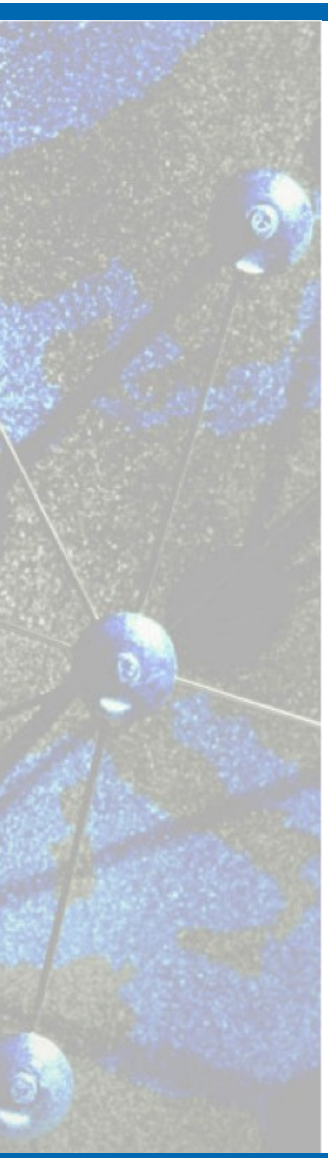


Earth System Grid and EGI interoperability

J. Raciazek (IPSL, FR), S. Denvil (IPSL, FR),
P. Kershaw (STFC, UK),
H. Schwichtenberg (SCAI, DE), A. Gemund (SCAI, DE),
M. Petitdidier (IPSL, FR)

