User communities of NDGF: specifics and requirements

Oxana Smirnova

EGI Technical Forum

Lyon, September 21, 2011

NDGF

Nordic DataGrid Facility

NDGF
Nordic DataGrid Facility

NDGF users

- NDGF was initiated by users – the LHC groups
  - Key VOs: ATLAS, ALICE, CMS
  - Hardware is contributed by these communities
    - Installed at different HPC centres
- Frequent suggestion from funding bodies: involve researchers outside LHC
  - They are often (much) smaller communities without own hardware
    - join forces (also with LHC) in hardware bids
    - apply for time on national Grid-enabled HPC resources

- NDGF is an <u>international</u> distributed facility by design
  - Each country/node has a great degree of autonomy w.r.t. allocating resources to users outside LHC

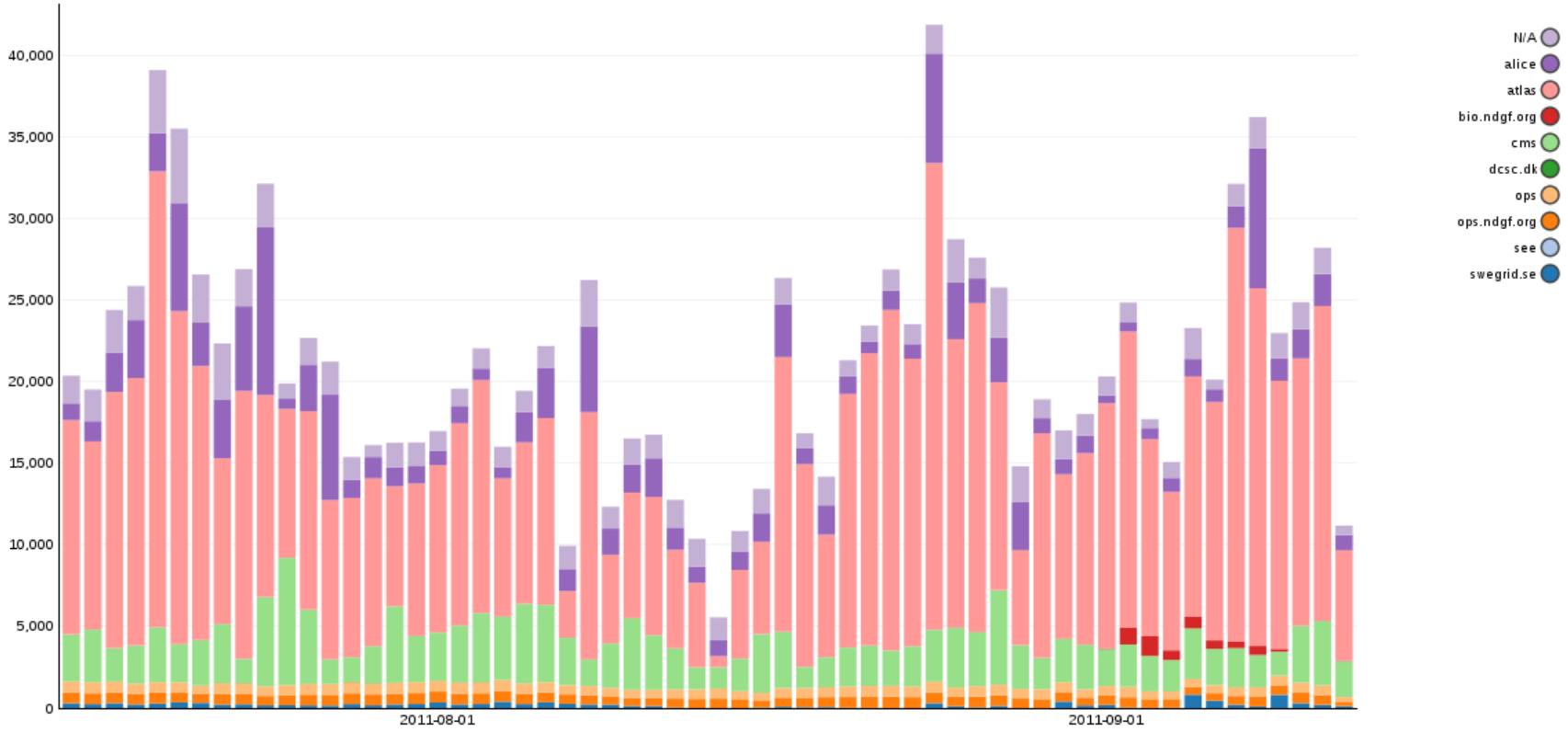
NDGF
Nordic DataGrid Facility

Existing user communities

- LHC (ATLAS, ALICE, CMS)
- Computer science
- Astroparticle physics (ICECube)
- Bioinformatics (BioGrid, ELIXIR)
- Environmental sciences (CO2 sequestration)
- Computational Chemistry
- Material Sciences

NDGF jobs per VO

- Clearly dominated by LHC (specifically, ATLAS)



New users

NDGF is open to new users ever since it was created in 2006

Potential new communities:

- More bioinformatics (also proteomics, genomics)
- More material sciences
- More computational chemistry
- Radioastronomy? Many close contacts, but so far they prefer not to share resources

Challenges:

- NDGF has no “free” resources to offer – only know-how
- Non-LHC user communities often have no massive resources to contribute
- After all, NDGF is LHC-tuned

NDGF
Nordic DataGrid Facility

Technical challenges

Each user community comes with own middleware

- ALICE – AliEn, ATLAS – Panda, BIO – Taverna, etc
- NDGF relies on ARC and dCache, but has also to deploy LFC, FTS, VOMS, site-BDII, VOBoxes etc to satisfy users

Difficult to streamline

- Users and sysadmins alike do not like to change tools without an extremely good reason
- In HPC, sites/queues are tailored for applications and other way around – on the Grid, all sites are different, and applications can not be tailored

Difficult to optimize load

- LHC – nearly constant (though fluctuating) load
- Most other communities – episodic activities

NDGF
Nordic DataGrid Facility

