



# EUDAT

PARALLEL TRACK 1: EUDAT SERVICES

VPH - Community integrated  
B2STAGE tools

**Stefan Zasada**, University College London, UK

**Giuseppe Fiameni**, CINECA, Italy





# Involved parties

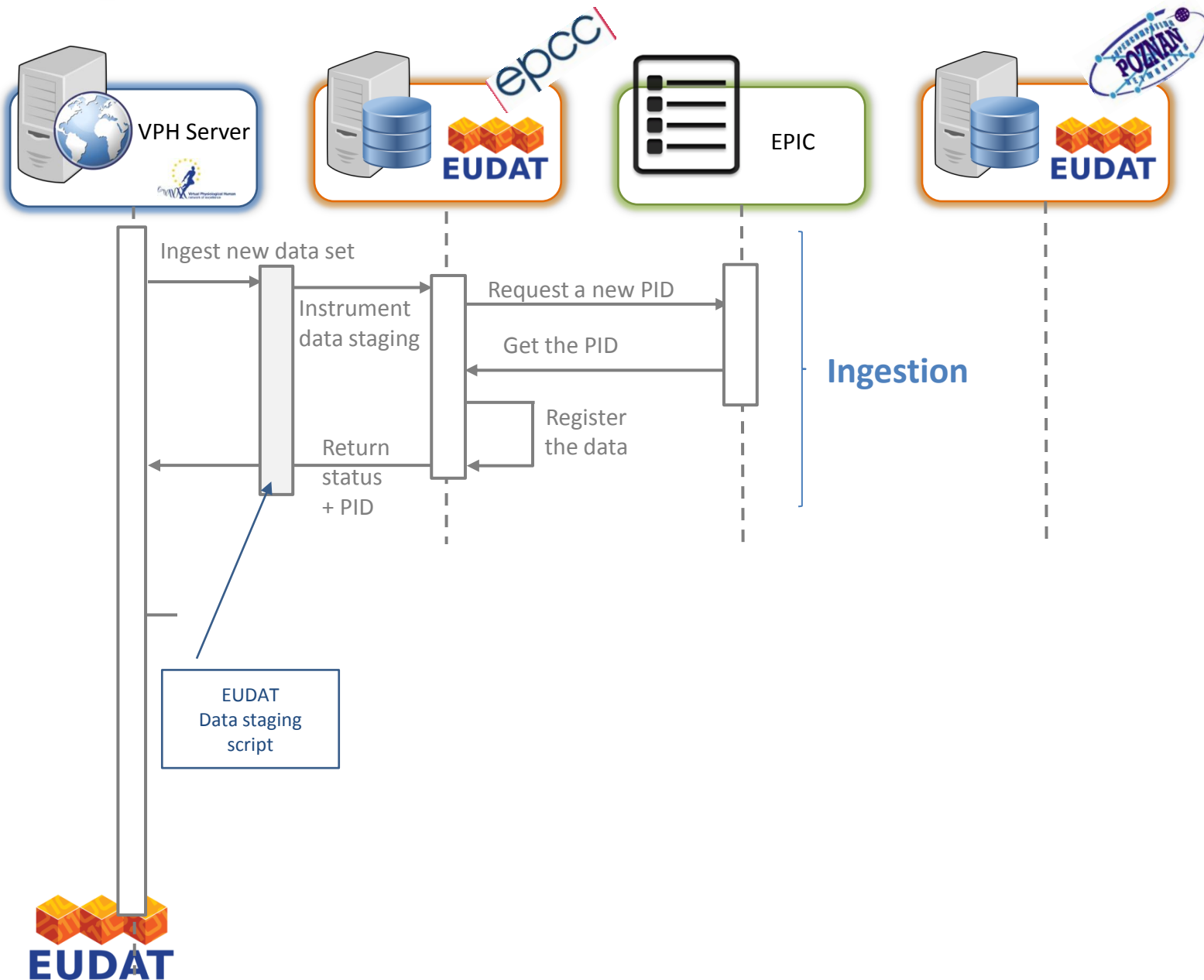
- VPH Community Portal
- EUDAT Node
- EPIC Server
- PRACE HPC Facility