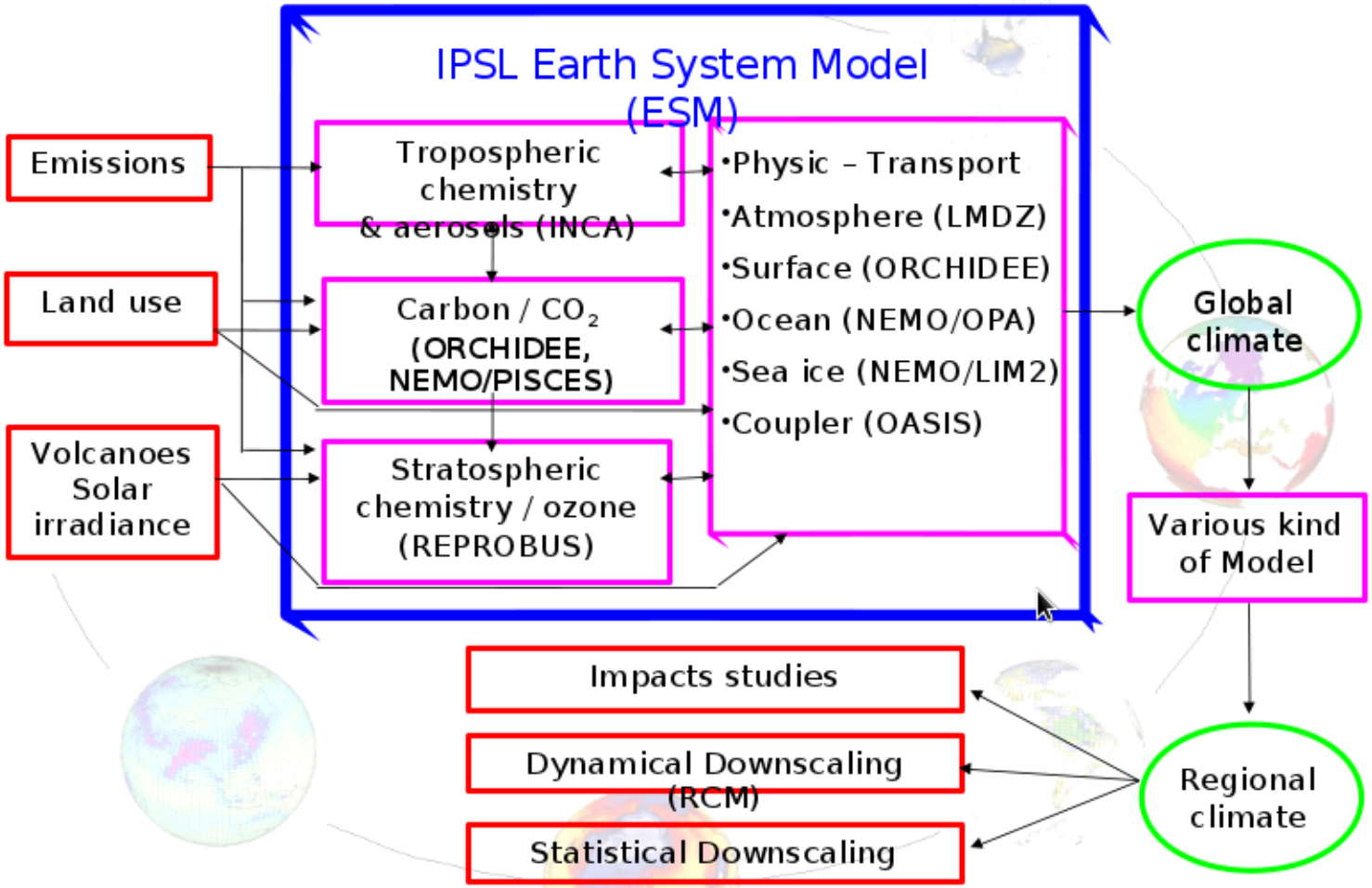
Overview

- Problem statement
- Climate model data
- Earth System Grid Federation (ESGF)
- ESGF security model
- ESGF and EGI security models comparaison
- Solutions

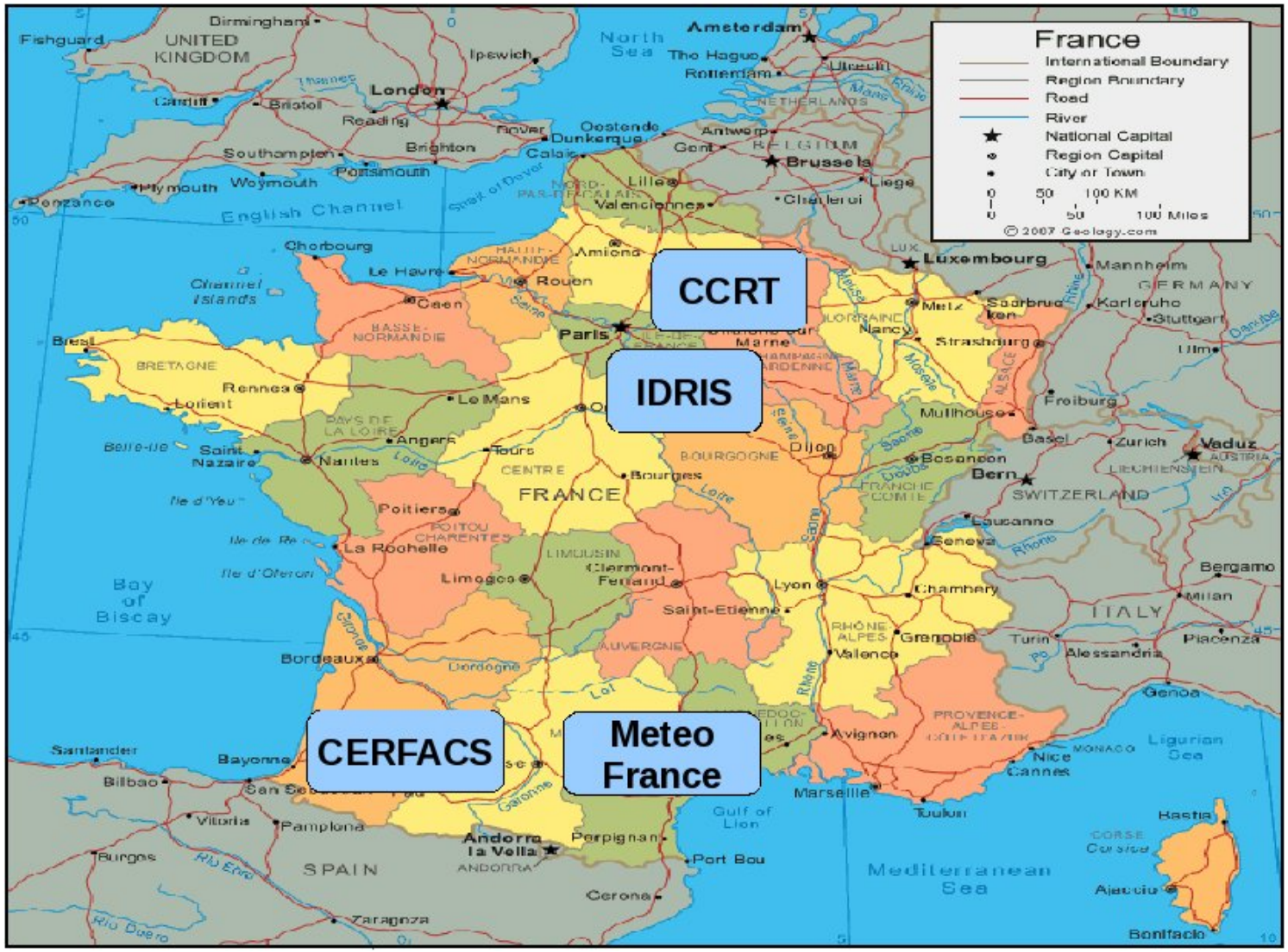
Problem statement

- EGI users need access to climate data
- Possible using two X509 certificates
- Not possible using only one certificate
- Makes data management difficult for end-users
- Hard-coded password / short-lived ESGF certificate

