



Science & Technology Facilities Council
Rutherford Appleton Laboratory

Earth System Grid Federation Proposal

ESGF Meeting 9 May 2011

Bryan Lawrence

Stephen Pascoe

Philip Kershaw



**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



ESGF and GO-ESSP

- GO-ESSP Mission Statement
 - About building a community of software developers
- ESGF:
 - initially an infrastructure to support CMIP5
 - Something to be proud of!
 - Could become a **de facto standard for distributed access to earth science data** BUT ...
 - To do so it needs to support *participation* and *contribution* from a diverse community ...



A Diverse Spectrum of Usage Profiles

- Entrant organisations with **no** existing systems
 - They want to deploy ESGF software to be part of ESGF
- Those with large **existing** systems
 - ESGF must integrate with the need to abandon these systems or duplicate infrastructure
- Contributors to the development of the ESGF software
 - perhaps for goals in projects beyond CMIP5
 - They need to be able to do so within their own project constraints
- Members from a broad range of communities who want to build their own systems
 - They need to know the interfaces to use CMIP5 data and metadata



Two things we MUST get right

- Working together towards **documented interface specifications** as well as software.
- Stable, controlled **software releases** that meet these specifications.



Baseline to Meet a Spectrum of Collaboration Needs

- ESGF should be manifest through the deployment of services which conform to specified **interfaces between organisations**
- The deployment of given software alone is *not* sufficient
- Further, s/w protocols and data formats in themselves are inadequate to define the interfaces required
- They must be constrained within agreed *profiles* to define
 - Exact use
 - Expected behaviour for any given use case
- What is the test of this?
 - It should not be necessary to deploy ESGF software to be part of the federation (but of course software development is encouraged!)
 - From this interface specification alone, I could write compliant software



Interface Specification

- This has worked for the security system:
 - ESGF Security ICD
 - Only example of two independent fully interoperable implementations
- We propose specifications for existing functionality:
 - Logical Data Model: specify datasets, metadata etc.
 - Data Publishing: define a profile for THREDDS catalogues, publishing workflow
 - Metadata Harvesting: production and consumption of Curator RDF and Metafor CIM
 - Registry: data model and service
- ... future functionality:
 - Dataset query service
 - Workflows and service chaining



Open Source, Open Development and Governance

- *Operational deployment* requires an operational mindset
 - What each commits to do *when*
 - How we rely on each other
 - What the fallback position if we fail on these commitments
- Open Source is a key tenant
 - No one institution has the resources to support the rest without shared development
 - **But it's not sufficient for ongoing shared development**
- An open development process is needed
 - Define common interfaces, goals, timescales
 - Allow contributors to assign IPR to the distributor, est. CLIs

