EUDAT EGI interoperability use case

Giuseppe Fiameni, Diego Scardaci (EGI), Peter Gille

EGI Conference 2015 - Lisbon

www.eudat.eu

Collaboration

- **Goal**
  - Enable researchers seamlessly access e-Infrastructures services pairing data and high-throughput computing resources together
- **Communities**
  - ELIXIR, EISCAT-3D, BBMRI, ICOS
  - EPOS, ENVRI+
  - …
- **Timeframe**
  - March 2014 – February 2016

Use Case "alpha"

- **Definition:**
  - Permit a user of both e-infrastructures to instantiate a VM on the EGI Cloud Federation for the execution of a computational job consuming data stored onto EUDAT resources

- **Assumptions:**
  - a) the user holds a valid and trusted security token (X.509 certificate)
  - b) he/she has been granted to access EGI resources
  - c) he/she has already deposited data within the EUDAT infrastructure

More detailed flow

1. The user authenticates to the EGI infrastructure
2. The EGI infra instantiates the environment and performs preliminary configuration (account, network, firewall, tools, libraries, compilers, applications, ...) → *contextualization*
3. The virtual environment is started up → *user's credentials are copied into the VM*
4. The VE stages data from EUDAT to a local storage area for performance and locality reasons, using authentication obtained in step 1
   → *data staging via GridFTP or HTTP (CDMI)*
   → *data discoverability via B2FIND*

More detailed flow (cont.)

5. The user launches the computational job
6. Any intermediate result is shared with the public via the EUDAT B2DROP service
→*either results or inputs*
7. Final results are ingested back onto EUDAT for being preserved in the long term (B2SAFE)
→ *registered via PID*

Technical Challenges

- **Authentication/Authorization**:
  - Credentials federation
  - Support for eduGAIN
    - →*EUDAT and EGI both member of the AARC project*
    - →***Iota** profile?*
  - Authorization?
- **Data access**
  - PID management
    - → Handling of data via PID
  - Data staging performance
  - POSIX access to data via FUSE
    - → AAI
  - Adoption of CDMI (?!)
- **Deposit of intermediate results via B2DROP or similar services**
  - AuthN/AuthZ
    - → *EUDAT Library*

Preliminary plan
(Phase 1 – August 2015)

- **Spec. Document**
- **Set up a Proof of Concept environment**
  - Services testbed
    - *EGI VM*
    - *EUDAT B2SAFE instance*
- **VM contextualization**
  - Define tools and services to be installed on the VM image
- **Streamline authentication process**
  - Make the VM obtain proper authentication tokens through the EGI authentication infrastructure using x. 509 certificates
- **Data staging/access**
  - Evaluate data access and staging from EUDAT resource to the EGI VM → *GridFTP, HTTP, FUSE*

Adoption

- **EPOS**
  - *"The MISFIT application, takes as input the synthetic seismograms generated from HPC simulations for a certain Earth model and earthquake and, after a preprocessing stage, compares them with real observations in order to foster subsequent model updates and improvement (Inversion)"*
- ??