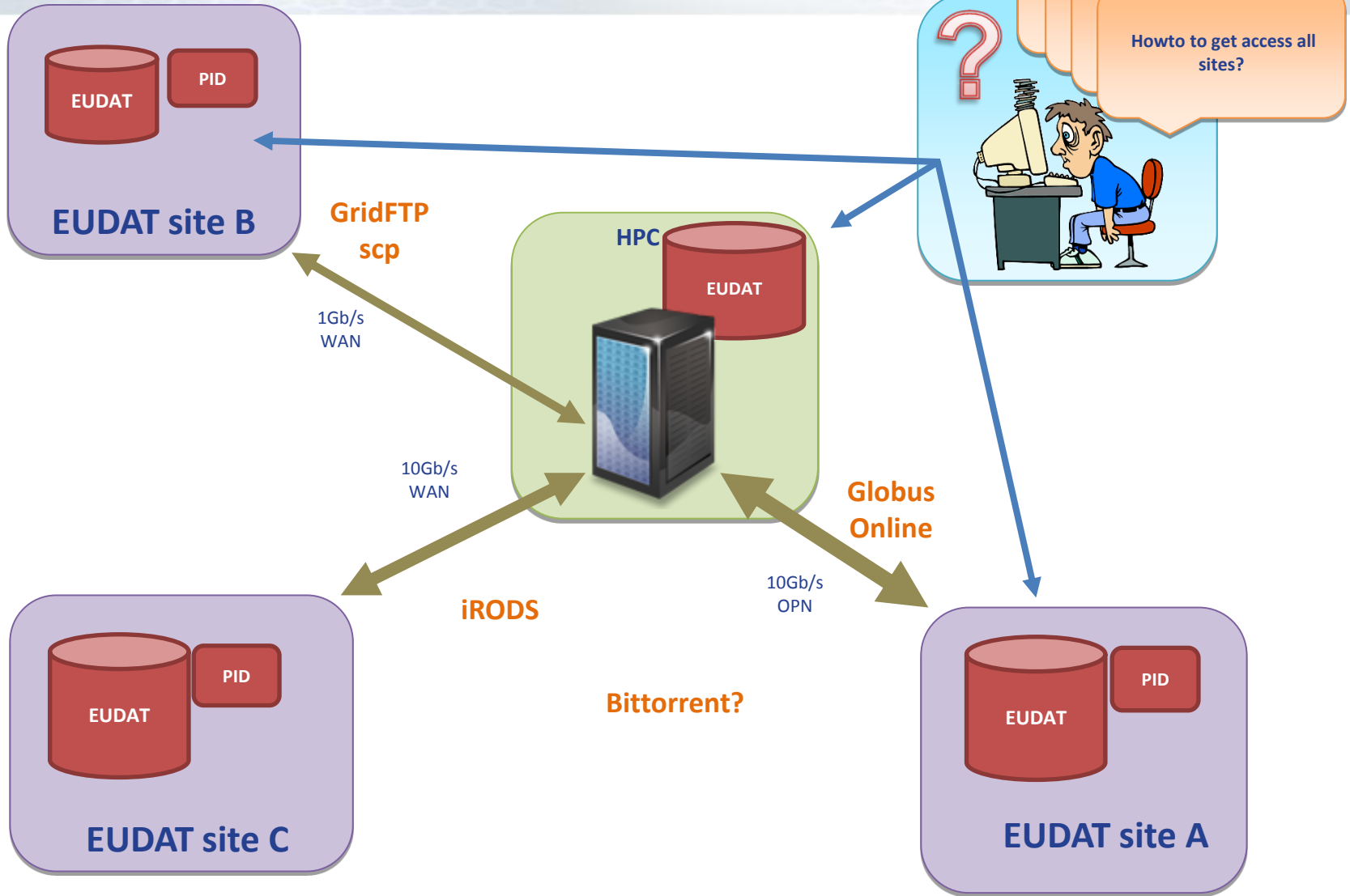




# B2STAGE

**Get data to computation!**

**Giuseppe Fiameni – [g.fiameni@cinca.it](mailto:g.fiameni@cinca.it)  
3<sup>rd</sup> EUDAT Conference – Sept. 25 2014  
Amsterdam**