Climate Model Example



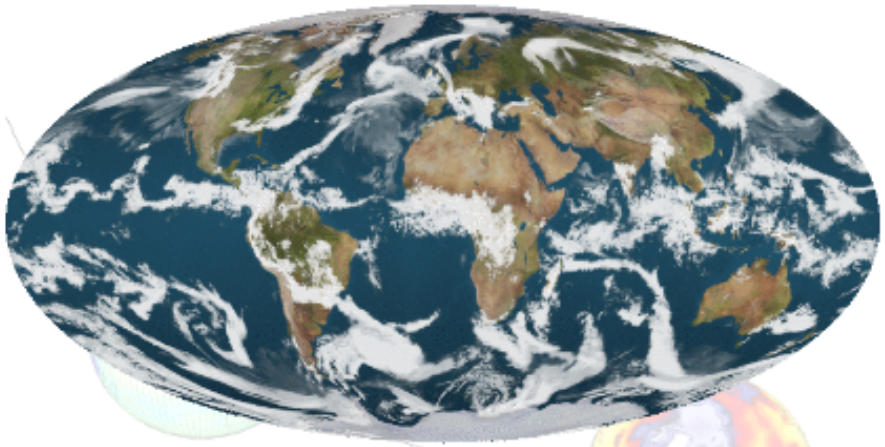
Core Computational Centers



Data volume

CMIP5 global data amount

Raw Data amount lower bound ➔ **565 TB**
Raw Data amount higher bound ➔ **1000 TB**
CMIP5 Distribution (50%) ➔ **280-500 TB**
Global Storage (Raw+Distributed) ➔ **800-1500 TB**

