LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
MNM TEAM
Munich Network Management Team
Maximilian Höb
hoeb@mnm-team.org
www.mnm-team.org

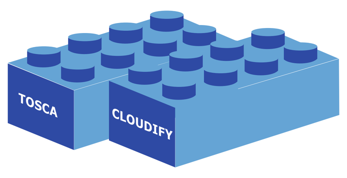
PROCESS

PROviding Computing solutions for ExaScale challengeS

www.process-project.eu

Abstract: Addressing the grand challenges of scientific research, health, engineering or global consumer services necessitates substantial changes in processing, especially for calculations needing massive performance and processing extreme scales of data volumes. In particular, the connection between data storage locations and actual computing centres needs to be adapted regarding the fast increasing demand in more data, huge amounts of data sets and increasing level of detail. Today's applications exploiting petascale clusters across the globe already pushed the bottleneck from simple computing power towards a hybrid barrier of performance and data management. This trend will be further noticed in more and more distributed applications, limiting their ability of scaling up despite the available computing power.

Virtualisation and Orchestration

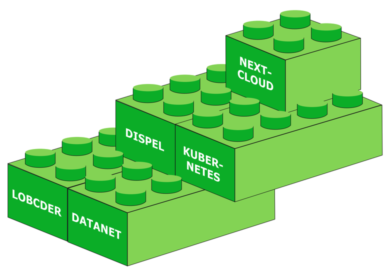


Any exascale system has to support :
- Functionalities required by user communities divert from
  - ◇ historical FORTRAN-code
  - ◇ legacy applications
  - ◇ **data intensive** applications
  - ◇ **compute intensive** applications

Containerized applications tend to be:
- **Flexible, scalable and reusable**
- Ready-to-use
- Not requiring special technical skills

Extreme data processing relies often on complex software tools involving expert knowledge about its management.
- Introduce **abstract level**
- **Hide implementation** from the users
- » **Increase number of exascale-ready applications**

