Nordic Data Grid Facility
NDGF – www.ndgf.org

What – Why – How

Paula Eerola, paula.eerola@hep.lu.se
1st Iberian Grid Infrastructure Conference
Santiago de Compostela, May 14-16, 2007

NDGF
Nordic DataGrid Facility

Nordic Countries

- Common history and culture
- Similar societies
- Long and productive tradition of co-operation (e.g. free mobility long before EU)

Vision of the NDGF

- The vision of the Nordic Data Grid Facility (NDGF) is to establish and operate **a Nordic computing infrastructure** providing researchers in the Nordic countries with a seamless access to computers, storage and scientific instruments.
- NDGF **coordinates** activities - NDGF does not own resources or middleware
- NDGF is the interface to e-Science in the Nordic countries.
  - Single Point of Entry for collaboration, middleware development, deployment, and e-Science projects
  - Represents the Nordic Grid community internationally

History of Nordic Grid Collaboration

- Collaboration initiated 2001-2003 (initial partners universities of Copenhagen (DK), Lund, Uppsala (SE), Oslo (NO), Helsinki (FI))
  - "Nordic Testbed for Wide Area Computing and Data Handling" → **NorduGrid**
  - Funding by Nordunet2 R&D grant
  - It became soon evident that the existing middleware prototypes at that time (2001-2002) were not suitable and/or mature enough → **development of own middleware, ARC** (Advanced Resource Connector) **necessary for getting the Testbed operational**
  - Testbed in operation since August 2002
  - Successful participation in Particle Physics (LHC-ATLAS experiment) Data Challenges already 2004
  - Forming of the NorduGrid middleware consortium (2005)
- NDGF Pilot Project 2004-2005 (joint Nordic funding)
  - A pilot production facility based on national resources
- NDGF Production Phase 2006-2010

NDGF basic facts

- NDGF – production phase – start 1.6.2006
- Co-owned by: DK, FI, NO, SE
- Hosted by NORDUnet A/S (Nordic IP backbobe network, fully owned by the Nordic Research Councils)
- Funding shared between the 4 Nordic countries
- Funding (2 MEUR/year) by National Research Councils of the Nordic countries
- Status: organization in place, integration of the national resources ongoing

NDGF Organization

Top management structure, administration and office shared with NORDUnet

CEO

CTO

CERN Coordinator

Technical Coordinator

Software Coordinator

System Integrator

Software Developer

National Node Coordinator

NDGF Board of Directors

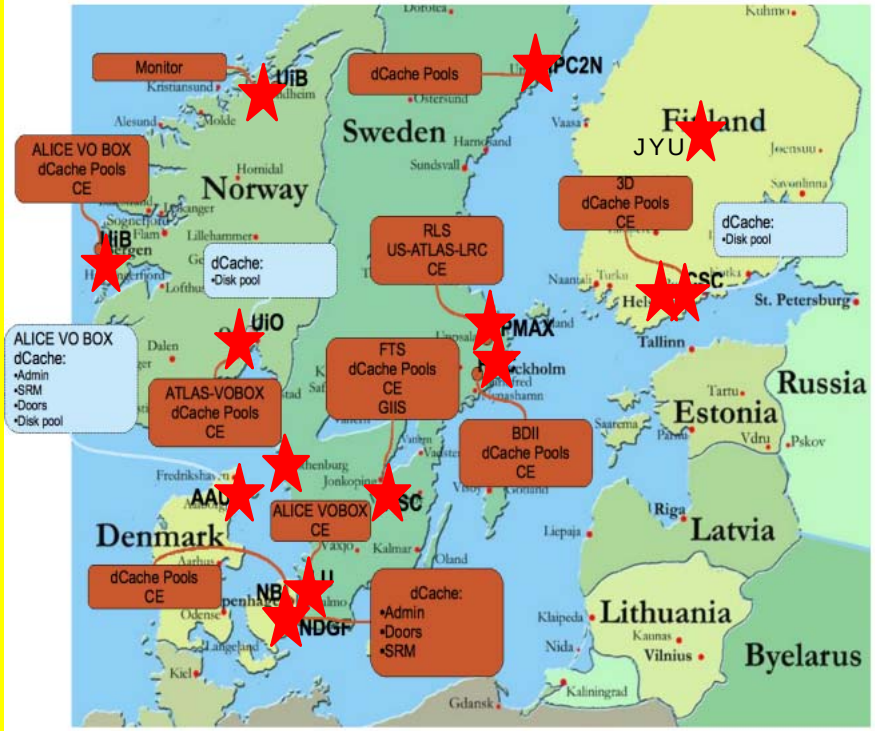
NDGF subcommittees: CERN Committee, ...

