B2STAGE

GridFTP and iRODS

Christine Staiger

SURFsara

Workshop, Utrecht, January 18th 2016

www.eudat.eu

B2STAGE

Facilitating communities to:

- **move large amounts of data** between data stores and high-performance compute resources
- **re-ingest computational results** back into EUDAT
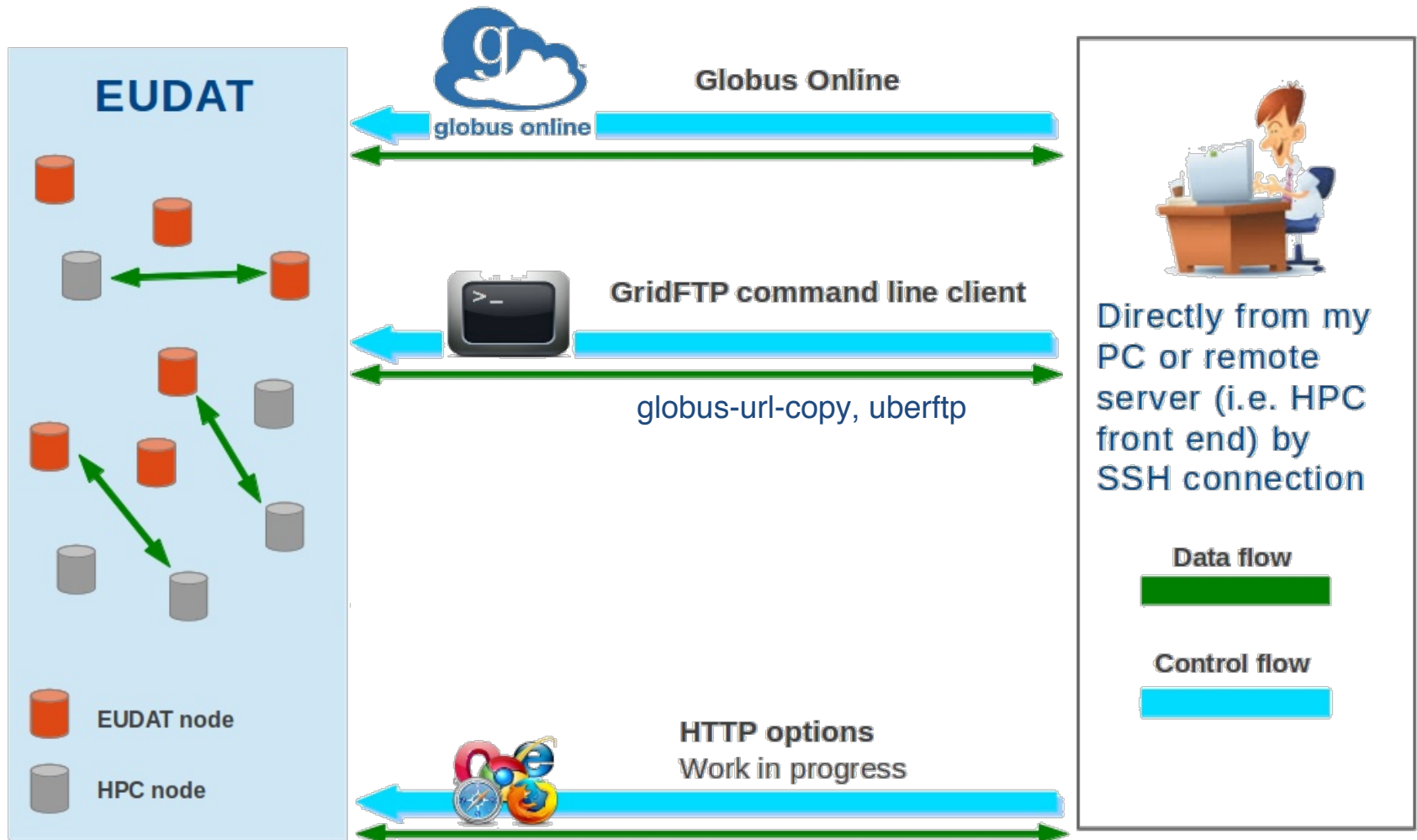- **deposit large data sets** onto EUDAT resources for long-term preservation

