Science & Technology Facilities Council
Rutherford Appleton Laboratory

Federated Identity in the Earth Science Domain:
the Earth System Grid Federation, EGI-Inspire and GENESI-DEC

Federated Identity System for Scientific Collaborations Workshop

CERN 9-10 June 2011

Philip Kershaw [philip.kershaw@stfc.ac.uk] (STFC Rutherford Appleton Laboratory, NCAS/BADC, UK)
Sébastien Denvil, Jérôme Raciazek, Monique Petitdidier (CNRS / IPSL France)

Horst Schwichtenberg, André Gemünd (SCAI, Germany)

L. Fusco, R. Cossu, (ESA / ESRIN, Italy)

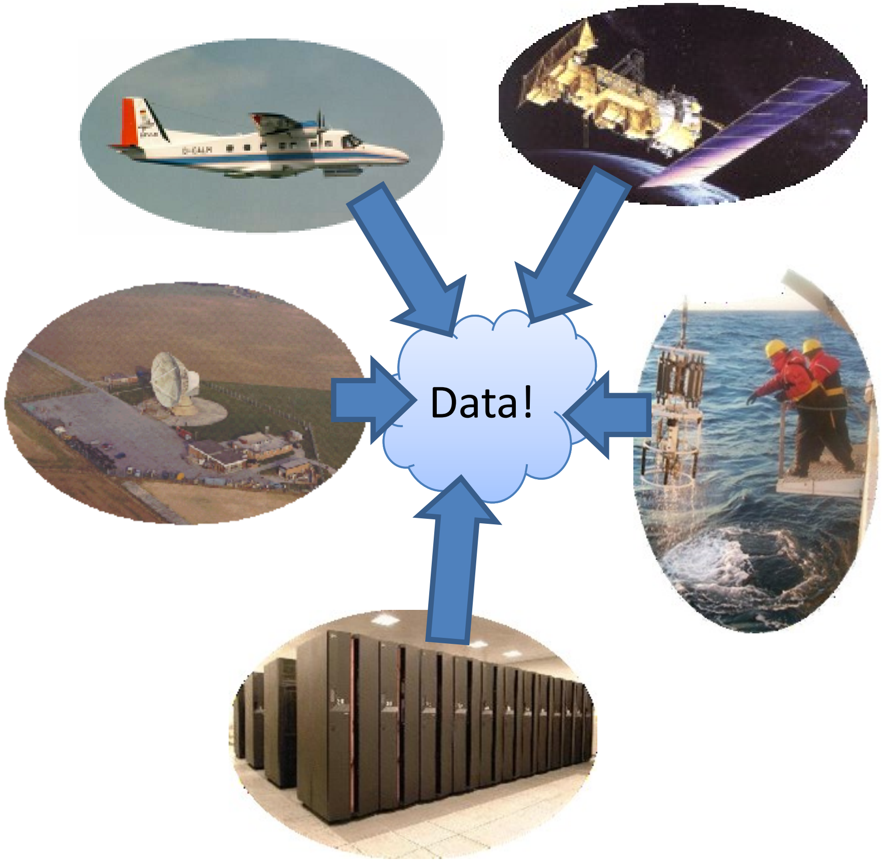
British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Overview

- Background and drivers for federated access control in the Earth Science domain
- the Earth System Grid Federation (ESGF)
  - A distributed infrastructure for the discover, access and analysis of Earth science data
  - CMIP5 as a motivator for the development
- Inter-federation trust
  - EGI-Inspire EU FP7 Project
  - ESGF ⇔ EGI
  - GENESI-DEC ⇔ EGI