Distributed Virtual File System



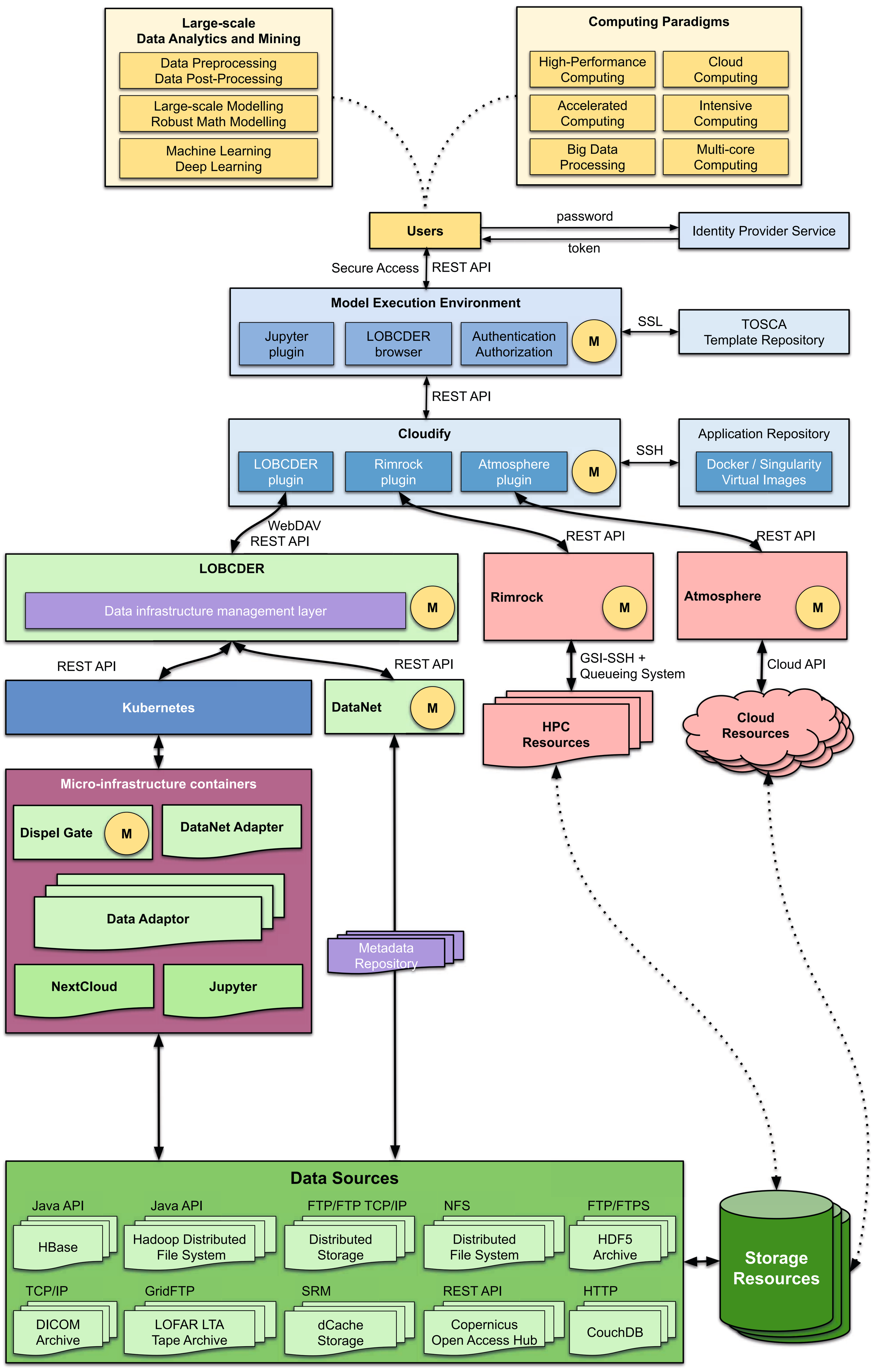
Scientific and commercial exascale application need to handle, process and work on more and more **complex and huge data sets**.

Ssuch systems need:
- To be capable of **complex workflows**
- European and inter-continental connections
- corresponding **data transfer mechanisms**
- » Disruptive **impact on entire scientific branches**, future computing and processing systems

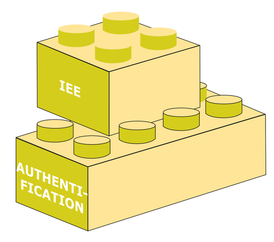
Today's and future high performance systems need to:
- Act as **distributed siblings**
- Offer **interoperability** functions
  - ◇ a **distributed file system** across local storages
  - ◇ and moreover **across geographical distant locations**
  - ◇ **Data Transfer Nodes** are mandatory

Processing enormous and complex data sets requires:
- Smart, scalable and efficient components:
  - ◇ DISPEL (**dataset pre-selection**)
  - ◇ DataNet (**meta-data handling**)
- » **Reduce transfer amount and optimize transfer**

An infrastructural service ecosystem is deployed through **micro-services**:
- Serve as adapters and connectors to any infrastructural service
- Are integrated into a **containerized micro-infrastructure**
  - ◇ customized according to the actual workflow requirements
  - ◇ connection to the distributed VFS



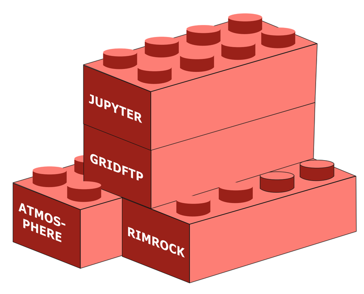
Users of exascale scientific applications



Interoperability of the data infrastructure and the computing clusters is a key critical requirement of any exascale system. Therefore:
- Architecture based on **containerisation** instead of virtual machines
- Platform **independent container**
  - ◇ comes with all its required dependencies
  - ◇ is able to access any distributed storage
  - ◇ does not rely on any local installations
- **Optimising the placement** of the required containerised components