eudat.eu/b2stage
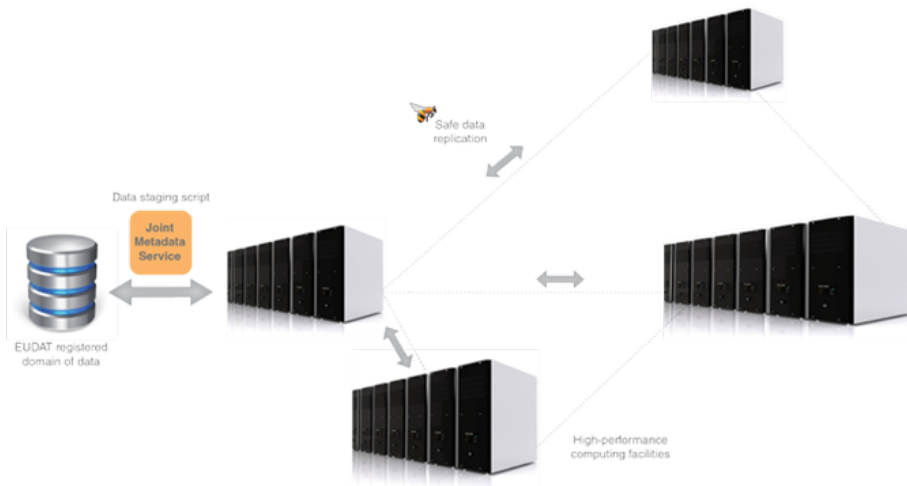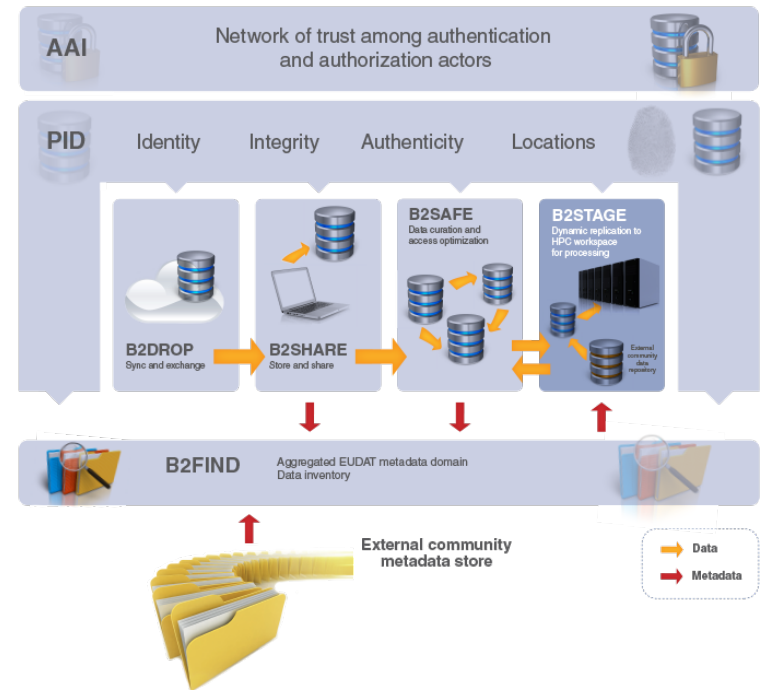
Features:

- high-speed transfer
- reliable and light-weight
- manages data with PIDs

How can you use B2STAGE?