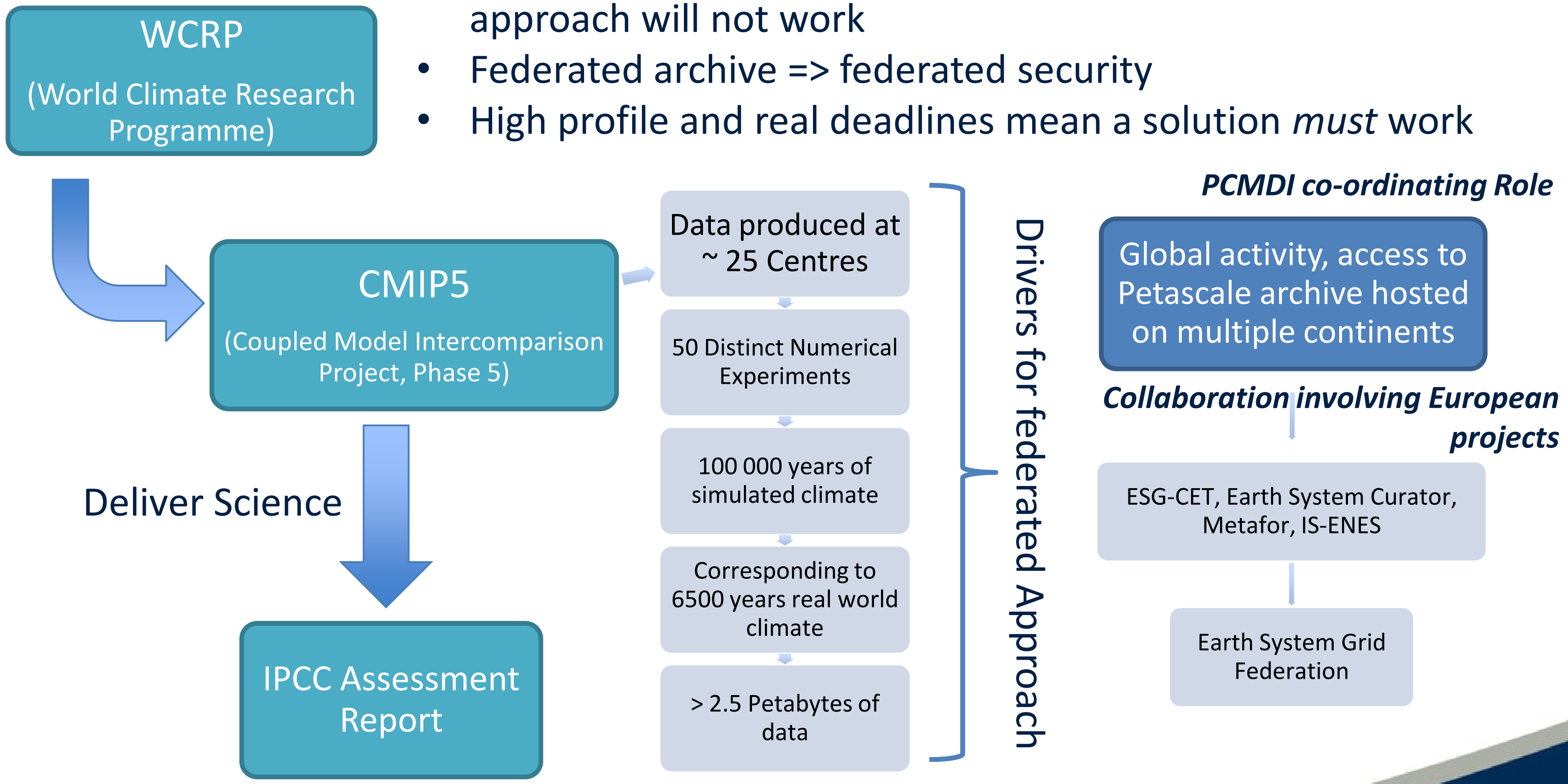
Data Challenge for the Earth Science Community



- Environmental scientists use numerous sources of data
- The ability to combine and compare **diverse** datasets is critical to furthering our understanding of the Earth system.
- *but* the integration of such datasets can be difficult, largely due to inherent technical complexities.
- Increasing data volumes necessitate distributed infrastructures
- **Organisational domains, trust, licensing, identity management**
- As a result, many valuable environmental datasets are underused.

CMIP5 and the Earth System Grid

- Size of prospective archive + distribution challenge => centralised approach will not work
- Federated archive => federated security
- High profile and real deadlines mean a solution *must* work



WCRP

(World Climate Research Programme)

CMIP5

(Coupled Model Intercomparison Project, Phase 5)

