Intro

Michael Grønager
Technical Coordinator, NDGF
System Administration and Integration
Aalborg, March 28th 2008

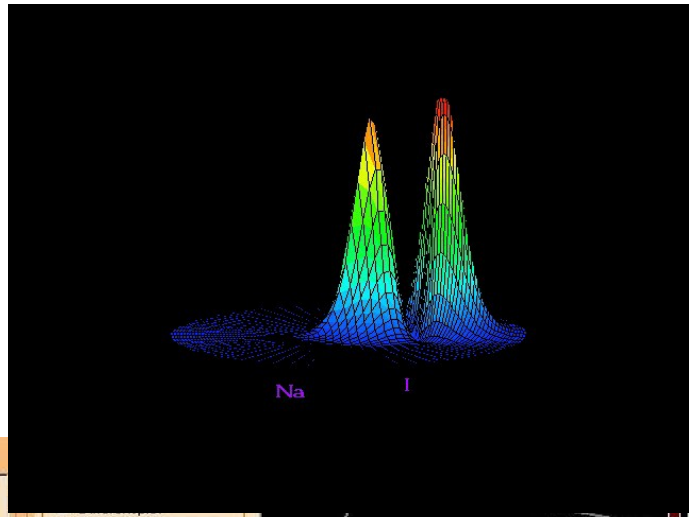
NDGF
Nordic DataGrid Facility

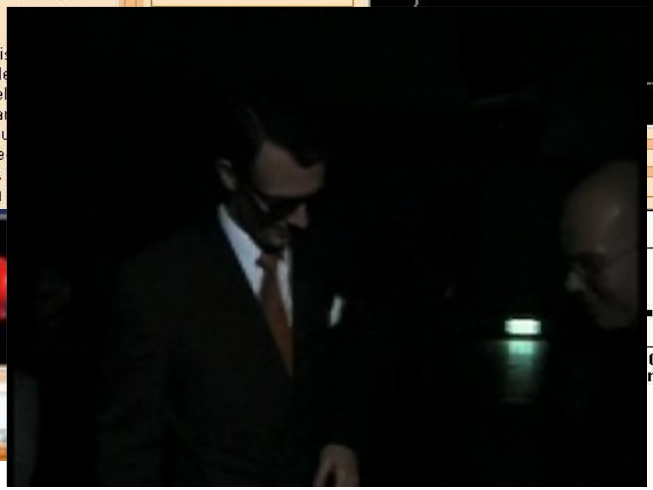
Outline

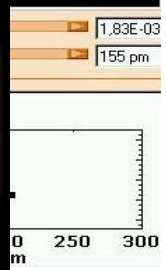
- Who am I ?
- The Grid and Big Science – NDGF intro
- Distributed Operation – across multiple countries
- Development on a running system

Who am I ?

- MSc in Chemical Engineering in 1995
- PhD in Quantum Mechanics in 1998
- UNI-C 1998-2004:
  - 3D programming for VisionQuest
  - Managing the VR Center
  - EU Projects
  - Research projects





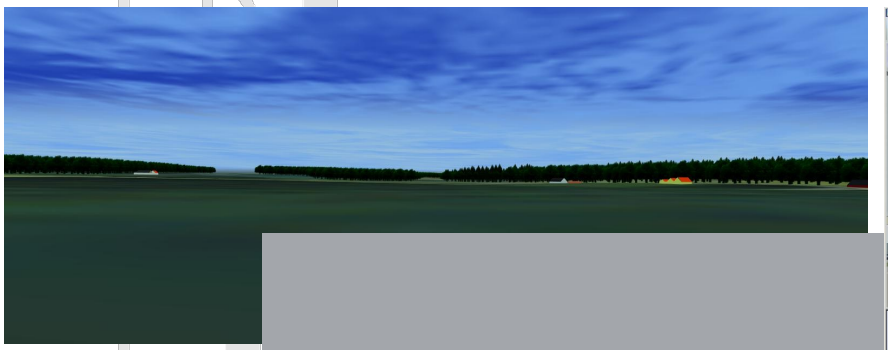


Integration

NDGF
Nordic DataGrid Facility

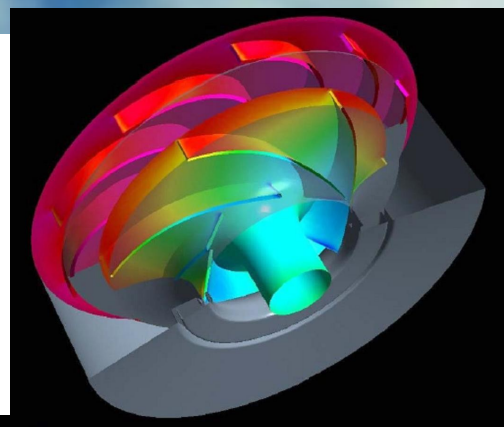
Who am I ?

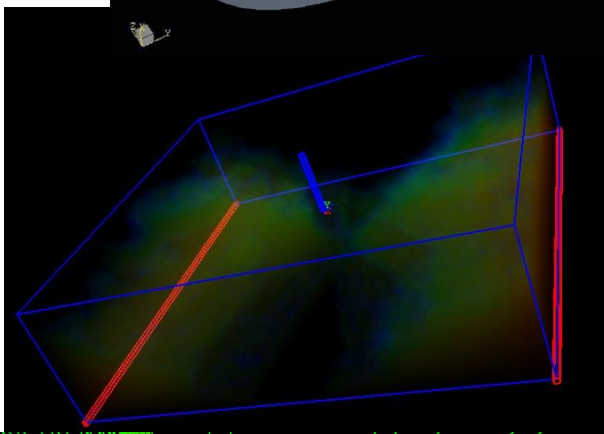
- Virtual Reality
  - Wind turbine visualization
  - Forrest growth visiualization
  - Construction visualization

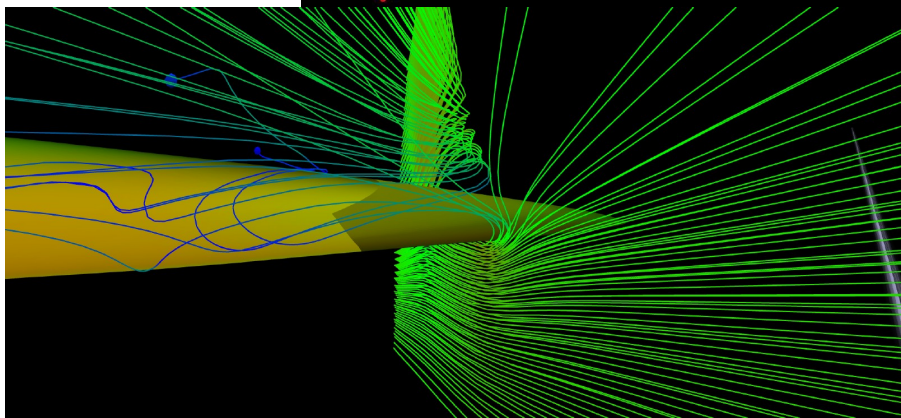


Who am I ?

- NBI 2004-2006:
  - Visualisation
  - Grid Computing
  - EU Projects
  - Open source projects:
    - OpenSceneGraph
    - VTK
    - ARC



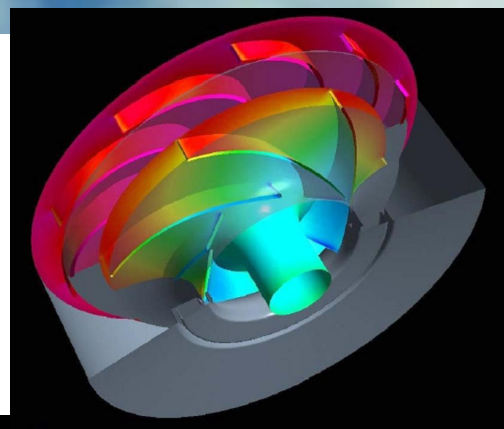


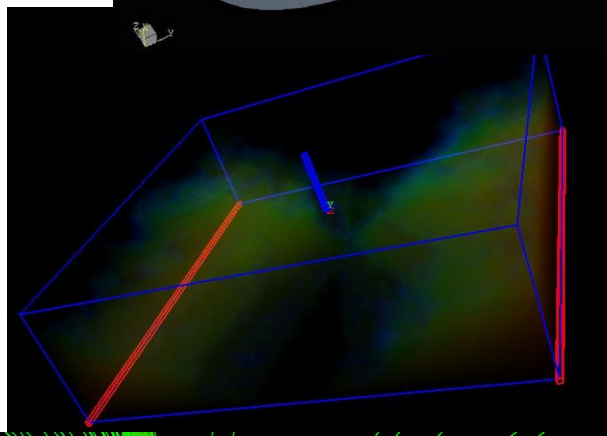


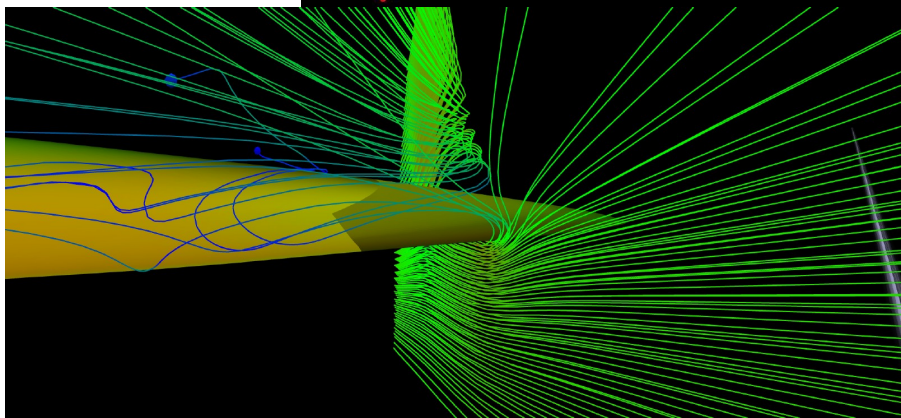
NDGF
Nordic DataGrid Facility

Who am I ?

- Today: NDGF
  - Operations
  - Management
  - International Relations







Big Science and the Grid
An Intro

Michael Grønager
Technical Coordinator, NDGF
System Administration and Integration
Aalborg, March 24th 2008

NDGF
Nordic DataGrid Facility

It's all about science!

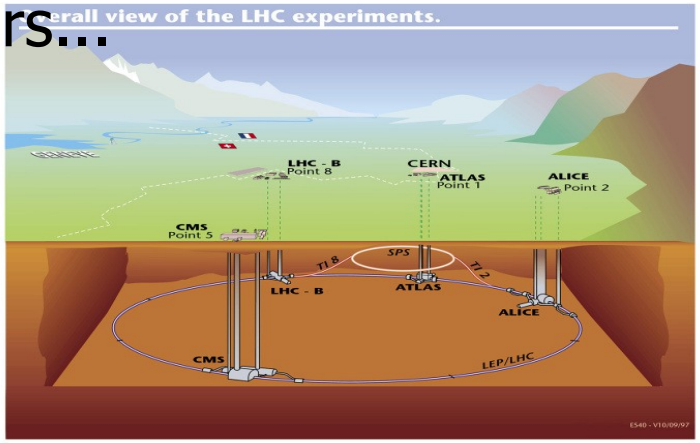
- Tycho Brahe made a sky survey late in the 16th century
- It listed planets with their orbit
- A huge effort with vast amounts of data

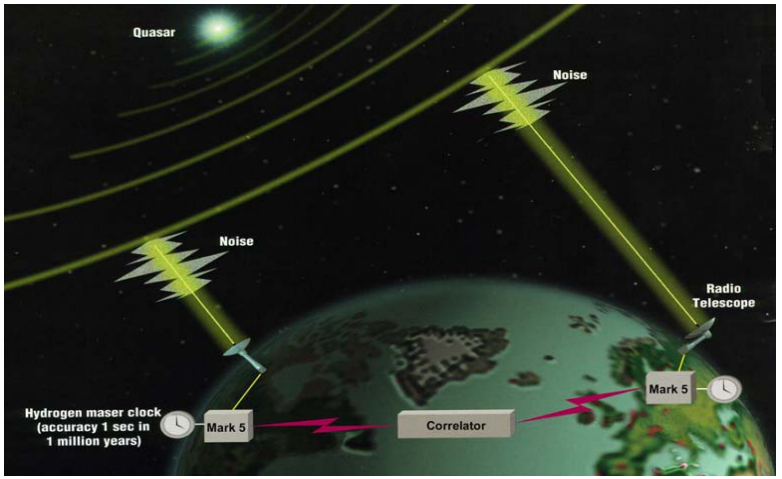
- Kepler did analysis on these data
- Found out about elliptic orbits