EUDAT B2STAGE options

B2STAGE

eudat.eu/b2stage

EUDAT

Globus Online

globus online

GridFTP command line client

globus-url-copy, uberftp

Directly from my PC or remote server (i.e. HPC front end) by SSH connection

Data flow

Control flow

EUDAT node

HPC node

HTTP options
Work in progress

Why GridFTP

- http://toolkit.globus.org/ and UberFTP
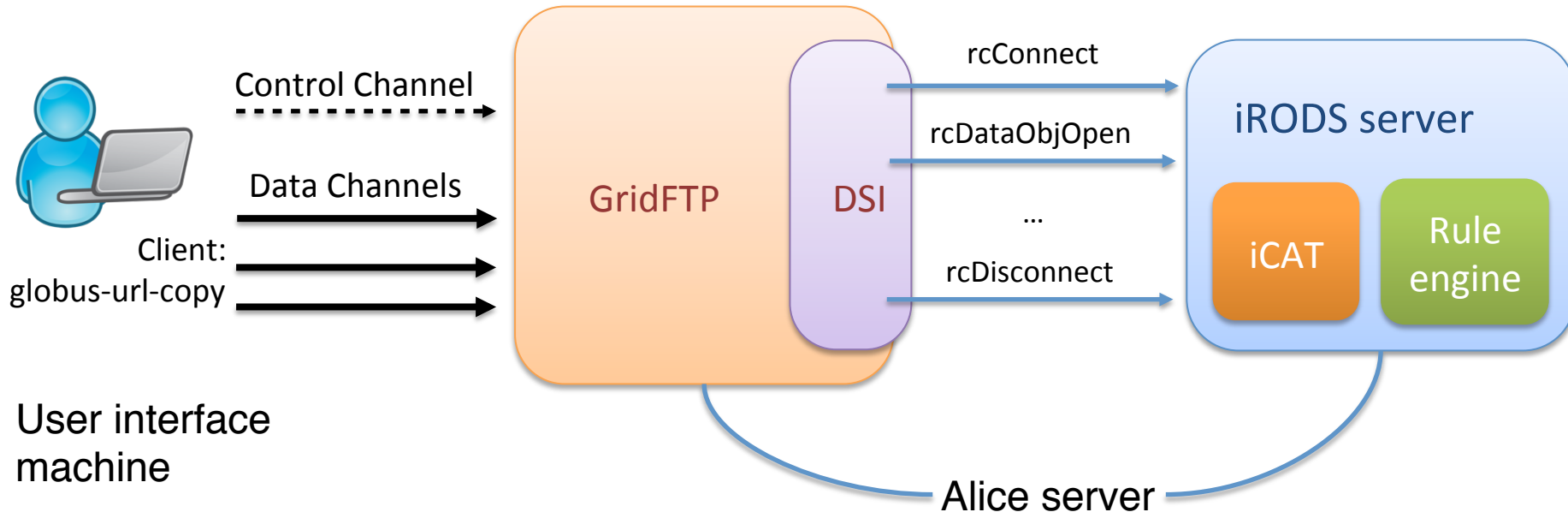
- Employs Grid Security Infrastructure (GSI)
    - → Certificates, chain of trust
- On the fly data integrity checks
- Resuming interrupted file transfers
- Parallel streams for transfers
- Can work easily with wildcards
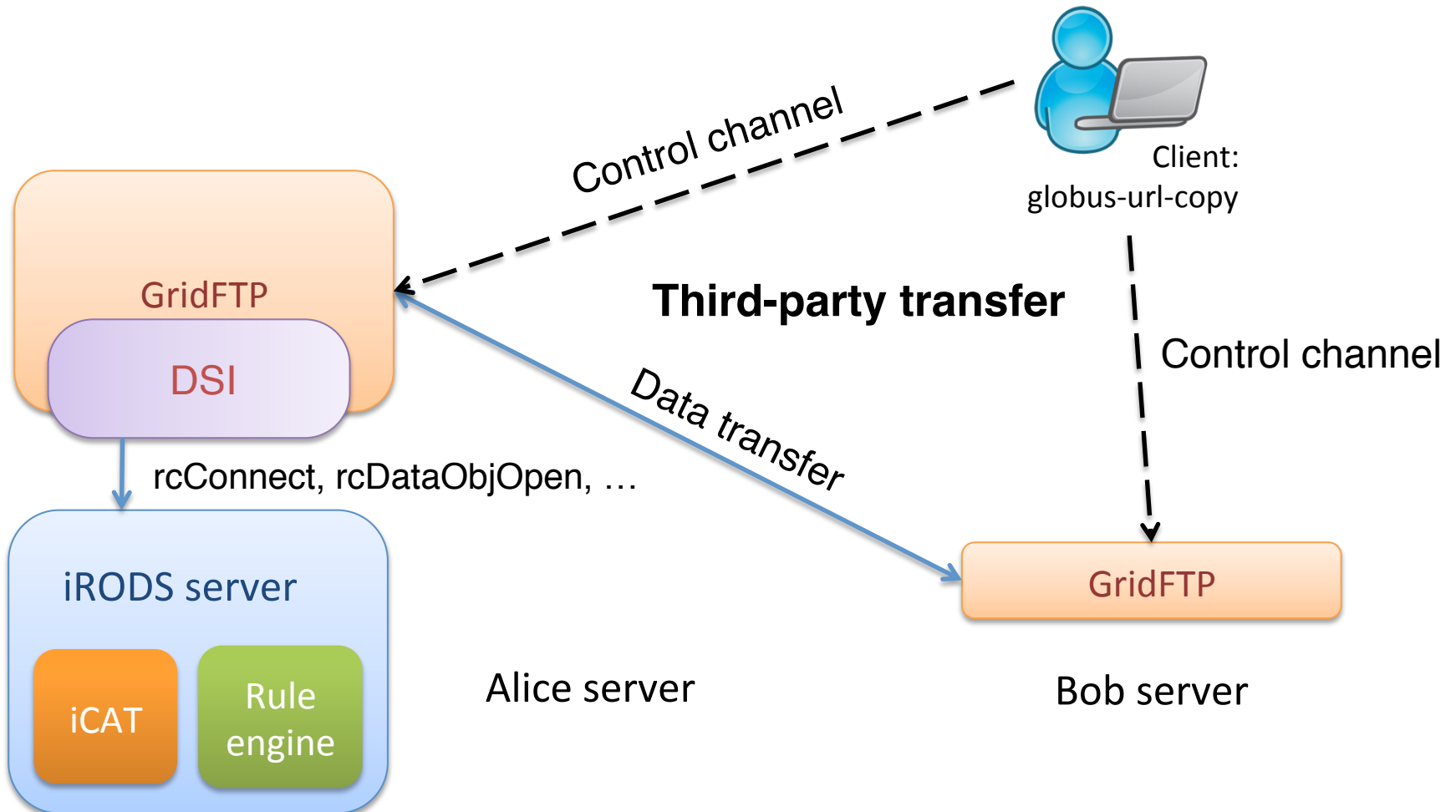
→ Powerful and secure data transfer protocol

Hands-on scenario 1

Data management

iRODS functions
C-library

User interface
machine

Code for the Data Storage Interface (DSI):
https://github.com/EUDAT-B2STAGE/B2STAGE-GridFTP

Hands-on scenario 2

EUDAT

B2STAGE

Control channel

Client:
globus-url-copy

GridFTP

DSI

Third-party transfer

Control channel

Data transfer

rcConnect, rcDataObjOpen, …

iRODS server

iCAT

Rule engine

GridFTP

Alice server

Bob server

Hands-on scenario 3

Control channel

Retrieve data by PID

Client:
globus-url-copy

GridFTP

DSI

rcConnect, rcDataObjOpen, …

Handle.Net®

iRODS server

iCAT

Rule engine

Alice server