# Case stories

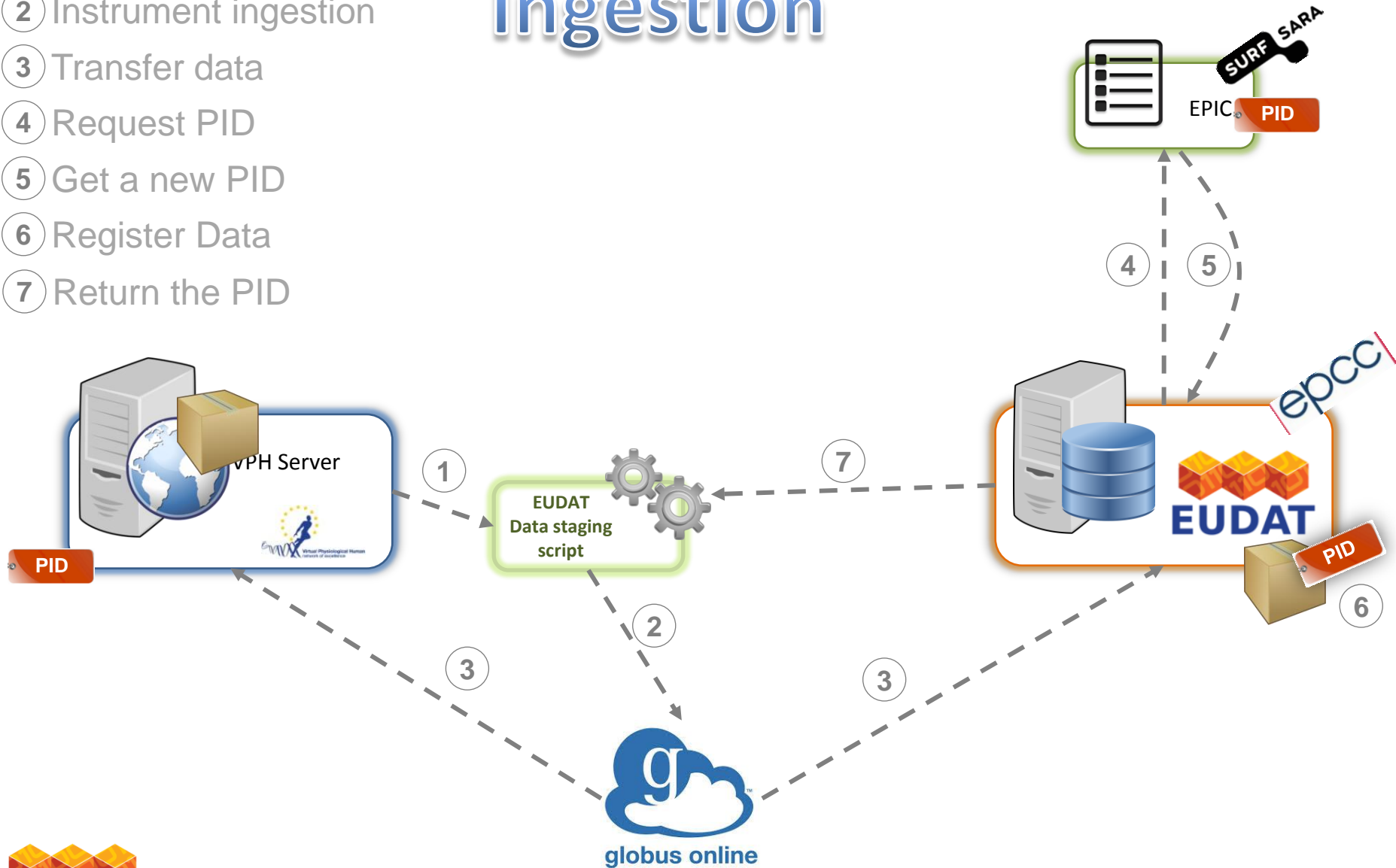
- **Data ingestion/replica** 
  - ingestion of VPH data sets onto EUDAT resources for preservation purposes
  - replica of data across EUDAT nodes for improving data accessibility
- **Data staging** 
  - staging of data from EUDAT resources to PRACE working area for processing
  - ingestion of produced results

