



Koninklijk Nederlands  
Meteorologisch Instituut  
*Ministerie van Infrastructuur en Milieu*

# The VERCE Platform





## *Virtual Earthquake and Seismology Research Community in Europe*

### *International structure*

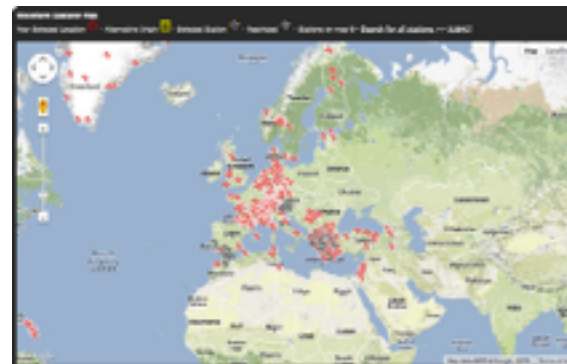
- **Global observations** and monitoring systems
- Integrated **Distributed Data Archives**
- **Data and metadata** formats

### *Supporting Scientific challenges*

Understanding **Earth's dynamics** and structures  
via **Earthquakes Simulations and Models Evaluation**

### *Impact on Society*

- **Natural hazard** and risk mitigation
- **Energy resources** exploration and exploitation
- **Underground wastes** and **carbon sequestration**





## *Virtual Earthquake and Seismology Research Community in Europe*

### *International structure*

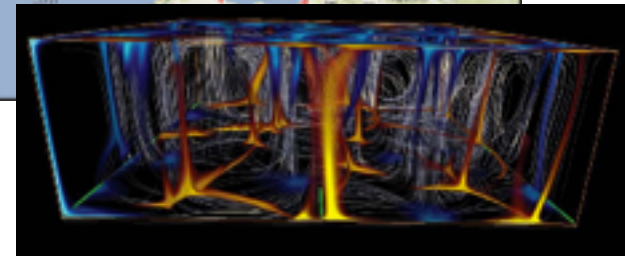
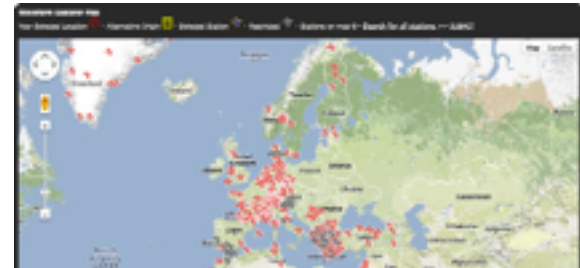
- **Global observations** and monitoring systems
- Integrated **Distributed Data Archives**
- **Data and metadata** formats

### *Supporting Scientific challenges*

Understanding **Earth's dynamics** and structures  
via **Earthquakes Simulations and Models Evaluation**

### *Impact on Society*

- **Natural hazard** and risk mitigation
- **Energy resources** exploration and exploitation
- **Underground wastes** and **carbon sequestration**