- Separation of:
  - Data Collection
  - Data Analysis

Big Science

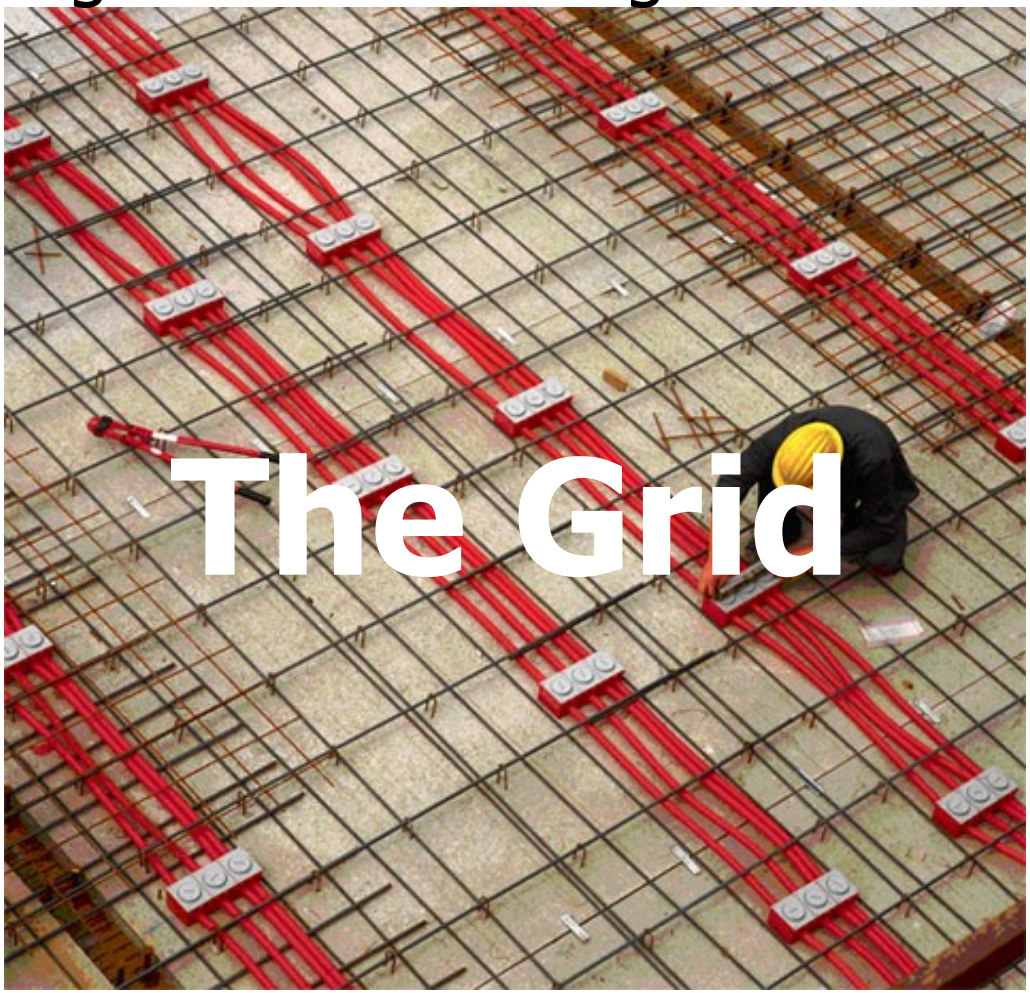
- Data Collection projects
  - Sloan Digital Sky Survey
  - Human Genome Project
  - The Icelandic screening of helth records vs genes
  - e-VLBI
  - CERN: ALICE, ATLAS, CMS and LHCb
  - Others…





Big Science

- Data Storage and Processing:

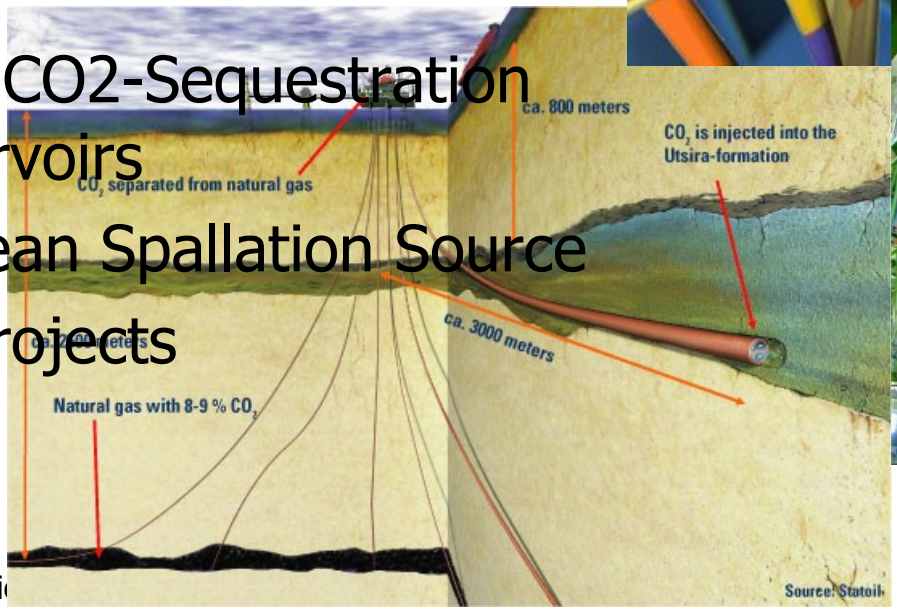


The Grid

- An Infrastructure for:
  - Storing huge amounts of scientific data
  - Accessing distributed data
  - Processing huge amounts of data
  - High availability
- Connecting:
  - Users from multiple organizations
  - Resources from multiple resource providers
  - Multiple user groups

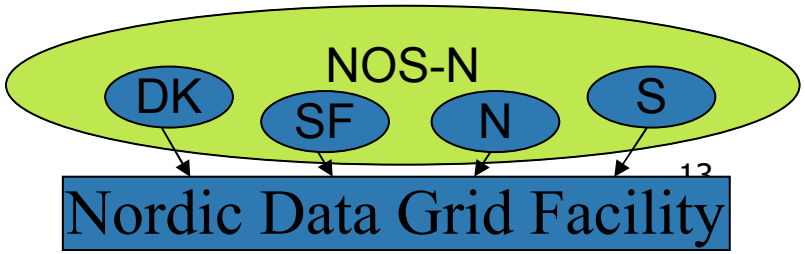
Nordic DataGrid Facility

- Nordic Participation in *Big Science*:
  - WLCG – the Worldwide Large Hadron Collider Grid
  - Gene databases for bio-informatics sciences
  - Screening of CO2-Sequestration suitable reservoirs
  - ESS – European Spallation Source
  - Astronomy projects
  - Other...



NDGF

- A Co-operative Nordic Data and Computing Grid facility
  - Nordic production grid, leveraging national grid resources
  - Common policy framework for Nordic production grid
  - Joint Nordic planning and coordination
  - Operate Nordic storage facility for major projects
  - Co-ordinate & host major eScience projects (i.e., Nordic WLGC Tier-1)
  - Develop grid middleware and services
- NDGF 2006-2010
  - Funded (2 M€/year) by National Research Councils of the Nordic Countries

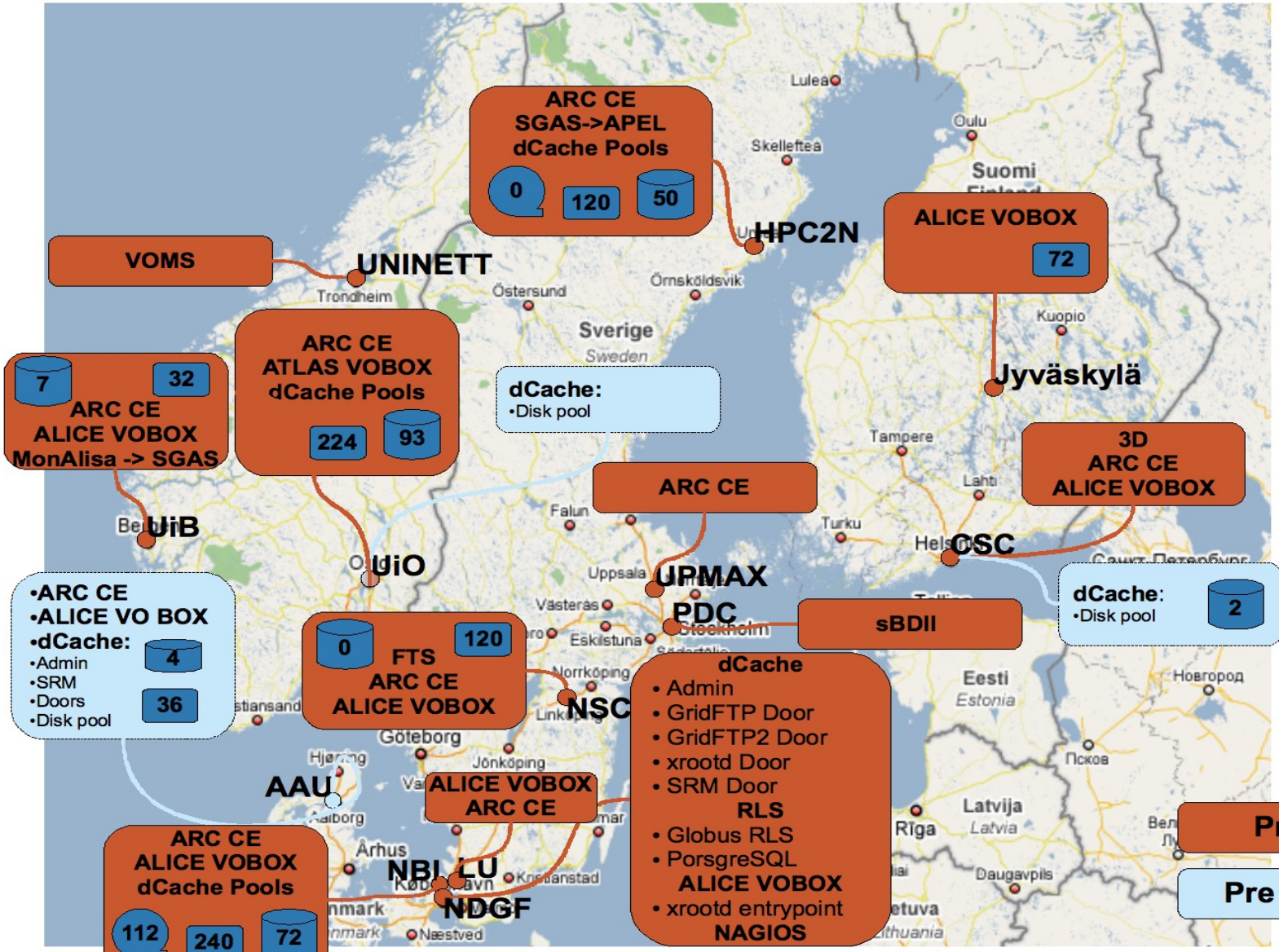


NDGF Vision

- "...to establish a Nordic data grid facility and to involve Nordic countries in European and global co-operation in data sharing in a variety of fields."
- To *coordinate* and *facilitate* the creation of a Nordic eInfrastructure sharing platform
- To enable Nordic researchers to participate in major international projects
- To optimize and standardize use of resources
- To optimize Nordic participation in international projects

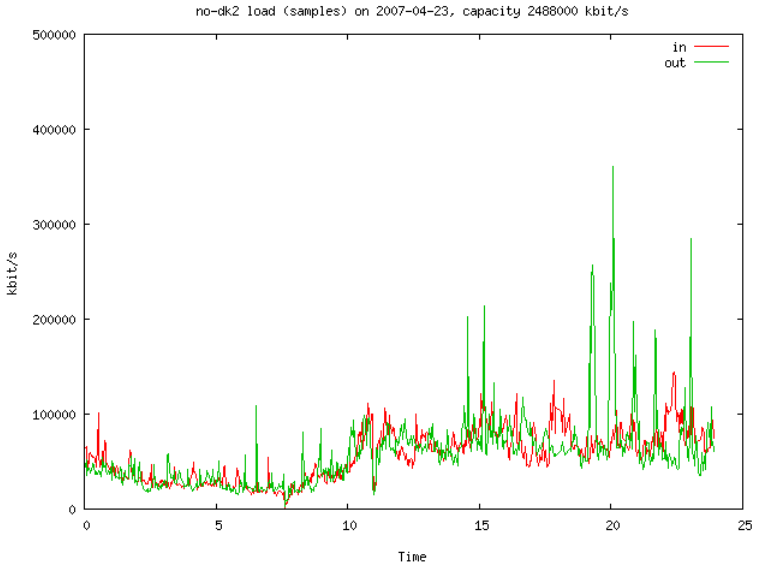
NDGF
Nordic DataGrid Facility

NDGF Facility - 2008Q1



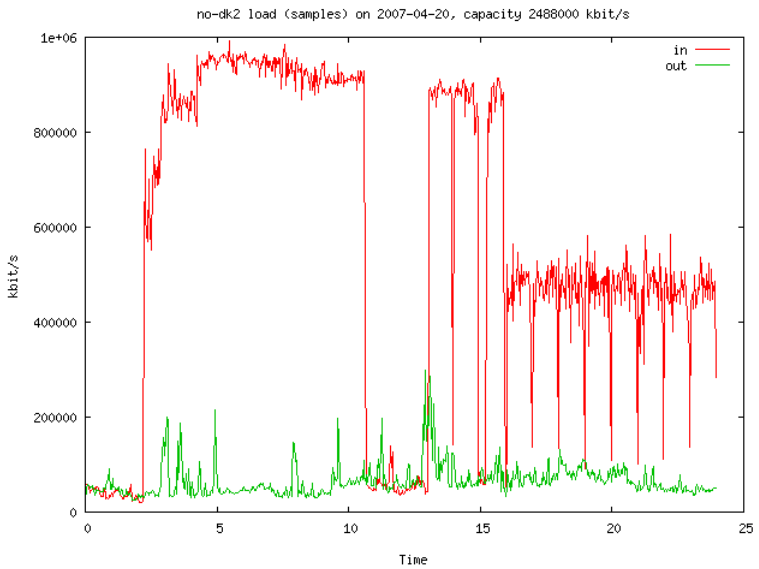
The Infrastructure: Networking

- Today NDGF is connected directly with GEANT 10GBit fiber to CERN
- Inter-Nordic shared 10Gbit network from NORDUnet
- A Dedicated 10Gbit LAN covering all dTier-1 centers next year