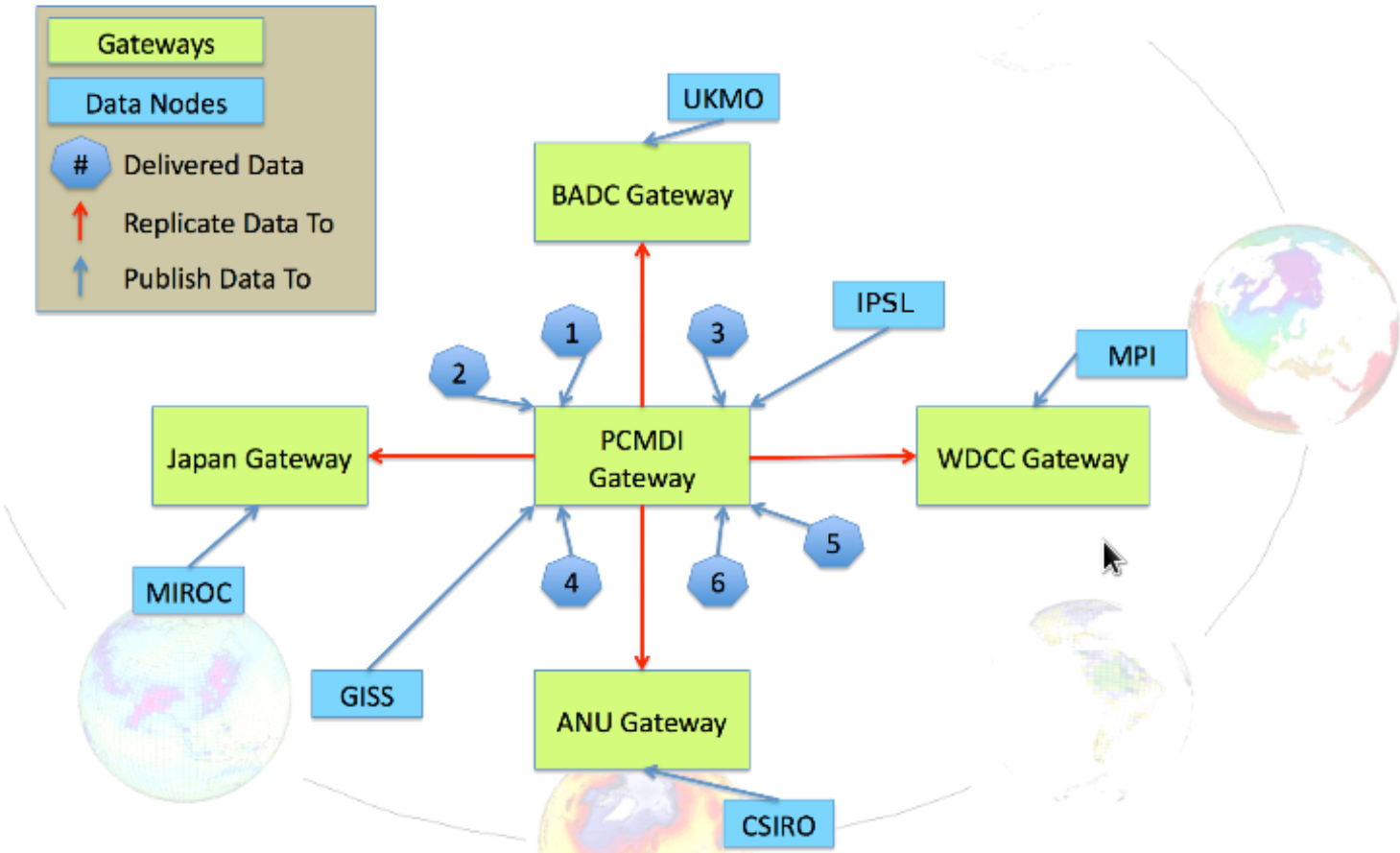


LMDz 0.5° (50 Km)

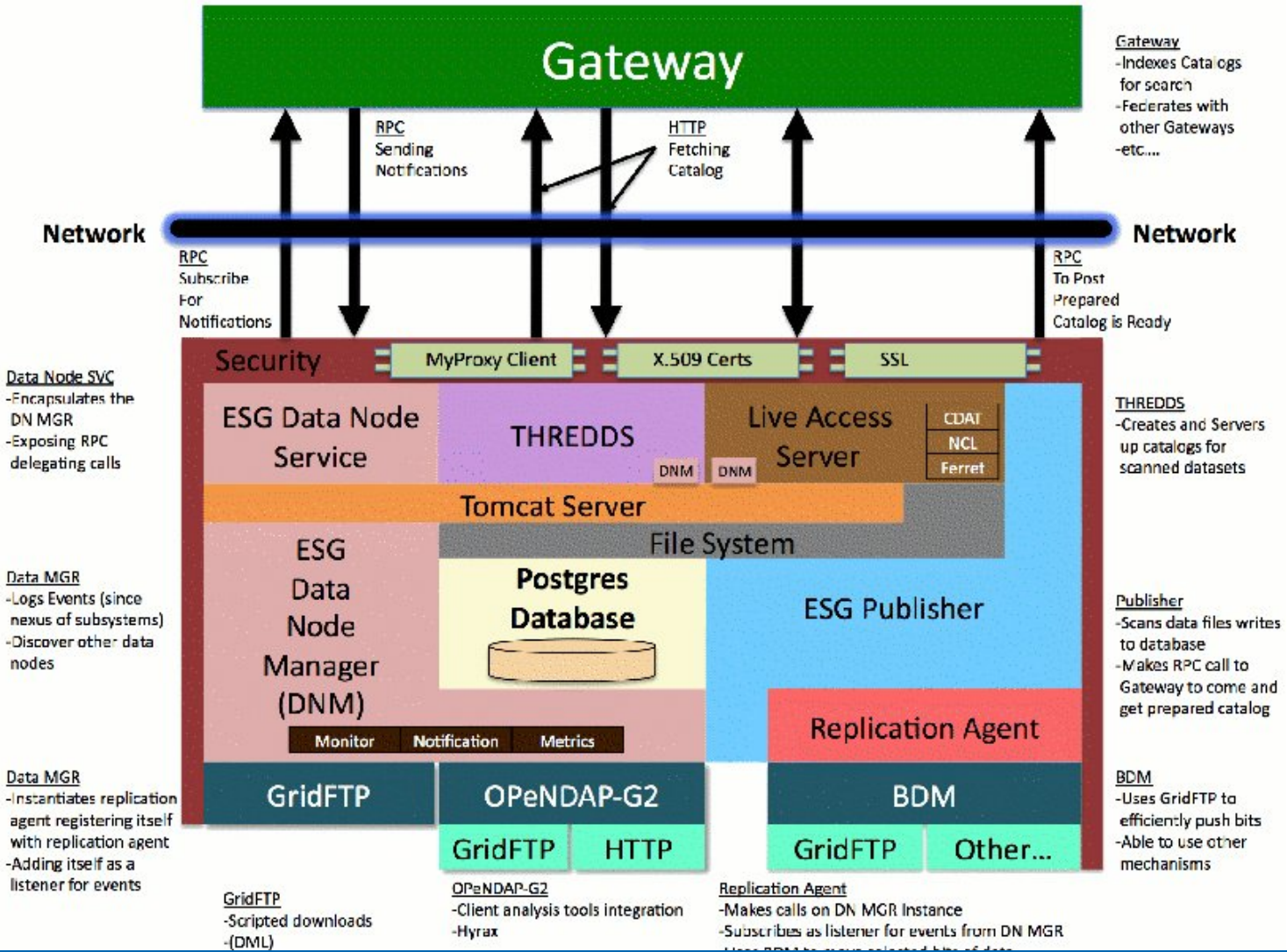
Earth System Grid Federation

- ESGF is an international collaboration providing access to CMIP5 data
- The ESGF software stack has been deployed at a number of site around the world
- Datanode and Gateway
- HTTP, OpenDAP, GridFTP
- Thredd catalog