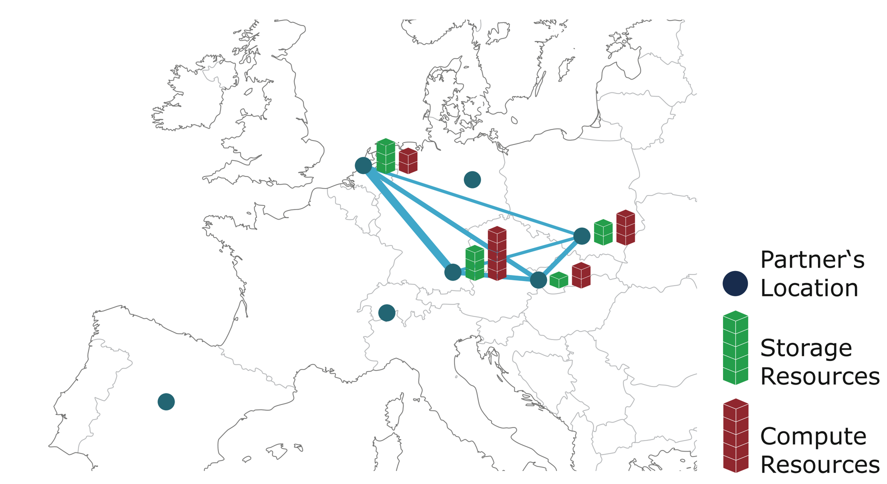
A containerised virtualisation layer **exploits the infrastructure resources in the most optimal way** and hence, supports the requirements from different communities.

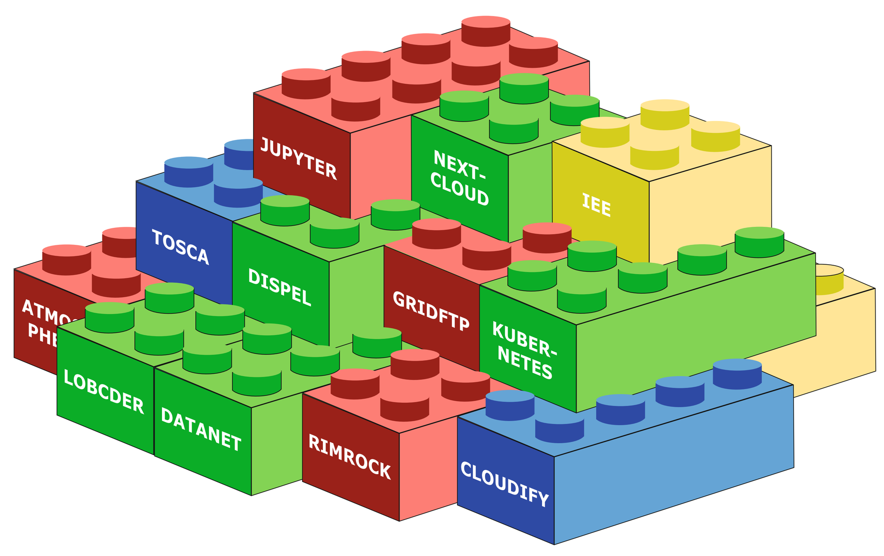
Interactive Computing Environment



To load and use the components of the infrastructure efficiently, the interactive computing environment takes care of the necessary operations:
- **Deployment** and **monitoring** of workflows on computing resources
- **Scheduling** and **load balancing**
- Include all possible **access** variations
  - ◇ communicate to the resources
  - ◇ **heterogeneous** European and worldwide ecosystem
- » **optimize utilisation of HPC or Cloud resource**

