European Commission

EOPEN Side Event for ESA EO Φ-WEEK 2020
28 September – 2 October 2020
Framework for delivering interoperable digital services

How EOPEN has tackled interoperability

Bernard Valentin
EOPEN Platform Lead
Space Applications Services, Belgium

Interoperability

Interoperability is about removing obstacles to facilitate the exchange of information between applications, databases, and other computer systems.

The easier the better

What is the impact of the lack of interoperability?

- Data is difficult to find and access (syntactic interoperability)
- We may not know how to use the data or the algorithms (semantic operability) => We thus need experts to interpret the data and develop applications
- Platforms hosting the data are complex and require teams of IT specialists to appropriately deploy and run applications
- As a consequence it takes time to develop, integrate and bring applications and services to market
- Once developed the applications are difficult to change (technology legacy)

Challenges in the EOPEN Platform

- The EOPEN Platform must have the capability to run on heterogeneous hardware infrastructures that may change in time.
- The Platform infrastructure must be able to integrate proprietary servers, cloud computing resources (incl. DIAS platforms) and HPCs.
- Processing capabilities must be scalable: It must be possible to react to increasing processing needs by deploying new processing nodes on-the-fly.
- The Platform must allow scientists and application developers to implement algorithms having specific needs in terms of programming language, third-party libraries, processing resources (RAM, Disk, GPU, …), data collections, etc.

EOPEN has tackled interoperability on three levels:

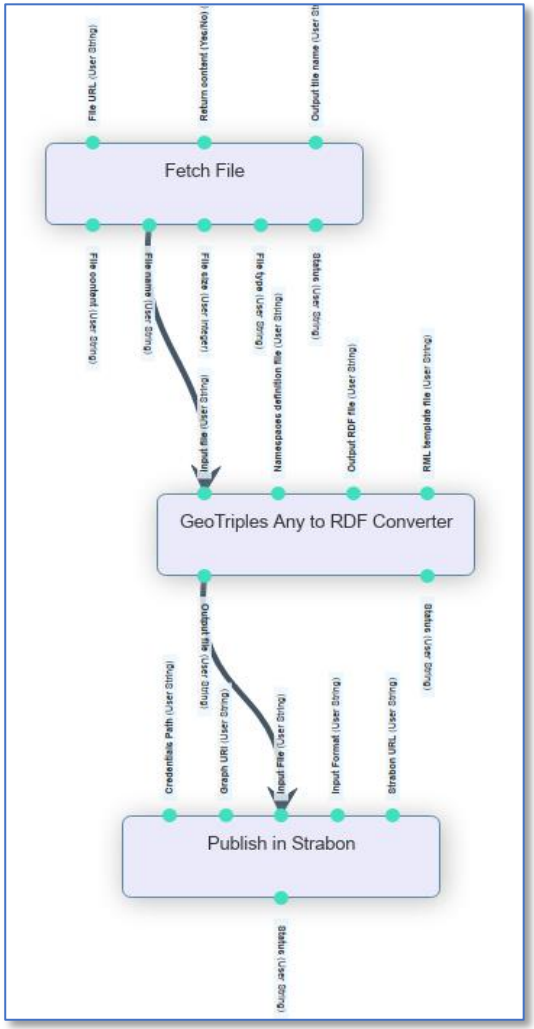
- Integration of user defined algorithms as re-usable processes
- Execution of processes in federated platforms
- Built-in processes for interacting with local and remote systems

Standards and Technologies

- Containerisation (Docker) used for core components, built-in processes and user defined algorithms
- Dockerised user algorithms are exposed via an OGC WPS service
  - A *process wrapper* function is used to integrate the algorithm within the service
- Dynamic orchestration of containers
  - Currently using Mesos / Marathon
  - Platform modularity allows integrating with other container orchestration frameworks
- Cloudify used to run algorithms in HPC environments

