Enabling Grids for
E-science in Europe

EGEE Infrastructure: How to Join as Users, Sites and Virtual Organisations

Steve Traylen
Rutherford Appleton Lab, UK
s.traylen@rl.ac.uk

Talk Outline

Enabling Grids for E-science in Europe

- Joining EGEE as a User.
  - The practice following on from the theory from earlier talks.
  - Obtaining a certificate and joining a VO.
- SA1 (EGEE Deployment) Structure.
- Joining EGEE as a Virtual Organisation.
  - Who to contact and what is needed.
- Joining EGEE as a Site.
  - Who to contact and what is needed.
- Support Structures.
- Thoughts on this talk.

Joining as a User: Find a CA

Enabling Grids for E-science in Europe

- Obtain a signed X509 key pair from an appropriate and approved certificate authority (CA).
- EGEE will maintain an approved CA list.
- Quality of CAs within EGEE is organised and maintained by the www.EUGridPMA.org group.
- The French CNRS CA can act as a catch all for users without a suitable CA.
- See also http://lcg-registrar.cern.ch/pki_certificates.html for current approved LCG CAs.

Joining as a User: Obtain a Certificate

Enabling Grids for E-science in Europe

- Certificate generation varies from CA to CA but usually one of:
  - openssl on its own is one option, e.g. CESNET CA.
  - grid-cert-request script from Globus, e.g. Cyprus CA.
  - Web browser based , e.g. GILDA or UK E-Science CA
  - Live KCA: Kerberos to short lived x509 credential.

Joining as a User: Obtain a Certificate

Enabling Grids for E-science in Europe

- Most CAs provide their own instructions for doing this.
- UK E-Science
  - **http://www.grid-support.ac.uk/ca/documentation.htm**
- Cyprus CA
  - **http://www.cs.ucy.ac.cy/cygrid-ca/cyca-howto.html**
- All other CAs have something similar.
- Now with a certificate you exist on the grid.

Joining as a User: Joining a VO

Enabling Grids for E-science in Europe

- A VO is just a list of users (DNs) maintained by the VO manager.
- This list is served as an LDAP or possibly HTTP service.
- Different VOs are populated by different methods.
  - Web form with client side X509 authentication, eg LCG, GILDA.
  - Any VOMS based VO can be web based but the existing API exists so could be anything.
  - Command line tools e.g. the now ageing EDG ones, earth observation.
  - Gridsite can manage VOs, a light weight solution, e.g.GridPP, MICE
  - Others, BaBar is just SLAC accounts replicated.
- In the medium future VOMS allows this to progress beyond just a list.
  - Users have roles, e.g. a production manager could examine jobs belonging to a VO jobs.
  - VOMS will be EGEE 1, VOMS aware software is ongoing.

Joining as a User: Joining a VO(2)

Enabling Grids for E-science in Europe

- VOs managers will normally require that you complete some membership details when joining.
  - Contact details such as a phone number and your home institution.
  - Sign something such as an acceptable use policy.
- This will say that will not abuse any computers, do anything illegal, etc.
- See LHC's VOs statement at http://lcg-registrar.cern.ch/
- These are important and allow sites to trust a VO.
  - VOs are not legal entities and cannot be sued by sites.
  - Sites must be able to trace individual data or jobs to a user.

Joining as a User: Access to a UI

Enabling Grids for E-science in Europe

- Finally you need access to the client software on a UI.
- This could be your laptop or desktop if you install the UI software.
- Tier1 centres often run a public access machine to allow people to start.
- GENIUS or some other application specific web interface such as the CMS Monte Carlo web interface can be used.
- Custom applications written against the APIs for grid submission such as LHCb's and Atlas's GANGA application.
- Asking your local sysadmin to install the UI across your department is excellent…. e.g. lxplus.cern.ch now has this.

