



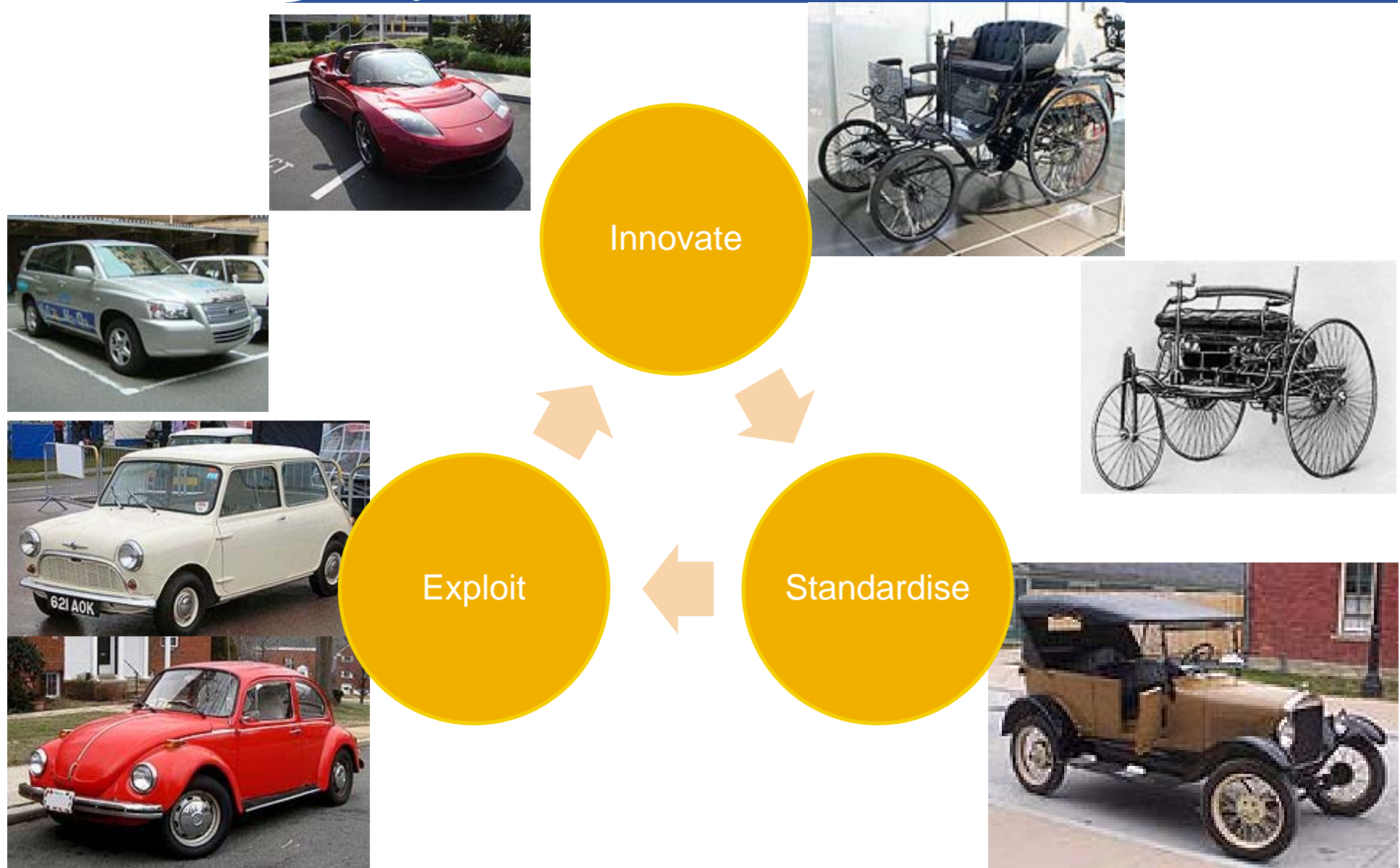
Enabling Grids for E-science

# Defining e-Infrastructure for the masses

*Steven Newhouse*  
*OGF 25, Catania*

[www.eu-egee.org](http://www.eu-egee.org)



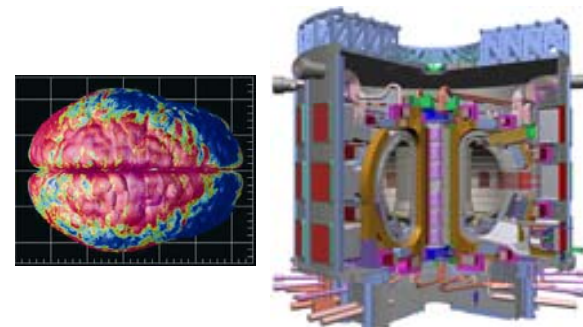


- **Innovate: The last century**
  - Software: Globus, Legion, Condor, ...
  - Experiments: I-Way, meta-computing, ...
- **Standardise: The naughty noughties**
  - Experience: European Data Grid, OSG, NAREGI, PRAGMA
  - Bodies: Global (→ Open) Grid Forum & Enterprise Grid Alliance
  - Specifications: Information, Compute, ...
    - The specification ‘flowers’ that bloomed while others did not...
- **Exploit: The coming terrific teens**
  - Reliability: Having something that works
  - Availability: Using something that’s there
  - Dependability: A foundation for multiple activities

**Enabling Grids for E-science** is the largest multi-disciplinary grid infrastructure in the world. It brings together more than **140 institutions**.

