PSIDA 2022

ESA's Planetary Science Archive: Present and Future

Ruben Docasal (Ruben.Docasal@ext.esa.int, Aurora Technology BV for ESA) and the PSA team

Background

- This presentation is intended to show the **current status** of the **Planetary Science Archive**, **what has been changed so far** and what is **expected** to be done for the **incoming weeks/months**. All of this from the **technical (not scientific)** point of **view.**
- This is based on a **small part** (since **2015**) of the long history of the archive, originally developed more than **12 years** ago.
- Also, some of the **features** have been gradually **started** from the pandemic in **early 2021** and more specifically from **Summer 2021**, after the new Frame Contract took place.
- This **all** has been made **possible** by the **fantastic work** and **huge effort** of **great people** ☺:


Jaime


Osi


Fran


Fernando


Ricardo

Joaquim

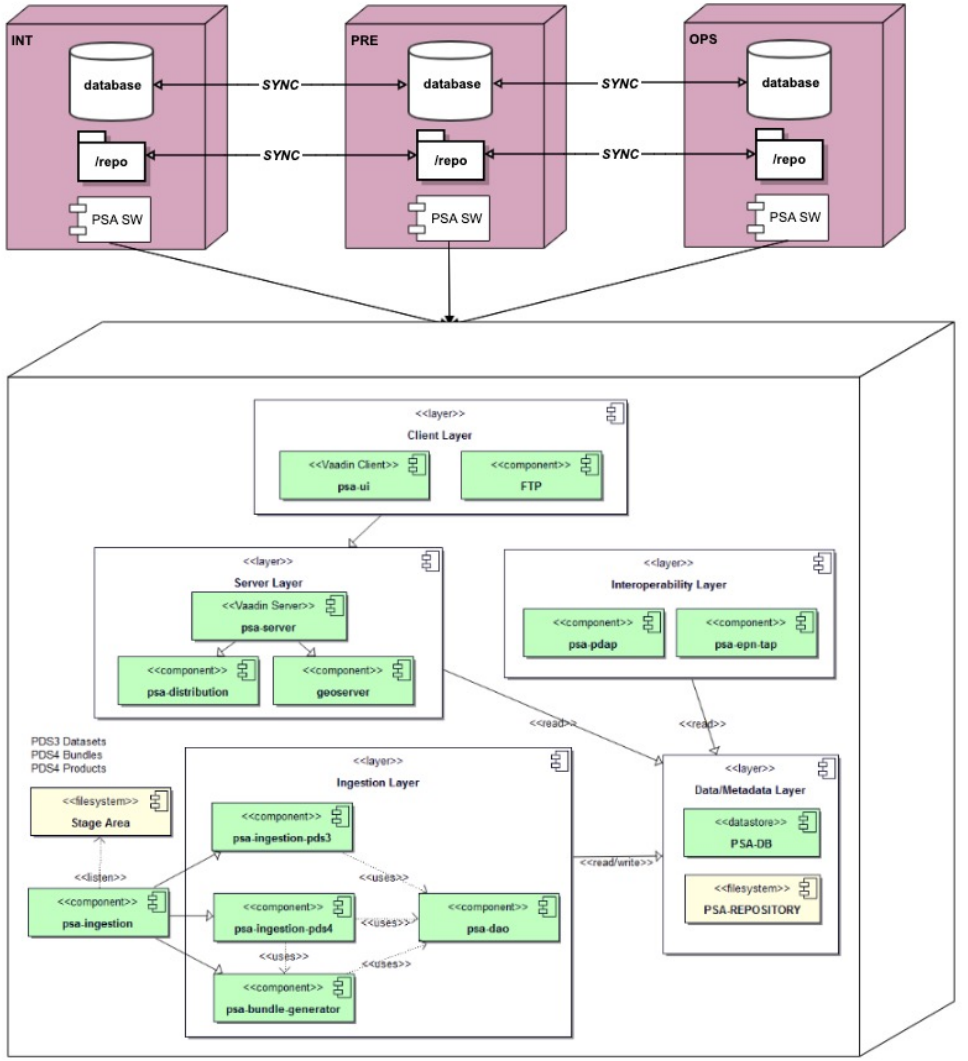

Pablo


Ruben

… And of course, although not shown here, the **former PSA mates** who contributed to the archive development.
A VERY BIG THANK YOU!

