LIFEWATCH USE CASE
All Hands Meeting (WP3-WP2)

eXtreme DataCloud

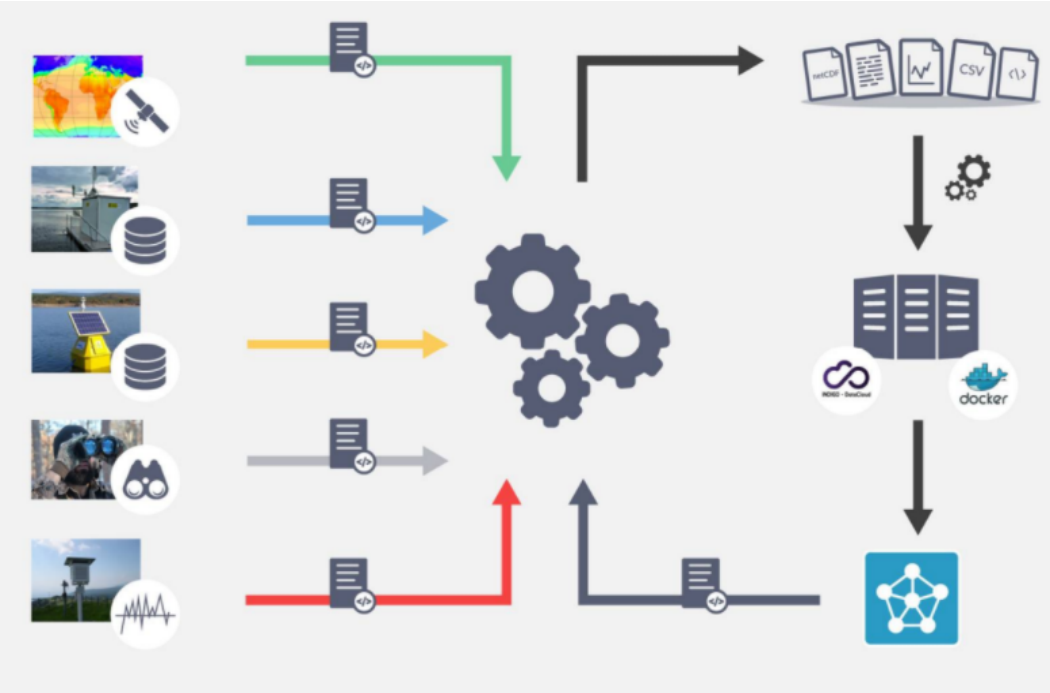
Data Management for extreme scale computing

Water Quality Forecast System

Daniel García & Fernando Aguilar
garciad@ifca.unican.es
aguilarf@ifca.unican.es

eXtreme DataCloud is co-funded by the Horizon2020 Framework Program – Grant Agreement 777367