Earth System Grid Federation



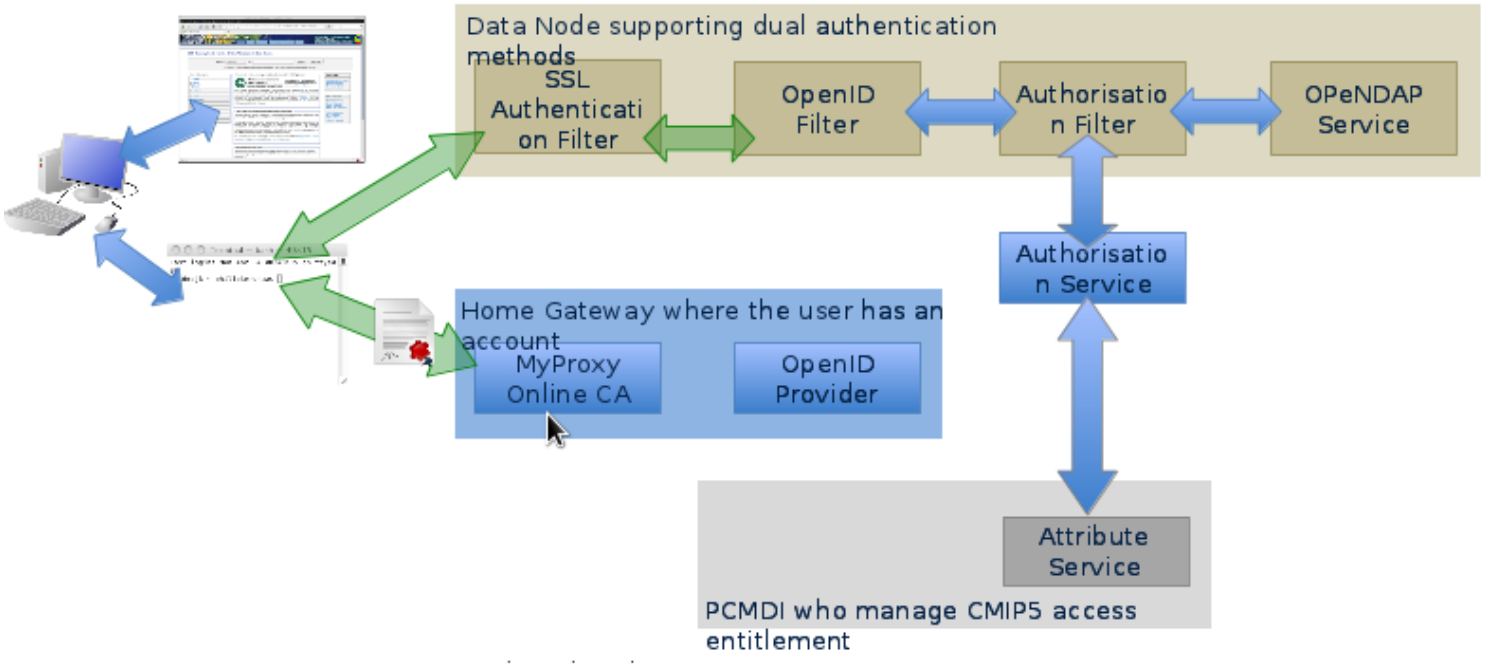
Earth System Grid Federation



ESGF Security Model

- Use a federated identity management system which is based on OpenID and PKI
- Attributes services
- PCMDI
- CMIP5-RESEARCH role
- Short-lived EEC
- No support for Proxy certificates
- Dual authentication mecanism link to the same credential