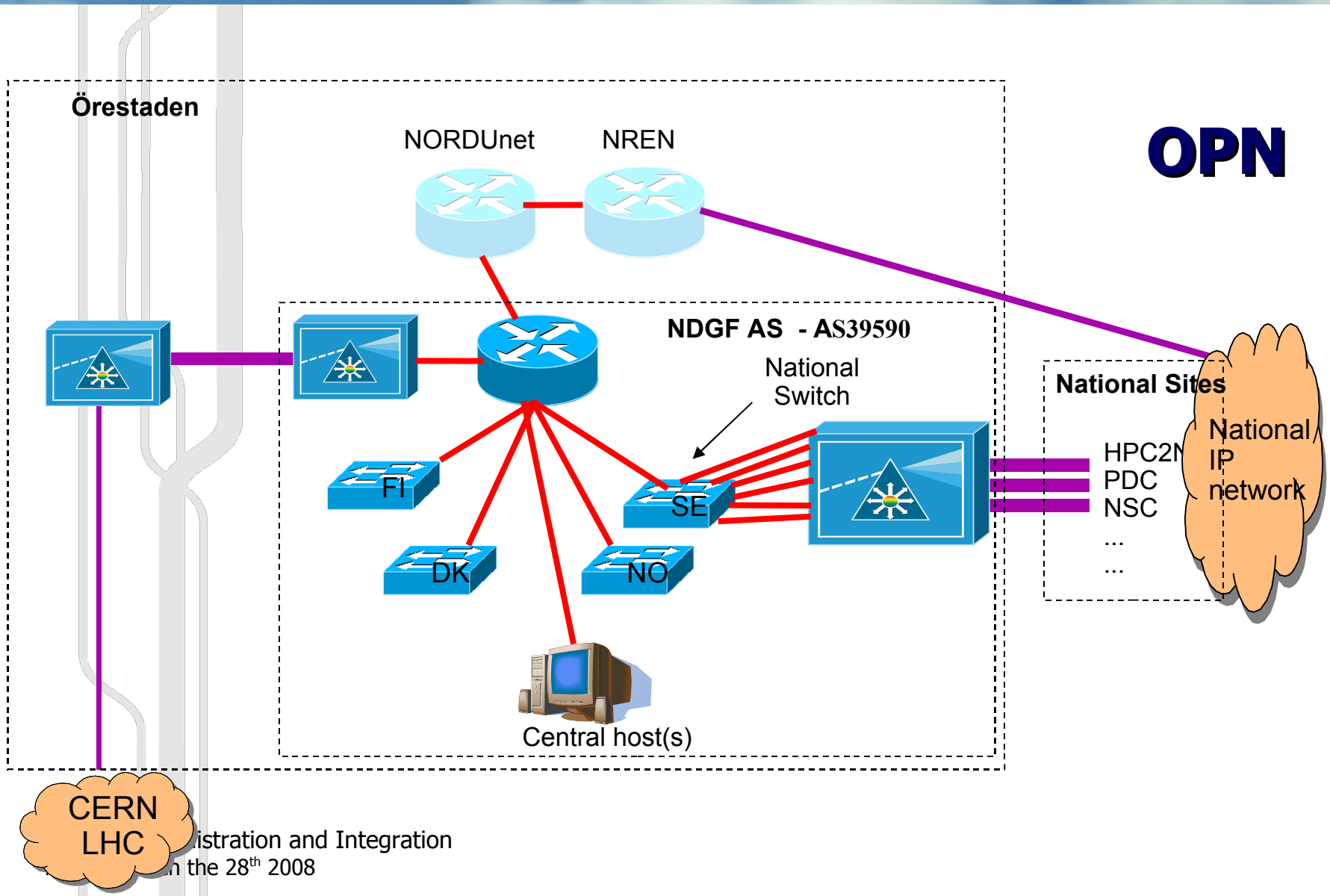


The Infrastructure: Networking

- Today NDGF is connected directly with GEANT 10GBit fiber to CERN
- Inter-Nordic shared 10Gbit network from NORDUnet
- A Dedicated 10Gbit LAN covering all dTier-1 centers next year

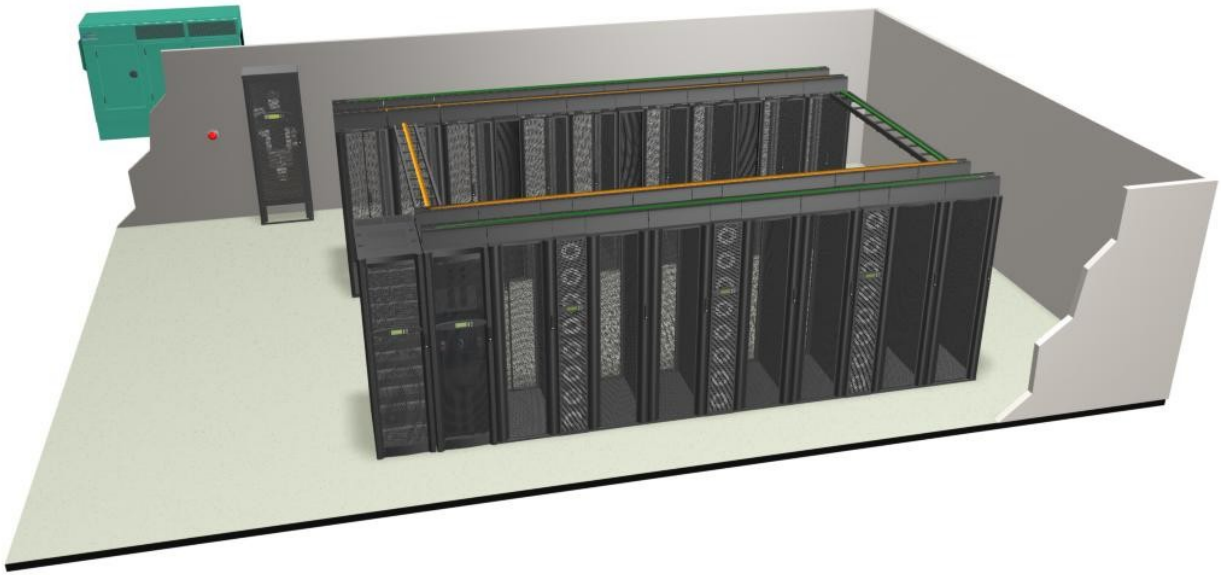


The Infrastructure: Networking



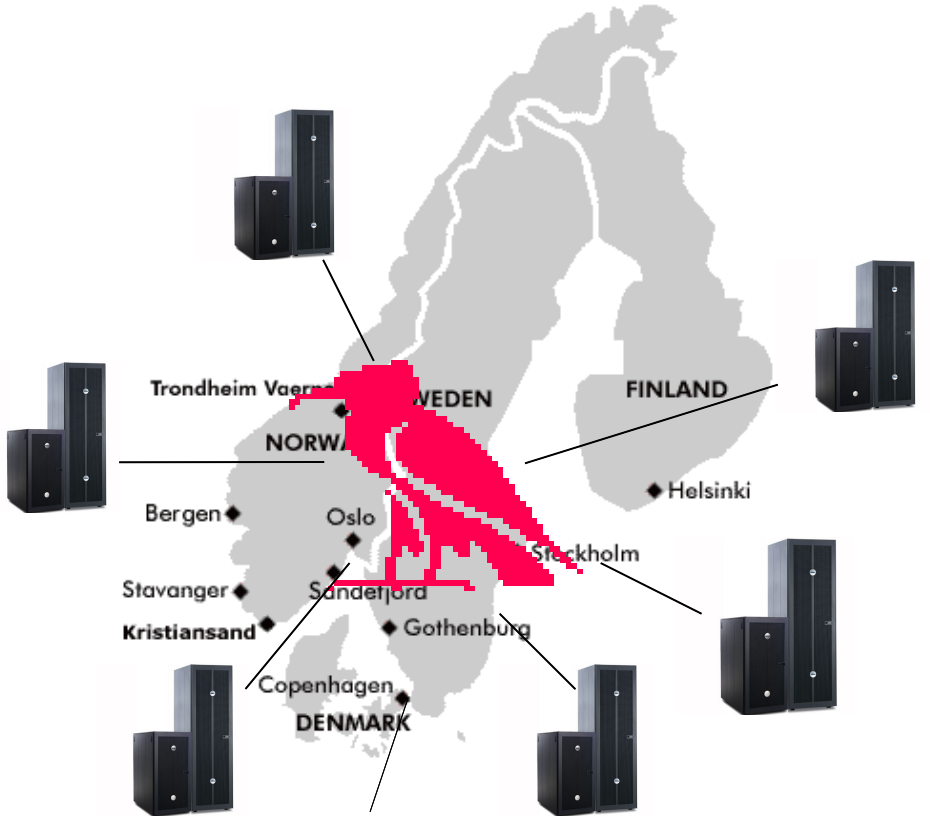
istration and Integration
n the 28th 2008

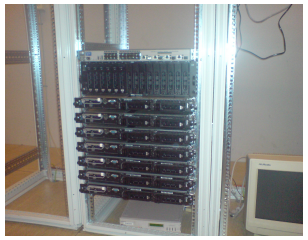
The Infrastructure: Storage



The Infrastructure: Storage

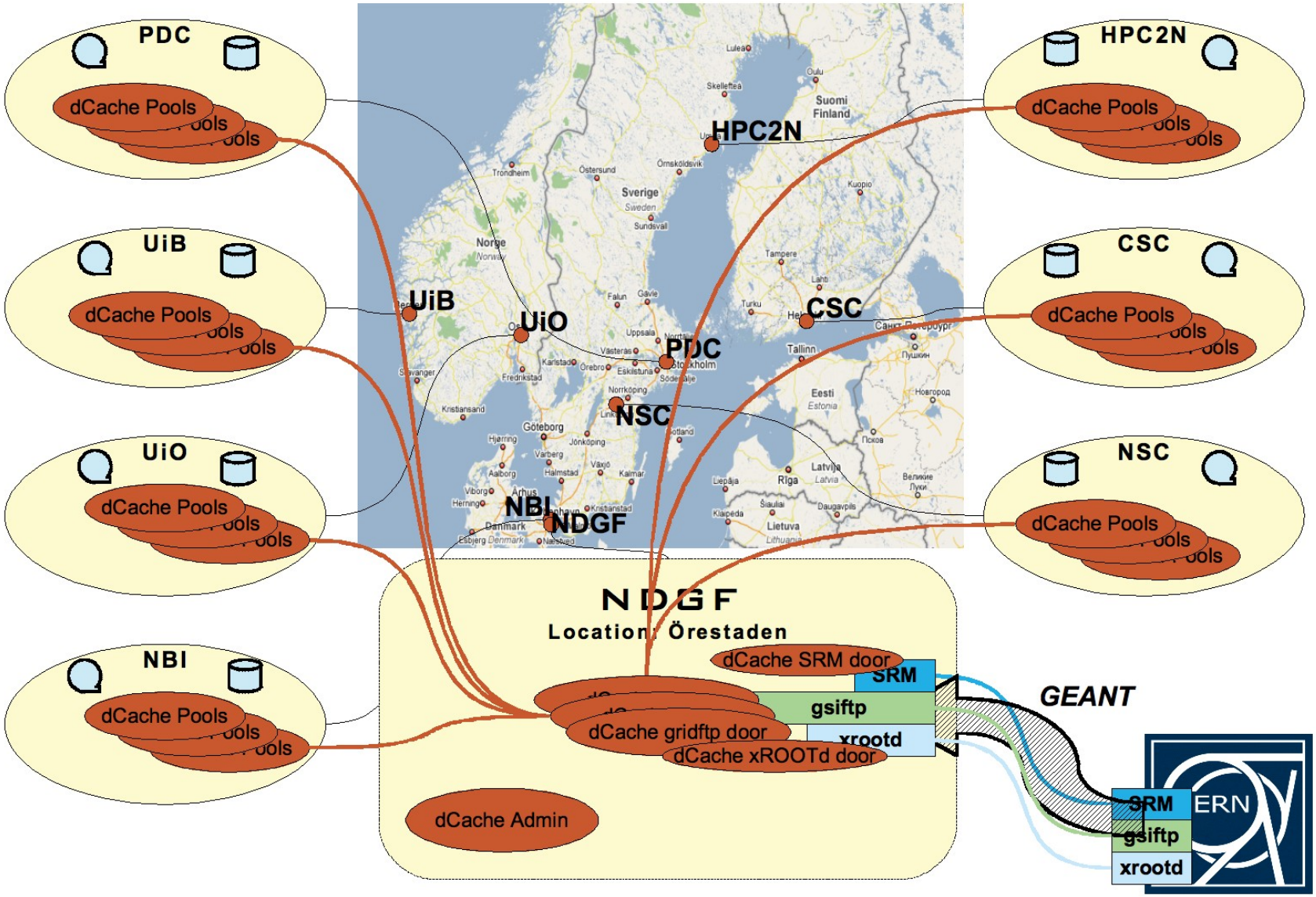
- dCache Installation
- Admin and Door nodes at GEANT endpoint
- Pools at sites
- Very close collaboration with DESY and FermiLab ensure dCache is suited also for distributed use