PSA current architecture

- 3 Environments: **INT**, **PRE** and **OPS**
- **Each** environment has its **own DB** and **repo**.
- **Syncing DB** and **repo** from OPS to PRE and INT
- PSA software is built on several tiers:
  - Client Layer:
    - **GUI, FTP**
  - Interoperability Layer:
    - **PDAP, EPN-TAP**
  - Server Layer:
    - **Data Distribution**
  - Data Layer:
    - **Database (DAOs/Hibernate, JDBC)**
  - Ingestion Layer:
    - **PDS3 Ingestor**
    - **PDS4 Ingestor**
    - **Bundle Generator**
    - **PSA Distpacher**



PSA (major) current issues

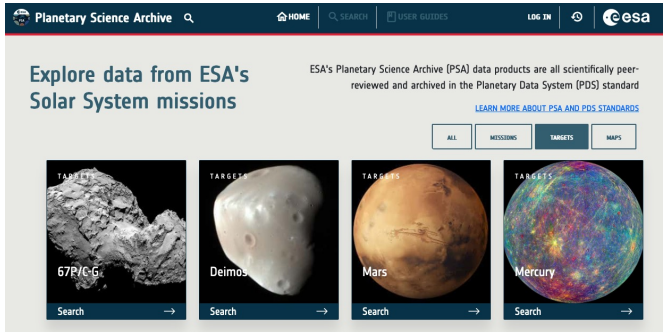
- Currently the PSA 6.2.2 installed in OPS (May 2022), a tough one (major version) with various iterations with SAS and solving E2E testing issues, being released after a long while (since April 2021) -> a need of releasing more frequently, being more agile, with less features tested during the sprints, resulting in technical and managerial changes.
- Based on a GUI (Vaadin 8) a bit of challenging to develop complex features from the client side (intensive use of wrappers). Also a ESDC constraint to use Angular as part of the new GUI generation of archives -> a need of using a more modern GUI web-based application.
- Dispersed interfaces to access the information (GUI, PDAP, EPN-TAP…) and different (obsoletes?) technologies of accessing the information (JDBC, Hibernate, store procedures, TAP…) -> a need of unify the access as much as possible by reducing the number of interfaces and technologies to be used.
- A Testing Framework (Cartographer & Selenium), for the E2E test cases, hard of maintain and strongly dependent on the data in some cases -> a need of using a more modern and reliable technology (Cypress along with Cucumber) to build the new PSA E2E test cases.

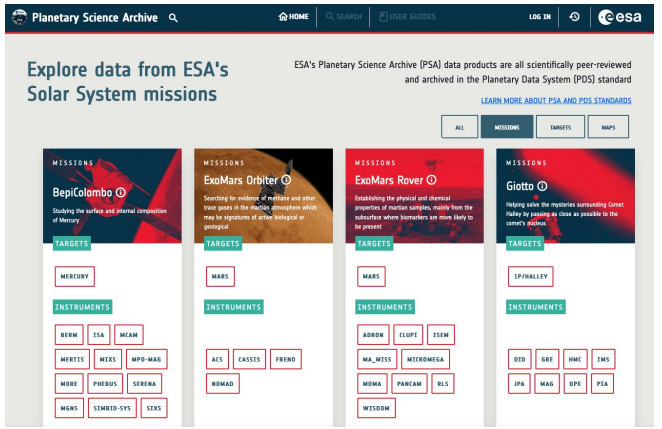
What have we achieved so far?

