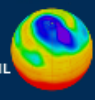
Access Control Architecture for the Earth System Grid Federation:
Building an Infrastructure of Secured Data Access Services for the Climate Science Research Community

Philip Kershaw [philip.kershaw@stfc.ac.uk]

Rachana Ananthakrishnan (Argonne National Laboratory, IL, USA),
Luca Cinquini (Jet Propulsion Laboratory, CA, USA),
Estanislao Gonzalez (Max-Planck-Institut für Meteorologie, Hamburg, Germany),
Dennis Heimbigner (University Corporation for Atmospheric Research, CO, USA),
Bryan Lawrence (BADC, Rutherford Appleton Laboratory, UK)

(Acknowledgements to the ESG PIs and development teams from all the partners)

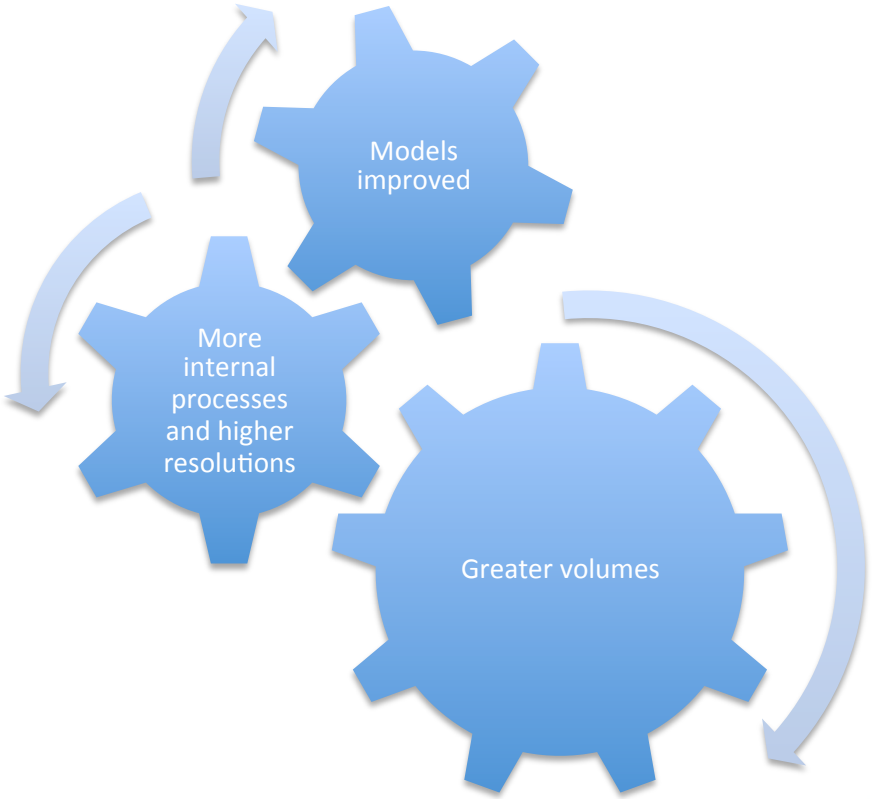


Overview

- Background Context to the Earth System Grid Federation
- Drivers for a Federated Approach to data access
- Access Control Requirements
- Divide and Conquer tactics with SOA, AOP and NetCDF
- Architectural Walkthrough
- Successes: securing OPeNDAP services
- Problems
- Future Work