ESGF Security Model



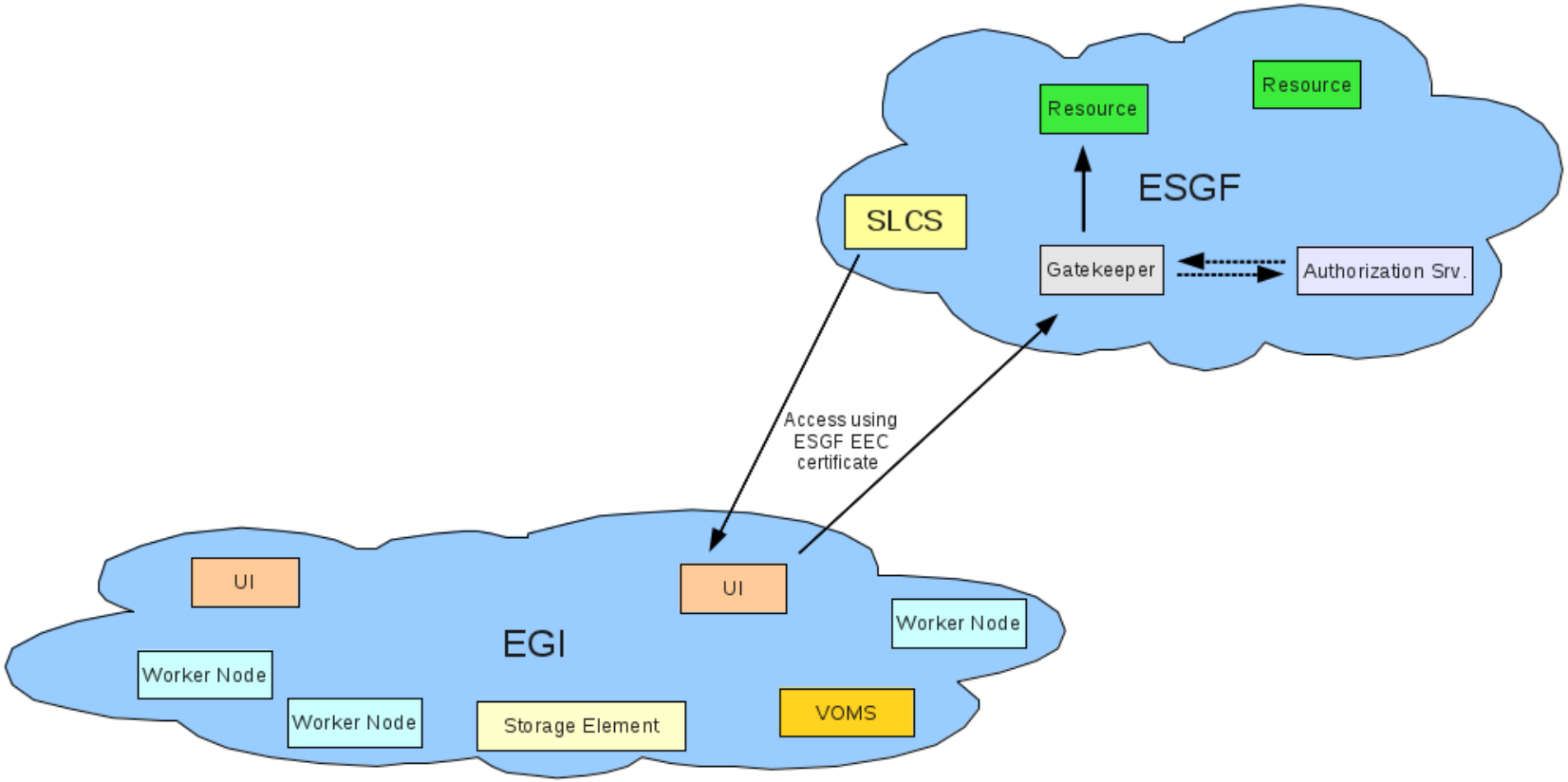
EGI Security Model

- ESR VO
- VOMS database
- Long-lived EEC
- Proxy certificates

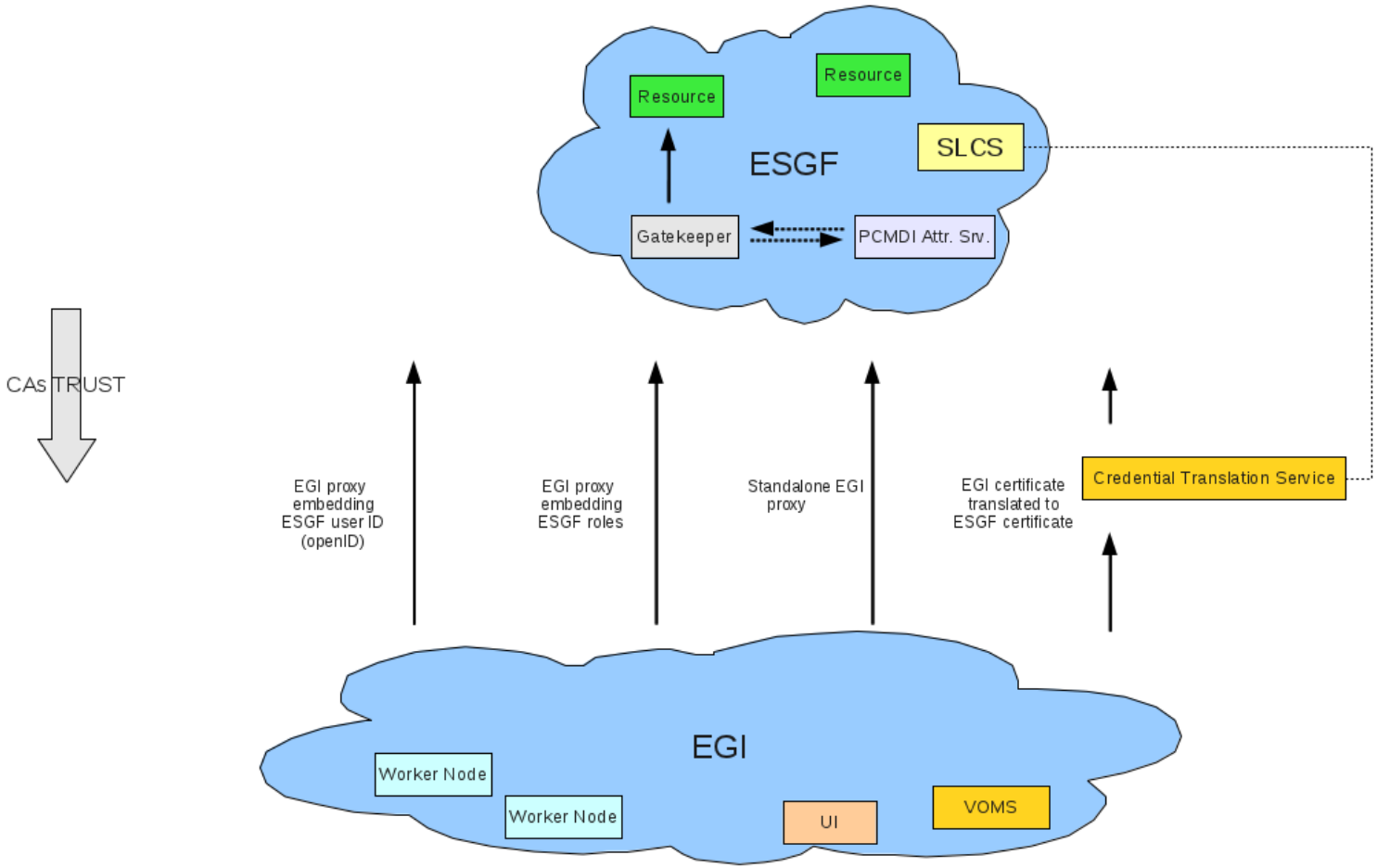
Security Models Comparaison

	EGI	ESGF
PKI / X509 architecture	yes	yes
User identifier	Certificate DN	OpenID
Certificate DN sample	O=GRID-FR, C=FR, O=IPSL, CN=Foo Bar	O=Grid, OU=GlobusTest, OU=simpleCA-pcmdi3.llnl.gov, CN=https://pcmdi3.llnl.gov/esgcet/myopenid/bar
Proxy certificate support	yes	no
CAs accreditation	IGTF	none
CAs list	http://repository.egi.eu/sw/production/cas/1/current	http://esgf.org/esg-certs/#ESG_Federation_Trust_Roots
Attributes propagation mechanism	PUSH	PUSH and PULL

Interoperability Scenarios



Interoperability Scenarios



Solution using STS

Preliminary steps (to be done only once):

- EGI user registers in ESGF (create an OpenID account)
- EGI user registers in STS and set up the mapping for his two accounts (with both sets of credentials)

Authentication and authorisation steps:

- EGI user creates proxy using voms-proxy-init
- EGI user connects to the ESGF STS
- STS authenticates the user using the EGI proxy

Solution using STS

- STS maps the EGI DN to ESGF OpenID
- STS creates an ESGF EEC
- STS send back the ESGF EEC
- EGI user connects with ESGF EEC to the ESGF data service
- ESGF gatekeeper allows or deny access to EGI user based on the ESGF EEC

Pros

- Remove need for ESGF to hold EGI trust roots
- Preserve separation between domains

Solution using EGI Certificate

Preliminary steps:

- EGI ESR VO is accredited to grant ESGF roles to ESR VO user
- An Attribute Service is set up at ESR VO (for ESR VO governed CMIP5 attributes)
- EGI user *explicitly* sign up to the CMIP5 usage policy during the ESR VO registration procedure

Solution using EGI Certificate

Authentication and authorisation steps:

- EGI user requests ESGF data using the EGI proxy
- the ESGF authorisation service retrieves authorisation roles from the ESR VO attribute service using EGI proxy
- ESGF authorisation service allows or deny access

Solution using EGI Certificate

Pros

- EGI user do not need to create an OpenID account

Changes in ESGF

- EGI CAs must be trusted by ESGF
- ESGF access control middleware must be modified, because ESGF currently expects an OpenID
- Support for Proxy certificates is also needed