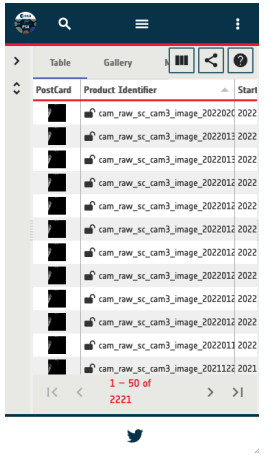
- A new **GUI based on Angular** – Started in February 2021 (ongoing, see further info next).
- A single access from the PSA GUI through **TAP+** – Started in March 2021 (ongoing, see further info next).
- **GIT** migration from SVN – October 2021.
- **Splunk** tool for FTP statistics via HTTP(s) and GSF statistics – October 2021
- **Cypress** Framework for the E2E tests (replacing Cartographer & Selenium) – March 2022.
- **User Stories** (US) applying in Scrum to be more *agile* – April 2022.
- **CI/CD** in the Integration Environment – April/May 2022.
- **PostgreSQL** updated to v13.6 – April/May 2022.
- **Quality Code improvements**, in Sonar by reducing blockers/critical issues and reaching above the 80% of code coverage in some of the PSA projects to keep the quality high – May/June 2022.

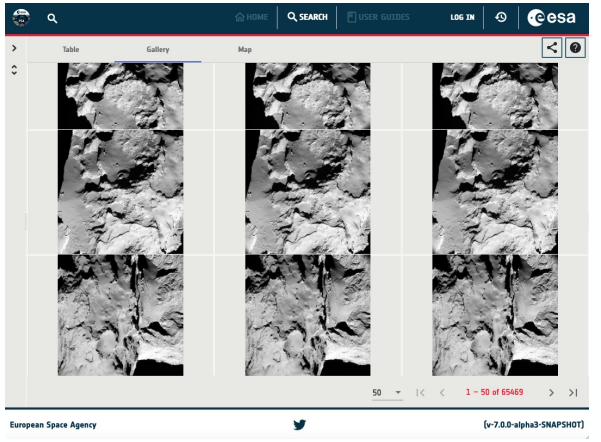
New GUI (part 1)

- Based on **Angular** v13 (**Material** for components design). Only in the INT env so far, soon in PRE.
- Refurbishment of the **Home** page (new L&F) based on the new **ESA branding** and the latest ESDC archives generation templates.
- Enhancing the User Experience (**UX**)
- Including the **visualisation** of the data via **Table**, **Gallery** and **Map** as the current PSA GUI.
- Implemented under a **Responsive Design** (tablets, mobile phones…).