## *Virtual Earthquake and Seismology Research Community in Europe*

### *International structure*

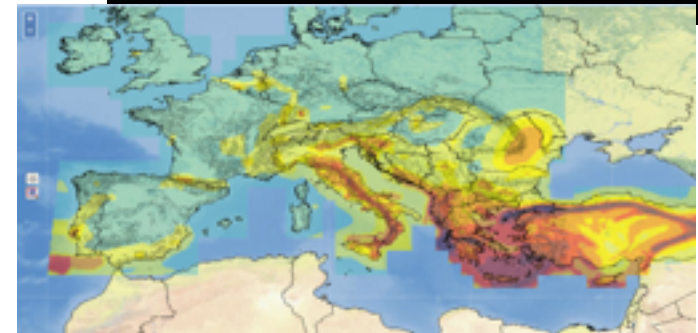
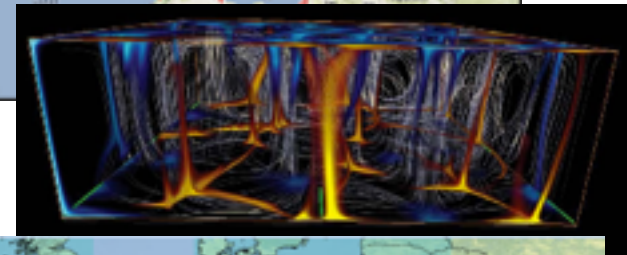
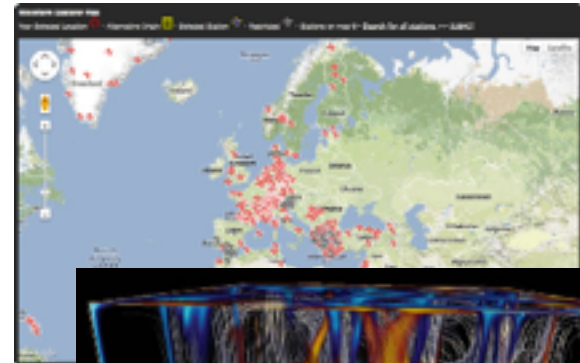
- **Global observations** and monitoring systems
- Integrated **Distributed Data Archives**
- **Data and metadata** formats

### *Supporting Scientific challenges*

Understanding **Earth's dynamics** and structures  
via **Earthquakes Simulations and Models Evaluation**

### *Impact on Society*

- **Natural hazard** and risk mitigation
- **Energy resources** exploration and exploitation
- **Underground wastes** and **carbon sequestration**





## Forward Modelling Use Case

1. Production of **synthetic seismograms** for various **Earth models and earthquakes** on a continental scale Requires the execution of **HPC simulation codes** called solvers (**Simulation**).
2. The synthetic data may be **compared with real observations** (**Raw Data Acquisition, MISFIT**)
3. **Model updates and improvement** (**Inversion**)



## VERCE Simulation Portal

### SEISMOLOGICAL INGREDIENTS

MESH

MATERIAL PROPERTIES  
(velocity model)

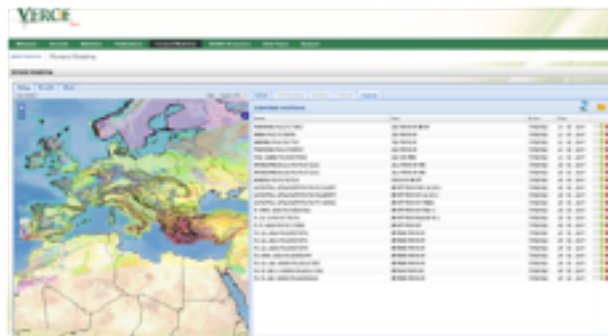
NUMERICAL CODE  
(SPECFEM3D)