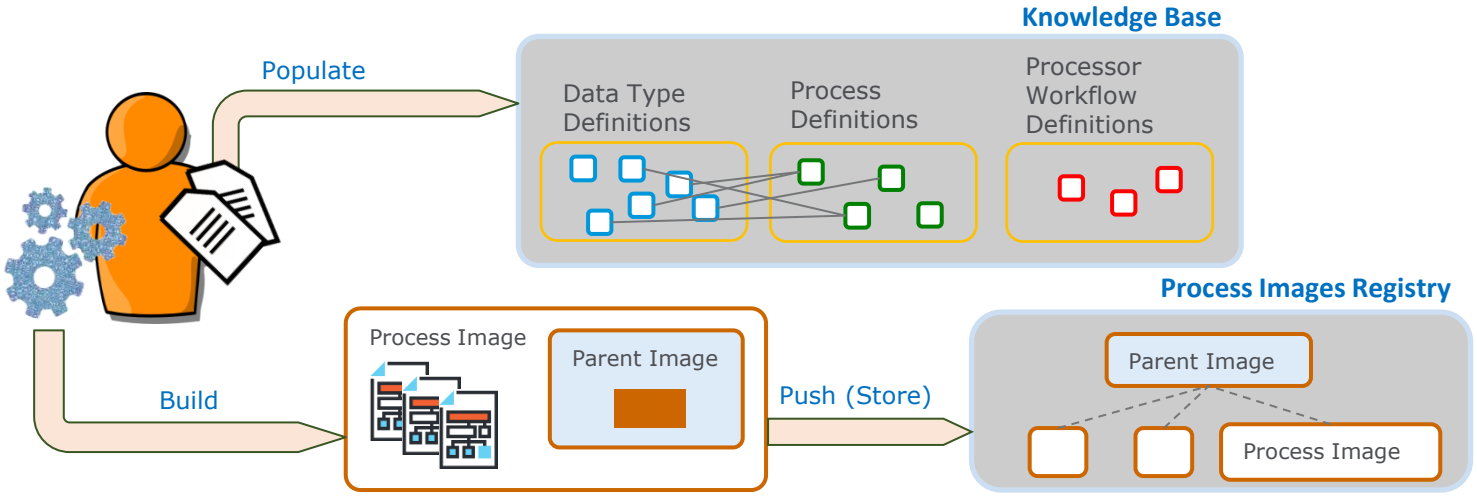
Processes and Workflows

- A **Workflow** is an application defined by inter-connected **Processes**.
  - The Platform includes a graphical **Workflow Editor** for creating workflows interactively.
  - **Workflows can be executed** on-demand, scheduled or externally triggered.
  - **Parameterization forms** are dynamically generated.
- A **Process** is a re-usable unit of execution.
  - A Process is a **Dockerised Algorithm**.
  - A Process has (typed) input and output parameters.
  - The Platform includes a tool that automates the packaging and the registration of user defined processes.



Dockerised User Algorithm

- Users provide the processing chain description and the algorithm files (artifacts) to add knowledge into the Platform: data types, parameters, processes and workflow definitions and build re-usable processes.