At present, it consists of approximately **300 sites in 50 countries** and gives its **10,000 users access to 80,000 CPU cores** around-the-clock.

The grid infrastructure currently processes up to **300,000 jobs per day** from scientific domains ranging from biomedicine to fusion science.

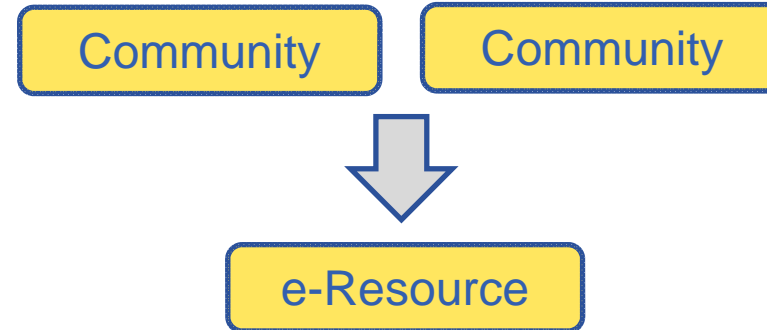


# What have we learned?

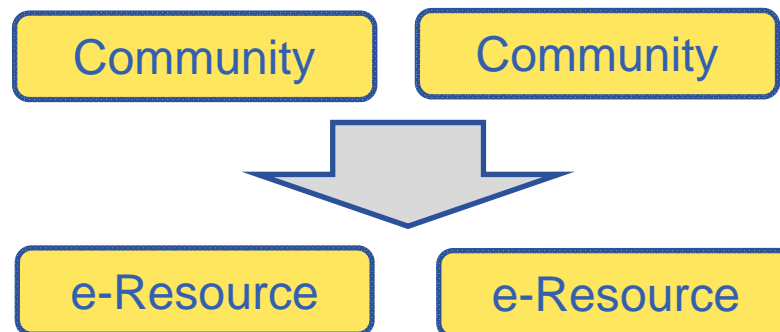
Life used to be simple....



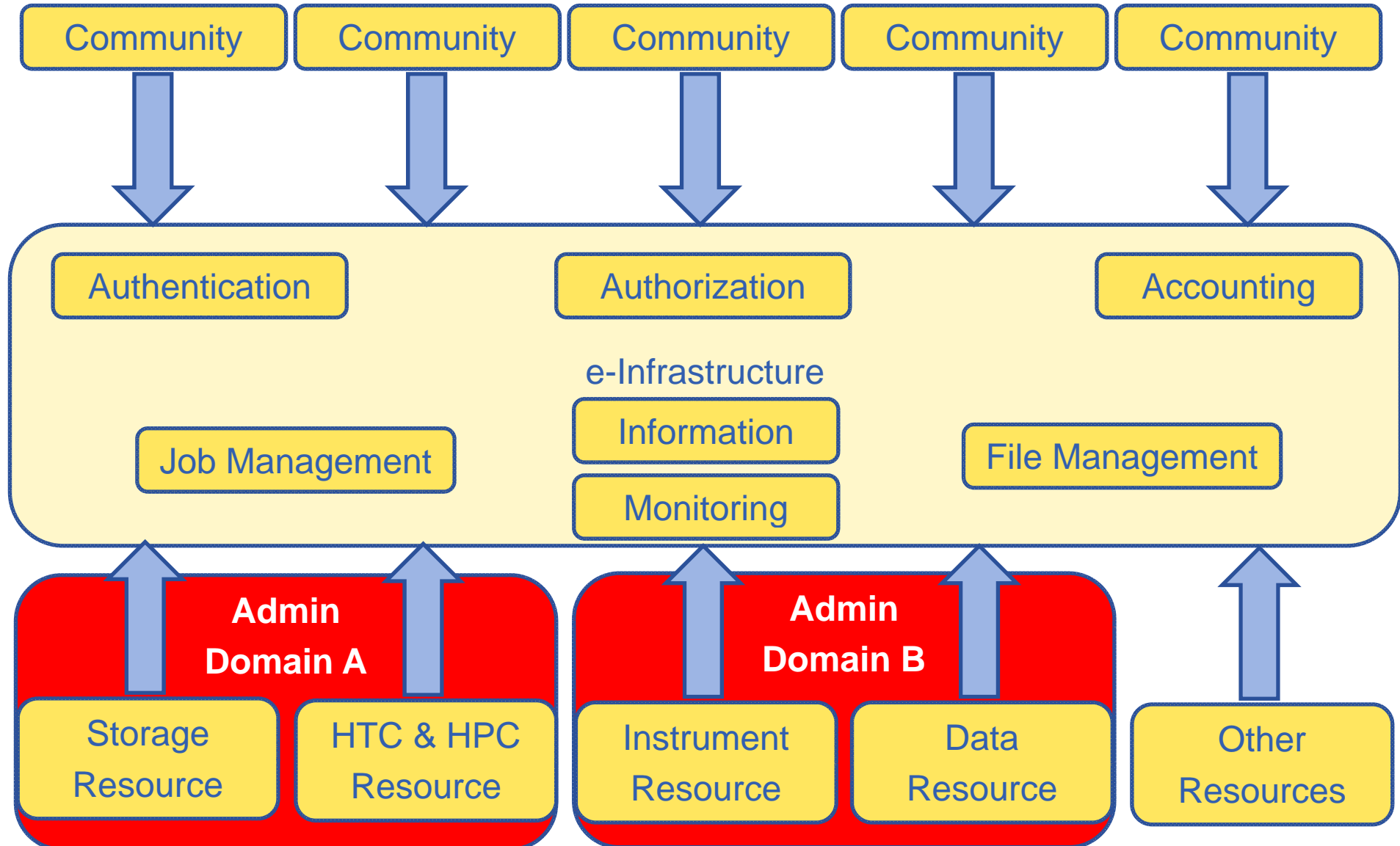
Then it got more complicated....