EVENTS

STATIONS

### HPC INGREDIENTS

HPC CONNECTION  
HPC CENTER POLICIES  
HPC CENTER



### SEISMOLOGICAL OUTPUT

SYNTHETIC SEISMOGRAMS

VOLUMETRIC DATA

Shakemap

Movie

SYNT - OBS MISFIT



## Earthquake Simulation and Misfit Calculation

### Data overview:

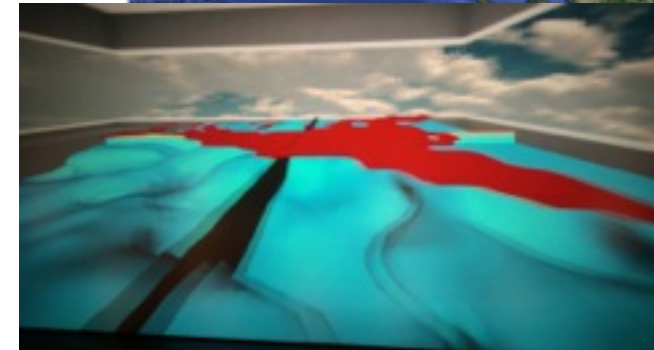
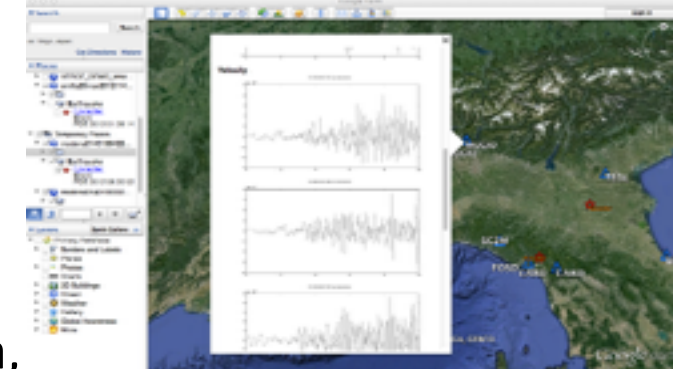
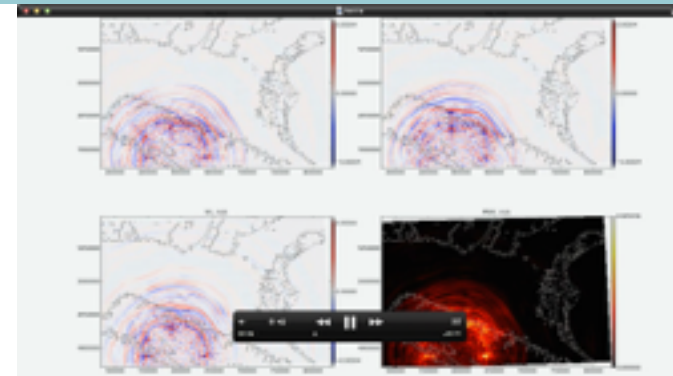
**Synthetic Data:** seismograms, plots, 3D Geometry, Videos, KMZ packages, **meshes and models.**

(100 stations = 900 products and metadata )

**6-10 GB for a SPECFEM3D simulation on 1000 cores**

**Raw Data:** on demand **access** and **staging** of **observational data** from **EIDA**: Earthquake Metadata, Sensors Metadata, waveform on regional scale.

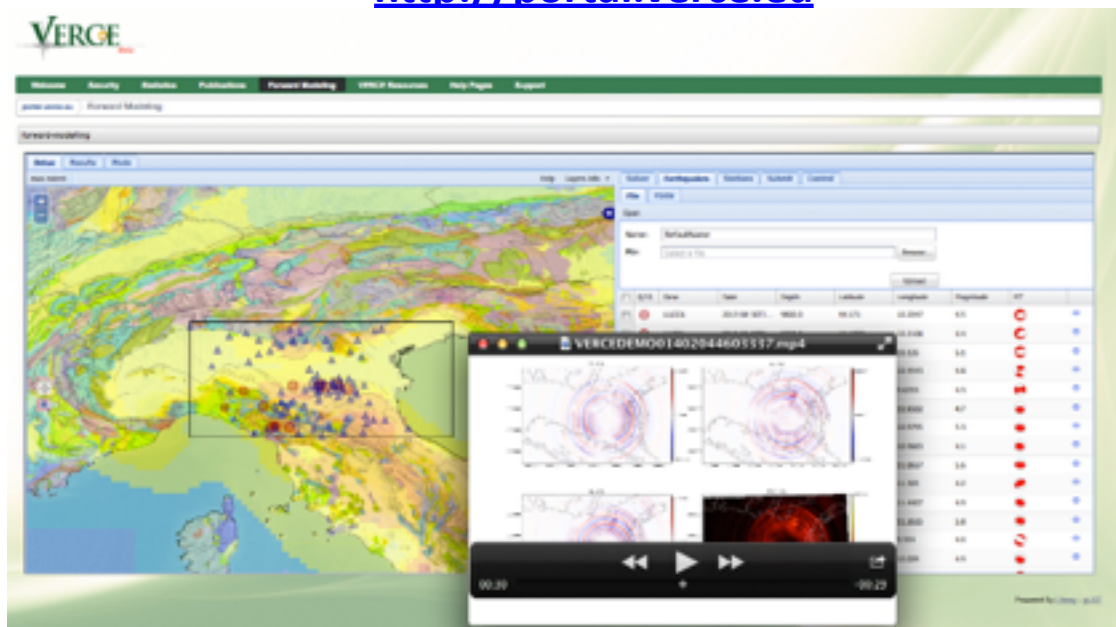
**(At the moment all via the FDSN WEB API)**