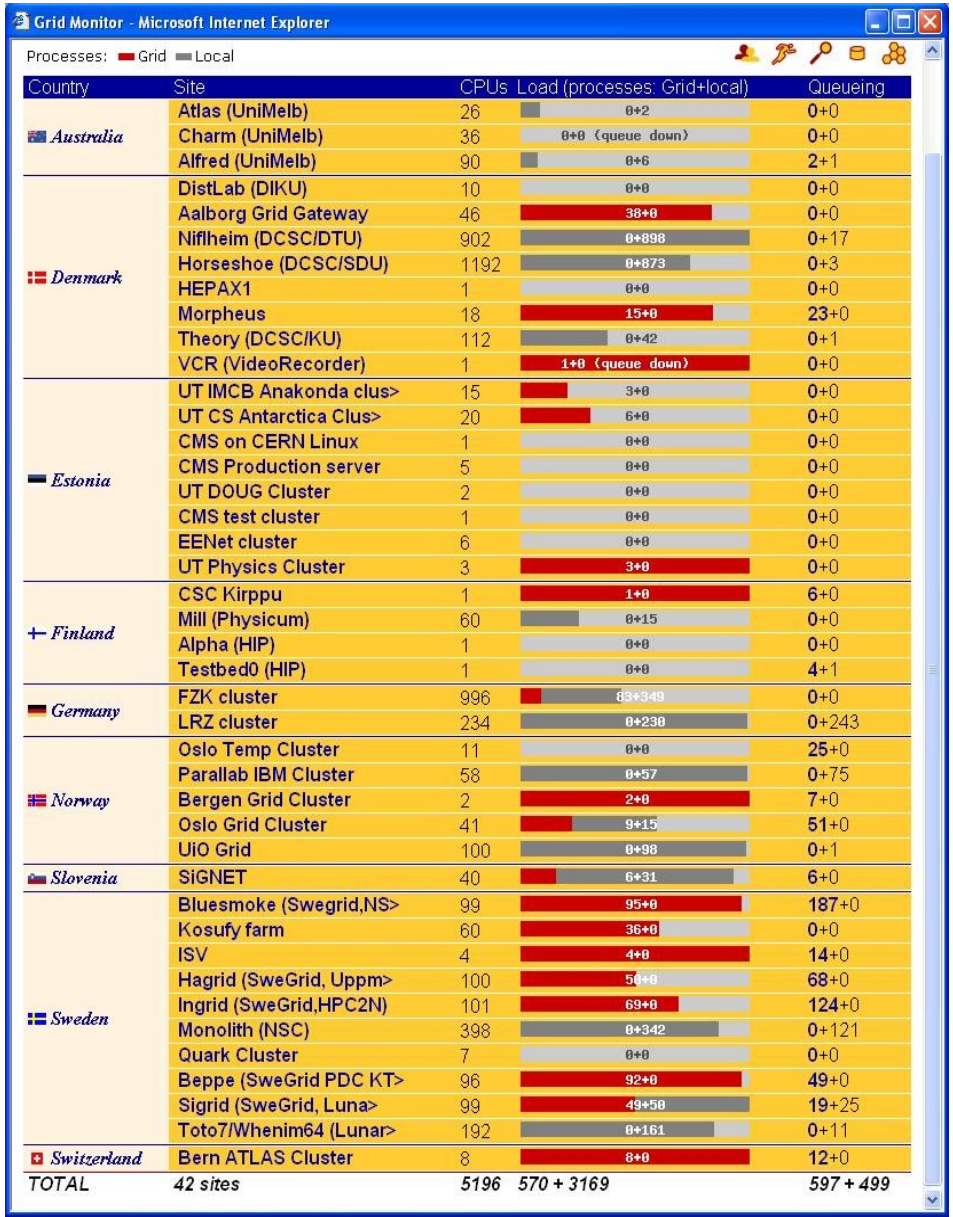
NDGF
Nordic DataGrid Facility

The Infrastructure: Storage



The Infrastructure: Computing

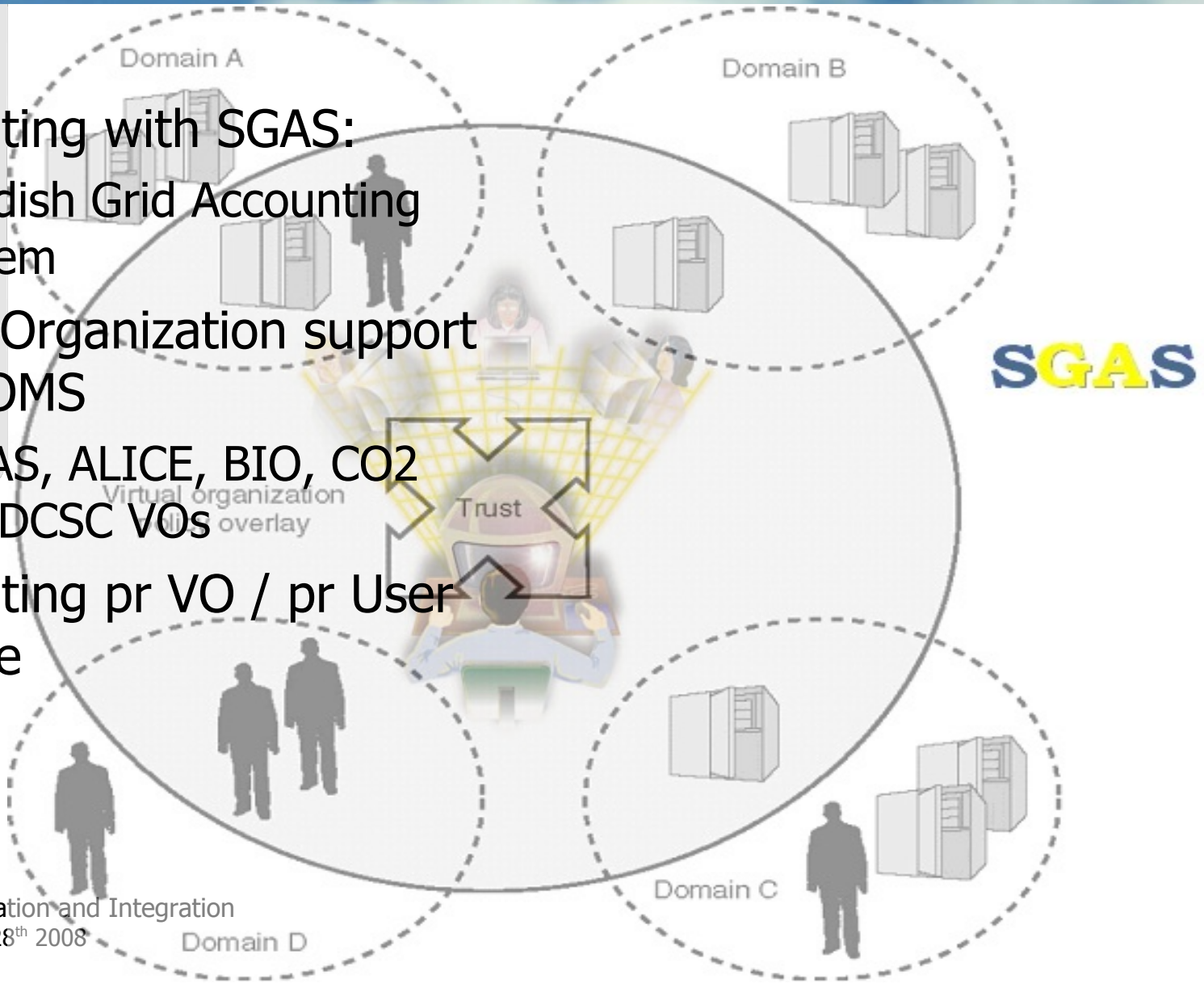
- NorduGrid / ARC middleware for Computing
- Used routinely since 2002 for e.g. ATLAS data challenges
- Deployed at all the dTier-1 sites



Country	Site	CPUs	Load (processes: Grid+local)	Queueing
Australia	Atlas (UniMelb)	26	0+2	0+0
	Charm (UniMelb)	36	0+0 (queue down)	0+0
	Alfred (UniMelb)	90	0+6	2+1
Denmark	DistLab (DIKU)	10	0+0	0+0
	Aalborg Grid Gateway	46	38+0	0+0
	Niflheim (DCSC/DTU)	902	0+898	0+17
	Horseshoe (DCSC/SDU)	1192	0+873	0+3
	HEPAX1	1	0+0	0+0
	Morpheus	18	15+0	23+0
	Theory (DCSC/IKU)	112	0+42	0+1
	VCR (VideoRecorder)	1	1+0 (queue down)	0+0
Estonia	UT IMCB Anakonda clus>	15	3+0	0+0
	UT CS Antarctica Clus>	20	6+0	0+0
	CMS on CERN Linux	1	0+0	0+0
	CMS Production server	5	0+0	0+0
	UT DOUG Cluster	2	0+0	0+0
	CMS test cluster	1	0+0	0+0
	EENet cluster	6	0+0	0+0
	UT Physics Cluster	3	3+0	0+0
Finland	CSC Kirppu	1	1+0	6+0
	Mill (Physicum)	60	0+15	0+0
	Alpha (HIP)	1	0+0	0+0
	Testbed0 (HIP)	1	0+0	4+1
Germany	FZK cluster	996	83+349	0+0
	LRZ cluster	234	0+230	0+243
Norway	Oslo Temp Cluster	11	0+0	25+0
	Parallab IBM Cluster	58	0+57	0+75
	Bergen Grid Cluster	2	2+0	7+0
	Oslo Grid Cluster	41	9+15	51+0
	UiO Grid	100	0+98	0+1
Slovenia	SiGNET	40	6+31	6+0
Sweden	Bluesmoke (Swegrid,NS>	99	95+0	187+0
	Kosufy farm	60	36+0	0+0
	ISV	4	4+0	14+0
	Hagrid (SweGrid, Uppm>	100	50+0	68+0
	Ingrid (SweGrid,HPC2N)	101	69+0	124+0
	Monolith (NSC)	398	0+342	0+121
	Quark Cluster	7	0+0	0+0
	Beppe (SweGrid PDC KT>	96	92+0	49+0
	Sigrid (SweGrid, Luna>	99	49+50	19+25
	Toto7/Whenim64 (Lunar>	192	0+161	0+11
Switzerland	Bern ATLAS Cluster	8	8+0	12+0
TOTAL	42 sites	5196	570 + 3169	597 + 499

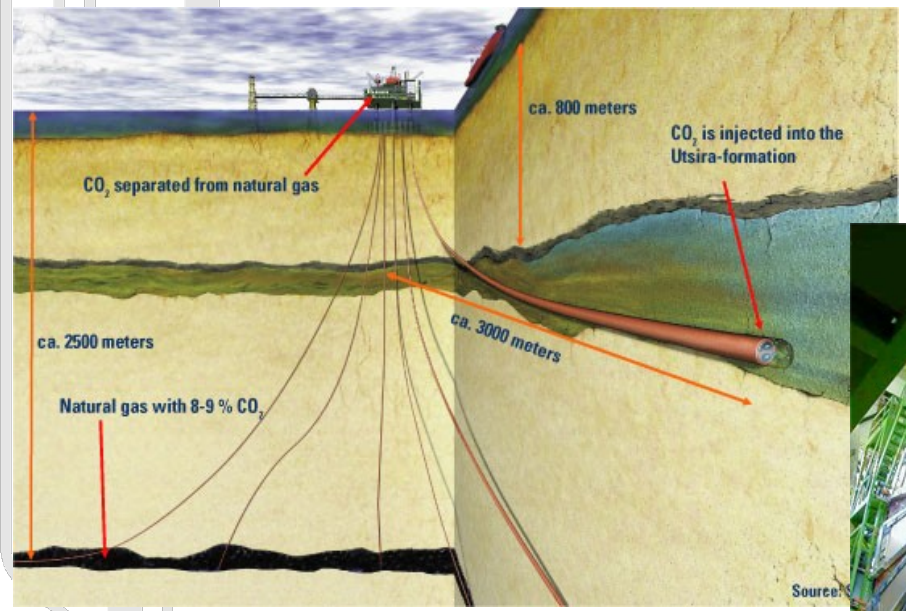
The Infrastructure: VOs & Account.

- Accounting with SGAS:
  - Swedish Grid Accounting System
- Virtual Organization support with VOMS
  - ATLAS, ALICE, BIO, CO2 and DCSC VOs
- Accounting pr VO / pr User / pr Site

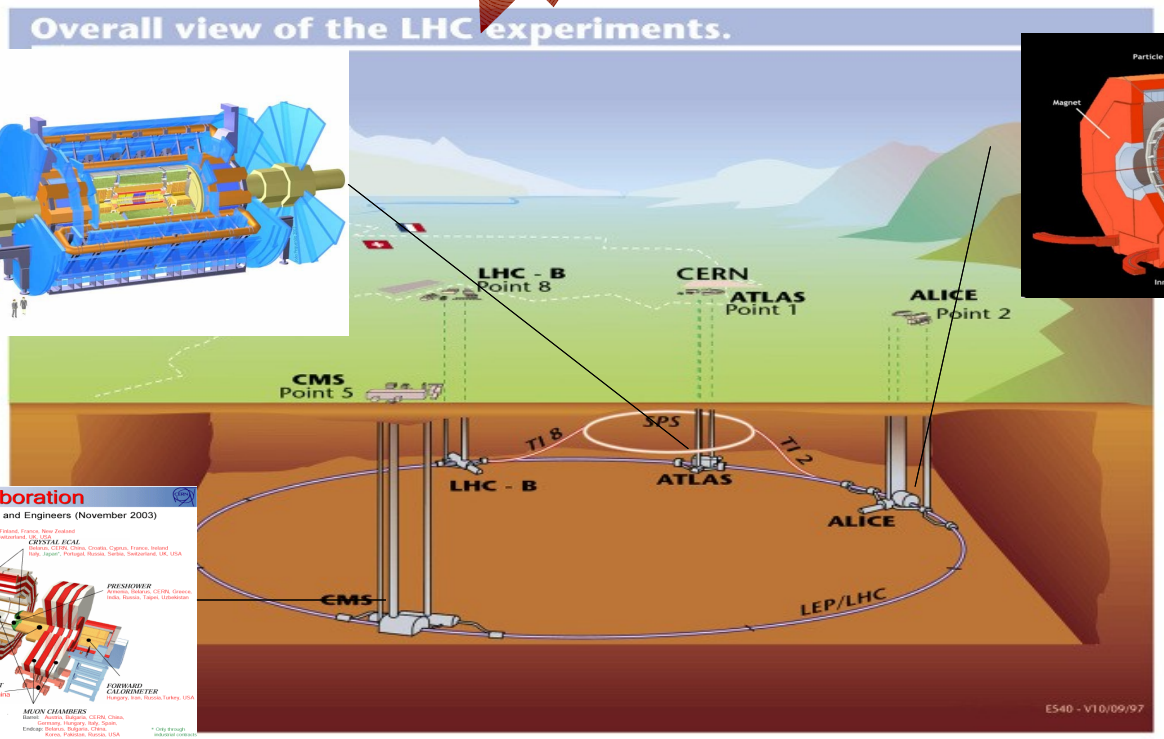


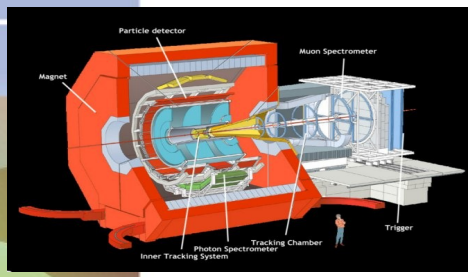
NDGF eScience projects

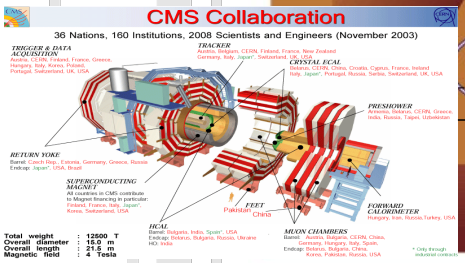
- BioGrid community grid
- CO_2 sequestration community grid
- Nordic Tier-1