Then it got **really** complicated....  
shared use of different types of e-resources



# We need some e-infrastructure!



# So why do we need standards?

- **Standards are a fundamental part of our lives**
  - Standard shoe and clothes sizes
  - Mobile phone networks
- **Why should e-infrastructure be any different?**
- **To enable an open world for flexible collaboration**
  - No central control of the infrastructure
  - Portability of applications and software
  - Encompass different perspectives on the 'best' solutions
- **Standards provide a route to independent collaboration**
  - Plugin and swap out components over time
  - A structured open mechanism to evolving common interfaces

- **GLUE: Grid Laboratory Uniform Environment**
  - Started in April 2002
  - Joint activity between EU-DataTAG, US-iVDGL and EDG
    - Focused on interoperability between US and EU HEP projects
  - Aimed to provide common schema to facilitate interoperations
  - v1.0 released Nov 2002
- **Work within the Open Grid Forum**
  - October 2006: First discussion of moving work into OGF
  - January 2007: Working group created and first meeting held
  - June 2008: Specification entered public comment
  - **March 2009: GLUE 2.0 approved as proposed recommendation**



- **Support from many Organisations & Projects**
  - EGEE, ARC, UNICORE, Platform, OSG, APAC, NGS, NAREGI, OMII-Europe, OGF-Europe
- **GLUE co-chairs**
  - Balazs Konya (Lund University)
  - Sergio Andreozzi (INFN)
  - Laurence Field (CERN)
- **Core GLUE Attendees:**
  - Stephen Burke (RAL), Maarten Litmaath (CERN)
  - Paul Millar (DESY), JP Navarro (ANL)