Key technical and project coordination functions

Currently 13+2 persons

NDGF Technical Platform

- NDGF is Nordic Production Grid giving access to national resources
- National resources are purchased and owned nationally
  - NDGF creates a single Grid from **existing resources (supercomputers, clusters, storage) in participating countries**
  - Provides grid project storage
- Critical issues:
  - Planning of available resources is more complicated than with a single-site centre
  - Funding of new resources (different national policies, funding sources, time scales), coordination of funding and purchases
  - Integration of existing resources, user policies



National resources *(for academic use)*

- SWEDEN: Major resources at **6 National Computer Centers** under a virtual metacentre, **SNIC**. **Swegrid** is the Swedish national Grid.
- NORWAY: Major resources coordinated by **UNINETT Sigma** metacentre. **NORGRID** is responsible for establishing and maintaining the national grid infrastructure. **eVITA** e-Science programme recently established.
- FINLAND: Major resources at **CSC**, Center for Scientific Computing. Material Sciences National Grid Infrastructure (**M-grid**), joint project between CSC and universities.
- DENMARK: Major resources coordinated by Danish Center for Grid Computing **DCGC**.

NDGF tasks –
what does it mean to ”coordinate”?

- **Facilitate, coordinate & host major e-Science projects** (e.g. Nordic LHC Tier-1 for Particle Physics Research)
- Operate Nordic storage facility for major projects
- Create a common policy framework for the Nordic Production Grid
- Facilitate joint Nordic planning and collaboration
- Contribute to development of ARC middleware
- Other potential future activities include e.g.:
  - Supporting Nordic HPC/Grid conference activities
  - Promotion and outreach of computational sciences in various ways

Supporting Nordic e-Science

- Provides a single point of contact for e-Science initiatives
  - A place to go for researchers initiating projects
  - Supports **all sciences** in need of compute resources
- Services for user groups and applications
  - Portals for user groups
  - Gridifying major applications
- Hosts and coordinates major e-Science projects
  - Coordinate technical implementation
  - Coordinate collaboration

- NDGF represents the Nordic Grid community internationally
  - Allows the Nordic Grid community to speak with one voice
  - Creates visibility for Nordic technology and solutions
  - Creates a single Nordic representation in major e-Science projects
- Single Point of Entry for international partners
  - To reach the Nordic Grid community, talk to us
  - Point of contact for e-Science projects, middleware development, and grid facility deployment

NDGF Users

- NDGF didn't formalize user policies yet, but the plan is to "transfer" the NorduGrid Virtual Organization (VO) to NDGF
- Currently, access to the NorduGrid resources is granted for the following users:
  - Members of the NorduGrid Virtual Organization, i.e. people who are affiliated with one of the Nordic academic institutions and have agreed to the Acceptable Use Policy document.
  - Members of other Virtual Organizations, who are accepted by the NorduGrid Steering Committee. Decision of acceptance is taken on case-by-case basis.
  - Users: physics, chemistry, computer science, bioinformatics, molecular medicine, production technology,...