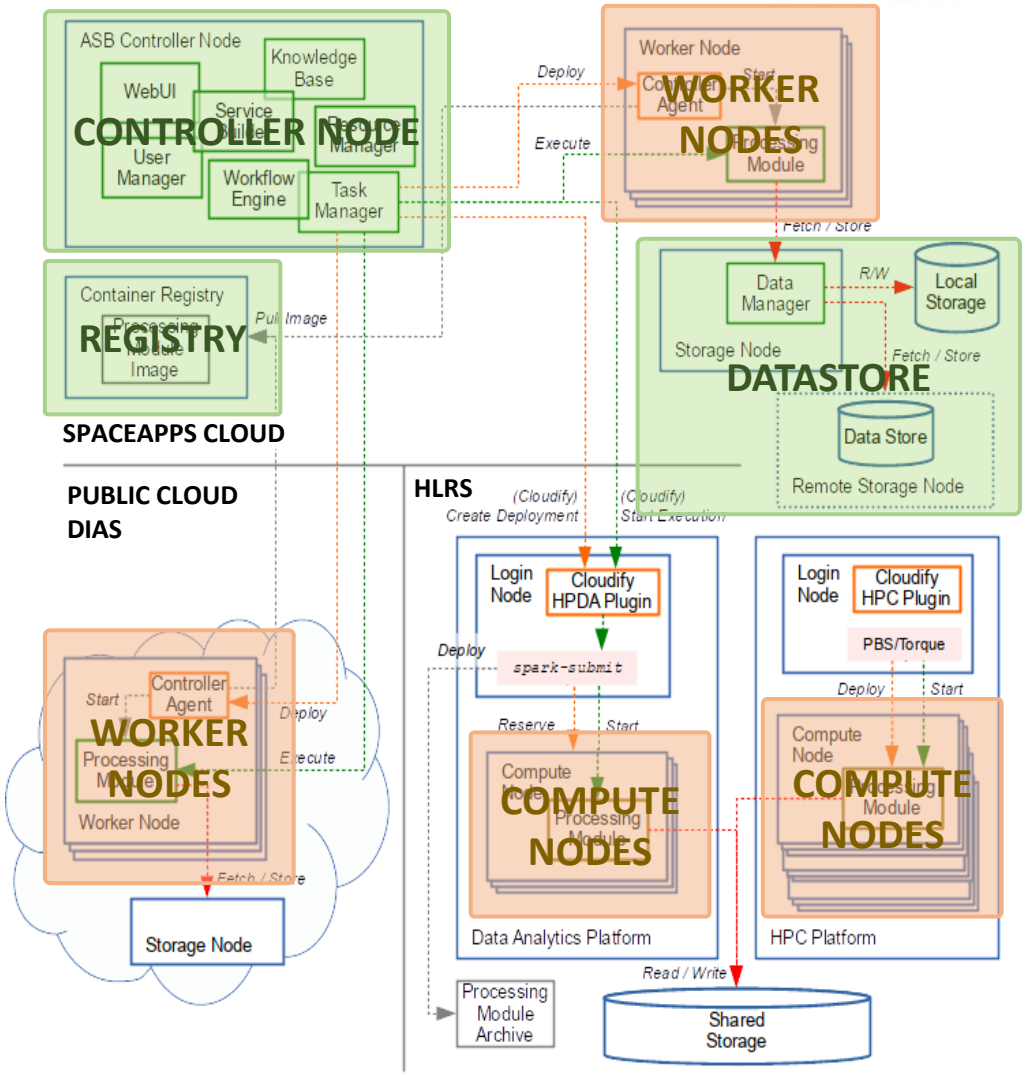


- EOPEN integrates a process import tool that automates these actions and generates consistent and standard conforming dockerised algorithms.

EOPEN Platform Designed for Interoperability

CORE LAYER:
- Controller node
- Process images registry
- Datastore

PROCESSES EXECUTED
- in Worker Nodes using Docker and
- in HPC Compute Nodes using Cloudify