# Ingestion



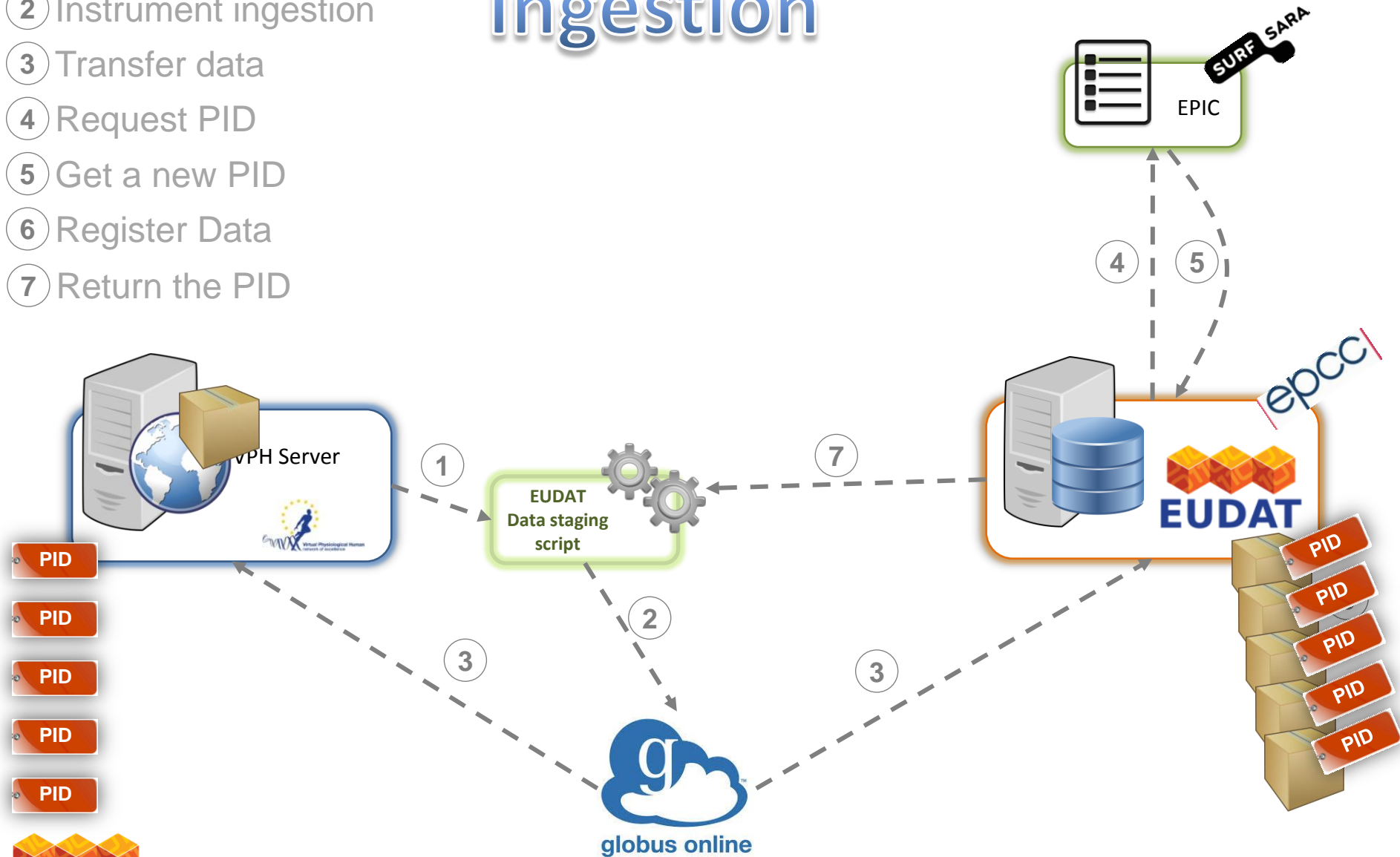
- ① Ingest new data
- ② Instrument ingestion
- ③ Transfer data
- ④ Request PID
- ⑤ Get a new PID
- ⑥ Register Data
- ⑦ Return the PID