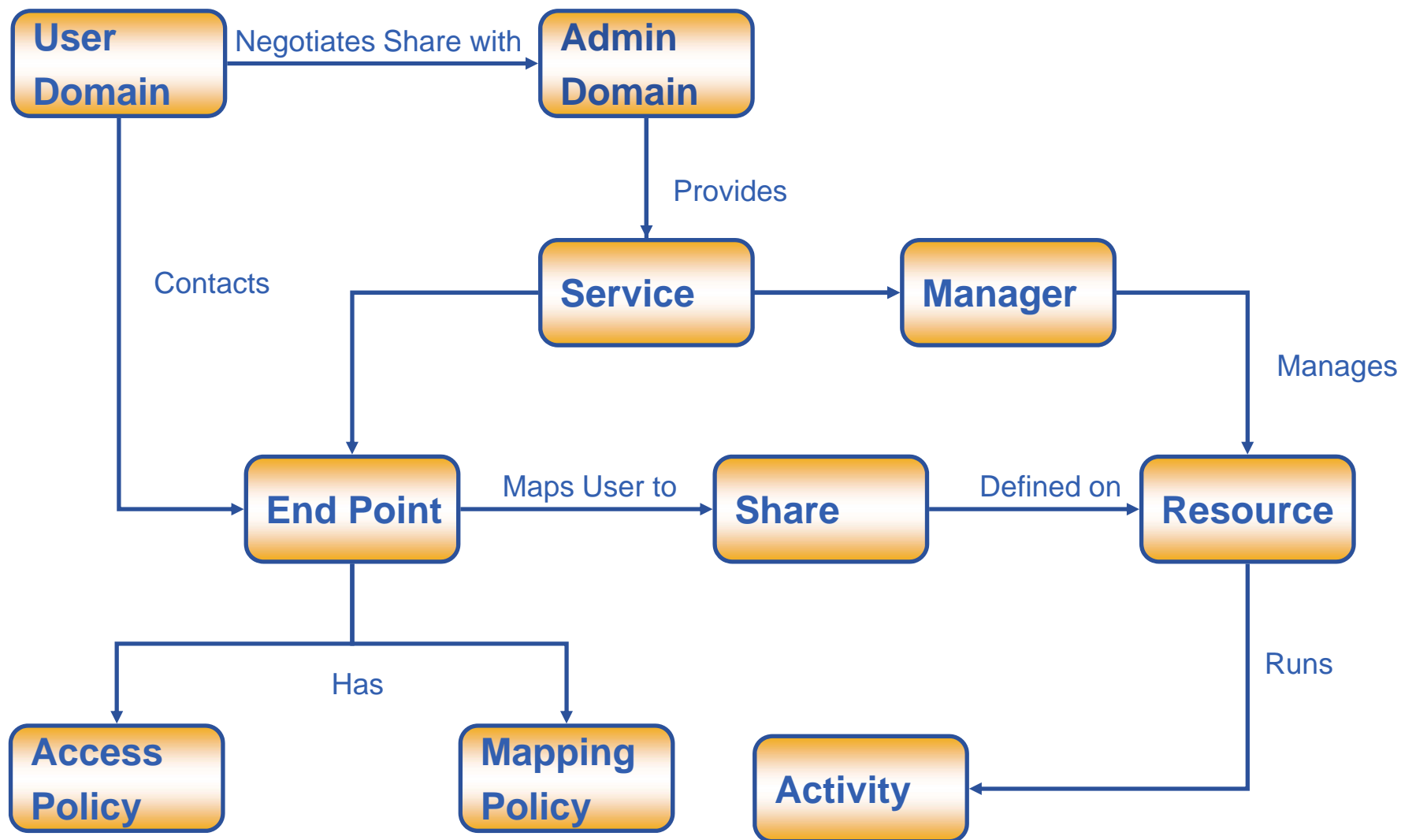


Sergio  
Andreozzi

### GLUE 2.0: Getting To Public Comment

- 347 days to produce the initial specification
- 45 phone conferences (~ 3 days)
- Core team of 5 people (~ 2 months FTE)
- 40 versions of the document (~ weekly updates)

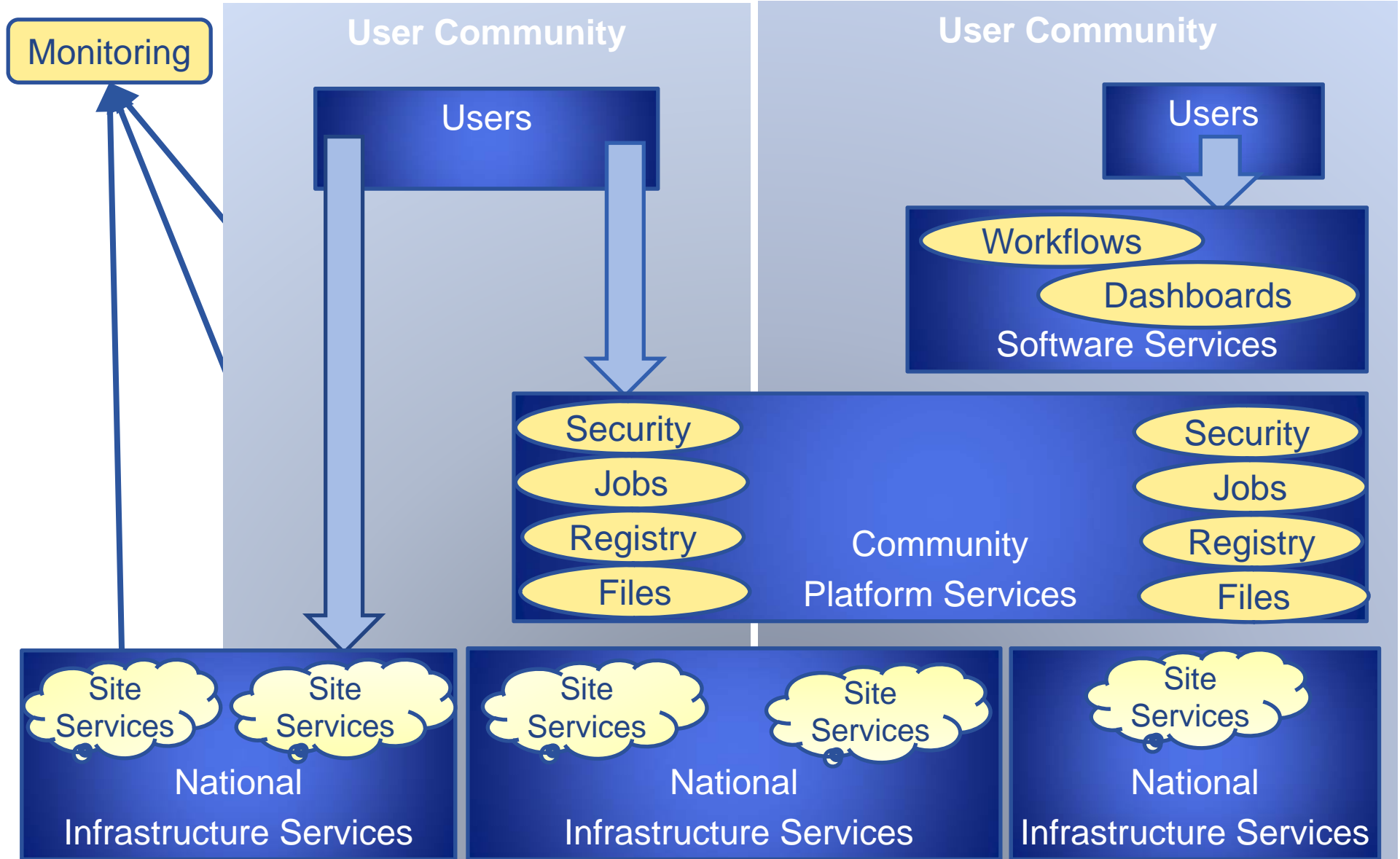
- ***Resources are provided and managed by Administrative Domains.***
  - Universities, research institutes and private companies are all examples of *Administrative Domains* that provide *Resources*.
- ***User Domains represent users who which to utilize a Resource.***
  - Individuals, projects, communities and Virtual Organizations are examples of *User Domains*.
- ***A User Domain negotiates with an Administration Domain to gain access to the Resource***
  - This may result in a *Service Level Agreement*.
  - Members of the *User Domain* are granted access and usage of a *Share of the Resource*.
  - May be implemented through a *Management* entity which manages the *Resource*.