Use Case Goals

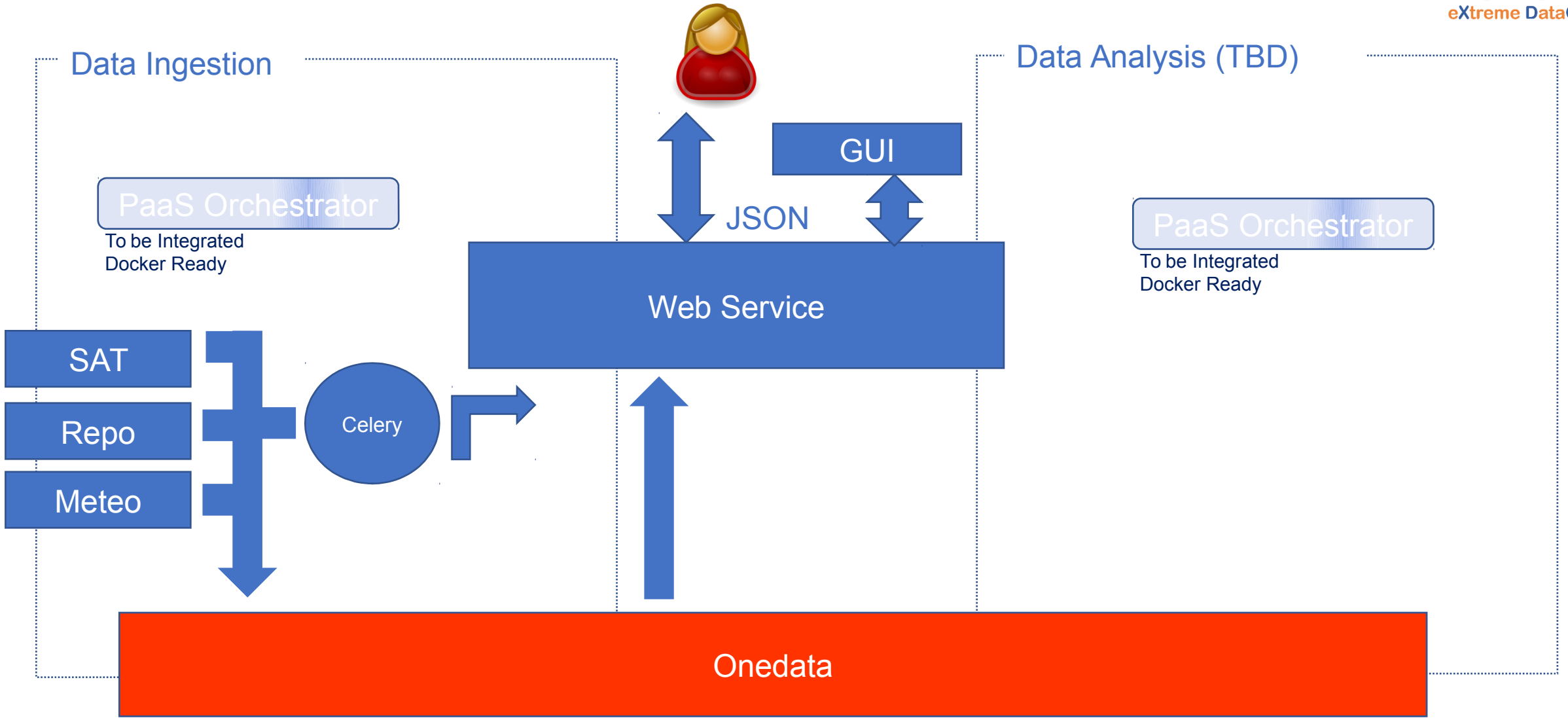


- **Objetives:** Integrate different and heterogeneous data sources: **satellite data, real-time monitoring system based on sensors, observations, and meteorological data** to feed the **hydrological and water quality models**, thus automating modeling and prediction of water quality.

- **General Requirements:**
  - Sufficient and accurate data to feed the hydrological and water quality models.
  - Good models to make the predictions.

- **Services Requirements:**
  - **OneData:**
    - OneData Attachment
    - OneData Discovery
  - **Orchestrator**

Use Case Architecture



Service Used

OneData:

- **Data Storage: Input, Ingestion, Output**
- **Metadata Attachment:**
  - Based on transformed EML(XML) (JSON)
  - Metadata attributes: admnistrative, content, structural.
- **Metadata Discovery:**
  - Defining attributes to be used by models/analisis
  - Finding data under certain constraind (geo, temporal, type, attributes).

Orchestrator: Not configured yet

- Ingestion: plugin to connect with external sources (e.g. Satellite data takes time).
- Manage workflows
- Data Analysis (output in onedata).
- Automatic ingestion (notifications required).

Integration issues

Onedata

- Potential problem with versions. Too many versions. Testbed is not the last one.
- Onedata docker: Ingestion docker based in onedata docker (Is it OK?).
- Privileged mode required?
- Map to specific user (LifeWatch – WP5 session to discuss).

Others Problems:

- PaaS Orchestrator: experience from INDIGO. Any important change?
  - Need to test.

Technical requirements

- PDDM#1: Notifications and Monitoring
- PP#1: Job-like deployment analysis
- PP#2: Data pre-processing based on software
- PP#3: Data Ingestion
- PP#4: Automated custom workflows
- DLC#1: Metadata Discovery and Data Access
- DLC#2: Metadata attachment
- DLC#3: PID minting

PDDM: Policy Driven Data Management
PP: Pre-processing, Processing & Ingestion
DLC: Metadata and Data Life Cycle Management

Testbed status

- **OneData:**
  - **~ 20 GB. More will be required to test (500GB-1TB).**

- **Planning to adopt a DevOps aproach:**
  - **Three components:**
    - **Web Service (Long-term running)**
    - **Data Ingestion (Job-like)**
    - **Data Analysis (Job-like)**
  - **API Documentation**
  - **Unit tests**