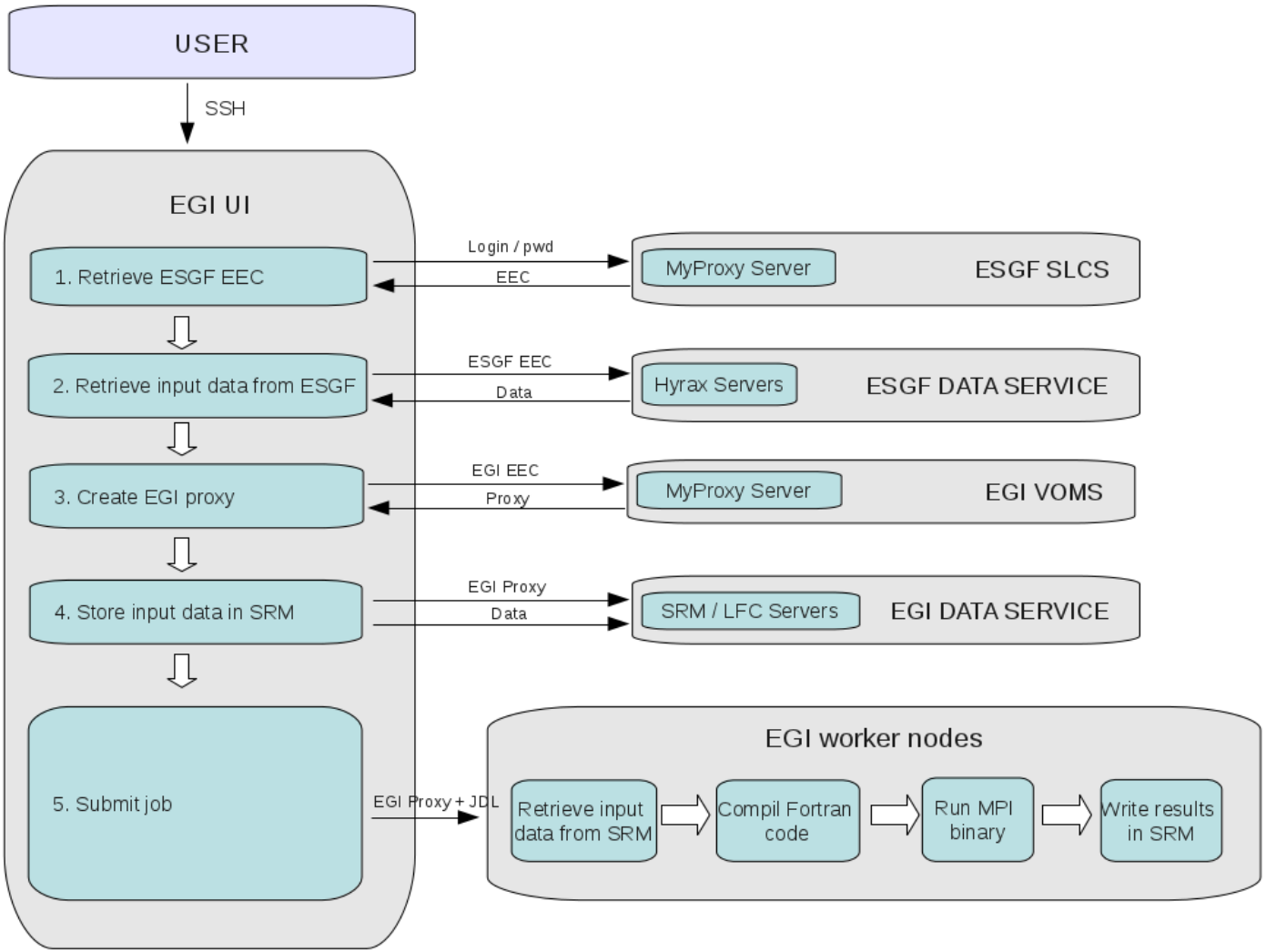
Changes In EGI

- ESR VO needs some means to register users where the user *explicitly* signs up to the CMIP5 usage policy.
- A SAML Attribute Service must run in conjunction with the ESR VO, so the ESGF access control could pull attributes from this service.

Testbed

- Application that needs access to ESGF CMIP5 data to study extreme events has been deployed on EGI
- Fortran parallel code (MPI)
- Already running in ESR VO (using SRM/LFC as intermediate storage area)

Testbed


USER
SSH
EGI UI
1. Retrieve ESGF EEC
Login / pwd
EEC
MyProxy Server
ESGF SLCS
2. Retrieve input data from ESGF
ESGF EEC
Data
Hyrax Servers
ESGF DATA SERVICE
3. Create EGI proxy
EGI EEC
Proxy
MyProxy Server
EGI VOMS
4. Store input data in SRM
EGI Proxy
Data
SRM / LFC Servers
EGI DATA SERVICE
5. Submit job
EGI Proxy + JDL
EGI worker nodes
Retrieve input data from SRM
Compil Fortran code
Run MPI binary
Write results in SRM

Future Related Work

A workshop on Federated Identity Systems for Scientific Collaborations has been scheduled for the 2-3 November 2011. The event will be hosted by STFC at Rutherford Appleton Laboratory in the UK.
https://indico.cern.ch/event/129364

To be continued...

Thank you