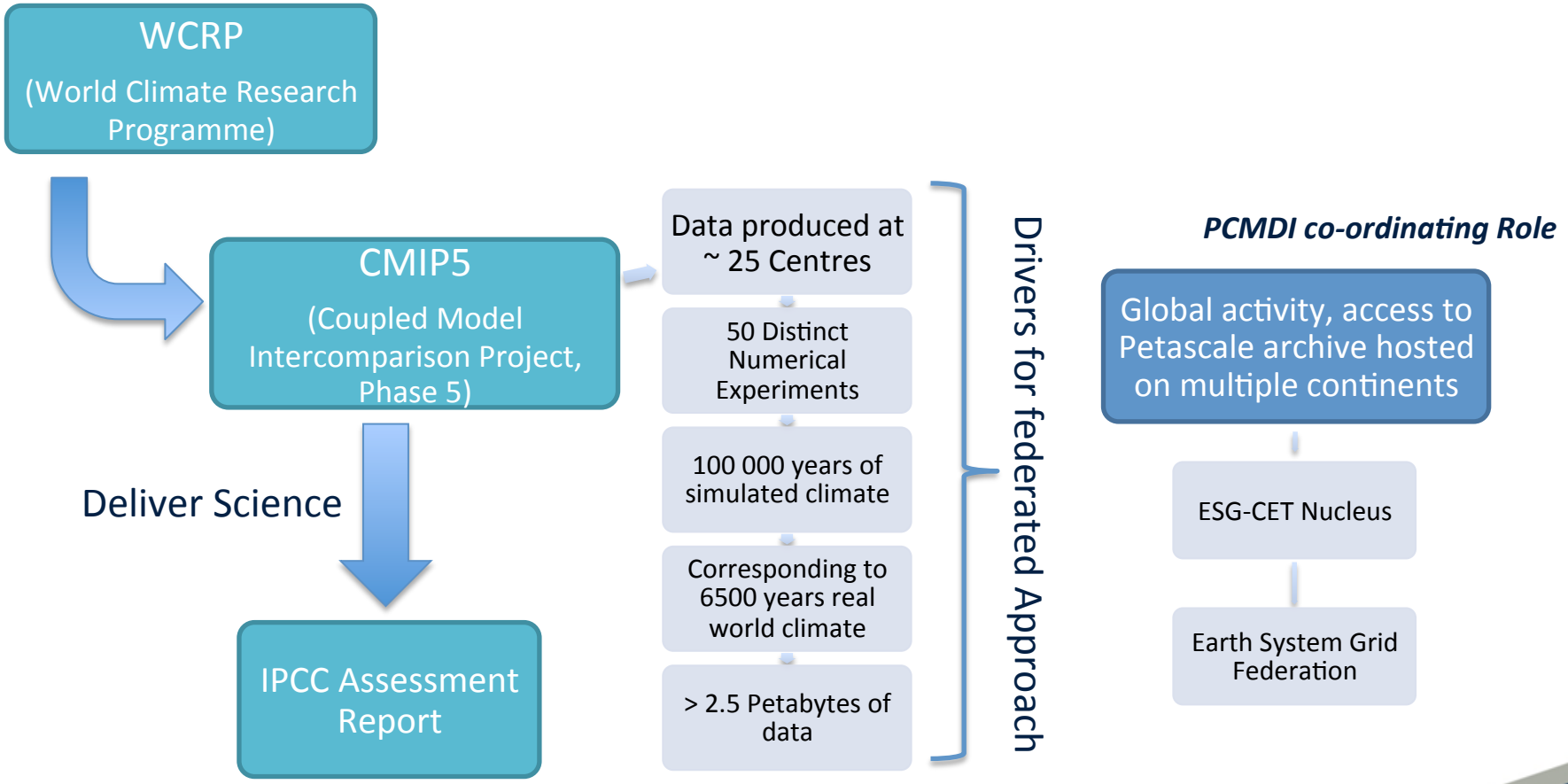
Background Context

- **Climate model simulations:** their production, evaluation and interpretation of is integral to earth system science
- They have always been on the leading edge of computing: high performance and high data volumes



Philip Kershaw [philip.kershaw@stfc.ac.uk]