# What is the B2STAGE?

- **It is a service to permit users:**
  - easily stage data between EUDAT resources and external data processing facilities, e.g. HPC machines
  - ingest computational results onto EUDAT resources
- **Leverages B2SAFE functionalities for PID management**
- **Comprises three different packages:**
  - **GridFTP Interface**
  - **HTTP Interface** compliant with the CDMI specification (**new**)
  - **Data Staging Script** to extend your existing system with staging functionalities
- **Enables interoperability with other e-Infrastructures**



## Who can use B2STAGE & Why?

- **Researchers** can transfer large data collections from EUDAT storage resources to HPC facilities for processing.
- **Community Data Managers** can replicate community data through a lightweight service and ingest data sets to EUDAT storage resources for long term preservation (in combination with the B2SAFE).

# EUDAT Services

**AAI**

Network of trust among authentication and authorization actors



**PID**

Identity

Integrity

Authenticity

Locations



**B2DROP**  
Sync and exchange



**B2SHARE**  
Store and share

**B2SAFE**  
Data curation and access optimization



**B2STAGE**  
Dynamic replication to HPC workspace for processing



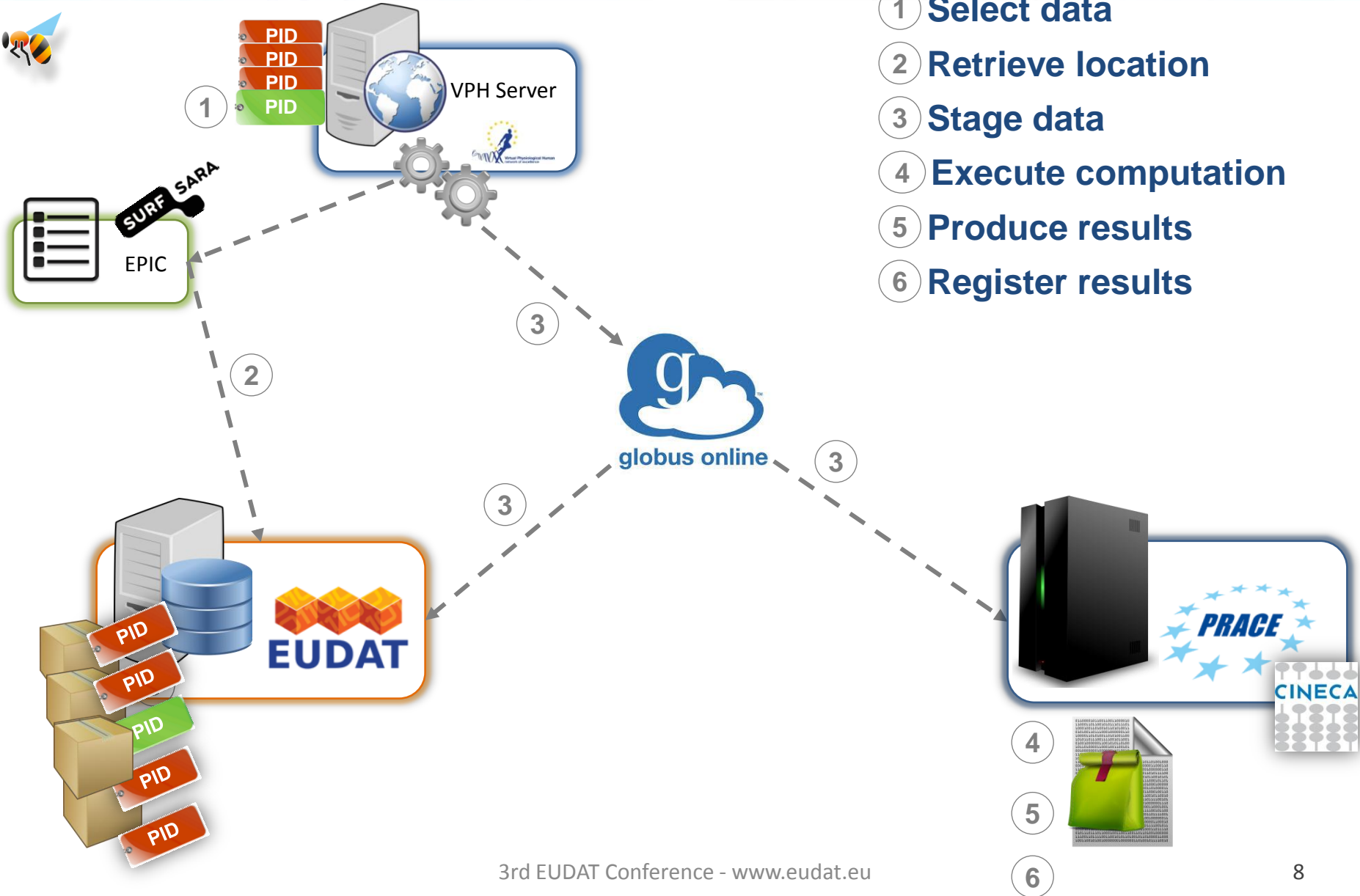
External community data repository



**B2FIND**

Aggregated EUDAT metadata domain  
Data inventory

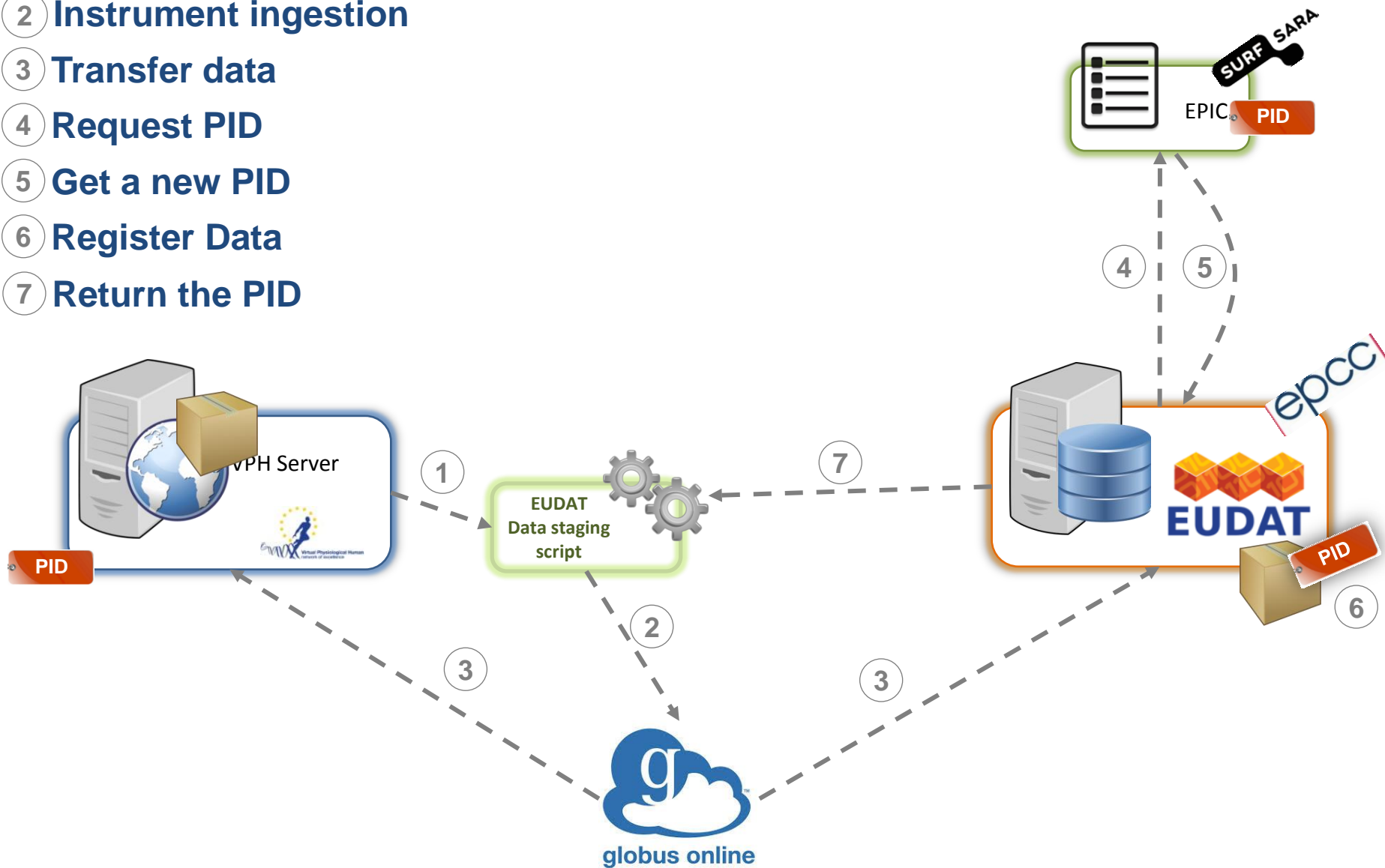




- ① **Select data**
- ② **Retrieve location**
- ③ **Stage data**
- ④ **Execute computation**
- ⑤ **Produce results**
- ⑥ **Register results**