NDGF Partnerships

- NDGF is a Facilitator and an Integrator, not a resource owner
  - NDGF does not own the network
  - NDGF does not own computing resources
  - NDGF does not own the middleware platform
- Resources are owned and managed by partners
- NDGF Partnerships
  - Network: **NORDUnet**
  - Grid compute resources: **National Grid Projects** (DCSC, SweGrid, NorGrid, M-Grid,...)
  - Other kind of compute resources (supercomputers, shared memory processors,...): **National Computer Centers**
  - Middleware and services: ***NorduGrid, KnowARC, EGEE/EGI,...***
  - Storage: dCache/DESY

NorduGrid Consortium

- Not limited to Nordic Countries any more
- Currently about 20 partners around the world, open to new partners
- The NorduGrid consortium:
  - Controls the ARC middleware
  - Contributes to development of Grid standards, e.g. via the Global Grid Forum.
- Currently 60 Sites, 6000 Processors, 60 TB storage, 1600 Grid users
- http://www.nordugrid.org

Current NorduGrid partners

Middleware

- One of the NDGF Core Tasks: Deliver the middleware for collaborative e-Science projects
  - Contribute to development of middleware(s)
  - Installation, testing and maintenance
  - Interoperability and co-existence of different middlewares
  - Development of tools and services not currently available

- **ARC – the Advanced Resource Connector –** developed for the Nordic distributed and heterogeneus computing and storage structure
- Does not require a specific hardware nor operating system
  - Open Source Grid Middleware platform, free download
  - Source code available for modification and incorporation in new projects
  - Open for international collaboration, contribution of code and ideas, and deployment
  - Interoperability and co-existence of different middlewares, e.g. ARC-LCG/gLITE
- A number of collaborative projects: e.g. KnowARC (EU-funded), New and Innovative Services for NorduGrid (Nordic-funded),...

NDGF
Nordic DataGrid Facility

e-Science case –

Particle Physics Research: NDGF Tier-1 for Large Hadron Collider (LHC) data processing and storage

LHC Computing Hierarchy

Experiment

~100-1500 MBytes/s

Tier 0

CERN Center
PBs of Disk;
Tape Robot

Tier 1

IN2P3 Center

NDGF

Tier 2

Tier2 Center

Center

nter

Center

Center

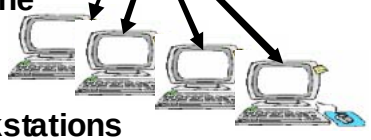
Institute

itute

stitute

Institute

Physics data cache



Workstations

Tier 0= CERN. Tier 0 receives raw data from the Experiments and records them on permanent mass storage. 5-8 PetaBytes of data/year, can grow up to 100 PB/year. First-pass reconstruction of the data, producing summary data.

Tier 1 Centres = large computer centres (12). Tier 1's provide permanent storage and management of *raw*, *summary* and other data needed during the analysis process.