Deliver Science

IPCC Assessment Report

Data produced at ~ 25 Centres

50 Distinct Numerical Experiments

100 000 years of simulated climate

Corresponding to 6500 years real world climate

> 2.5 Petabytes of data

Drivers for federated Approach

PCMDI co-ordinating Role

Global activity, access to Petascale archive hosted on multiple continents

Collaboration involving European projects

ESG-CET, Earth System Curator, Metafor, IS-ENES

Earth System Grid Federation

ESGF Requirements and Challenges

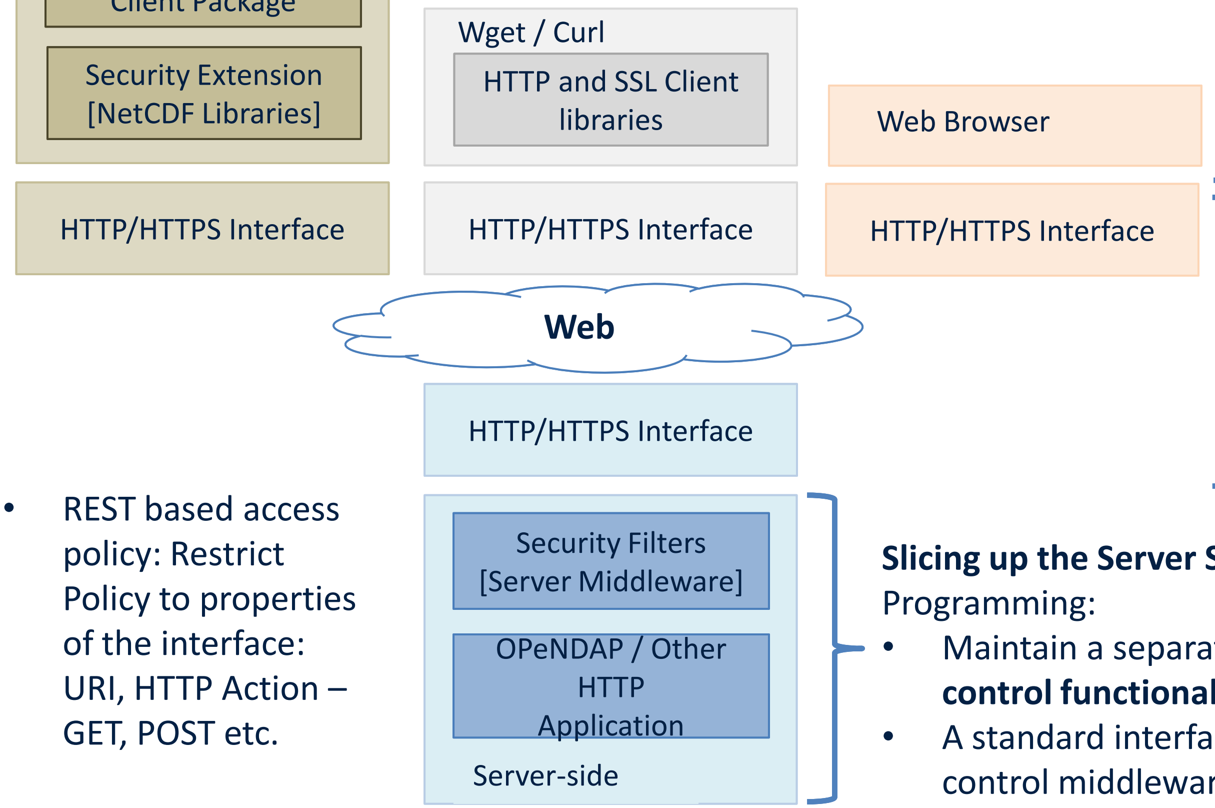
- Requirements
  - Low level of assurance required for CMIP5 access
  - PCMDI (Lawrence Livermore National Laboratory) to administer registration of access rights
  - Audit access
  - Register users to keep them up-to-date with changes to data and services
  - Protect finite resources at service providers (compute, bandwidth …)

- How to apply access control in a heterogeneous environment of data access services and tools in this domain?
  - OPeNDAP, Live Access Server/Ferret, OGC web services, GridFTP, CDAT, Matlab, IDL, Ferret …

- Multiple technologies in the field of access control and security
  - Grid, Shibboleth, SAML, OpenID, OAuth, Kerberos …

A Solution in Modular Design Principles

- Divide and Conquer approach applying SOA, AOP, REST and NetCDF


NetCDF-based Client
NetCDF dependent Client Package
Security Extension [NetCDF Libraries]
Wget / Curl
HTTP and SSL Client libraries
Web Browser
HTTP/HTTPS Interface
HTTP/HTTPS Interface
HTTP/HTTPS Interface
Web
HTTP/HTTPS Interface
Security Filters [Server Middleware]
OPeNDAP / Other HTTP Application
Server-side

Slicing up the Client Side:

- Simple interface suited to Wget and Curl
- Rich clients: Security integrated at a base level in the client software stack: into NetCDF client libraries

Organisational Boundaries: SOA (Service Oriented Architecture):