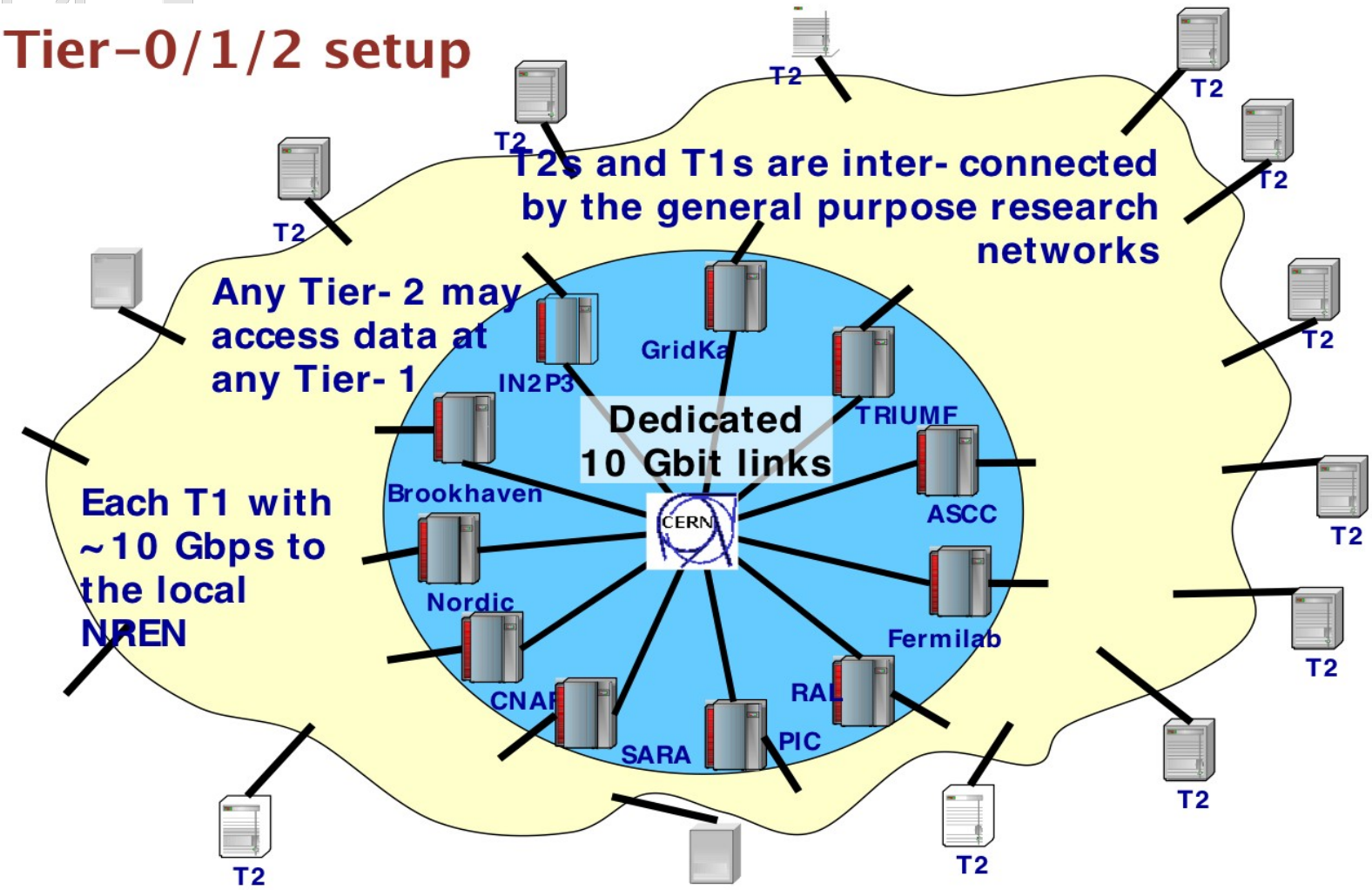
NDGF a Tier-1 for WLGC







The WLCG Infrastructure



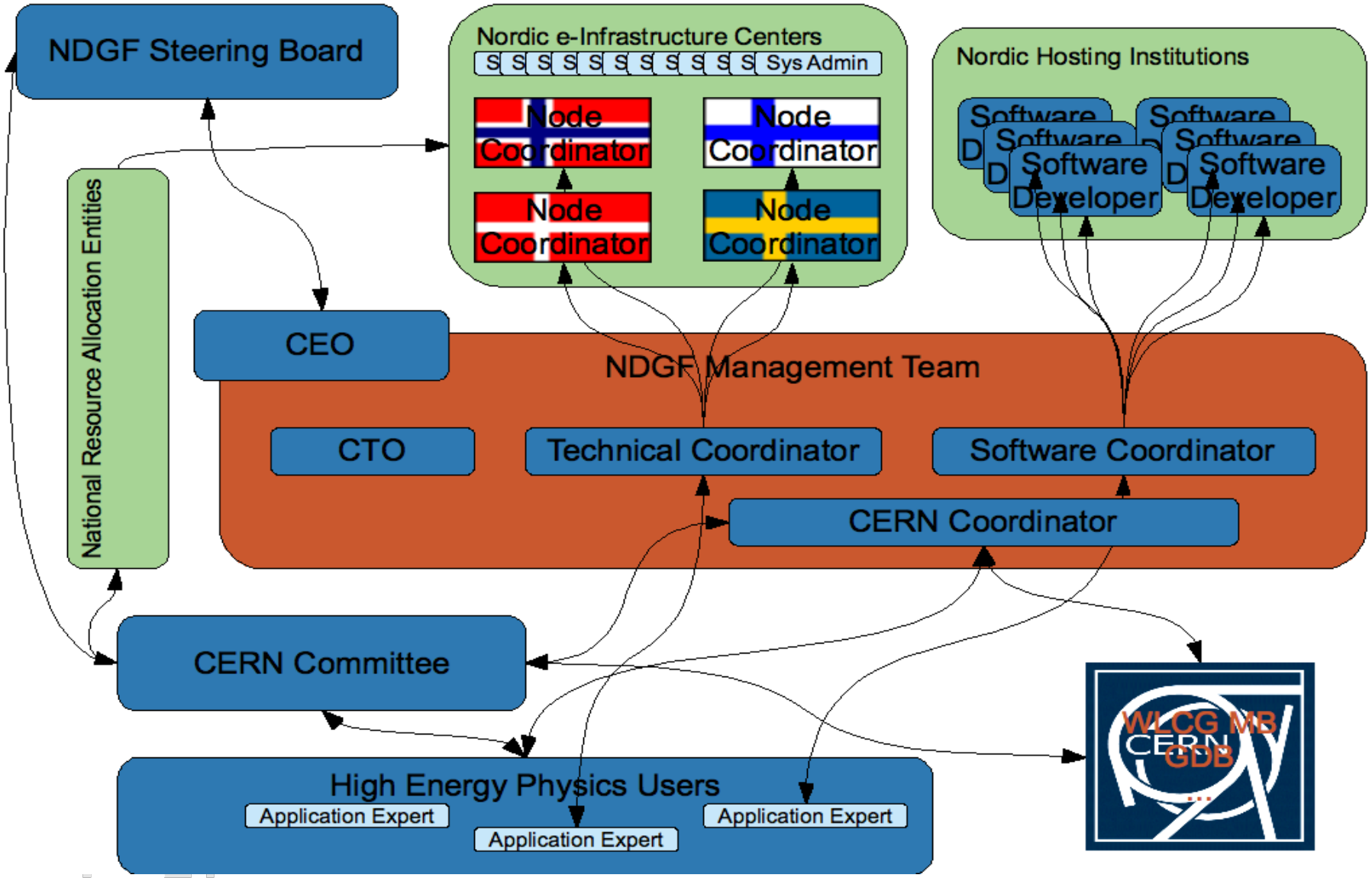
NDGF
Nordic DataGrid Facility

NDGF a Tier-1 for WLGC

- Organization / Governance
- Tier-1 Services:
  - Computing
  - Storage
  - ATLAS
  - ALICE
  - Accounting
  - Monitoring
  - Operation

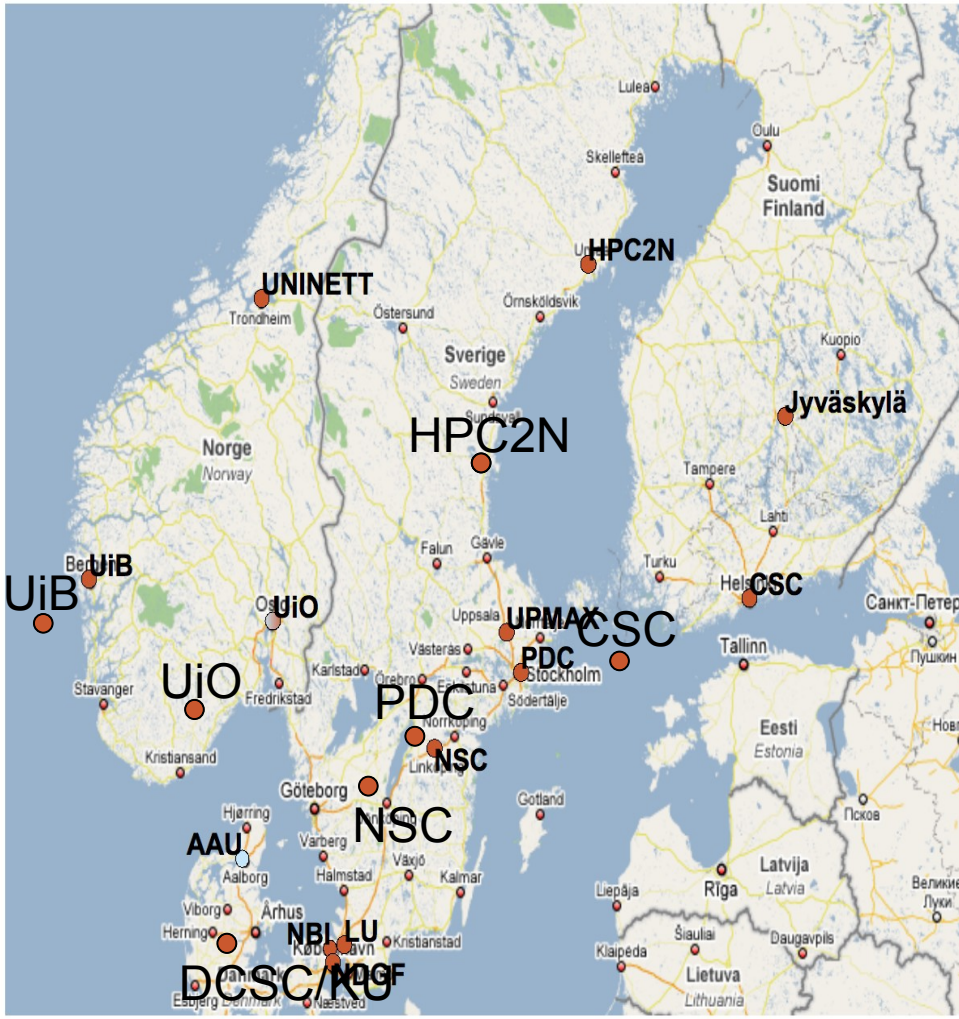
NDGF
Nordic DataGrid Facility

Organization – CERN related



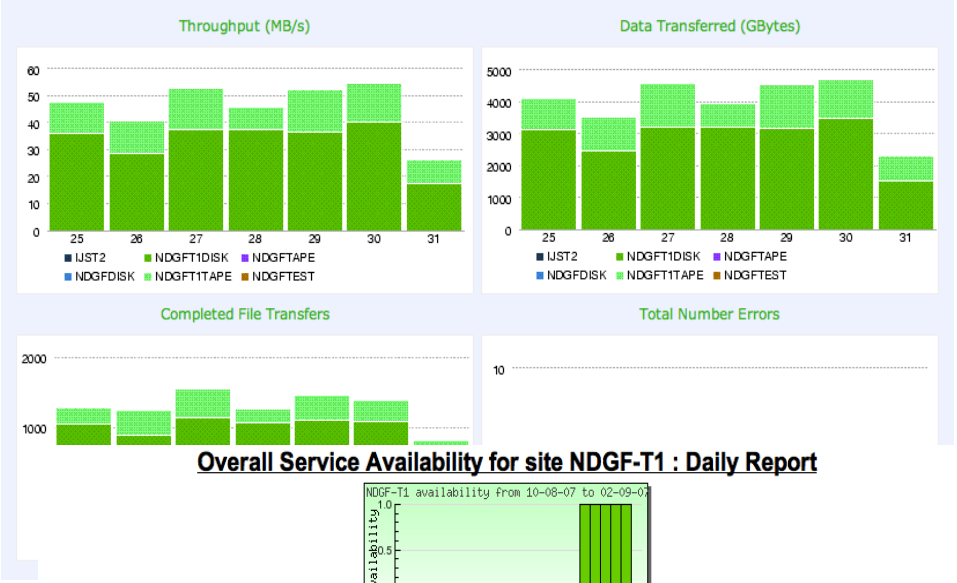
NDGF Tier-1 Resource Centers

- The 7 biggest Nordic compute centers, dTier-1s, form the NDGF Tier-1
- Resources (Storage and Computing) are scattered
- Services can be centralized
- Advantages in redundancy
- Especially for 24x7 data taking



Other Tier-1 Services

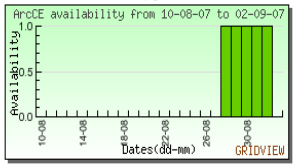
- FTS – File Transfer Service
- 3D – Distributed Database Deployment
- Accounting export to APEL
- Service Availability Monitoring – via WLCG SAM sensors



Overall Service Availability for site NDGF-T1 : Daily Report

Individual Service Availability for site NDGF-T1 : Daily Report

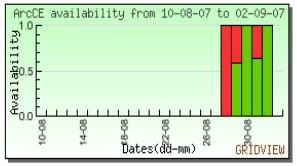
ArcCE



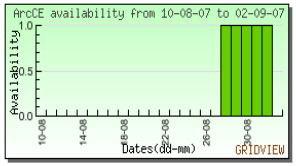
Service Instance Availability for site NDGF-T1 : Daily Report

Site Services

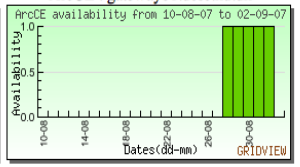
ArcCE : ce01.titan.uio.no



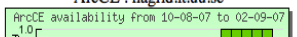
ArcCE : condor.titan.uio.no



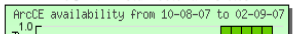
ArcCE : gateway01.dcsc.ku.dk



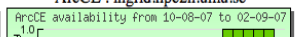
ArcCE : hagrid.it.uu.se



ArcCE : hive.unicc.chalmers.se



ArcCE : ingrid.hpc2n.umu.se



NDGF
Nordic DataGrid Facility

Looking ahead...

- Mature Nordic Infrastructure for *Big Science*
  - State of the art network
  - State of the art storage and compute resource sharing
  - Proofed by the now fully functional distributed Tier-1
- Partnering with EGEE on:
  - Operation (taking part in CIC on Duty)
  - Interoperability (submitting jobs from gLite to ARC)
- More e-Science projects are welcomed!
  - Big as well as small...