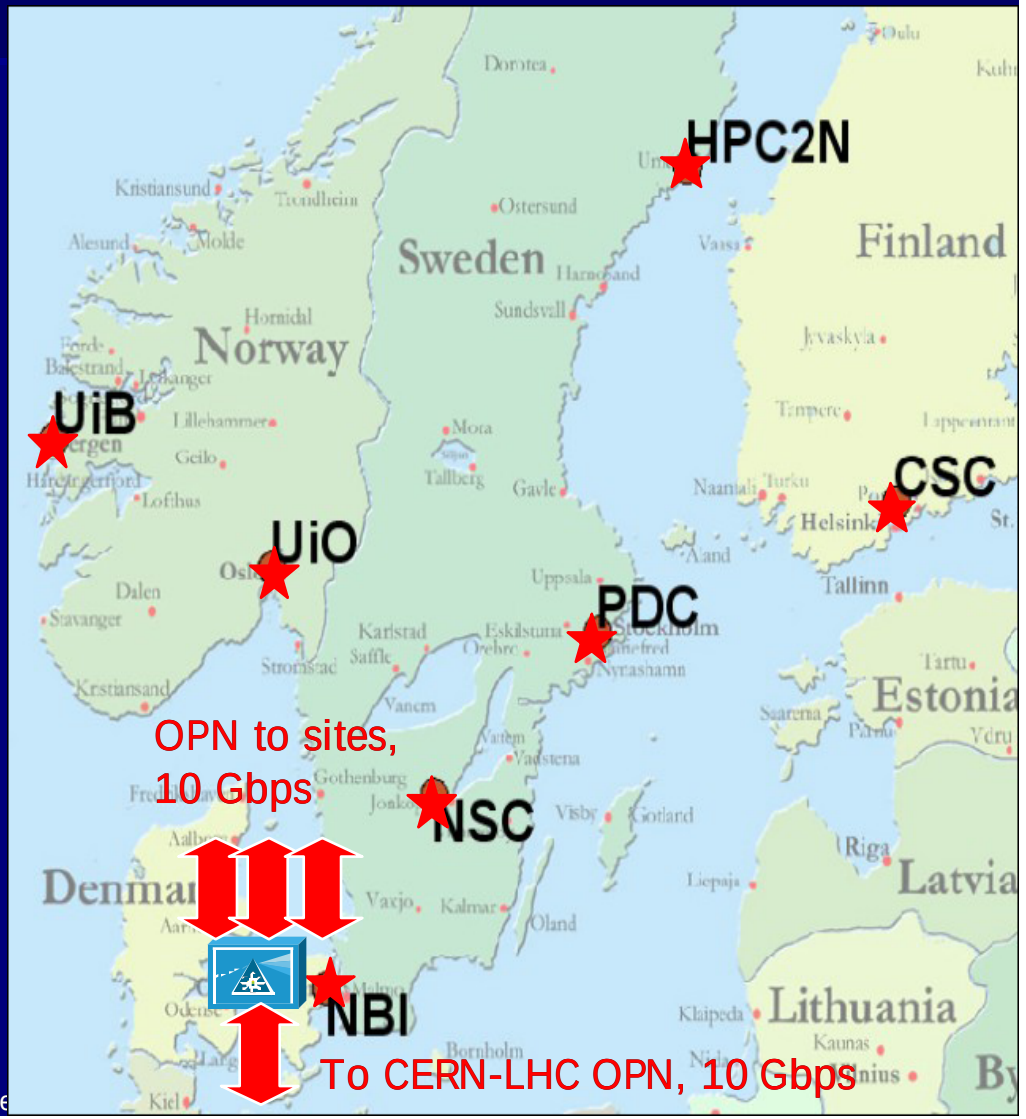
Tier 2 Centres = smaller computer centres (several 10's). Tier 2 Centres provide disk storage and concentrate on simulation and end-user analysis.

NDGF Tier-1 for LHC

- The Nordic High Energy Physics community prefers a Tier-1 of it's own, but...
- None of the Nordic countries is big enough to host a Tier-1 alone
- NDGF will
  - Host the Nordic Tier-1 for ATLAS and ALICE experiments
  - Co-ordinate network, storage, and computing resources
  - Co-ordinate communication towards CERN and LHC project partners
- NDGF creates ***one*** **technical facility** to host the Tier-1

NDGF Tier-1 Sites

- Distributed Tier-1 with 7 sites
- 24/7 operation
- Appears as a *single* site from the outside world
- Has one interface towards CERN (via GEANT2)
- A ”long-reach” LAN (Optical Private Network, OPN) connects the 7 sites
- The computing resources and storage are distributed at national computer centres
- Most resources run ARC
- LCG-ARC interoperability



NDGF Tier-1 Resources

Nordic Data Grid Facility	Pledged	Planned to be pledged				Comment
	2007	2008	2009	2010	2011	
CPU (kSI2K)	690	2110	3230	4280		NDGF LHC Tier-1 status April 2007: 70% of CPU, 34% of disk, 0% of tape in place
Disk (Tbytes)	390	1080	1820	2180		
Tape (Tbytes)	270	930	1900	2260		
Nominal WAN (Mbits/sec)	5000	10000	20000	20000		

NDGF Tier1	2007	2008	2009	2010	2011	Split 2008	ALICE	ATLAS	SUM 2008
CPU (kSI2K)	688	2107	3228	4280		Offered	1037	1070	2107
						% of Total	10%	6%	7%
Disk (Tbytes)	385	1079	1824	2180		Offered	523	556	1079
						% of Total	10%	6%	7%
Tape (Tbytes)	273	930	1898	2261		Offered	517	413	930
						% of Total	7%	5%	6%
Nominal WAN (Mbits/sec)	5000	10000	20000	20000					