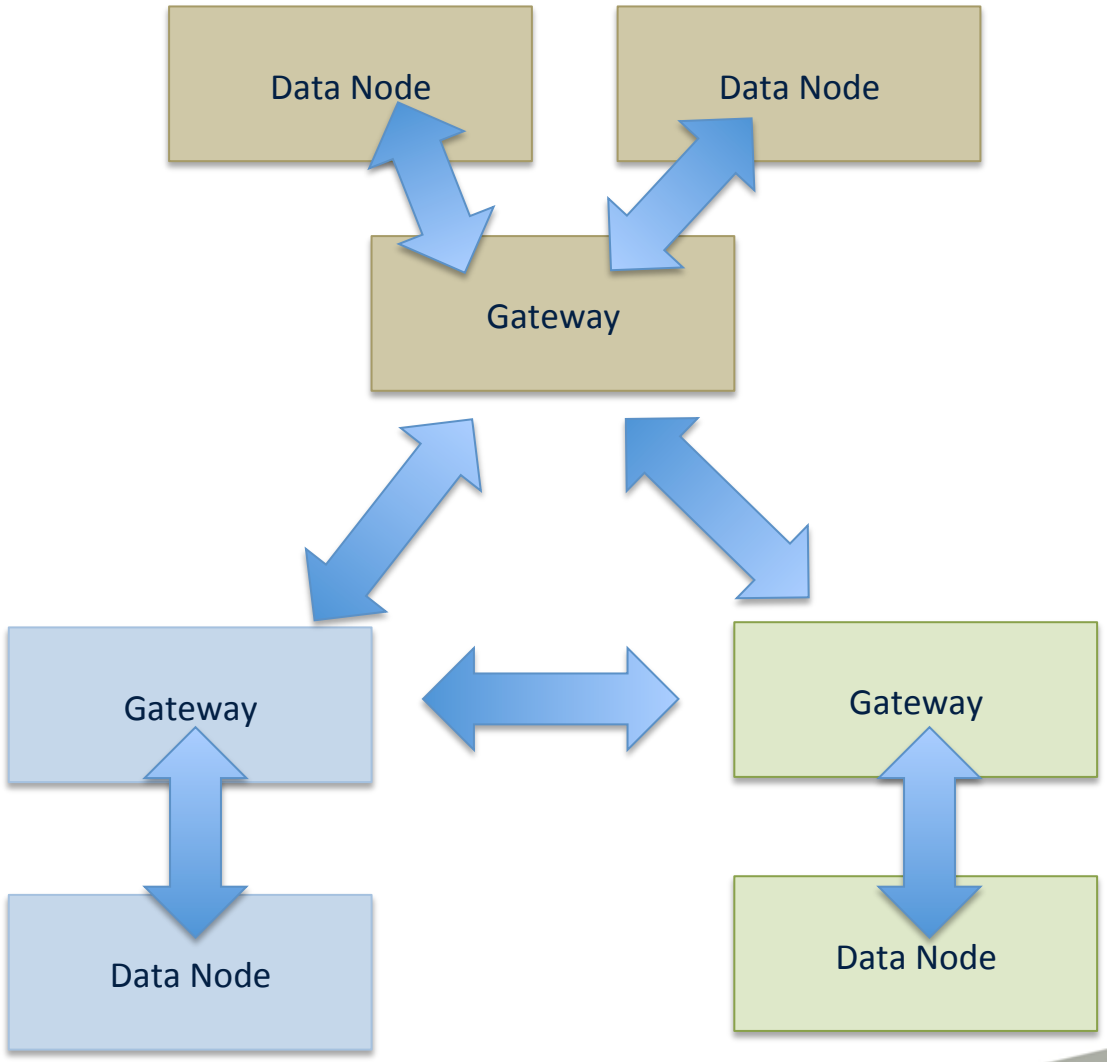
CMIP5 and the Earth System Grid



(A centralised approach is **not** *going to work!)*

Philip Kershaw [philip.kershaw@stfc.ac.uk]

ESGF High Level Architecture



Philip Kershaw [philip.kershaw@stfc.ac.uk]

Access Control Requirements

1) **Access Policy**: A mechanism to set policy on restricting access to chosen datasets, per dataset on a case by case basis

2) **PCMDI** (Lawrence Livermore National Laboratory, CA) are the **source of authority** for CMIP5 access entitlement

3) **Notification**: The ability to notify users of changes to data and services.