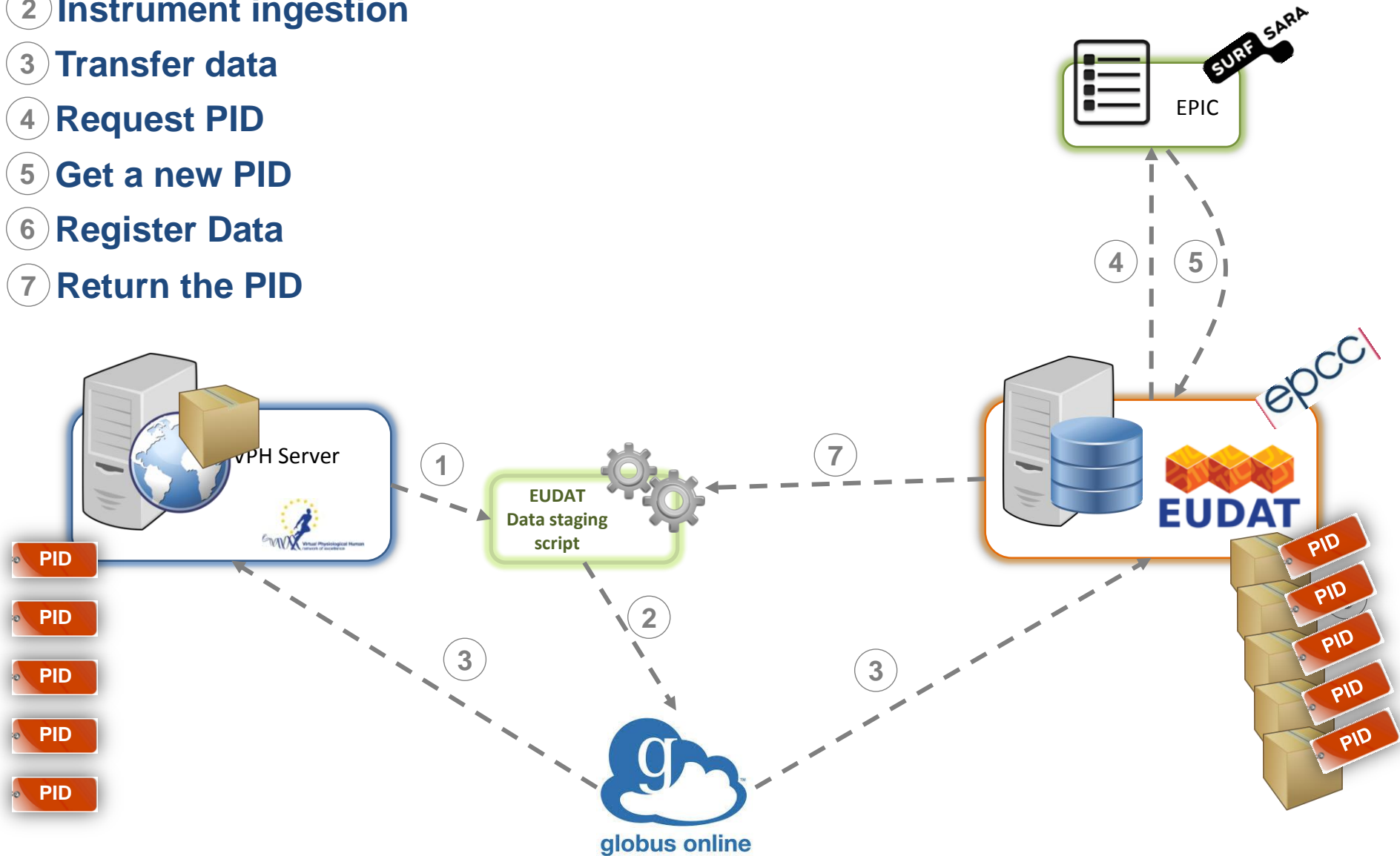


- ① Ingest new data **DAT**
- ② Instrument ingestion
- ③ Transfer data
- ④ Request PID
- ⑤ Get a new PID
- ⑥ Register Data
- ⑦ Return the PID





- 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







# GridFTP Interface

- **Lightweight ingestion and access** to data through the GridFTP protocol
- Meant for **high performance, high reliable** data transfers
- Developed internally as extension of the **iRODS data management solution**
- Compliant with any client available outside (i.e. guc, uberFTP, Globus On Line, FTSv3)
- Tuned to work with Globus OnLine





# RESTful HTTP Interface

- **Management of data through the HTTP protocol**
  - GET, PUT, DELETE objects
- **Based on the CDMI (Cloud Data Management Interface) specifications**
  - access in a standardized, extensible way
  - compatible with other e-Infrastructures
- **Programmable interface to access EUDAT storage resources**