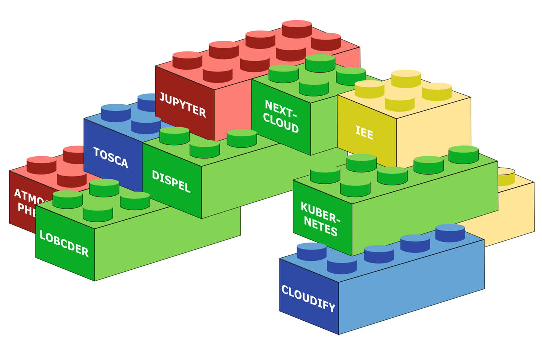
PROCESS Storage and Compute Centres

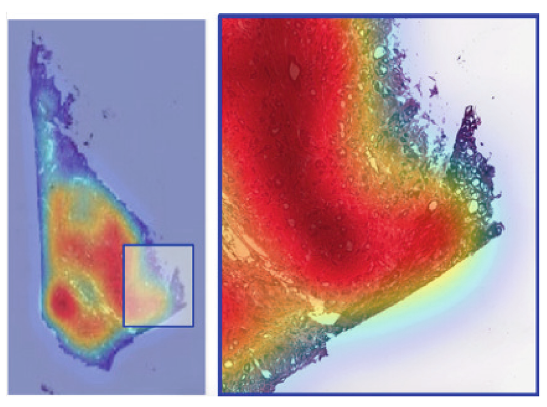


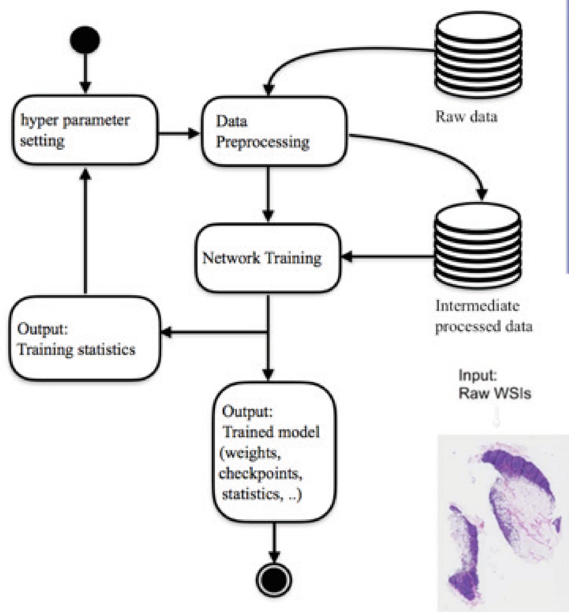


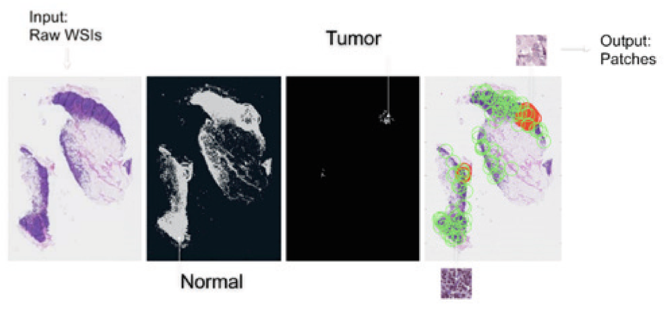
UC#1: Exascale learning on medical image data

Improve the performance of automated cancer diagnostics and treatment planning:
- Develop more powerful tools for **cancer detection**, localisation and stage classification
- Support the decisions of physicians during the **diagnostic process**
- » **Reduce time-consuming diagnostics and high disagreement rates between pathologists**
  - ◇ Visualisation and interpretation of the network decisions
- » **Increase the level of objectivity**