Running a multi country distributed Infrastructure

Michael Grønager
Technical Coordinator, NDGF
System Administration and Integration
Aalborg, March 24th 2008

NDGF
Nordic DataGrid Facility

NDGF
Nordic DataGrid Facility

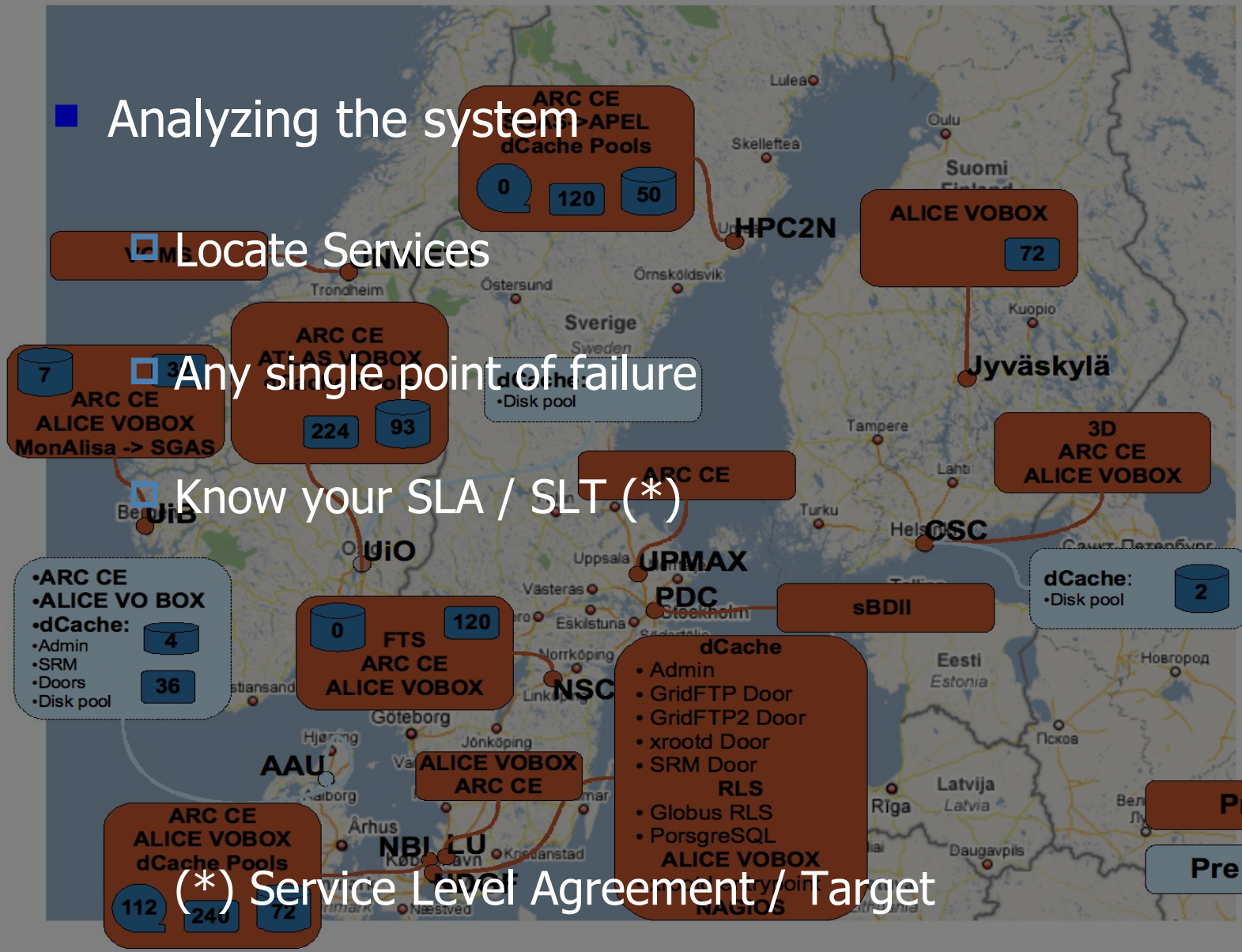
Outline

- Services
- Network
- Operation
- Procedures
- Motivation

NDGF
Nordic DataGrid Facility

Service breakdown

- Analyzing the system
  - Locate Services
  - Any single point of failure
  - Know your SLA / SLT (*)

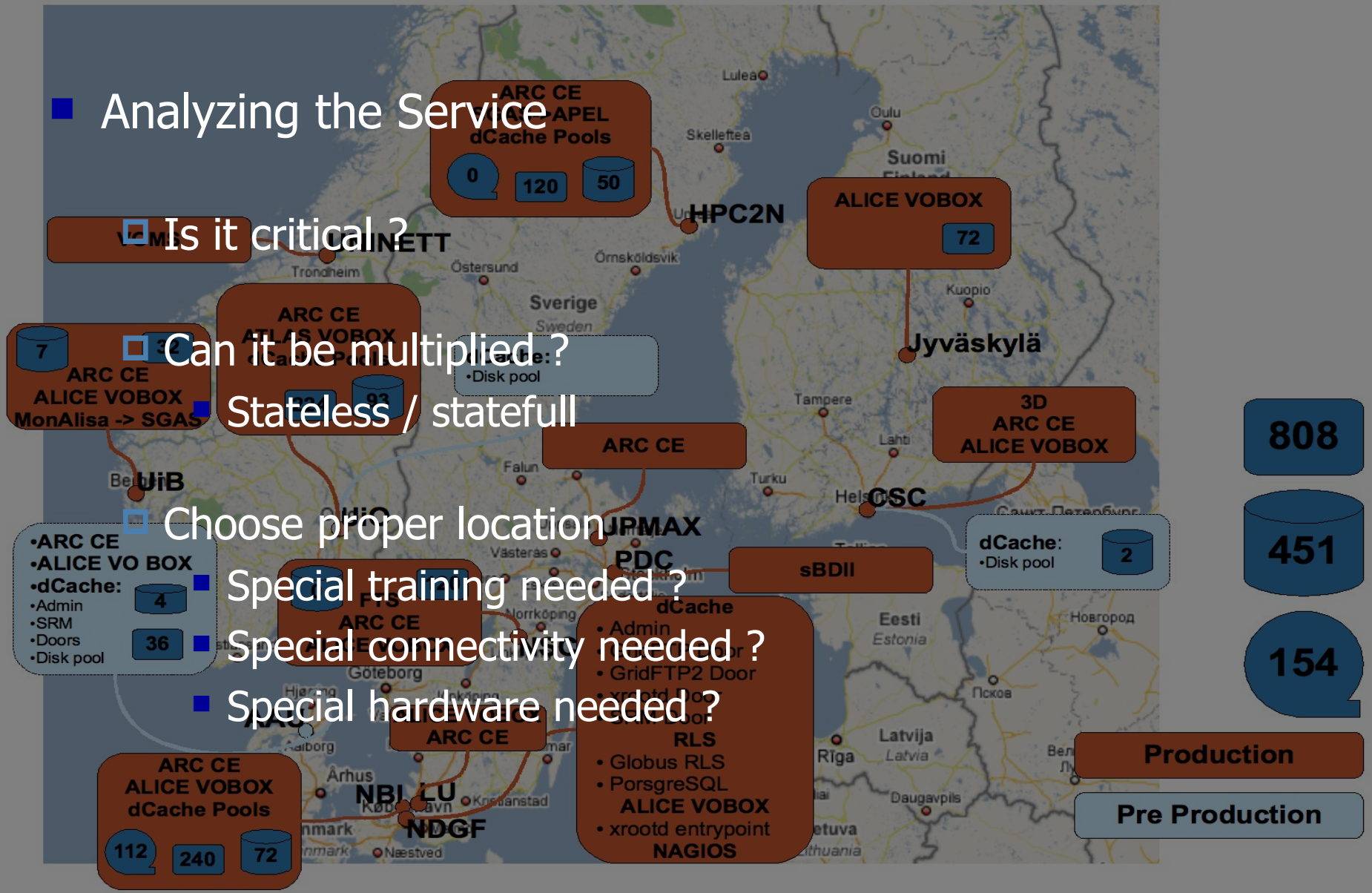


(*) Service Level Agreement / Target

NDGF
Nordic DataGrid Facility

Service Analysis

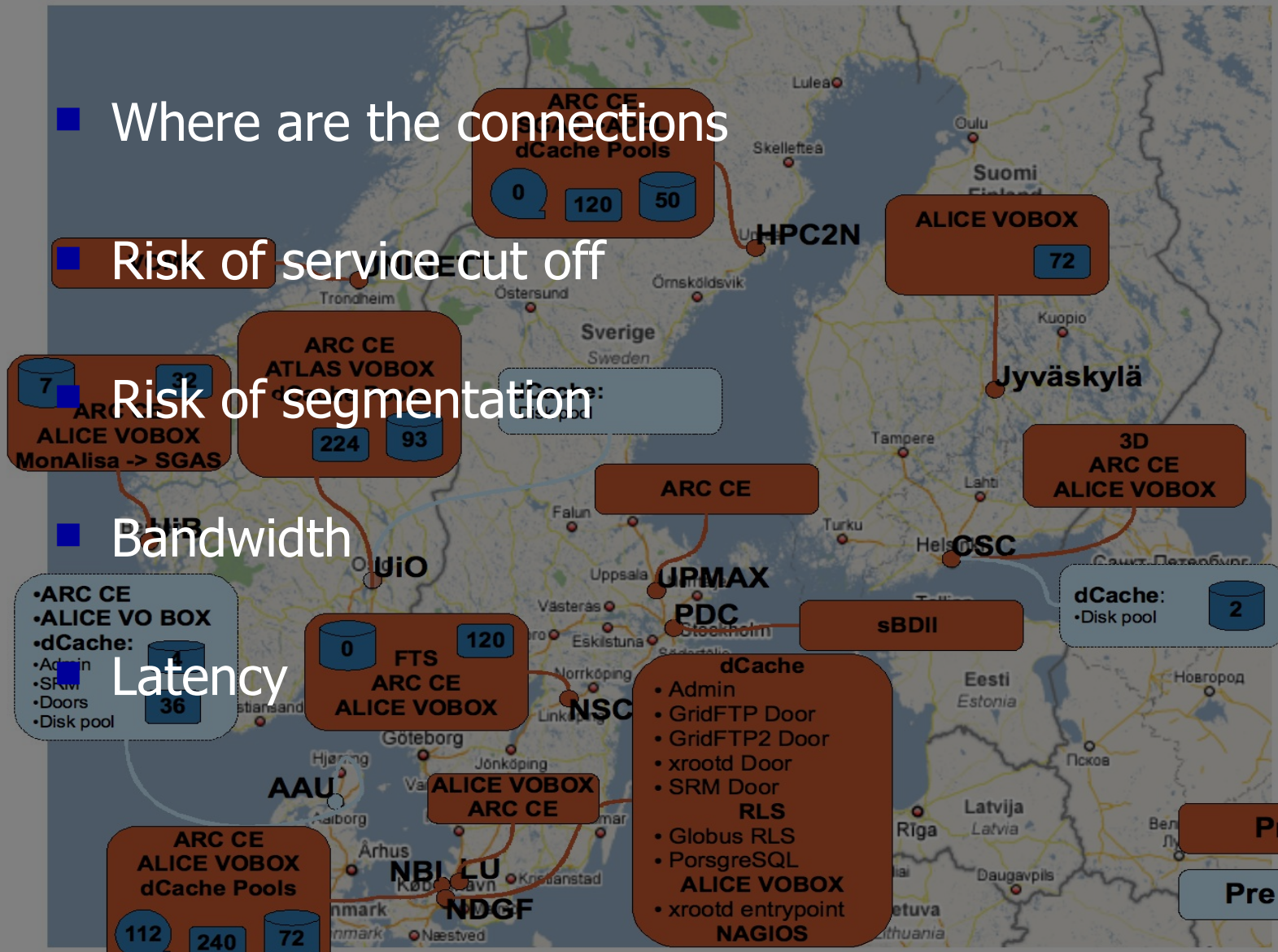
- Analyzing the Service
  - Is it critical ?
  - Can it be multiplied ?
  - Stateless / statefull
  - Choose proper location
  - Special training needed ?
  - Special connectivity needed ?
  - Special hardware needed ?