# Data Staging Script

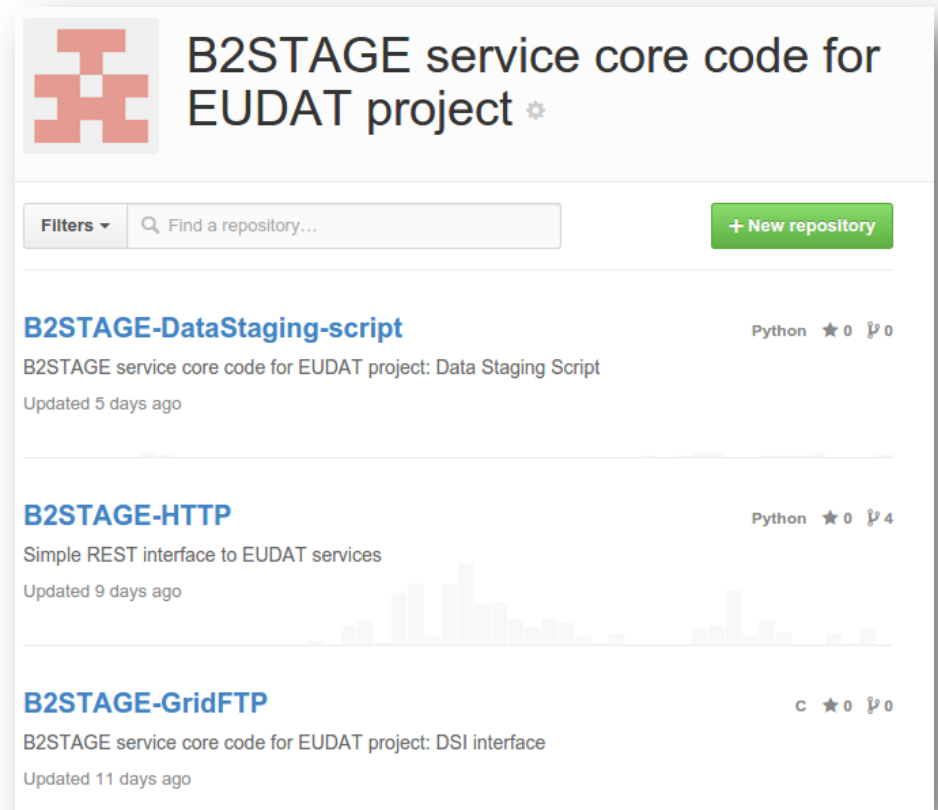
- A simple python modular script to help communities integrate data staging functionalities within their exiting system
- Based on Globus Online API and iRODS rule mechanism for data selection

```
prompt$ ./datastager.py -p /home/irods/data/archive -y 2004 -n MN  
-s AQU -c BHE -u cin0641a --ss ingv --ds GSI-PLX -dd  
/shared/data/userprace/tmp
```





All the code is  
available open  
source on **github!**



The screenshot shows the GitHub repository page for the B2STAGE service core code for the EUDAT project. The repository is titled "B2STAGE service core code for EUDAT project" and is marked as a service. Below the title, there is a search bar with the text "Find a repository..." and a "+ New repository" button. The repository list shows three repositories:

- B2STAGE-DataStaging-script**: Python, 0 stars, 0 forks. Description: B2STAGE service core code for EUDAT project: Data Staging Script. Updated 5 days ago.
- B2STAGE-HTTP**: Python, 0 stars, 4 forks. Description: Simple REST interface to EUDAT services. Updated 9 days ago.
- B2STAGE-GridFTP**: C, 0 stars, 0 forks. Description: B2STAGE service core code for EUDAT project: DSI interface. Updated 11 days ago.



# Future works

- **Consolidate and extend the HTTP interface.**
  - Interested communities are welcome to test and adopt it!
- **Extend the GridFTP interface to literally support PIDs**
  - `gsiftp://10.9281/192830192710`
- **Transform the Data Staging Script** from a simple tool to a more complete toolkit
  - Incorporate examples for interacting with other services



# B2STAGE SESSION

- \* **Collaboration with WLCG: WebFTS** - Andrea Manzi, CERN
- \* **Collaboration with PRACE: gtransfer - A tool for WAN transfers** - Frank Shiner
- \* **New features in the light of EUDAT2020: HTTP REST interface** - Shaun de Witt, STFC
- \* **Validation of the EUDAT Data Infrastructure and Preservation for the ALEPH experiment** - Marcello Maggi - INFN



# Get in touch

Email:

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

URL:

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



The screenshot shows the B2STAGE website interface. At the top, there is the EUDAT logo and the B2STAGE logo with the tagline "Get Data to Computation". A navigation menu includes links for Home, About, User Forum, Services, Training, Working Groups, Conferences, News & Publications, Contact, and Search. The main content area features a "What is B2STAGE?" section with a detailed paragraph explaining the service's purpose in addressing data management challenges in research. Below this is a bulleted list of key features: supporting large data transfers, providing reliable tools, enabling result back-ingest, and ingesting data for long-term preservation. A diagram at the bottom illustrates the data flow between EUDAT storage, HPC resources, and the B2STAGE service. The footer of the page asks "Who benefits from B2STAGE?"