4) **Metrics**: The ability to collect metrics about data download, the number of unique downloads

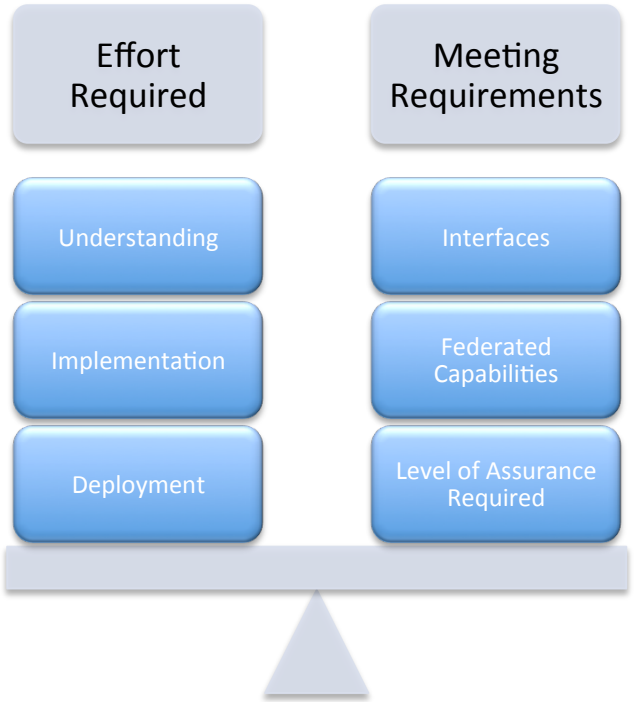
5) **Seamless access for users** to data hosted by all organizations in the federation: **single sign-on**

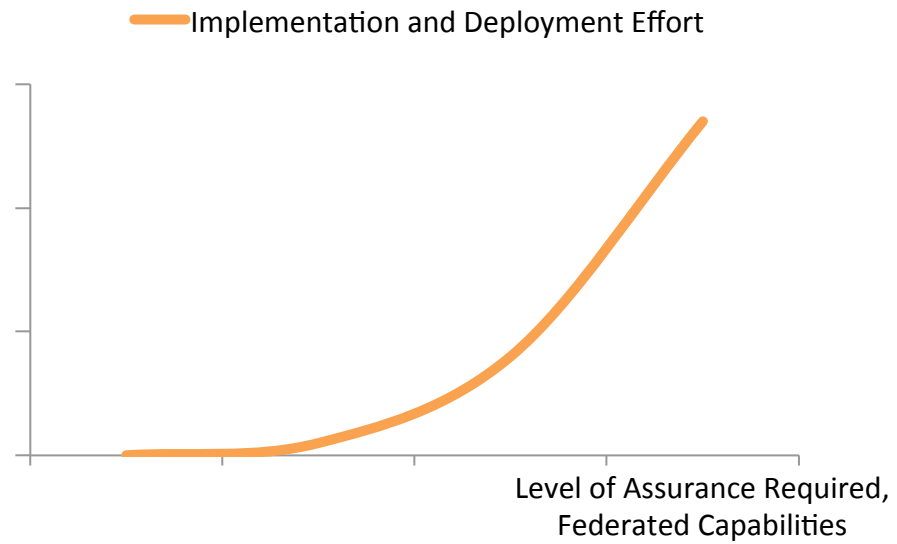
6) **Clean integration** with services and tools that scientists commonly use (browser and thick client access).

7) **Protection of Resource Providers** - their finite computing assets - from malicious or unintended requests

Philip Kershaw [philip.kershaw@stfc.ac.uk]

Requirements vs. Effort





- How much are the requirements *really* worth?
- Can I leverage existing technology and infrastructure?

Philip Kershaw [philip.kershaw@stfc.ac.uk]