NDGF
Nordic DataGrid Facility

Network Analysis

- Where are the connections
- Risk of service cut off
- Risk of segmentation
- Bandwidth
- Latency



Tools

- Nagios
  - Alerts
  - Reports
  - Uptime analysis
- Ganglia
  - System analysis
  - Better understanding
  - Problem solving
- Tickets vs. mail vs. IM vs. phone

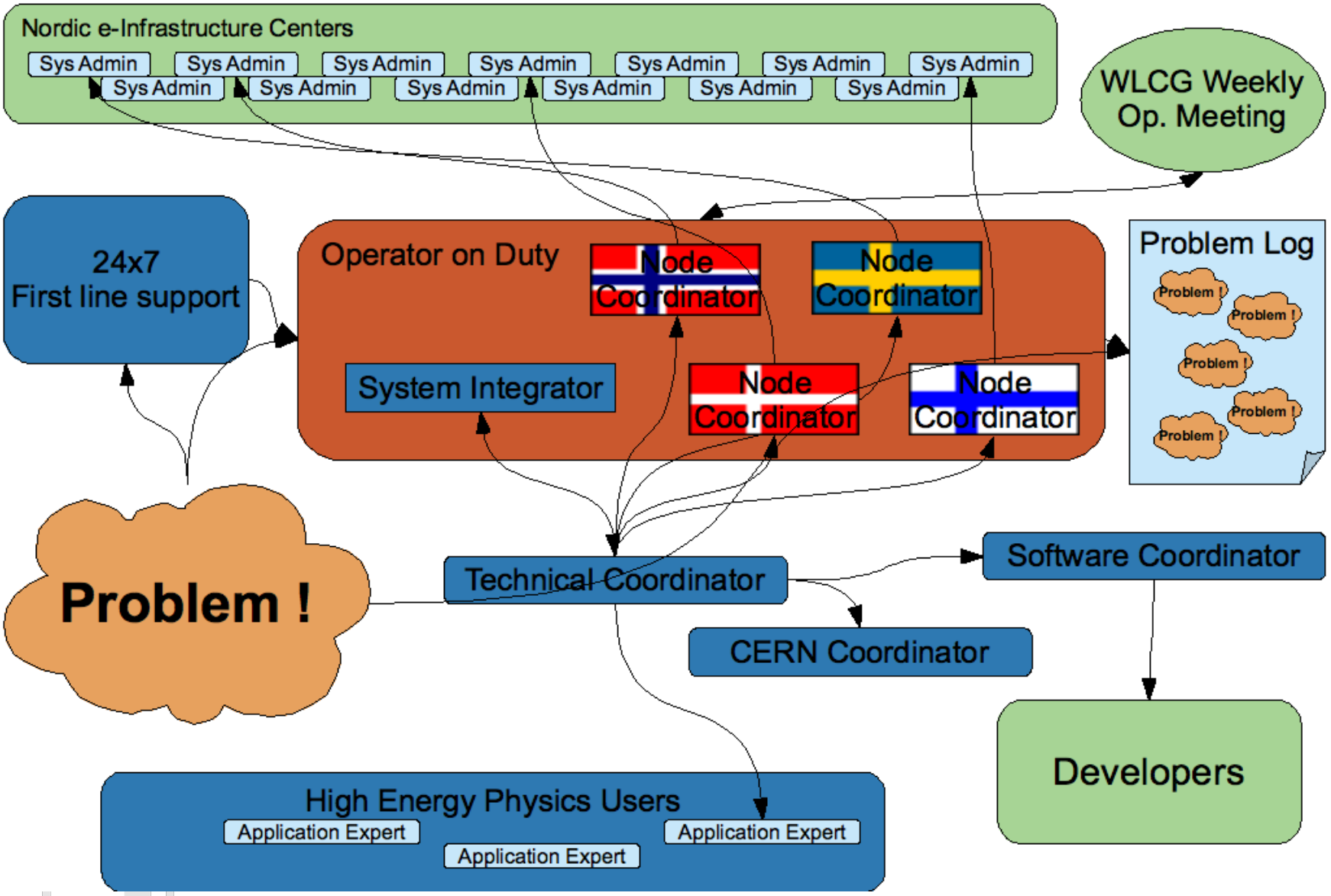
NDGF
Nordic DataGrid Facility

Operations

- What is the SLA / SLT ?
- Who are in the team ?
  - Dedicated
  - Others
- Escalation procedures
- Administrative contacts
- Technical contacts

NDGF
Nordic DataGrid Facility

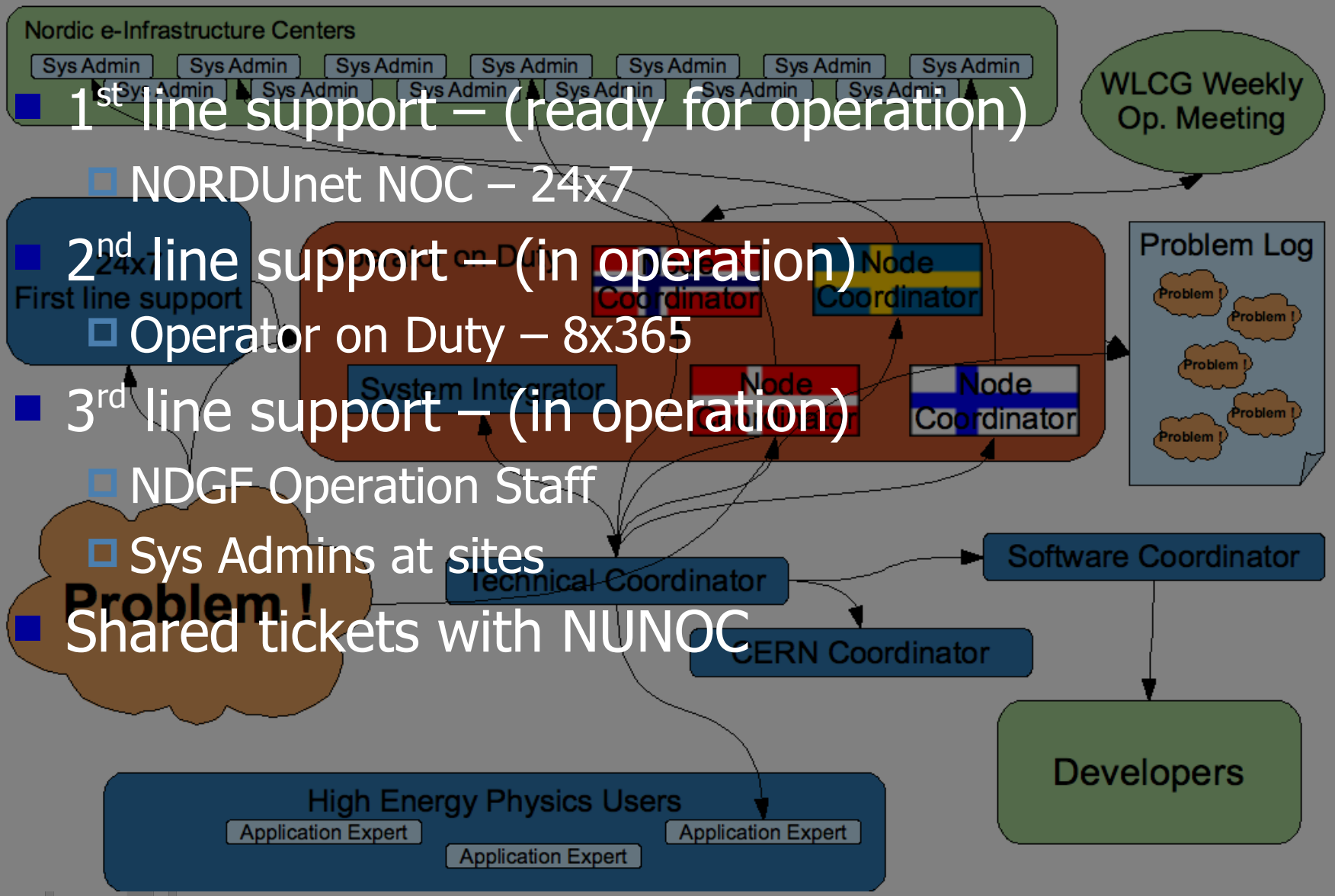
Operation



NDGF
Nordic DataGrid Facility

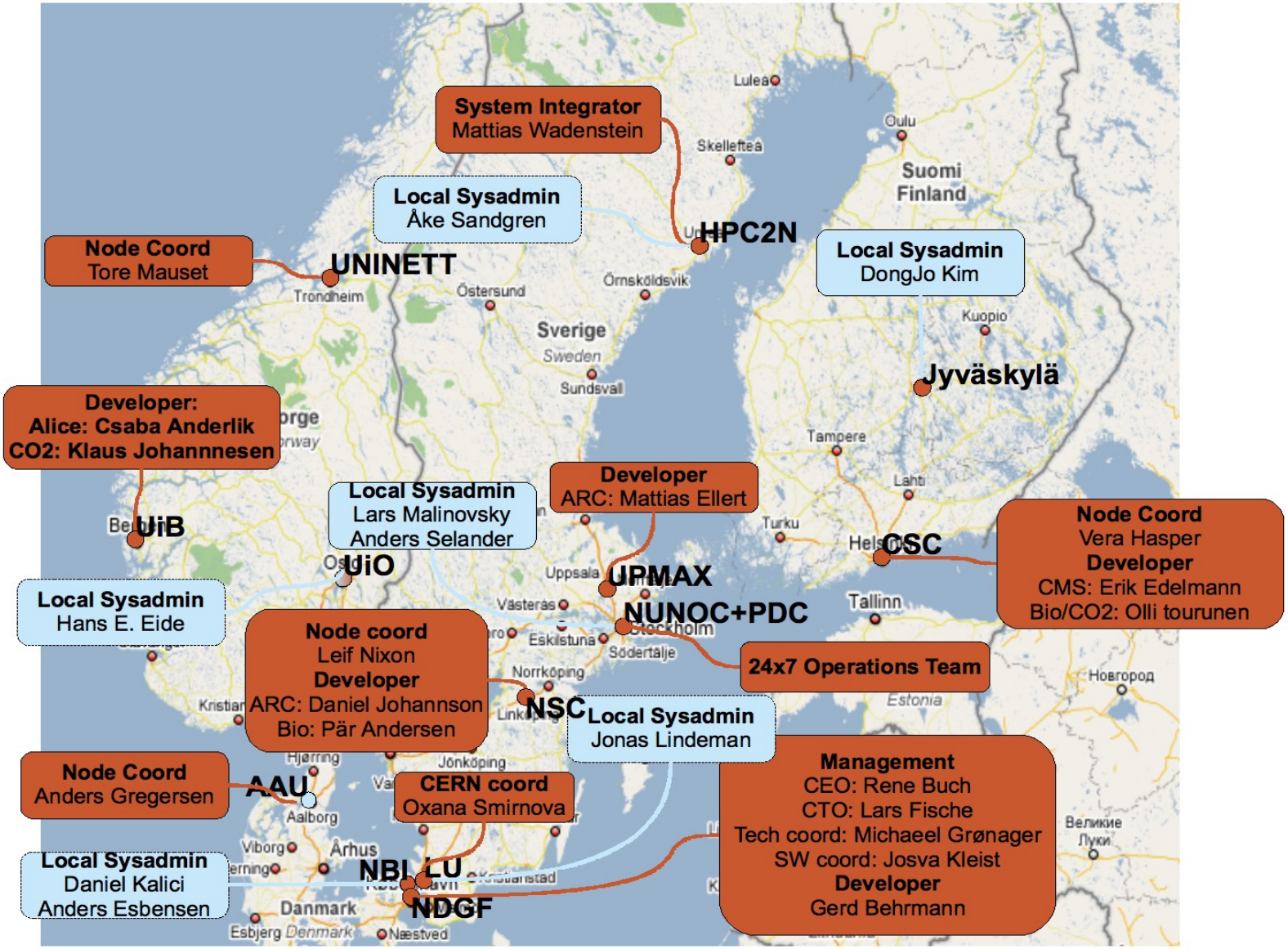
Operation

- 1st line support – (ready for operation)
  - NORDUnet NOC – 24x7
- 2nd line support – (in operation)
  - Operator on Duty – 8x365
- 3rd line support – (in operation)
  - NDGF Operation Staff
  - Sys Admins at sites
- Shared tickets with NUNOC



NDGF
Nordic DataGrid Facility

People



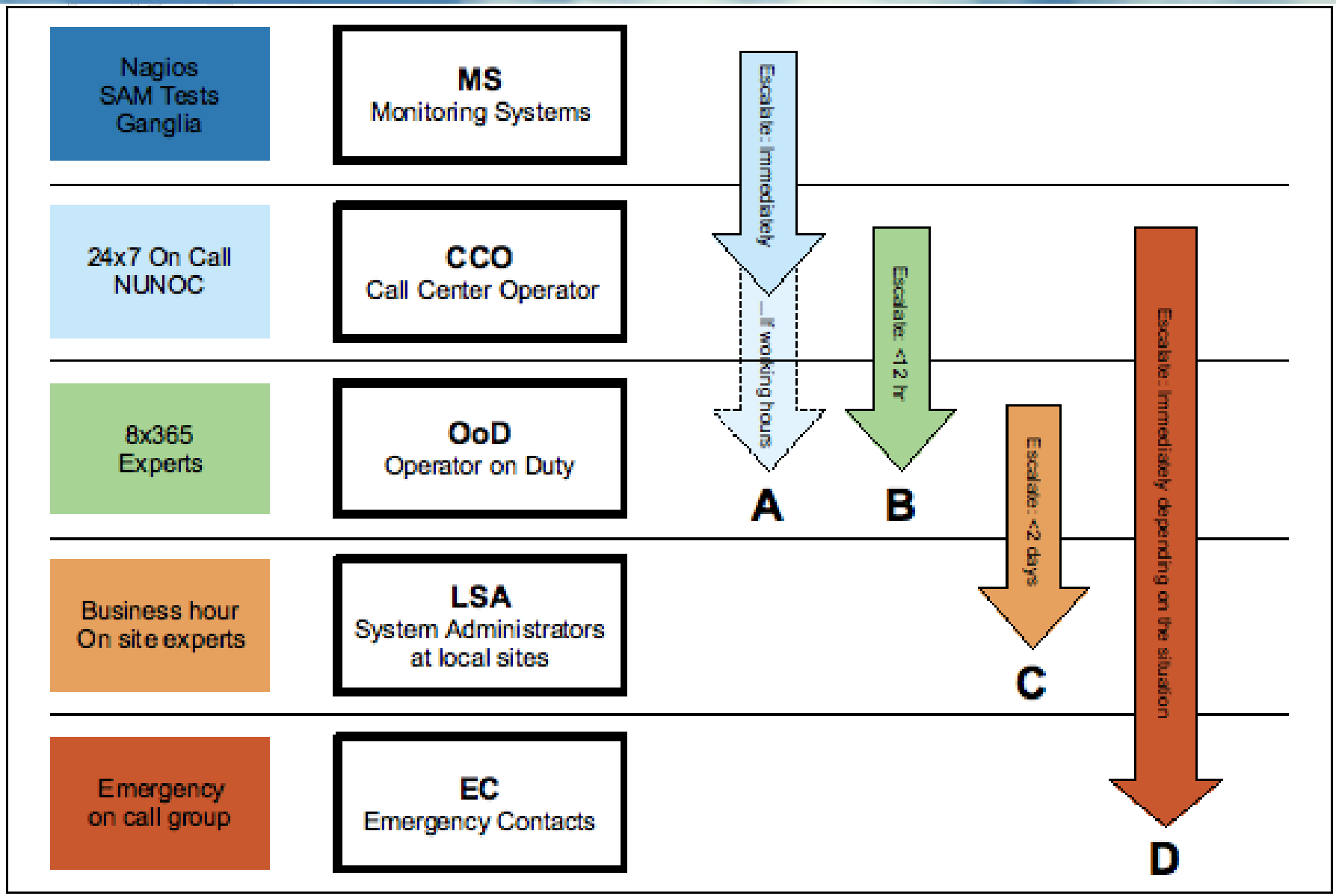
Syste
Ålborg

Procedures

- Who is "on-duty" ?
- Who to escalate to ?
- When are people available ?
- Normal procedures
- Special procedures