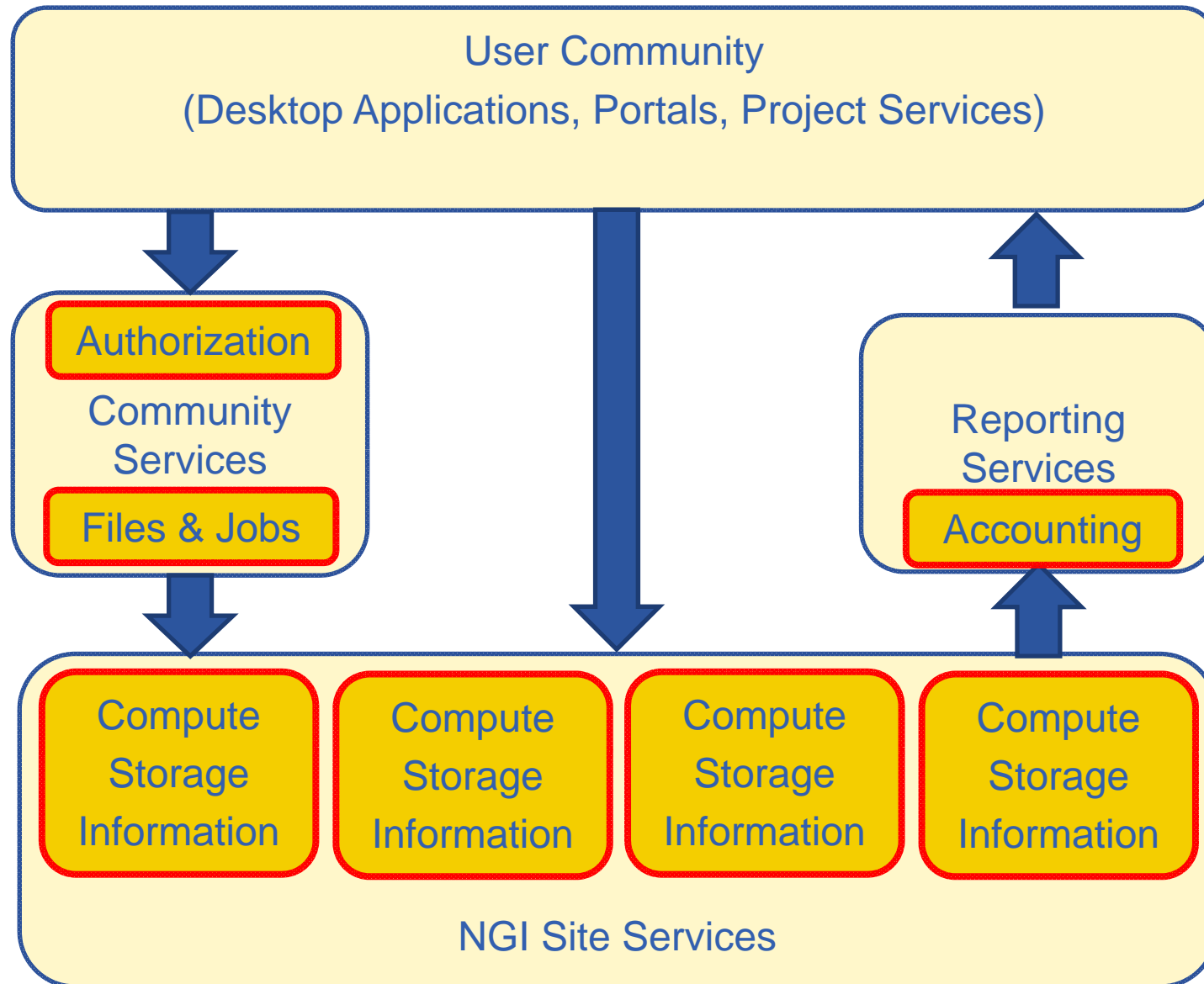
- **Defines an extensible abstract usage model**
- **Concrete usage models in the specification:**
  - Compute resources
  - Storage resources
- ***User Domains* access resources via a *Service End Point*.**
  - This *End Point* requires an *Access Policy* and *Mapping Policy* to ensure the *User Domain* is mapped to the correct *Share*.
- **The *User Domain* is then able to run an *Activity* on the Resource which is accounted to correct *Share*.**