Divide and Conquer with SOA, AOP, REST and NetCDF

Slicing up a cake, but can I place a standard interface between the slices?!! ☺

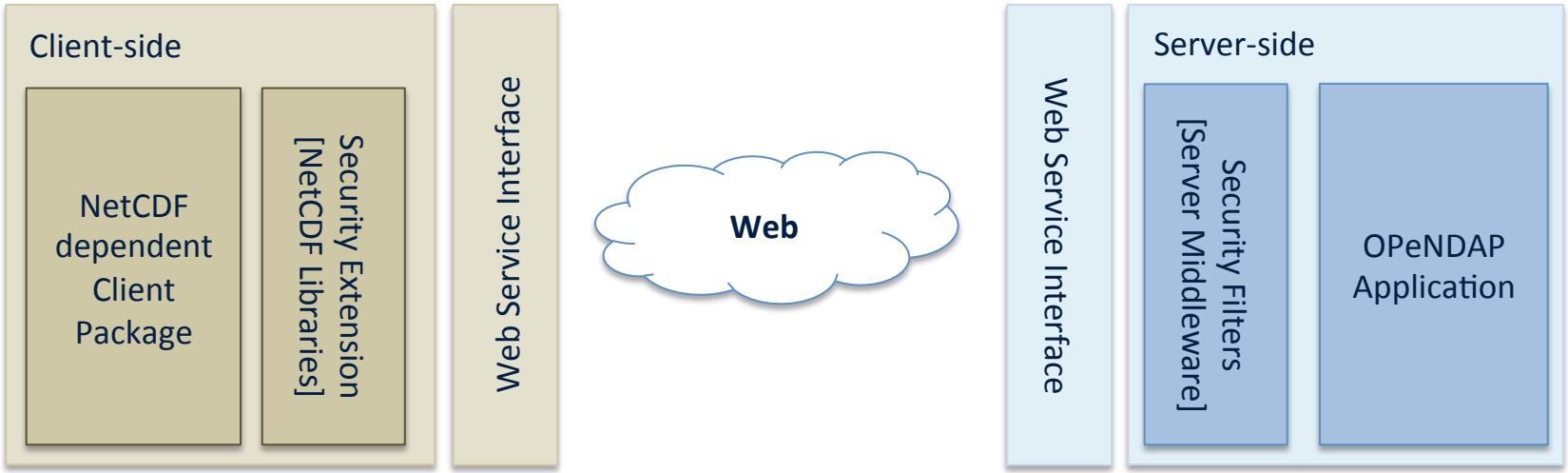
- The problem:
  - multiple distributed services deployed in mixed environments at host institutions
- Solution:
  - target along lines in the system to divide up the problem and simplify it
- Slice up and place standard interfaces between the slices
- SOA, AOP, REST and NetCDF?? …

Philip Kershaw [philip.kershaw@stfc.ac.uk]

Divide and Conquer (contd.)

Organisational Boundaries: SOA (Service Oriented Architecture):

- Defined interfaces with web services => interoperability and peer reviewed protocols: OpenID, SAML, PKI



Slicing up the Client Side: Security hooks integrated into NetCDF client libraries