NDGF Tier-2 resources

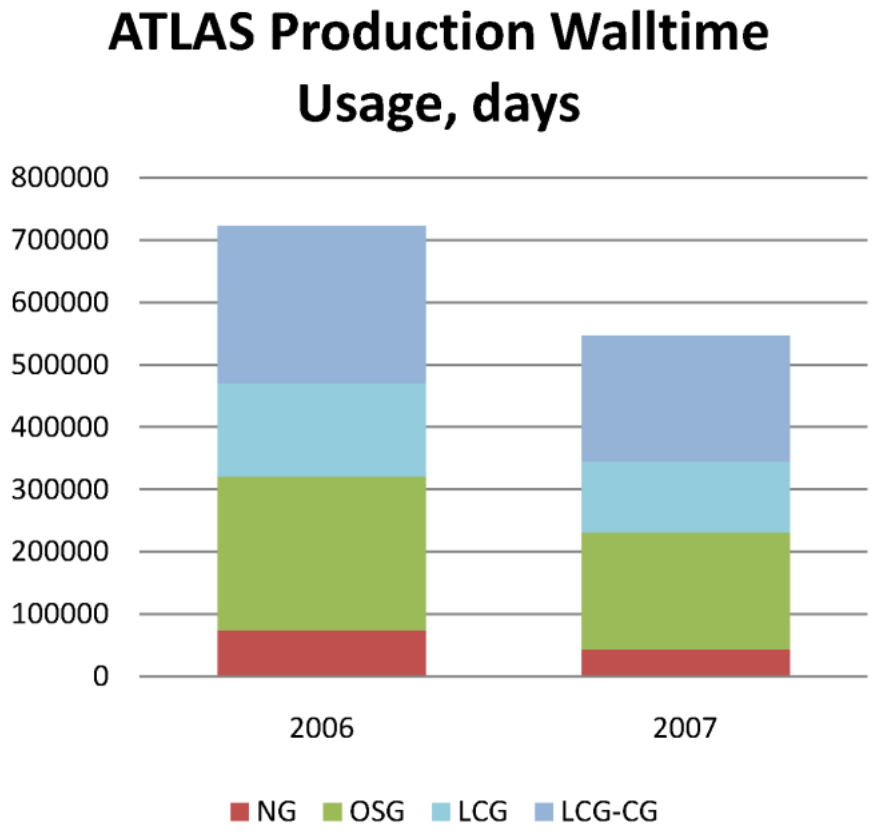
Finland, NDGF/HIP Tier-2	Pledged	Planned to be pledged				Comment
	2007	2008	2009	2010	2011	
CPU (kSI2K)	190	410	670			
Disk (Tbytes)	20	100	210			
Nominal WAN (Mbits/sec)						

Norway, SIGMA Tier2	Pledged	Planned to be pledged				Comment
	2007	2008	2009	2010	2011	
CPU (kSI2K)		110	330	450		
Disk (Tbytes)		60	90	130		
Nominal WAN (Mbits/sec)		10000	20000	20000		

Sweden, SNIC Tier2	Pledged	Planned to be pledged				Comment
	2007	2008	2009	2010	2011	
CPU (kSI2K)	240	980	1550			
Disk (Tbytes)	80	360	600			
Nominal WAN (Mbits/sec)	5000	10000	20000	20000		

Tier-1 performance

- ATLAS Data Challenges = processing of Simulated Data
  - In 2007, NorduGrid/NDGF has processed appr. 8% of all ATLAS production with overall efficiency of 89%. In 2006, appr. 10% share with 74% efficency.
- Critical issues:
  - Hardware purchases
  - Interoperability and accounting needs more work
  - Incorporating ALICE
  - 24/7 in practise



NorduGrid/NDGF (NG) in red

Walltime: Days equivalent on a single processor

SUMMARY

- NDGF is a collaborative Nordic Grid Production Facility
- NDGF Production Phase 2006-2010
- Status 2007: central organization with 15 persons in place and operational, integration of resources ongoing
- NDGF LHC Tier-1 with 7 sites, hardware status April 2007: 70% of CPU, 34% of disk, 0% of tape in place
- Other e-Sciences to follow
- ARC middleware development goes ahead with success
- Challenges:
  - Hardware purchases (national resources)
  - 24/7 Tier-1 operation: not yet tested

NDGF
Nordic DataGrid Facility

Thank you!