- **Increasing focus on interoperability**
  - Driven implicitly by the user community
  - Their need to use resources within different e-Infrastructures
    - EGEE with Open Science Grid and the Nordic Data Grid Facility
- **Capturing our intellectual knowledge**
  - A lot of time and effort has been invested in middleware
  - Standards are one approach to expressing know how
  - Provides a return on the EU investment
  - Knowledge transfer to EU business community
- **A middleware exit strategy**
  - Define its functional behaviour – the standard interface
  - Define its non-functional behaviour – performance characteristics

- **Opening up middleware provision to other providers**
  - Work with software providers in the EU, worldwide & industry
- **NGIs have subsidiarity in the software they deploy**
  - EGI provides a solution that works across common environments
    - NGIs are free to substitute alternatives
    - Comply with same interface & minimal performance capabilities
- **Emerging need to independently assess software**
  - One of the original goals from the OMII-Europe project
  - Is this a role for OGF-Europe to move into in?



# Where do we need standards?





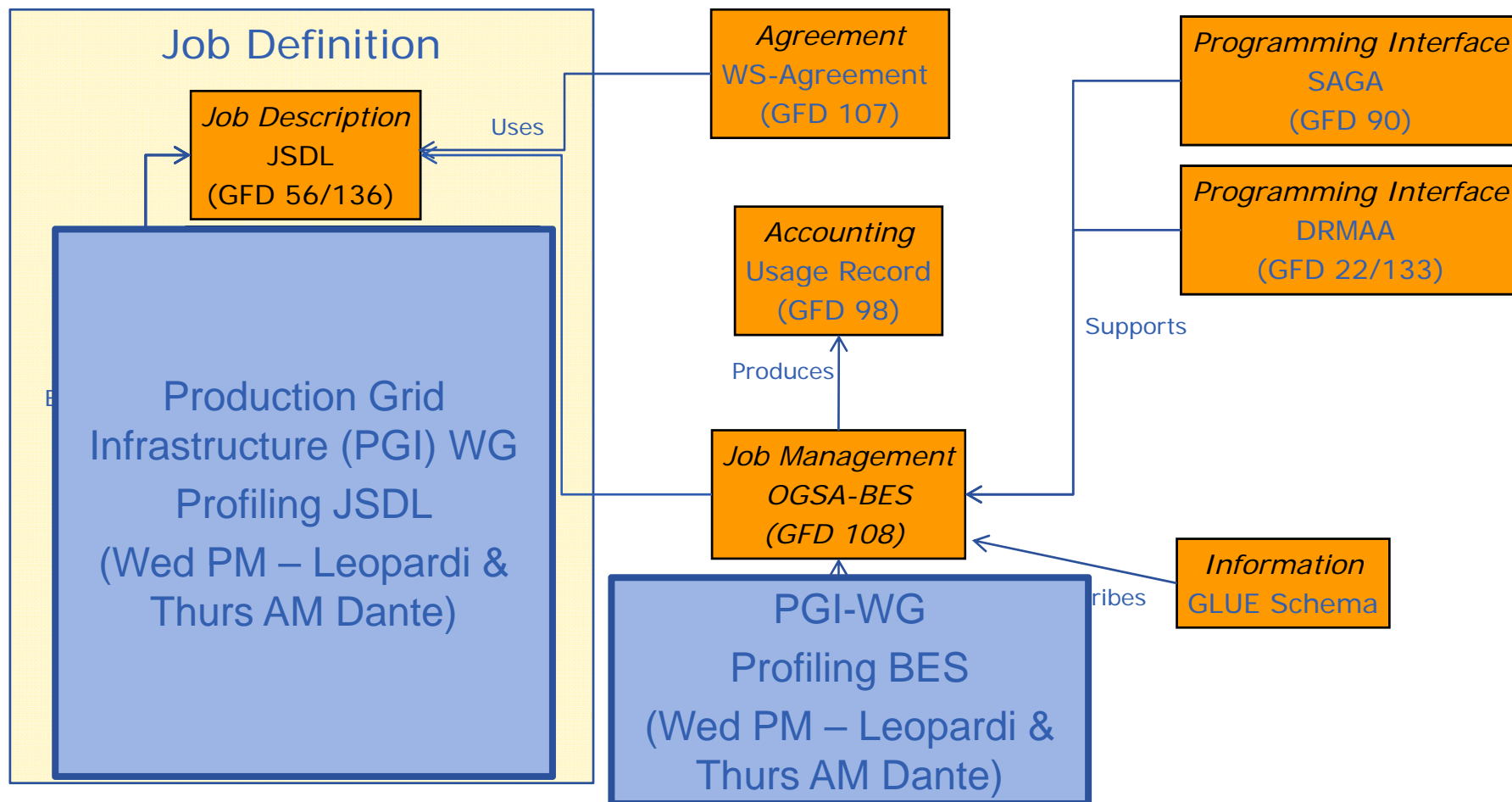
- **Authorization Services:**
  - X.509 Certificates, SAML, XACML
- **Information & Reporting Services:**
  - Grid Laboratory Uniform Environment (GLUE)
  - Usage Records (Describing Accounting Activity)
  - *Resource Usage Service (Exporting Accounting Records)\**
- **Compute Services:**
  - Basic Execution Services (Job Submission Interface)
  - Job Submission Description Language
- **Storage Services**
  - GridFTP (Data Transport) & Storage Resource Management
  - *Database Access Integration Services (DAIS-WG)\**
  - *Data Movement Interface (File Transfer)\**