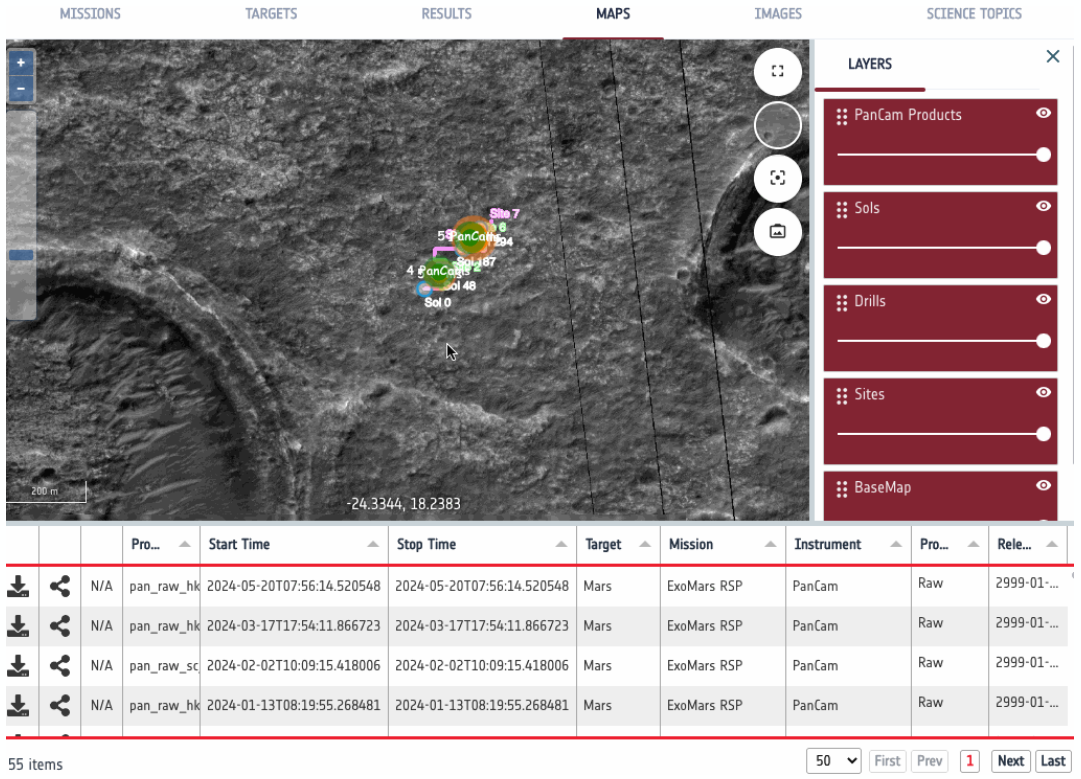




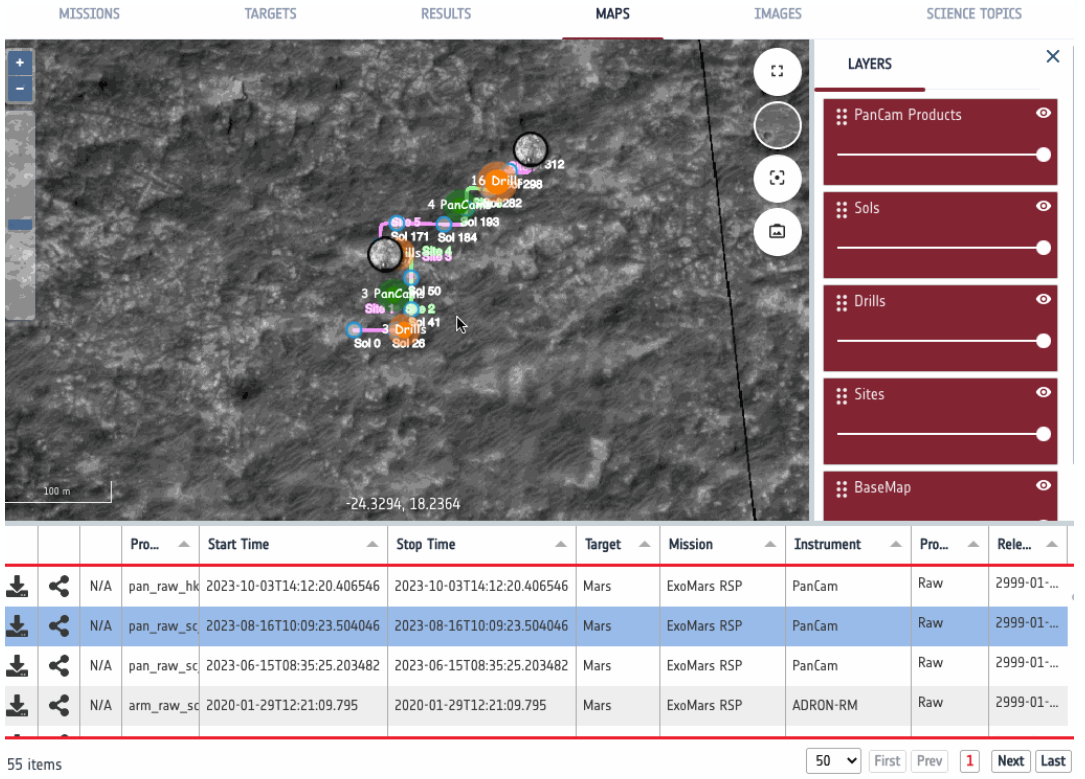
New GUI (part 2)

Progress on the **ExoMars RSP** mission (Traverse Map View):

- Display of the **Sites**, **Sols** and **Drills** of the Rover, as well as **RMC** information
- **CaSSIS** Raster Basemap and **PANCAM browse** products visualisation.