Operations Infrastructure

Enabling Grids for E-science in Europe

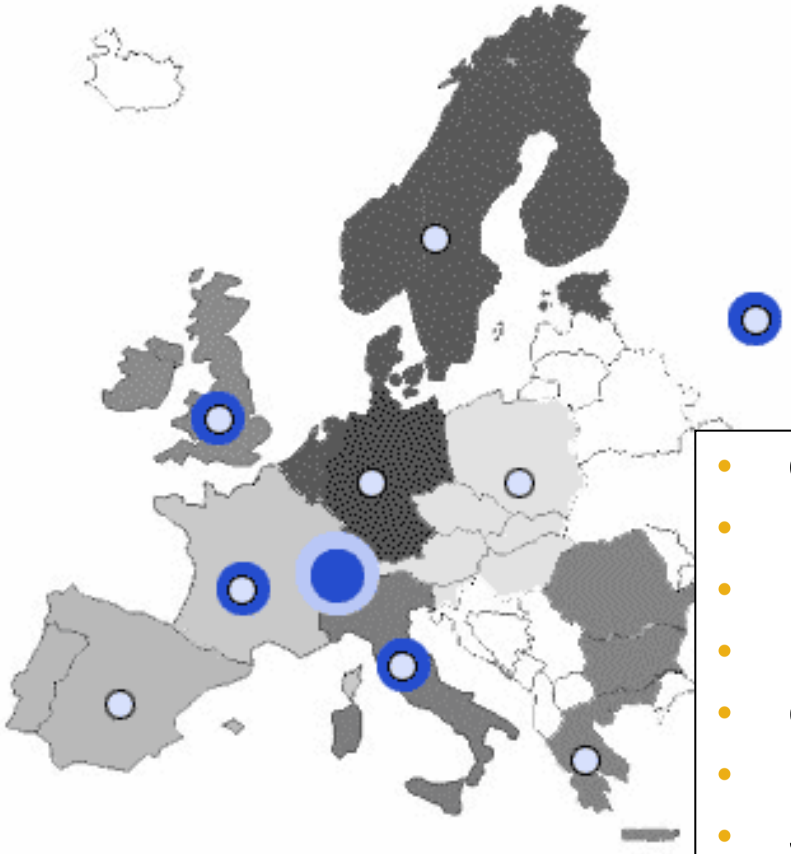
- The SA1 activity within EGEE is responsible for deployment.
- SA1 represents half of the effort within EGEE.
- Deployment is broken into three levels.

1. Operations Management Centre. Only one at CERN.
   - Runs grid level services, e.g. Middleware software repository.
   - Defines contents of the EGEE production release.
2. Core Infrastructure Centre.
   - Runs VO level software, e.g. VO or VOMS servers.
3. Regional Operation Centre.
   - Supports the operation of sites within their region.

- In reality there is overlap between all of these.

Operations Infrastructure

Enabling Grids for E-science in Europe



Operations Management Centre

Core Infrastructure Centre

Regional Operations Centre

- CERN (OMC, CIC)
- UK+Ireland (CIC,ROC)
- France (CIC, ROC)
- Italy (CIC, ROC)
- Germany+Switzerland (ROC)
- Northern Europe (ROC)
- South West Europe (ROC)
- South East Europe (ROC)
- Central Europe (ROC)
- Russia (CIC – M12, ROC)

- 48 Partners involved in SA1
- ROC's in several regions are distributed across many sites

From Ian Bird.

Joining EGEE as a VO

Enabling Grids for E-science in Europe

- Procedure is defined but has never happened yet.
- For any Virtual Organisation:
  - A VO Manager is a requirement.
  - A VO Server is a requirement.
  - An RLS instance or perhaps some other VO specific Middeware, eg Atlas's Magnor file catalogue, BaBar's SP lookup or CMS's using the Storage Resource Broker.
- All contact for a VO should be with their local Core Infrastructure Centre (CIC).
- NA4 will the be helping new VOs through the whole process.