UC#3: Supporting innovation based on global disaster risk data

UC#4: Ancillary pricing/airline revenue management

UC#5: Agro-Copernicus

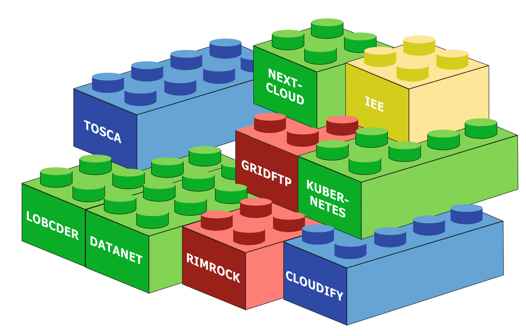
UC#2: Analysis of Radioastronomy Observations

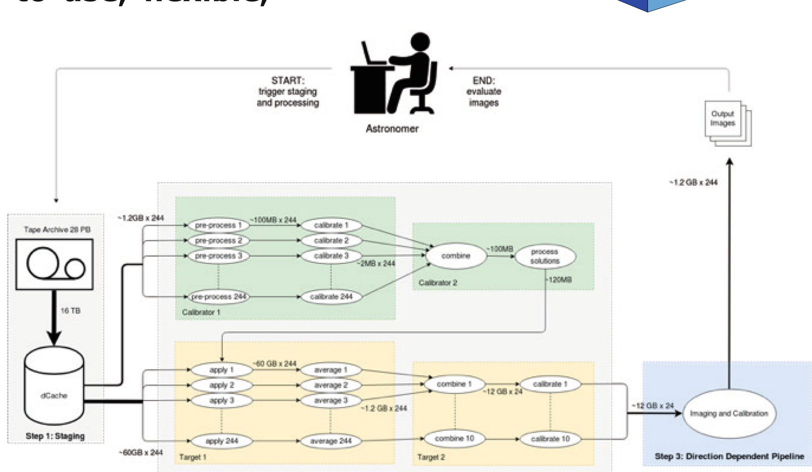
The goal of this use case is to **simplify and improve the processing of archived data of LOFAR** (and later SKA) observations.

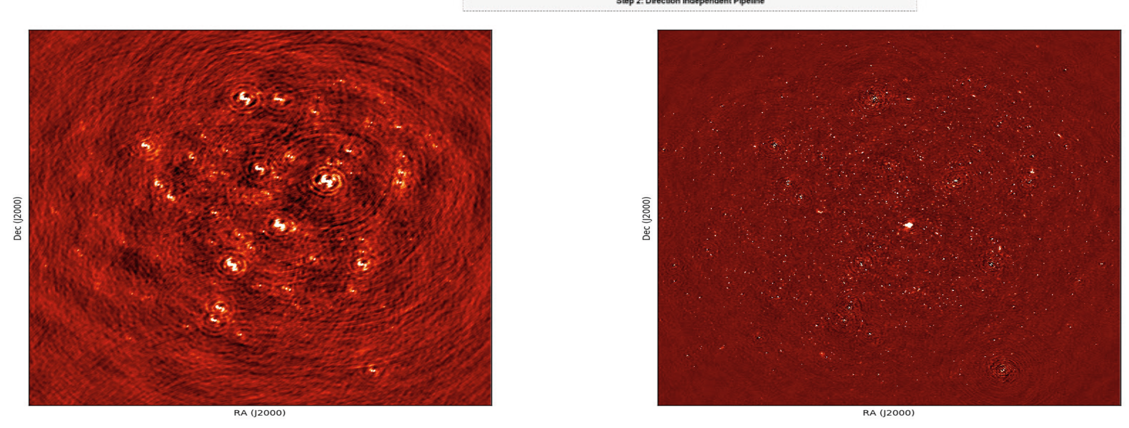
Astronomers should be able to
- select a dataset (~16TB) on a portal,
- select a workflow,
- launch the processing pipeline.

This needs an **easy to use, flexible, efficient and scalable workflow infrastructure** for processing of extremely large volumes of observation data.

PROCESS will **unlock the LOFAR long-term archive** and increase its scientific output.







This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777533.

University of Amsterdam
netherlands eScience center
Hes·so
Lufthansa Systems
Ústav Informatiky Slovenská akadémia vied
inmark europa
AGH
CYFRONET