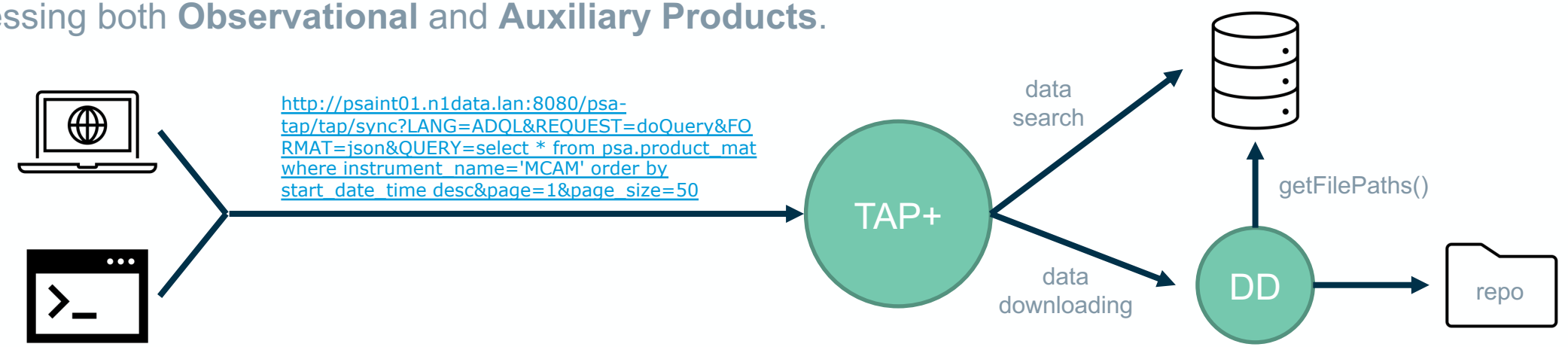
CaSSIS basemap



PanCam browse

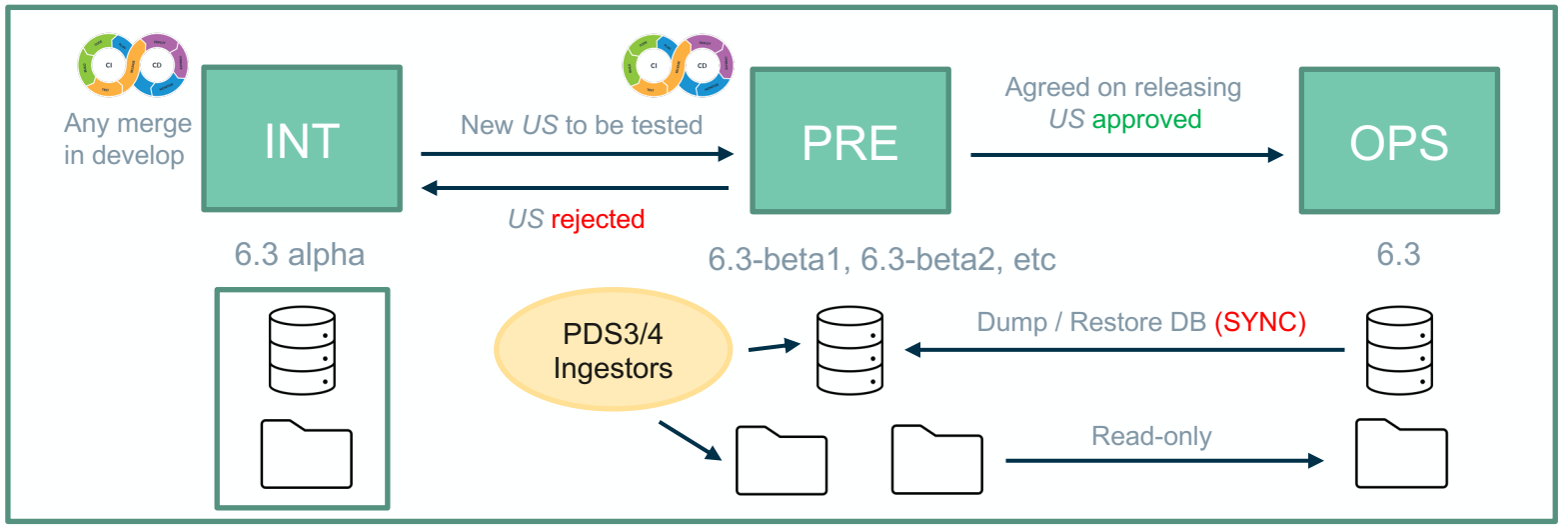
New (PSA) TAP+

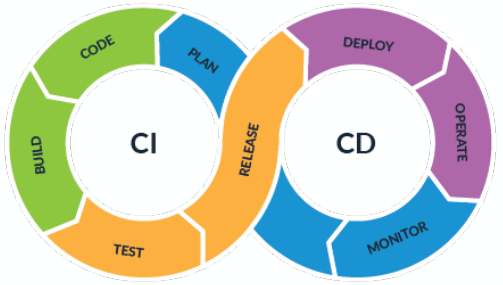
- Based on the **ESDC TAP+** (Table Access Protocol).
- Relying on **authentication** mechanism.
- Accessible via **GUI** or **programmatic.**
- Use of **ADQL** (Astronomical Query Data Language) to query the data.
- This will be the **single point of access** for the new PSA **GUI** to **search and download the data,** to **homogenise** the **interfaces** and **technologies.**
- Connected to the current Data Distribution (DD) module to allow the **data downloading** (for both **public** and **private** data) being the TAP+ to delegate the **authorisation** of PDS4 data to the Data Distribution.
- Accessing both **Observational** and **Auxiliary Products**.



Ongoing Development (part 1)

- New ways of **synchronising** the **repository** and the **database** among the different environments (**INT, PRE, OPS**):
  - **No** data **sync** in **INT** from OPS to **avoid overridden** development and **missing features**
  - Only data **sync** in **PRE** from OPS to ensure:
    - **SAS** have the **latest data** in **PRE** from OPS for **each sprint**/month (**no IT dependency**).
    - **SAS** can **ingest/delete** their own data in **PRE** to test the User Stories if needs be
- Applying the **CI/CD** in the **PRE** environment so that the **SAS** can **test quickly the User Stories (*US*)** and then releasing in **OPS** in a **monthly basis**, **speeding up** the **feedback** and the **iteration** between **SAS** and the **dev team**.





Ongoing Development (part 2)

- **GEOGEN**:
  - **New** refurbishment of the GEOGEN software from the PSA side, being developed in **python**, and improving the performance of the Datasets computation. Also creating **automatic pipelines** to compute the geometry right after the PDS3/4 ingestion.
  - PDS3 and PDS4 (**EM16/CaSSIS**) data onboard to calculate the geometry **operationally** with the most recent GEOGEN version
- Implementation of **FUSE** (Filesystem in USErspace), to access mission data as a **virtual volume**, structuring the folders and files from the database and the repository, being then made public accessible via FTP or browser.
- New **declarative partitioning** mechanism in **PostgreSQL** for some specific (and large) tables to **save space** and improving the **performance**.
- New **volumes** being created for the **legacy missions** to be used via **DATALABS**.

What's next? (incoming weeks/months)

- We aim to release the new **PSA GUI based on Angular** and the access via **TAP** this year at some moment (future **PSA 7**, stay tunned!).
- Keep **improving** the **TAP+ performance**.
- Keep **migrating** some of the **current functionalities** from the **PSA 6.X** to the new one and **improving/redesigning** others (e.g. refurbishment of the **Data Download Manager**)
- **Redesign** of the **current data model** to be more **efficient** for **searching** and **downloading**.
- **TAP+** using its **own Data Distribution** module for the data dissemination.
- Use of **Matomo** to handle the statistics along with the current **Splunk**.
- Use of **Cucumber** along with the current Cypress, making the E2E tests friendlier to write.
- Migration to **OpenJDK.**
- Use of **Docker** and **Kubernetes** for scalability. Also **microservices** will be assessed.

Questions?

THANKS FOR LISTENING!