# Ingestion



- ① Ingest new data
- ② Instrument ingestion
- ③ Transfer data
- ④ Request PID
- ⑤ Get a new PID
- ⑥ Register Data
- ⑦ Return the PID

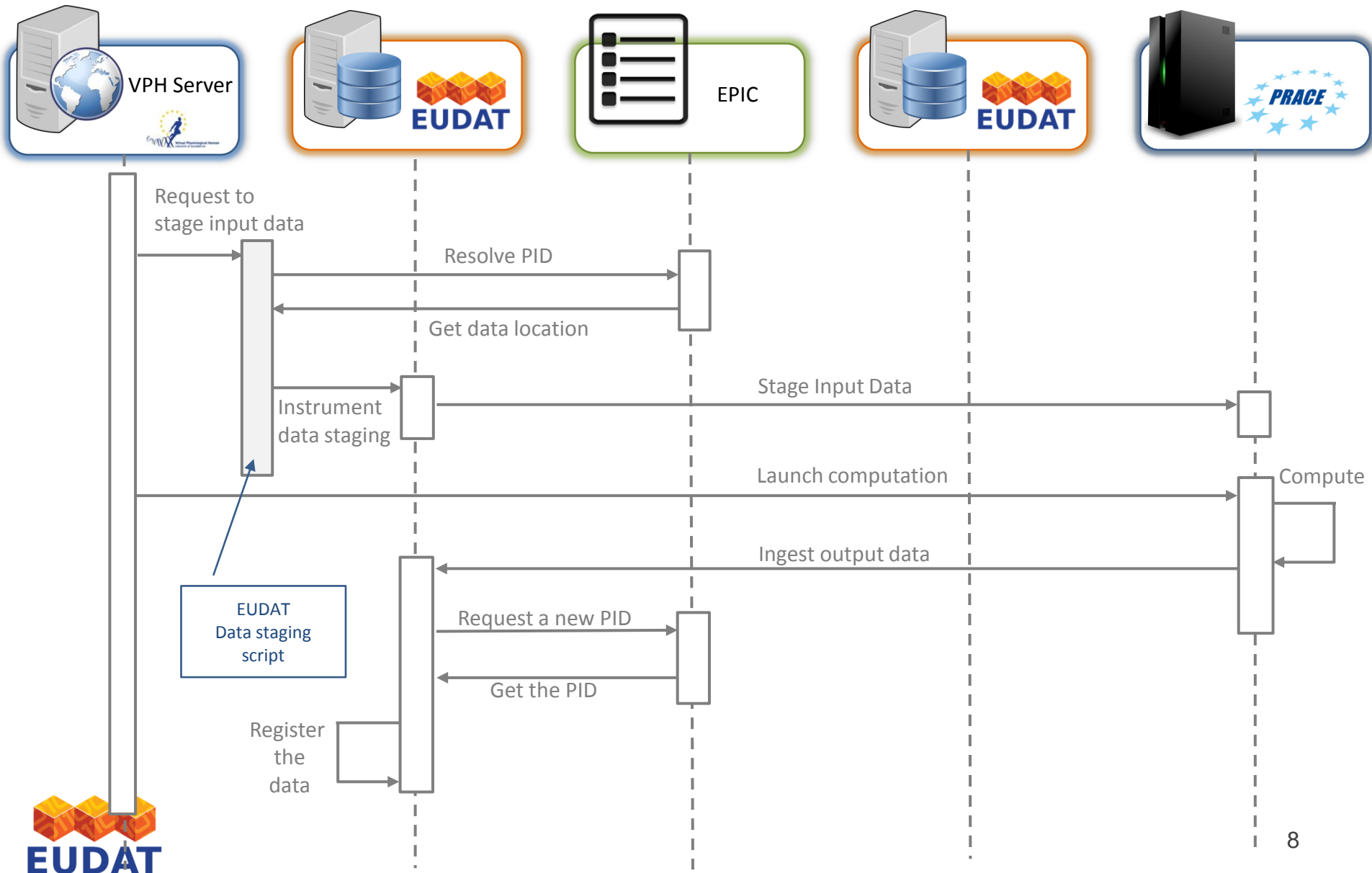
# Ingestion





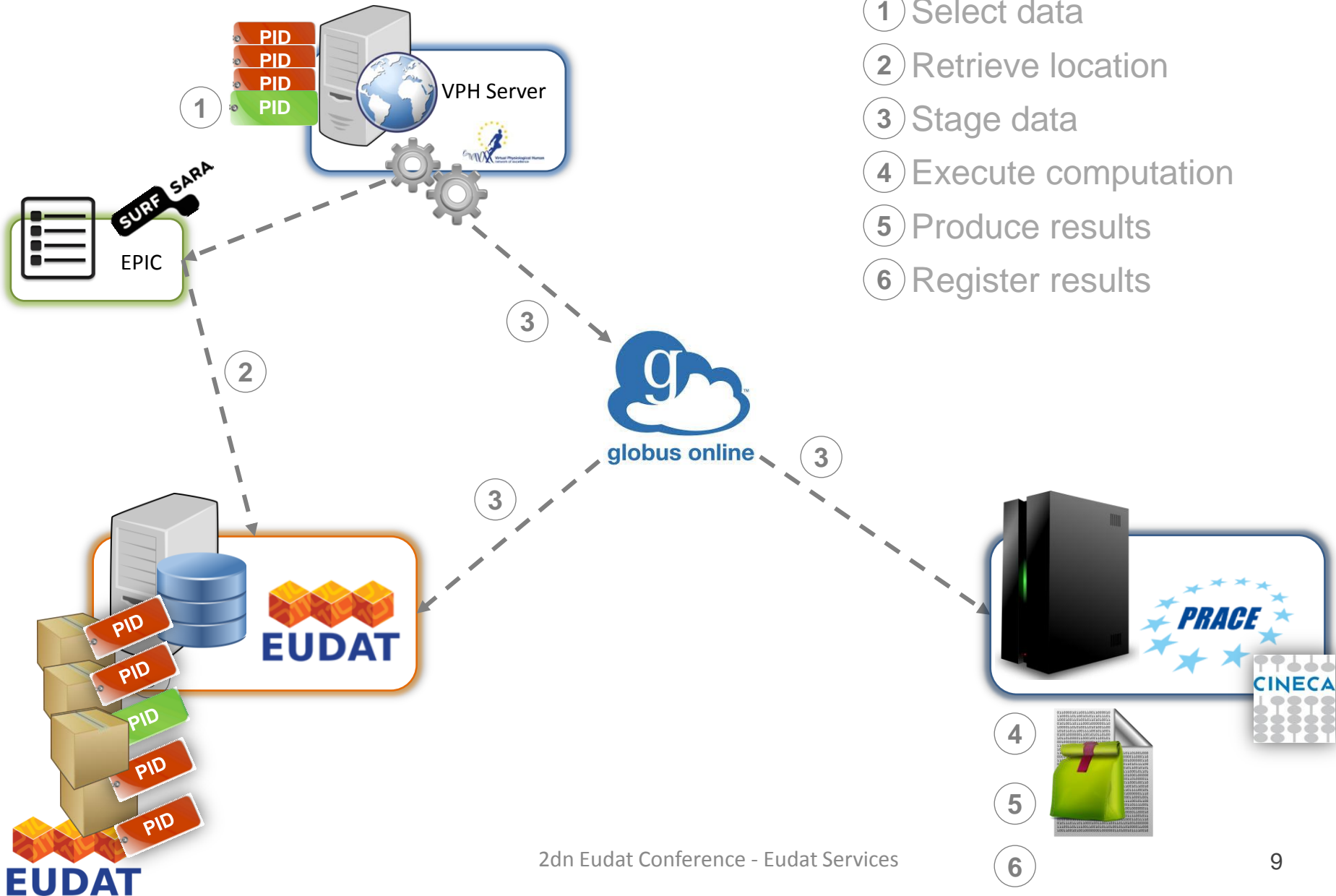
# LIVE DEMO

# Staging





# Staging

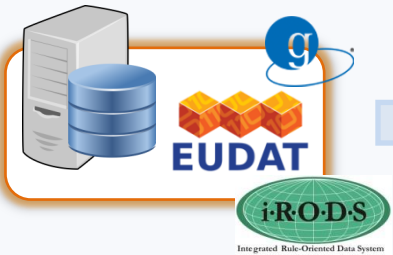


- 1 Select data
- 2 Retrieve location
- 3 Stage data
- 4 Execute computation
- 5 Produce results
- 6 Register results

# Involved Technologies



- **VPH Portal**
  - Moteur Workflow
  - Application Launching Engine (AHE)



- **EUDAT Node**
  - iRODS for data management
  - EUDAT microservices for PID management and registration
  - GridFTP-iRODS DSI component for data staging



- **EUDAT Data Staging Script**
  - Integration script for data ingestion/staging handling via PID
  - Python
  - Globus Online API



- **PRACE Node**
  - UNICORE for job submission and management
  - GridFTP for data staging