## Software as a service via the **VERCE Science Gateway**

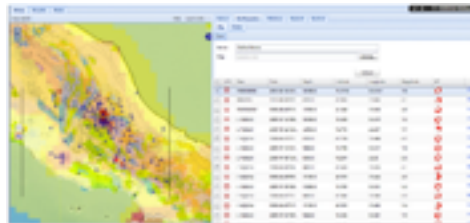
<http://portal.verce.eu>



- Support for Workflow technologies
- Data Management and Provenance System
- Combine HPC/DI computing infrastructures
- Synergy with the European Infrastructures (PRACE, EGI, EPOS)
- Dedicated User Interfaces



# VERCE Platform, Components Interaction



*Science Gateway  
Models and custom  
dataset upload.  
Simulation and  
MISFIT WF Setup*



External  
Data Services

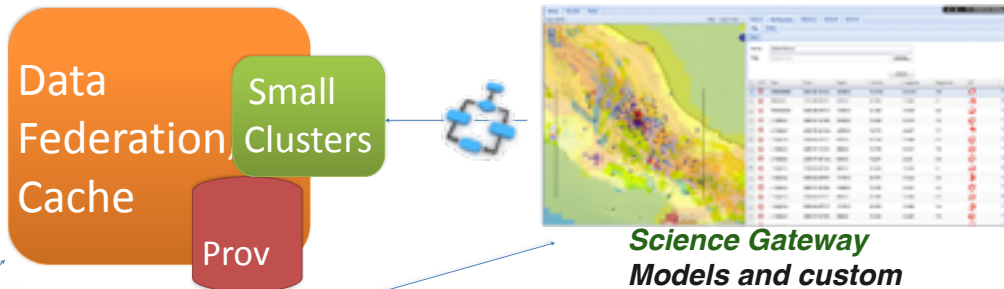
OGC®  
Open Geospatial Consortium, Inc.



# VERCE Platform, Components Interaction

*1 - Raw data acquisition*

*3 - MISFIT*



External  
Data Services

OGC®  
Open Geospatial Consortium, Inc.

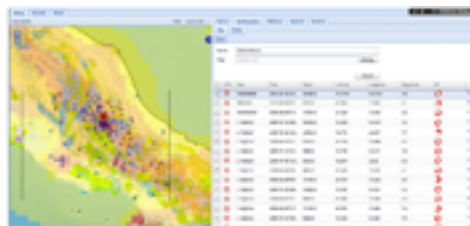
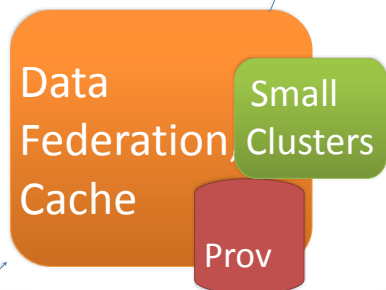


# VERCE Platform, Components Interaction

*2 - HPC Simulation  
(model stage-in)*



*1 - Raw data acquisition  
3 - MISFIT*



*Science Gateway  
Models and custom  
dataset upload.  
Simulation and  
MISFIT WF Setup*