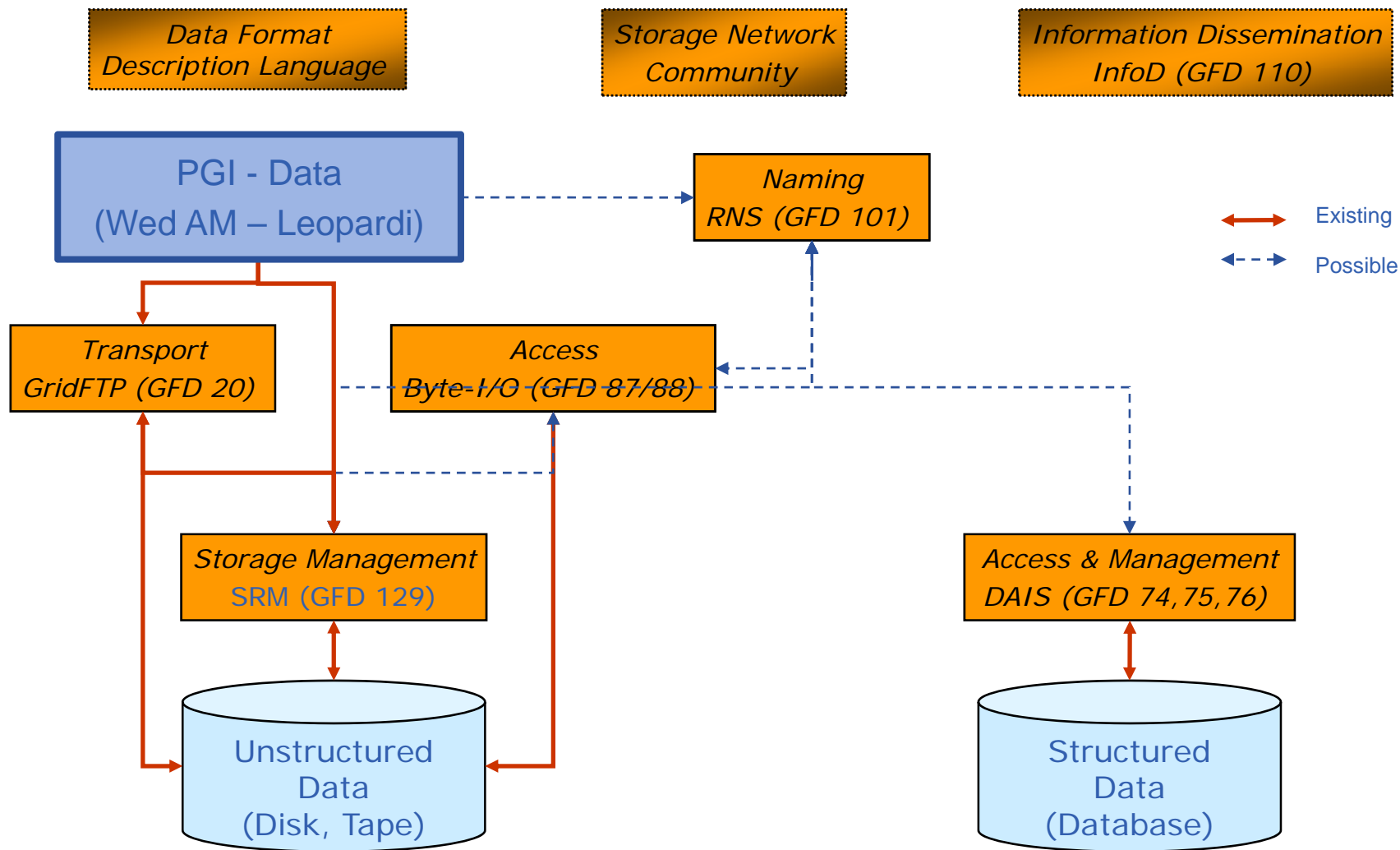
**\* Specifications potentially relevant to EGEE**