# Get in touch

Email: [eudat-b2stage@postit.csc.fi](mailto:eudat-b2stage@postit.csc.fi)

URL: <http://www.eudat.eu/b2stage>

Many thanks for your attention!

**EUDAT**  
European Data Infrastructure

Home About User Forum Services Training Working Groups Conferences News & Publications Contact Search

## B2STAGE

**B2STAGE**  
Get Data to Computation

**What is B2STAGE?**

Many researchers are not only struggling to manage increasingly large volumes of data, but are also finding they need more computational power to process them. As research challenges get larger and more complex - full-Earth climate simulation, coupled simulations of multiple organs in the human body, seismic analysis of earthquakes at continental scale - researchers' data and compute demands are rising fast. Furthermore, in distributed computing, where resources are geographically dispersed, efficient shipping of data to high performance computer (HPC) workspaces is essential.

To address this need, EUDAT has developed the B2STAGE service, to allow users to easily ship large amounts of data between EUDAT storage resources and workspace areas of high-performance computing systems. Besides transferring data across distributed resources, B2STAGE can be used to simply ingest community data onto EUDAT resources using a high performance protocol, like GridFTP. This utilization mode represents an extension of the B2SAFE service and offers communities a light approach to ingest and replicate data. Data ingested through B2STAGE is registered through a PID using the mechanism adopted by B2SAFE.

**B2STAGE:**

- supports researchers in transferring large data collections from EUDAT storage resources to HPC facilities;
- offers reliable, efficient, easy-to-use tools to manage data transfers;
- provides the means to re-ingest computational results back into the EUDAT infrastructure;
- ingests data sets onto EUDAT resources for long-term preservation.

**Who benefits from B2STAGE?**