Providing Federated Platforms

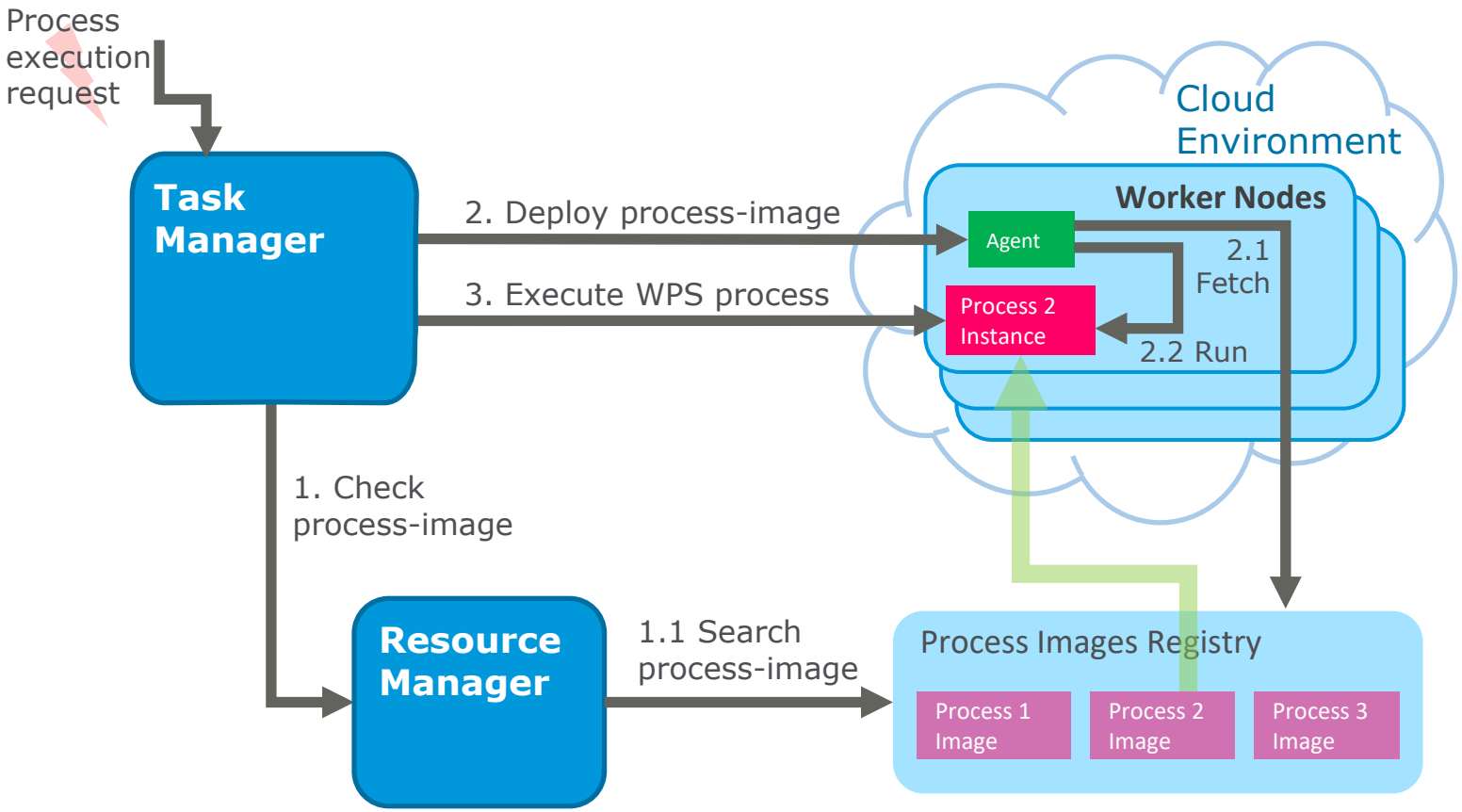
- Goal: Create a pool of processing resources located in various environments that may be used in a uniform manner to deploy and run processes.

- **Preparatory work** has required to deal with platform specificities:
  - Registration and booking for processing resources: Virtual Machines, Storage Space
  - Preparation and configuration of VMs: connectivity / networking, orchestration agent, monitoring tools
  - Configuration of the access to the data available in the platform from the VMs and containers

Using the Federated Platforms

- **Worker selectors** give the possibility to workflow developers to directly or indirectly select a target platform or a processing node having certain characteristics (e.g. GPU)
- Built-in Process for **searching for products available locally**
  - Receives search criteria
  - Identifies the hosting platform
  - Searches for products in the platform catalogue
  - Returns products metadata with local path(s)
- **Umbrella API** is an independent service that:
  - Harvests (and keeps up to date) a series of Copernicus Hubs and CollGS
  - Exposes Sentinel products metadata through a catalogue service API
  - Provides in search responses access URLs in the remote catalogues

Process Execution in a Federated Platform



A particular case: HPC/HPDA Integration

- Need to deal with provider-specific technical and non-technical methodologies and rules.
- Container orchestration (deployment, execution) may not be allowed.
- EOPEN opted for a Cloudify-based solution: blueprints describe all the steps for preparing, executing and wrapping up a process
- Generic, re-usable process integrable in workflows acting as Cloudify client.

Conclusion

EOPEN has implemented several measures to support interoperability. It provides a flexible platform for federating heterogeneous resources for users.

But, there is a constant need to respond to technology advances and user demands. Examples are:

- Connectors for other platform technologies such as Yarn, Kubernetes, Docker Swarm, etc.
- A decentralised datastore using distributed object stores or an IPFS network.
- Front-end APIs compatible with the EO Exploitation Platform Common Architecture, and in particular an ADES and/or EMS interface.

Thank you

Any questions?

Bernard Valentin
bernard.valentin@spaceapplications.com