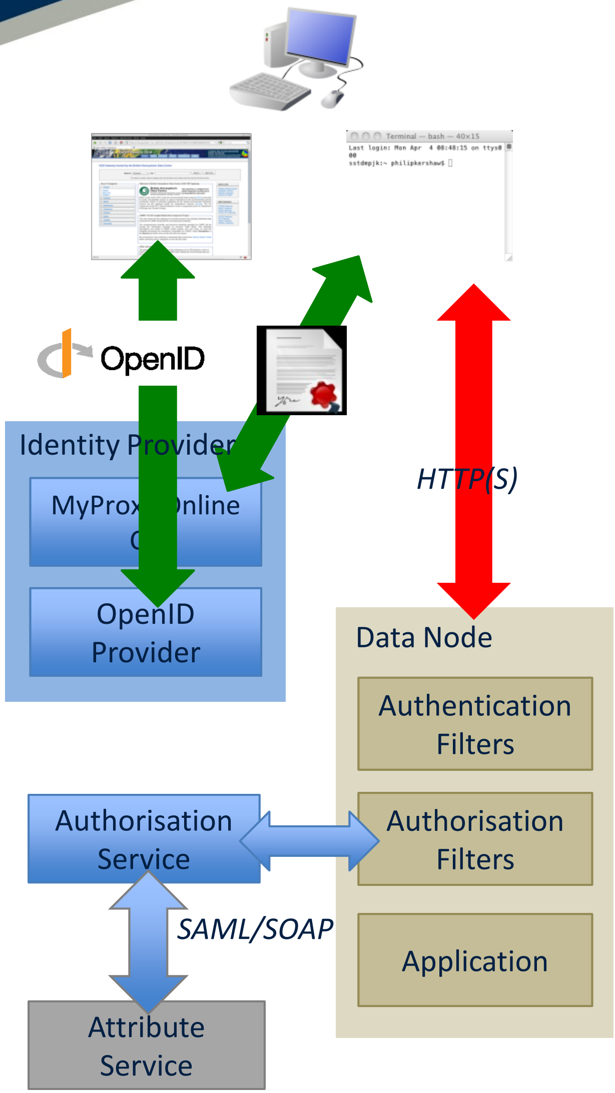
- Defined interfaces with web services with *profiled* specs : OpenID, SAML, PKI => interoperability

- REST based access policy: Restrict Policy to properties of the interface: URI, HTTP Action – GET, POST etc.

Slicing up the Server Side: AOP – Aspect Oriented Programming:

- Maintain a separation of concerns between **access control functionality** and **application** to be protected
- A standard interface between the two enables access control middleware to be configured to protect any app which supports that interface

