delivery** : As the PIC is considered as one complete set of data, there should not be deliveries of singular tables. When an update of the PIC is delivered, it shall always comprise the full set of PIC tables to prevent keeping track of versions of single tables.
- **Documentation of PIC** : A documentation package shall be included in the PIC delivery. This documentation shall at least include information on:
  - name, datatypes and units of columns
  - whether a column can have null value
  - a description of encoding, if a column is using encoded data (e.g. flags or bitmasks)

WHAT WE THINK SHOULD BE COVERED BY REQUIREMENTS ON THE PIC (CHECKED BY P. MARESE)

- **Updates to the PIC** : Each version of the PIC needs to be consistent on its own. Differential updates would maybe save some disk space, but as each PIC delivery is only ~23 GB (current estimate) this is not an issue.
- **Release numbering** : A new unique identifier (e.g. version number) shall be given to each release and included in the delivery.
- **Primitive data type usage** : For direct ingestion into the database, any data must comply with primitive data types (that are defined uniquely across the PDC).
- **Consistent data types per column** : The database won't allow the mixture of data types within the same column.
- **Handling of non-existing values** : If a value is not available for an entry, this value shall be encoded with "\N".
  WP34 should also make sure that an empty field means that no value is available (i.e. no processing error or such).

DISTRIBUTION OF THE PIC

- **PIC delivery via PDC-DB :** The PDC-DB will be the central data exchange hub for all PLATO related data within the PDC. Since it is still being designed, no details can be given on the implementation of this interface as of now.
- **PIC delivery without PDC-DB** : Until the PDC-DB exists or in case of unavailability, the PIC shall be delivered in csv format.
- **PIC for SOC** : PDC-DB shall deliver the validated PIC to the SOC no later than 10 months before launch. PDC-DB shall deliver the validated PIC to the SOC no later than 9 months before the start of each sky field observation.

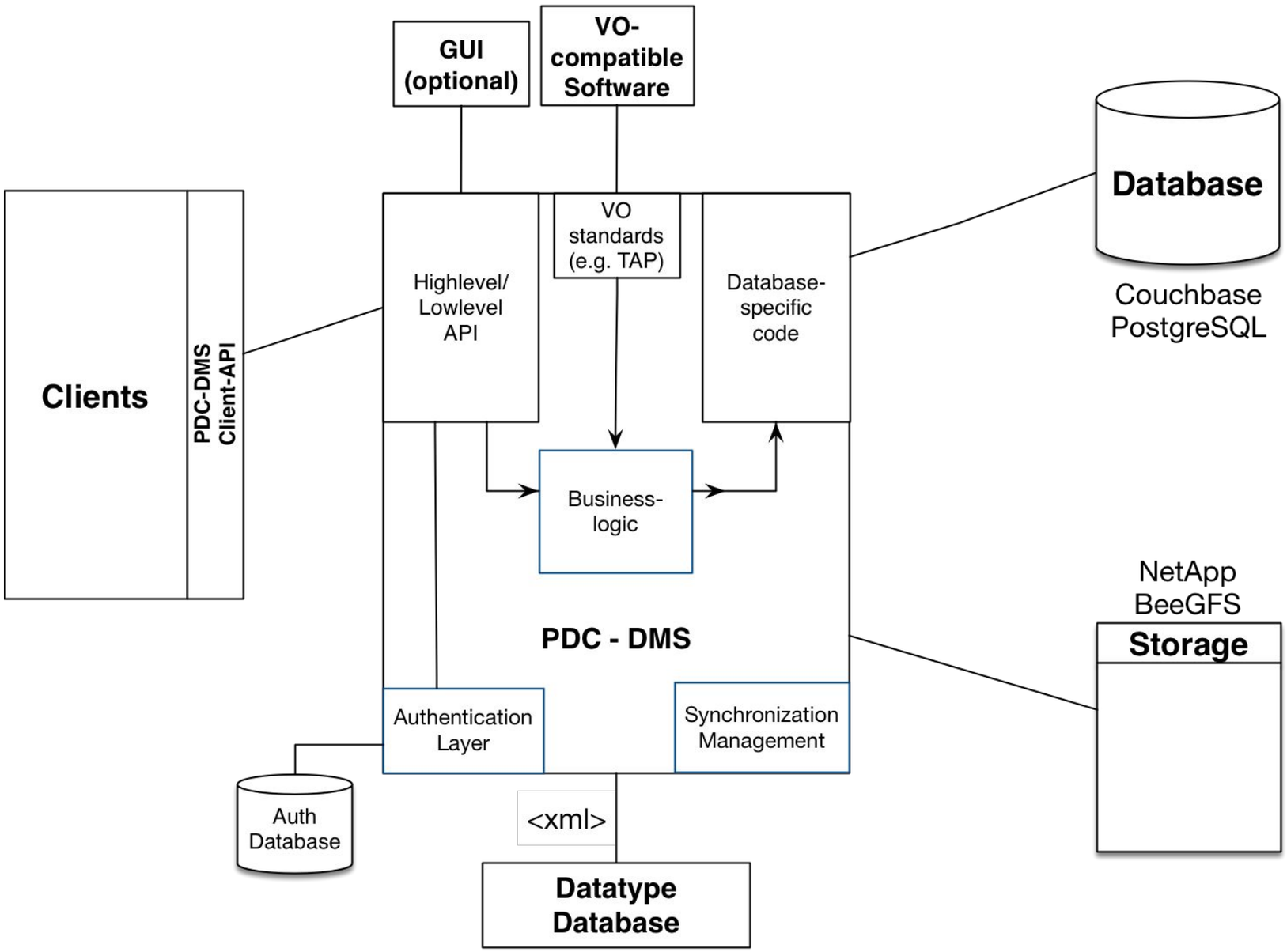
BACK TO PSM?

- **According to SIP:**
  - L+0.5 years: PIC final release for First Field
  - L+1 year: PIC preliminary release for Second Field
  - L+2.5years: PIC final release for Second Field
  - L+4 years: PIC final release for Step-and-Stare Fields
- **In the meantime**: the current version of the PIC will be accessible as any other piece of data in the database (after authentication)

A short definition of terms

Database	the “real” database, running in the background, existing software package, not written by us, examples: PostgreSQL, MySQL, …
PDC-DMS	data management system, software developed at MPSSR, “middleware”, offers API (programming interface) to the clients, translates to database specific commands, planned to support VO standards, lots of internal logic, specially designed for PLATO needs
PDC-DB	the whole thing: PDC-DMS + Database + data type definitions + (optional) Tools, GUI
API	Application Programming Interface - set of commands the PDC-DMS offers to read/write data - you probably used something like this already
VO	Virtual Observatory - set of standards for seamless integration of datasets

PDC-DB design concept - “simple”



- meta data will go into database
- “real” data will be stored in filesystem

- VO-Software connects to dedicated interfaces
- all other software accesses via API
- authentication and synchronization are dedicated processes

CONCLUSION

- The PIC is a PDC-internal interface (WP31 <-> WP34)
- Topic will be covered later by the PMC SGS User Requirements Document (SGS URD), Interface Requirement Document (IRD) and Interface Control Document (ICD)
- A GUI access to the PIC is at this stage optional, and not a responsibility of WP 31