Middleware requirements

- Generic middleware must not be tied to a single operating system
- Generic middleware must not have “vendor locks” w.r.t. third party dependencies
- Generic middleware must follow worldwide community practices and standards (FSH, licensing, packaging, monitoring, configuration, auditing etc)
- Generic middleware must provide API and development libraries to build application-specific tools and utilities
- Middleware shortcomings should not be getting solved by operational means
- Documentation, documentation, documentation

Operations

Now

- Operations team comprising of
  - Systems integrator – primarily general WLCG support
  - Dedicated ALICE/CMS/ATLAS support experts
  - Storage specialist – specific WLCG storage requirements
- Operator on Duty (OoD) team, one country on a 4 week rotation basis
- OoD handles EGI ROD duties and WLCG monitoring support

Future challenges

- More user communities – more dedicated experts needed – more organisational complexity?
- More people are trying ARC
  - operational procedures still are often gLite-specific (YAIM this, APEL that etc)
  - more ARC support specialists may be needed

Sustainability thoughts

- NDGF is a strictly research facility
  - Much like NORDUnet or other NRENs
  - Very little EU-funded effort (below 10%)
- Sustainability relies on ever increasing need for computing and storage capacities in academic research
  - Public funding
  - Business model: states invest as much as they can, researchers use as much as they need
    - Works as long as we are much cheaper than commercial providers

- Cloud technologies look appealing
  - If we can offer services comparable to commercial providers for smaller costs
    - How?

Summary

- LHC community is by far the largest user, and will remain such in foreseeable future
- If we want to attract new communities, both middleware and operations must become less LHC-specific (obviously)
  - Reduce overhead for very small VOs
  - Learn how to share resources
  - Accommodate more different use cases and scenarios
  - Get rid of single-OS dependency and other in-house specifics

- NDGF is an international effort by several HPC centres
  - Experience is good, as long as coupling is sufficiently loose
  - Key benefit: we share experts, we are a community