Joining EGEE as a VO: Procedure

Enabling Grids for E-science in Europe

- New application will nominate a VO manager.
- VO must find a CIC to run VO server and an RLS server if an RLS is required.
- Discussion will follow to understand demands and supplies for VO. (Funding, Political Constraints,…..)
- For middleware it must be understood what implication this has on the system at large.
- VO will enter a registration process to check how the VO operates, e.g. do the users make any agreements when they join.

Joining EGEE as a VO: Procedure

Enabling Grids for E-science in Europe

- Sites configure support for VO if and only if they want to.
- CIC or VO may themselves run extra RBs or UIs if required.
- The VO can now start running jobs.
- Software installation for a VO is still unresolved. At a educated guess.
  - It will likely be shipped around by the replica manager.
  - It will then be available as a cacheable copy.
  - Your software will be everywhere in effect.
  - VOs themselves can publish values into the information system.
    - This allows them to mark sites as good or bad.

Joining EGEE as a Site

Enabling Grids for E-science in Europe

- Within EGEE production and pre-production options exist.
  - Production suitable for large resources.
    LCG maps onto this testbed.
  - Pre-Production suitable for small dynamic resources with interest in new middleware.
    This will absorb any remnants of the EDG testbed.
- New sites will operate under the guidance of their local ROC.
  - The ROC will provide support, initial testing and validation.
  - The ROC will collect security contact information.

Physical Requirements for a Site

Enabling Grids for E-science in Europe

- Service Nodes.
  - A gatekeeper node is required to your present batch system as a computing element.
  - A storage element node to serve disk space or mass storage.
  - In the not so distant future an extra “Service Node”. Largely a tomcat server with servlets is likely to appear.
    - R-GMA servlets
    - Local Replica Catalogue.
  - Large sites may run extra core services such as resource brokers, user interfaces.

Special Requirements for Sites

Enabling Grids for E-science in Europe

- Support for Specialised Batch System
  - Requires job manager for the gatekeeper to exist. PBS, LSF and SunGrid Engine exist already.
  - Information providers to publish to the GLUE schema required. Only PBS and LSF exist in any maturity.
  - With accounting arriving then filtering your batch logs may also be required.
- Specialised Storage System
  - LCG and EGEE have converged on Storage Resource Manager as **the** interface to storage.
  - To add mass storage an SRM interface is needed. HPSS, Castor, EnStore exist now but for new ones such as DMF this is non trivial.

Joining as a Site Today.

Enabling Grids for E-science in Europe

- The EGEE production facility can not be joined today or for some months yet but joining LCG is possible and encouraged.
- Predominantly for LHC and a few HEP experiments extra VOs could exist.
- Installation is very well documented and has been completed now by many sites.
  - LCFGng installation is well supported.
  - A manual documented installation is supported and is in use successfully by some sites.
  - <<Add URL>>
- To join LCG today contacting your local Tier1 centre is the way proceed, this could be your ROC.