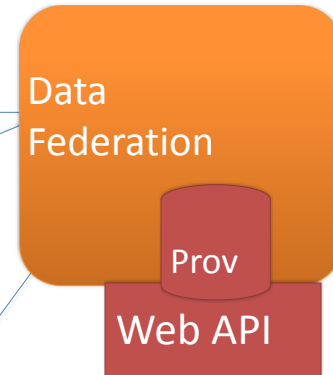
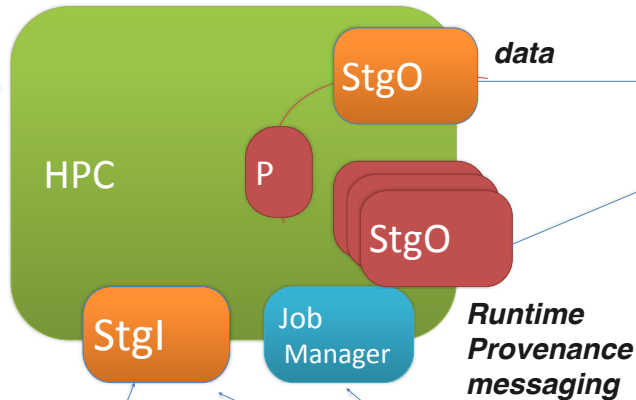
External  
Data Services



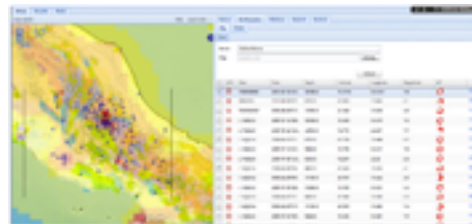
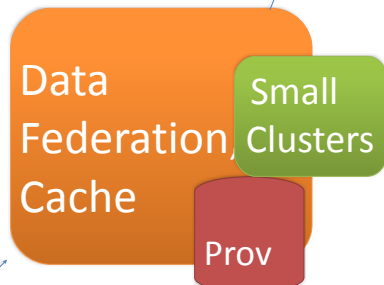
# VERCE Platform, Components Interaction

**2 - HPC Simulation  
(model stage-in)**

**Results and provenance  
management**



**1 - Raw data acquisition  
3 - MISFIT**



**Science Gateway  
Models and custom  
dataset upload.  
Simulation and  
MISFIT WF Setup**



**External  
Data Services**

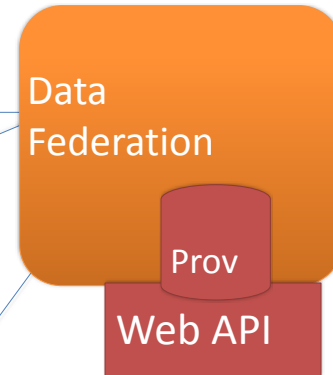
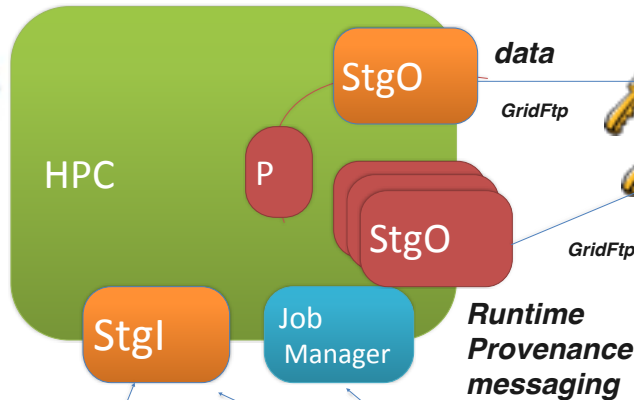




# VERCE Platform, Components Interaction

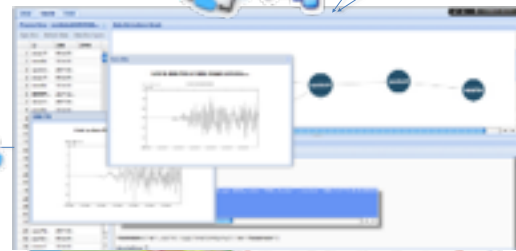
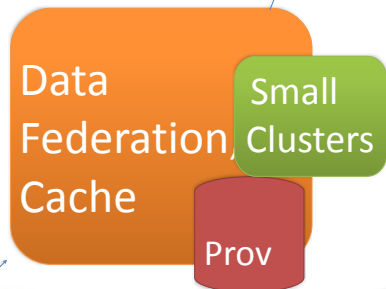
**2 - HPC Simulation  
(model stage-in)**