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

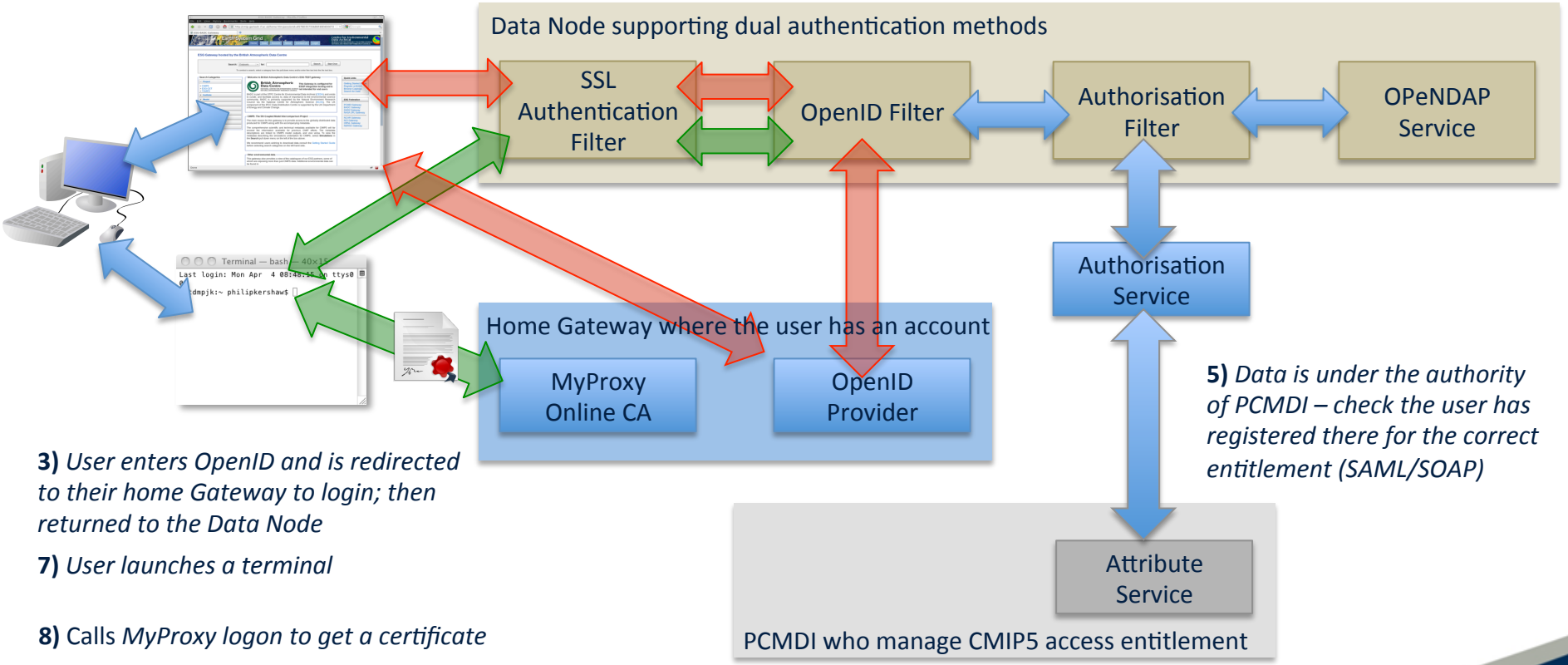
Access Control Architecture

1) *Access via Browser*

2) *SSL Filter intercepts request (HTTPS) - no certificate passed – pass on to OpenID Filter*

4) *The user is authenticated – now check their authorisation (SAML/SOAP)*

6) *Access granted – pass control to app to serve data*



3) *User enters OpenID and is redirected to their home Gateway to login; then returned to the Data Node*

5) *Data is under the authority of PCMDI – check the user has registered there for the correct entitlement (SAML/SOAP)*

7) *User launches a terminal*

8) Calls *MyProxy logon to get a certificate*

9) *User launches rich client app which authenticates over HTTPS with the certificate*

10) *Authorisation and access proceeds as before*

Philip Kershaw [philip.kershaw@stfc.ac.uk]

Successes

- A Standard Solution for Securing OPeNDAP Services
  - Access for simple HTTP clients: **Wget**
  - SSL-based authentication enables access via Grid based services
  - Delegation capability for securing workflows
  - Integrated into new **NetCDF 4.1.2** release
    - filters down to all the dependent packages: CDAT, Ferret …
- Highly configurable Access Control Middleware
  - Easy to support multiple security paradigms e.g. OpenID and Grid based
- Security is built on trust – relationships between organisations
  - The close collaboration required has in turn fostered more partnerships
  - ESGF Open Source development effort: Python and Java implementations

Problems

- Security is inherently complex
  - PKI (Public Key Infrastructure), *PKI*, **PKI**!
  - A fundamental building block to anchor trust but difficult to manage and administer
- Does the level of security required justify effort needed?
- Federation management, SLAs must not be overlooked
- Remember who are the stakeholders
  - Users: do they understand Single sign-on?!
  - Organisations – deployers
  - **Developers:** A need to pass on knowledge and expertise to Developers
  - Them and us?!

I've got this great Idea!

<deep intake of breath> There's **no** *way* that will be secure

Future Work

- MashMyData Project (Poster XL234, Thurs 17:30-19:00)
  - Intercomparison of environmental data in web-style Mash-up
  - Processing of data in situ with OGC Web Processing Service (WPS)
  - Proxy Certificate based Delegation in workflow with WPS and OPeNDAP services
  - OAuth alternative solution
- EGI Collaboration
  - Enable access for Grid services to CMIP5 Data through ESGF OPeNDAP services
- IS-ENES (InfraStructure for the European Network for the Earth System Modelling) EU FP7
  - Delegation use case
- Prodiguer – IPSL (Institut Pierre Simon Laplace)
  - Leverage ESGF access control architecture