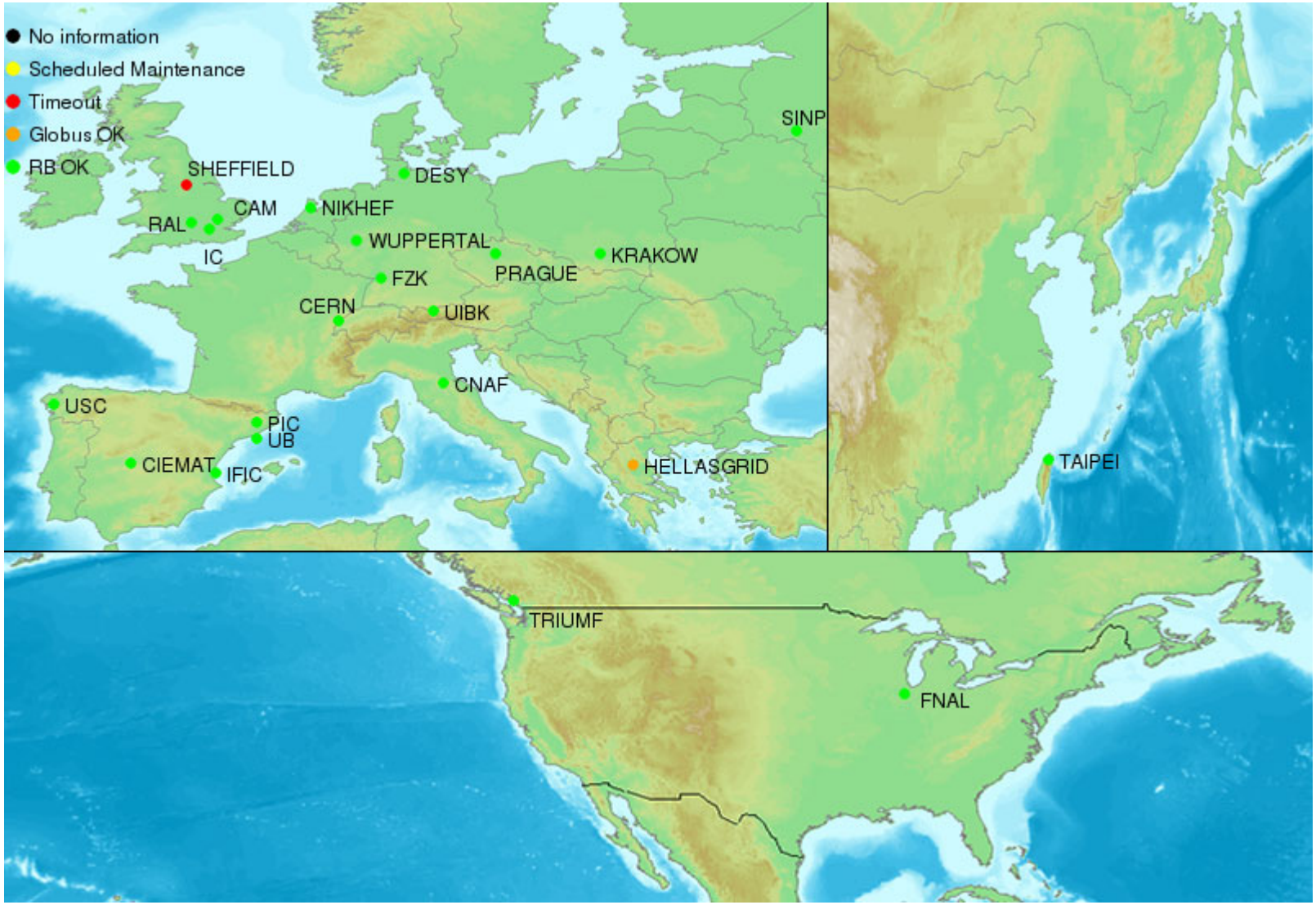
Grid Operations Centre

Enabling Grids for E-science in Europe

- The GOC exists to maintain the smooth running of the GRID.
- It maintains a database of sites, resources and site contacts.
- It performs monitoring of sites.
- It has the power to add and remove any problem sites.
- Sites initially appear in a test zone before being moved to a core zone. The GOC can move sites backwards and forwards.
- End users can choose the view they want to see.

Current State of LCG Test Zone

Enabling Grids for E-science in Europe



Support

Enabling Grids for E-science in Europe

- First line informal support for:
  - users within the virtual organisation.
  - sites is available from their local ROC.
  - virtual organisations will be from their supporting CIC.
- Formal support will be from the global grid user support system. (Today it is not clear to me if this is a good idea.)
  - http://gus.fzk.de/
- Longer term problems, requests or requirements that require fixing will be registered with Savanah and then passed on to a developer or whoever.
  - http://savanah.cern.ch/

Discussion and Thoughts on this Talk.

Enabling Grids for E-science in Europe

- A case study for a site and VO once there is one.
- In reality a lot of this formal procedure does not exist or will change.
- Some overlap with concerning addition of VOs and the NA4 talk yesterday.
- Much of the existing/imminent middleware mentioned deserves a slide elsewhere, e.g. VOMS, SRM and GFAL.