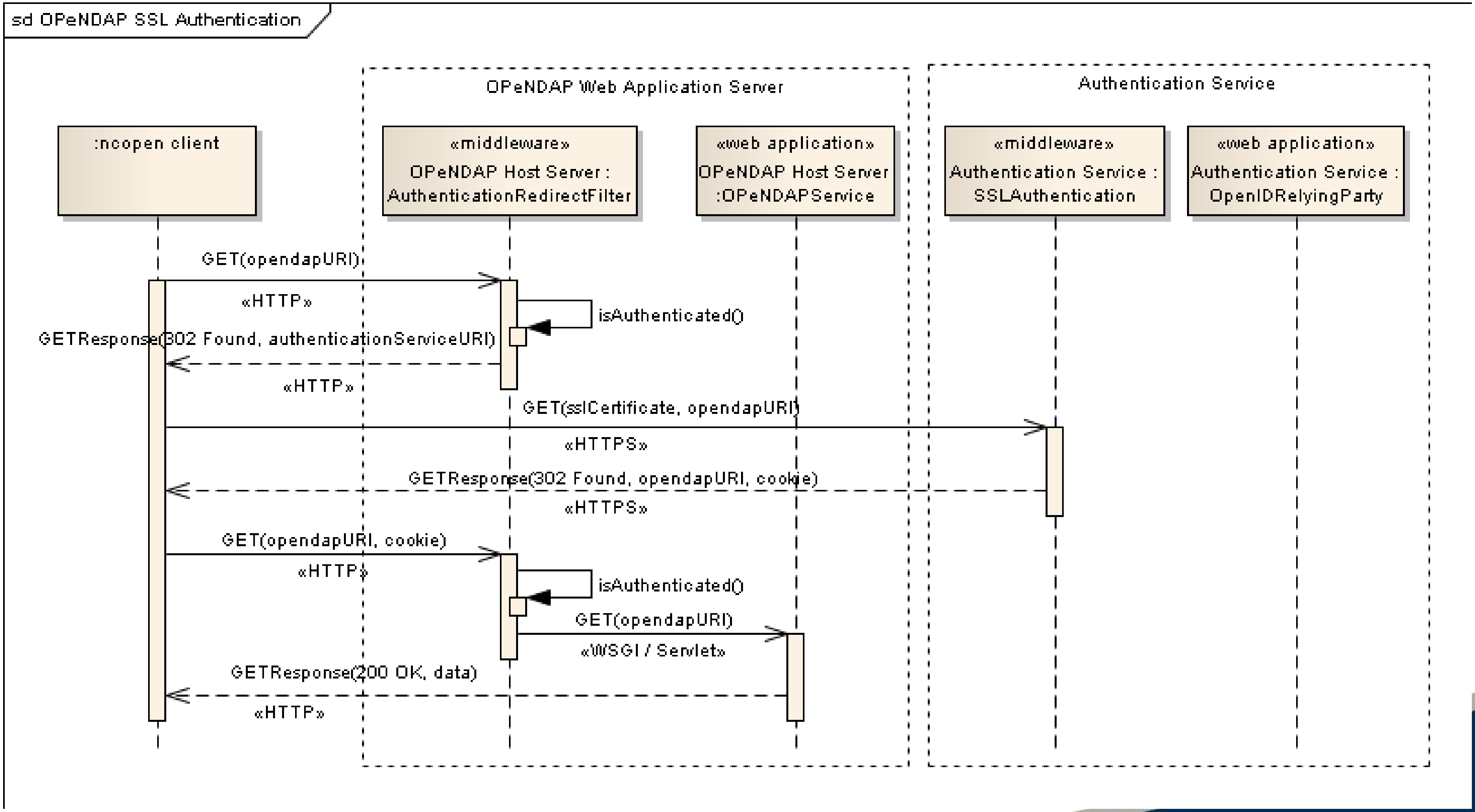
Federated Security Architecture



```
Terminal — bash — 40×15
Last login: Mon Apr  4 08:48:15 on ttys0
00
sstdmpjk:~ philipkershaw$ 
```

- Single sign on with OpenID and MyProxy
  - Dual authentication mechanisms link to the same credentials
- Applications *simultaneously* support the dual authentication methods
  - The server-side access control layer is agnostic to the authentication approach employed by the client
  - The underlying application is kept independent
  - Access control filters can be assembled in different configurations to suit different application scenarios
- Attribute propagation:
  - Pull-based model with SAML Attribute Services
  - Push-based also supported with OpenID AX (Attribute Exchange) and embedding of SAML assertions in user certificates
- Authorisation:
  - Authorisation Service with SAML interface
  - Can accept queries from a range of PEPs fronting services – GridFTP, OPeNDAP
  - XACML Policy engine used in Python implementation at the BADC/CEDA
- Service Discovery with Yadis protocol: XRDS over HTTP(S)
  - Introspect IdP services from a user's OpenID URI
  - Discover Attribute Service and MyProxy server endpoints from a user's OpenID

Multiple Authentication Methods



Successes

- A standard solution for securing OPeNDAP and other HTTP-based services
  - Access for simple HTTP clients: **Wget/Curl**
  - Integrated into new **NetCDF**
    - filters down to all the dependent packages: CDAT, Ferret …
  - Access for Grid based infrastructure: SSL-based authentication
  - Delegation capability for securing workflows
- Interface Control Document
  - Python and Java implementations
- Highly configurable access control middleware
  - Easy to support multiple security paradigms e.g. OpenID and SSL based
- Security is built on trust – relationships between organisations
  - The importance of a strong common goal in CMIP5
  - The close collaboration required has in turn fostered more partnerships
  - ESGF Open Source development effort