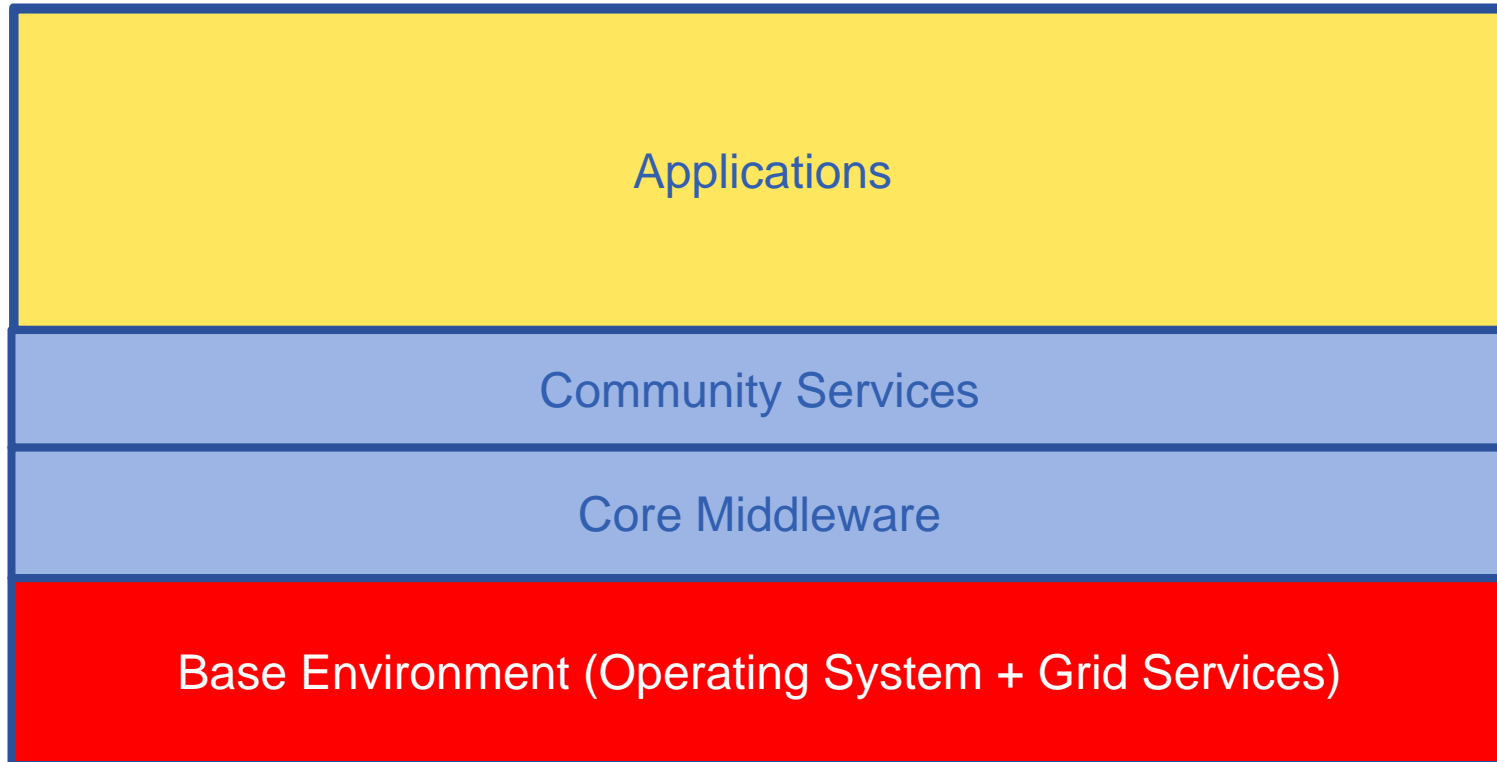
*Architecture*  
OGSA EMS Scenarios  
(GFD 106)

*Use Cases*  
Grid Scheduling Use Cases  
(GFD 64)

*Education*  
ISV Primer  
(GFD 141)







- **The first innovation phase of e-infrastructure is over**
  - The EDG and EGEE series of projects are coming to an end
- **Moving through the standardisation phase**
  - But ‘the only colour you can have is **black**’!
- **EGI is start of the e-infrastructure exploitation phase**
  - A defined minimal foundation that can be built upon
  - Enable further innovation, standardisation & exploitation
- **Generating collaborative specifications is **hard** work**
  - OGF adds negligible overhead to this collaboration
- **OGF’s portfolio of standards is continuing to expand**
  - EGEE’s and its collaborators are driving those critical to us!
  - Resulting specifications have greater value than having been developed in private

- The GLUE team & Laurence Field
- Those working on standards in EGEE (& other projects)
- The e-infrastructure user community within OGF
- Further standards related information:
  - <http://www.ogf.org/standards>
    - Overview of the OGF standards area
  - <http://www.ogf.org/>
    - Follow the documents link for the detailed technical specifications
  - <http://www.gridtalk.org/>
    - Jargon free 'Grid Briefings' available under Documents