**Results and provenance  
management**



**Metadata and Provenance  
Archive and  
Services**

**1 - Raw data acquisition  
3 - MISFIT**



**Interactive Validation and  
Visualisation throughout the  
process**

**Science Gateway  
Models and custom  
dataset upload.  
Simulation and  
MISFIT WF Setup**



**External  
Data Services**



VOMS X.509



# Multi Layered Workflow specification

Comprehensive pre-post processing framework across DI-HPC models

1

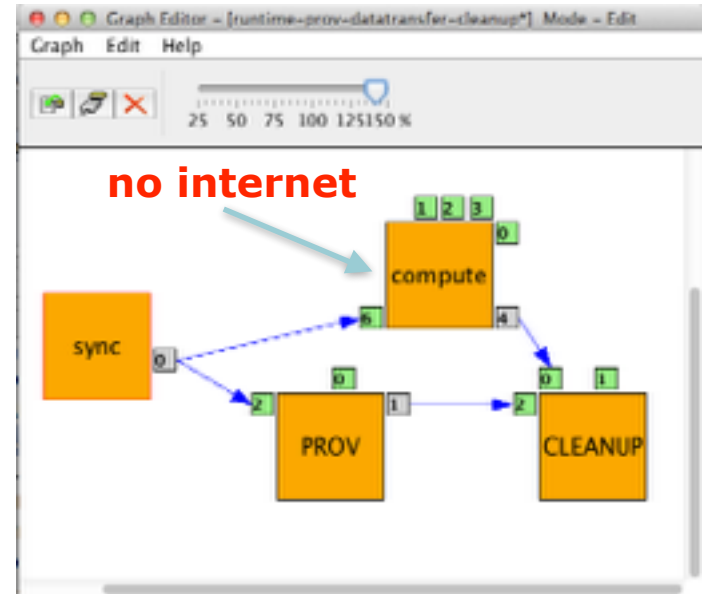
## Platform Control Workflow:

**sync:** metadata preparation and staging

**compute:** actual computation

**prov:** reads metadata, updates prov repository, intermediate data stage-out based on prov (**runtime**)

**cleanup:** full data stageout and cleanups





# Multi Layered Workflow specification

Comprehensive pre-post processing framework across DI-HPC models

1

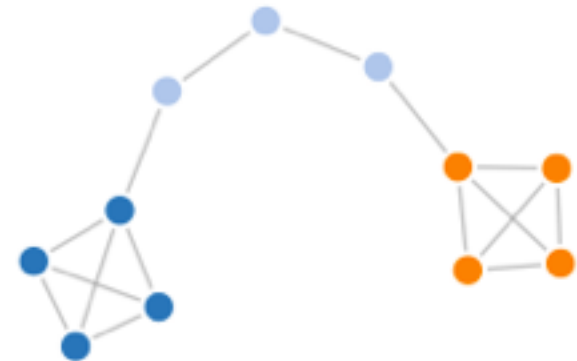
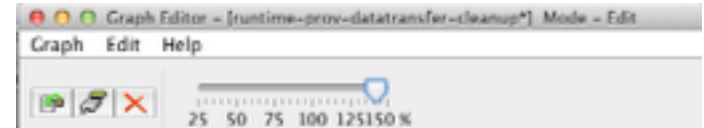
## Platform Control Workflow:

**sync:** metadata preparation and staging

**compute:** actual computation

**prov:** reads metadata, updates prov repository, intermediate data stage-out based on prov (**runtime**)

**cleanup:** full data stageout and cleanups



**compute** receives and runs scientific workflows

2

## Science Case Workflow (According to community preferences)

Extraction of **user defined metadata** and **fine-grained Provenance**





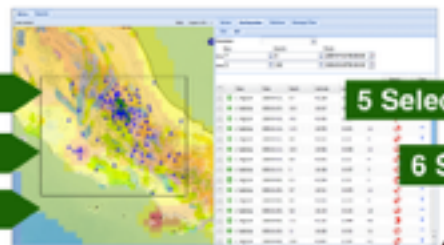
# Integration in Community Science Gateways



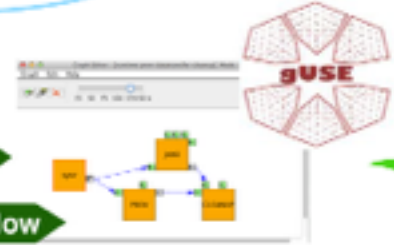
Integrated view of community webservices  
via the Forward Simulation GUI



- User's Actions**
- 1 Select Solver
  - 2 Select Mesh
  - 3 Select Velocity Model
  - 4
    - Upload Input
    - Retrieve Input
    - Reuse Input
  - 5 Select Workflow
  - 6 Submit Workflow
  - 7
    - Access Provenance
    - Visualise Results



**VERCE Science Gateway**



Provenance Production

**MyProxy**  
Credential Management Service



Data storage within  
the VERCE Federation



Many Tasks  
Simulation workflows  
in HPC  
resources



## Support from PRACE and EGI partners

### **1. Grid Middleware: Globus, UNICORE.**

- gLite not used in VERCE

### **2. Software modules available across resources.**

- homogeneous management of software modules, including naming convention is desirable

### **3. X.509 VOMS enabled certificates and web-based proxy generation tool!!**

- EC requested to Evaluate a simplified access to limited resource, without having to apply for a X.509

### **4. Special need for reserved resources could be pre-booked according to site policies and special situations (Training, Reviews).**

What is guaranteed after the end of the project?



## VERCE BYOB Scenarios (Brew your own Bottle)

- **Federated AAI System and user profiling** that can be used across **data and resource providers, data management systems, science gateways**:
- **BYOB: Account activation, validation and authentication** mechanisms are required to **authorise and trigger allocation of resources** to users, **across the providers** of the consortium.
- **Scalability evaluation based on available resources and outreach** for the supported HPC software.
- **BYOB: Exposing software as a service** via a **Science Gateway** requires somehow to provide clear **information on the scale and the possibilities** offered by the platform to the users.