Issues

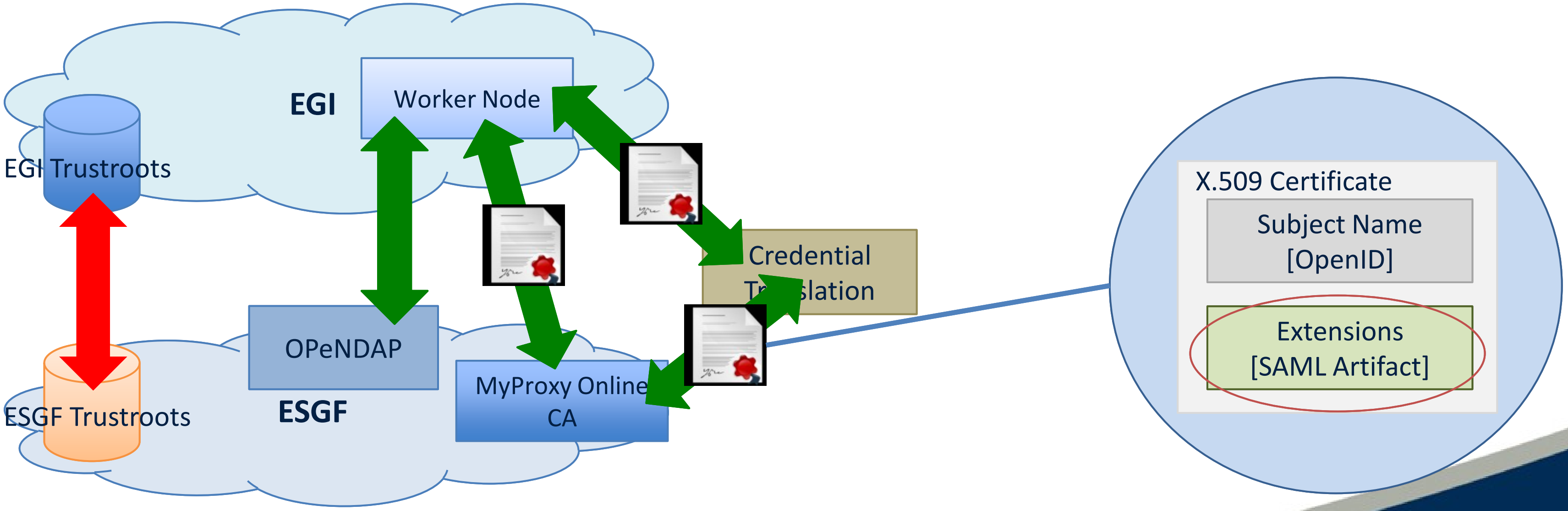
- Security is inherently complex
  - PKI (Public Key Infrastructure), a fundamental building block to anchor trust but difficult to manage and administer
- Does the level of security required justify effort needed?
  - Need to support Levels of Assurance for authentication mechanisms
- Federation management, SLAs must not be overlooked
- Remember who are the stakeholders
  - Users: do they understand Single sign-on?!
  - Deployers: can organisations easily deploy?
  - **Developers:** A need to pass on knowledge and expertise

ESGF Integration with EGI

- Objective: enable access for Grid services to CMIP5 Data through ESGF OPeNDAP services
- How to solve trust between grids: *exchange of PKI trust roots* or *credential translation*?

- Exchange of trust roots
  - ESGF credentials trusted in EGI
  - Make ESGF MyProxy Online CAs subordinate to IGTF trust roots
  - Add respective EGI trust roots to ESGF infrastructure
  - Proxy certificate support
  - Certificate lifetime and CRLs
  - OpenID in X.509 Subject Name

- Credential translation
  - Convert from ESGF to EGI certificate
  - Removes need for EGI to hold ESGF trust roots
  - Preserves separation between domains
  - But, reverse mapping needed EGI -> ESGF?
  - ESGF certificates hold a SAML assertion, if signed this could be used as the authentication credential and passed in different certificate 'containers'



GENESI-DEC

- Goal: harmonisation of satellite data access
  - FP7 project successor to GENESI-DR
  - Completes in 2012
  - Portal and search services to interface with other applications
  - Standards work with OGC (Open Geospatial Consortium) and OpenSearch
  - Uses Grid-based security model – PKI/Proxy certificates
- Handling of VOs in federated infrastructures
- Cloud IaaS (Infrastructure as a Service) and OGC WPS (Web Processing Services)
- More developments underway with security model …

Future and Related Work

- OAuth for delegation
  - MashMyData Project
    - Proxy Certificate based Delegation in workflow with WPS and OPeNDAP services
  - ExArch Project
    - G8 funded collaboration, US, Canadian and European partners
    - WPS: keep the processing near the data
- IS-ENES (InfraStructure for the European Network for the Earth System Modelling) EU FP7
  - Delegation use case is important e.g. Portal ⇔ WMS client ⇔ WMS access control
- ISIC (International Space Innovation Centre), RAL
  - Create critical mass for space-related activities
  - Earth observation hub
  - ESA also now located at RAL
- Shibboleth support?

Any Questions?

- More information on ESGF security:
  - ESGF Security paper for GCA2011, Las Vegas, July
  - http://philipkershaw.blogspot.com/