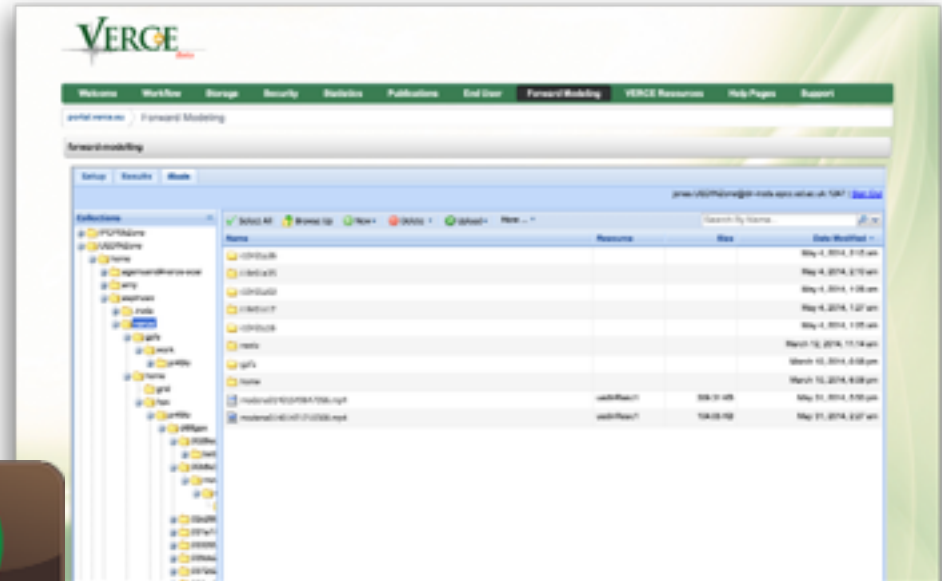


## VERCE BYOB Scenarios (Brew your own Bottle)

- Some **PRACE** sites **restrict** incoming and outgoing **traffic** to GridFTP. This prevented us to perform call to external HTTP webservices (GET/POST).
- **BYOB**: Exchanging small amounts of data with **WEB APIs** requires **staging phases** through external systems **talking GridFTP** and hosting **mediation services towards HTTP** end-points (FDSN, PROV).
- Need of an **Homogeneous and Web Friendly** Data Management System (WebDav interfaces, Front-end GUI, **integrated DOI services**)
- **BYOB**: **Rapid and interactive** access to **intermediate results** and metadata (**runtime provenance**) required to implement a solutions based on **external Data Management Systems, NoSQL tech and User Interfaces**.



Browse provenance  
and workflows output



Access Data Products shipped from clusters  
to iRODS  
(Integrated Rule-Oriented Data System)



## VERCE BYOB Scenarios (Brew your own Bottle)

- **Access to results Data should be sharable across users**
- **BYOB:** We have secure folders assigned to each user in iRODS. Access can be done via **GridFTP** (X.509) and **Front-end GUI** (login/pwd). We have to implement a way to allow a **controlled and shareable** access to these resources.
- **Work Practice: Promote and support co-development**
- **BYOB: Sharing responsibility** on implementation progress across partners. Identification of **Task Forces**, progress meetings with **chair turn-over (IT and Researchers)**, adoption of **proper tools (eg. Trello)**.



Boards 🔍 Trello + AS Alessandro Spinuso 🔔

**Misfit Calculation** VERCE ☆ Public ◀ Show sidebar

### To Do

As a User I want to be able to select stations from different providers for a run so that I don't have to do multiple runs for that. 🗨️ 1

Misfit Frontend Submit 🗨️

Upgrade gUse? 🗨️ 1


Add a card...

### Doing

reorganize iRODS storage backend 🗨️ 4 📄 0/4 AG

Misfit processing workflow - 1 dispel4py: identification and production of relevant PEs plus parametrisable graph. 🗨️ 10 📄 2/4 AS FM RF EC

Data Download gUSE workflow: to submit dispel4py job to any cluster with internet access on compute nodes. 🗨️ 8 📄 7/7 AS AG IK



Misfit front-end 1 🗨️ 11 📄 1 📄 0/8

Add a card...

### Waiting

Data download + quality check: User want ask for data download as he/she fires the simulation. User should be allowed to choose between a predefined timewindow and a custom one. REQUIRED DISCUSSION: Default Frequency can be provided from the mesh metadata or users specify a priority list. All the components must be available 🗨️ 6 FM LK EC

Possible priority selection: 🗨️ 1 LK

Misfit script which consumes a Json File as an input file. This script has to read the trace's information from the JSON file, cut the traces and apply the misfit function. A previous version of the Misfit script, where the traces are directly inside the script, has been tested and shared. As soon as the new Misfit script is

Add a card...

### Done

Data management for downloaded data: 🗨️ 5 AG

Data Download in dispel4py: Evaluating integration of Lion script as dispel4py wf or single PE. To be executed by a gUSE workflow. Workflow gets in input a conf-file a stationxml, with station list and bounding boxes and produces in output the location of the files and downloaded stations metadata 🗨️ 1 AK LF RF

Dispel4py workflow for post-processing data. This workflow is able to apply Misfit and Pyflex functions, and it gets two merged figures (pyflex and misfit) with the 3 components (with the labels in each subfigure). 🗨️ 1 AK FM RF EC

Add a card...

### Future Tasks

Misfit front-end 2 - Py dispel4py: Macro PE V 🗨️ 3

Misfit processing workflow dispel4py: Macro PE V code 🗨️ 1

Add a card...