NDGF
Nordic DataGrid Facility

Procedures



Demo

- The NDGF infrastructure

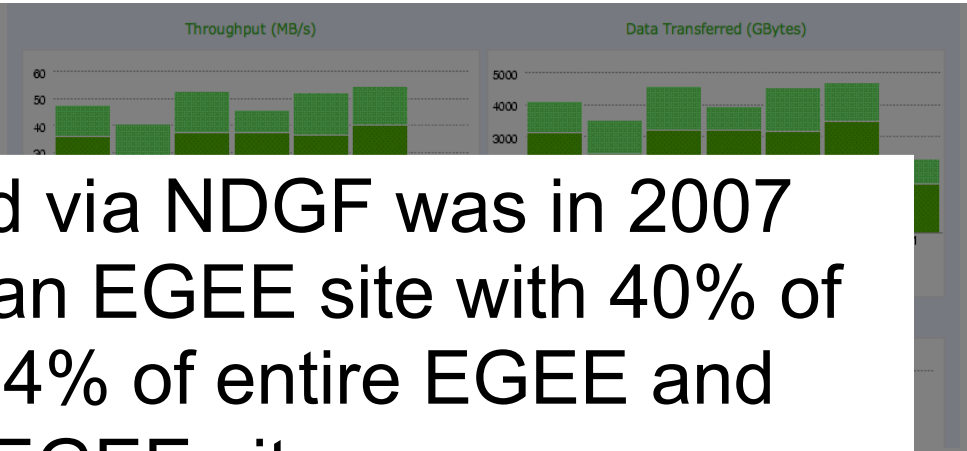
Motivation

- More important than SLAs!
- Ensure a team spirit
  - Meet often
  - Communicate often
    - Phone
    - IRL
    - Mail
    - Tickets

NDGF
Nordic DataGrid Facility

Distribution pays off

- FTS – File Transfer Service

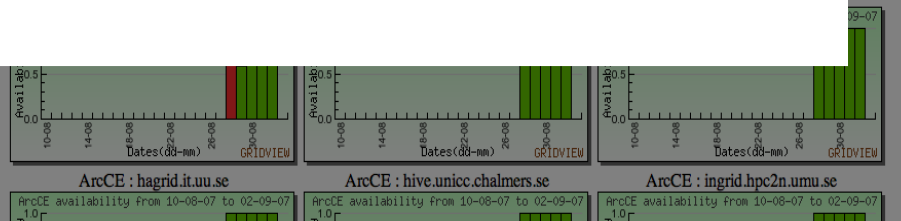


The resources contributed via NDGF was in 2007 the biggest North European EGEE site with 40% of all computations. - that is 4% of entire EGEE and the 5th biggest European EGEE site.

In 2007 NDGF was the most reliable Tier-1 in the world.

In 2007 NDGF was the Tier-1 with highe utilization.

Enabling Grids for E-sciencE



NDGF
Nordic DataGrid Facility

WARNING !

- How to be a bad system administrator :

- Make yourselves irreplaceable
- Don't document
- Don't ask questions
- Adhere the “not invented here” clause
- Attract more task than you can solve
- Never tell other how to solve things – do it for them
- Don't answer your mails
- Stay too busy

Thanks!

Developing a running system

Michael Grønager
Technical Coordinator, NDGF
System Administration and Integration
Aalborg, March 24th 2008

NDGF
Nordic DataGrid Facility

Outline

- Prerequisites
- Definitions
- NDGF examples:
  - ARC
  - dCache

Prerequisites

- Development means changes
- Changes means restart
- non-redundant system:
  - (short) system downtime
- redundant system:
  - no system downtime

Prerequisites

- What is a Grid?:

*A Grid refers to an infrastructure that enables the **integrated, collaborative use** of high-end computers, networks, databases, and scientific instruments owned and managed by **multiple organizations**. Grid applications often involve large amounts of data and/or computing and often require **secure resource sharing across organizational boundaries**, and are thus not easily handled by common Internet and Web infrastructures.*

Prerequisites

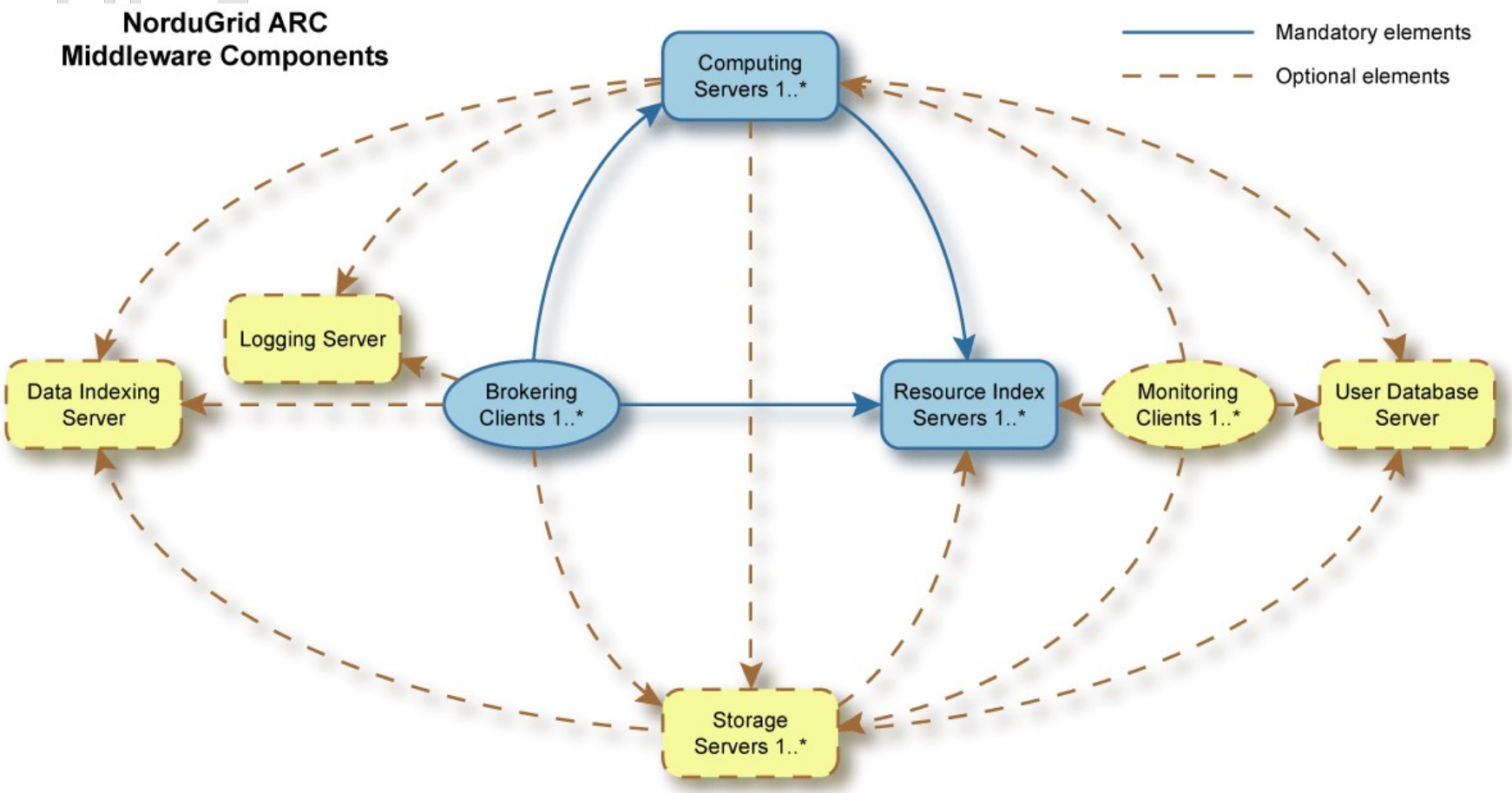
- What is a Grid?:

*A Grid refers to an infrastructure that enables the **integrated, collaborative use** of high-end computers, networks, databases, and scientific instruments owned and managed by **multiple organizations**. Grid applications often involve large amounts of data and/or computing and often require **secure resource sharing across organizational boundaries**, and are thus not easily handled by common Internet and Web infrastructures.*

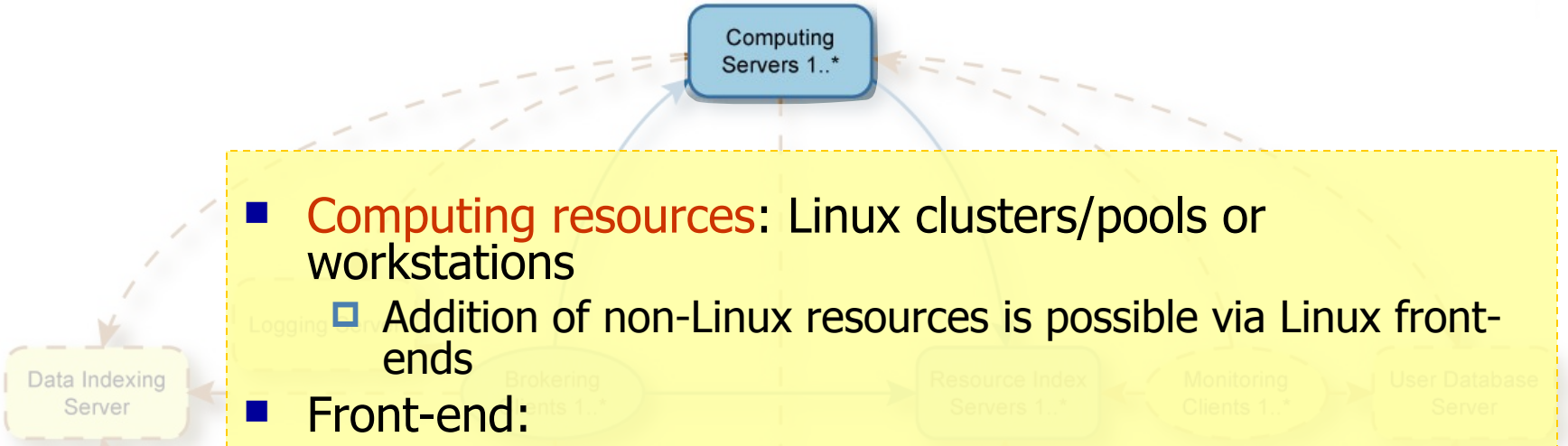
Prerequisites

- Developing in a grid
  - strict interface compatibility
  - several versions must be able to co-exist
    - support for several versions
    - compatible versions
  - Only loose organizational control
- Developing in a tightly coupled distributed infrastructure:
  - allow for non-compatible upgrades
  - upgrade windows with downtime
  - synchronization
  - Tight organizational control

Example: ARC

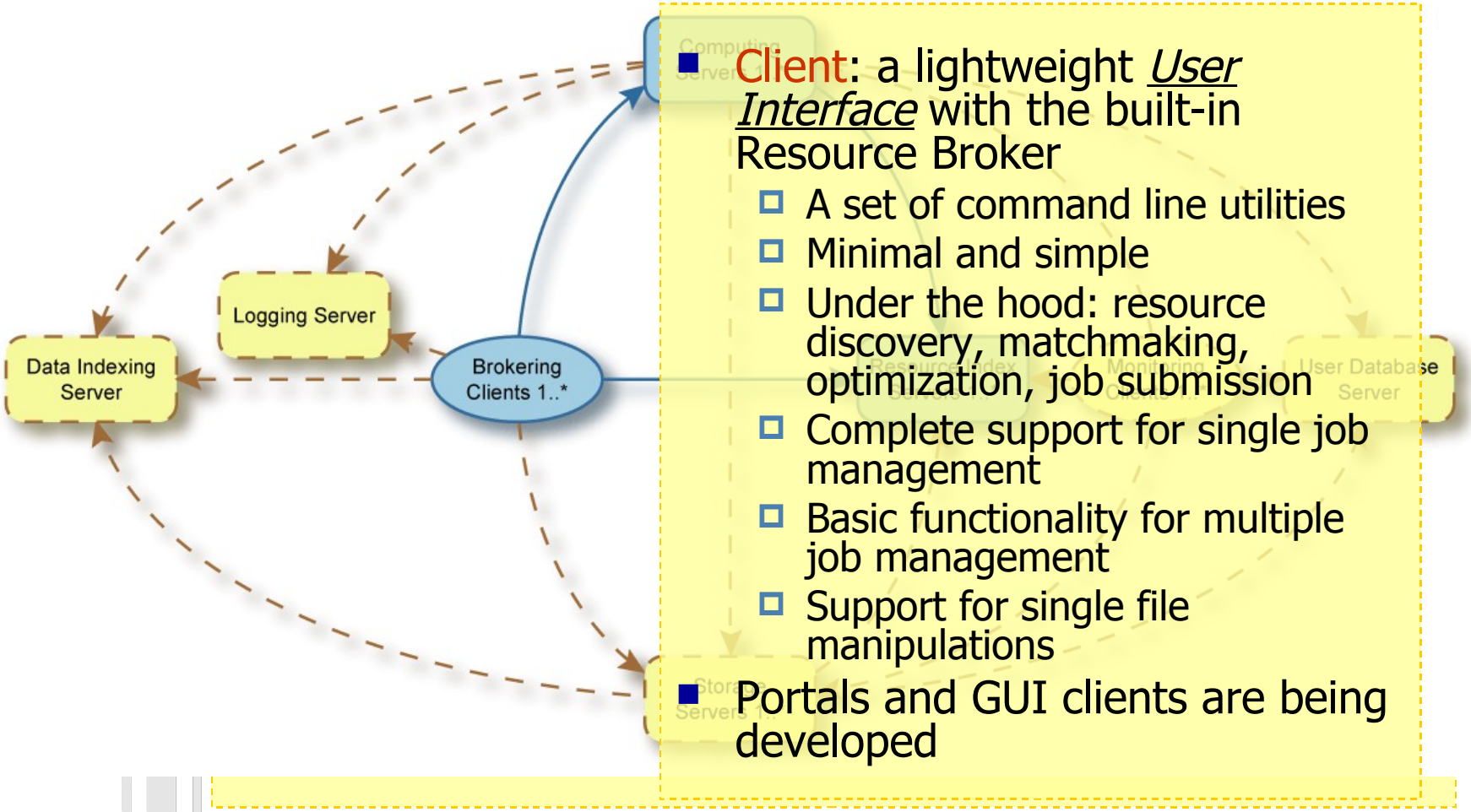

NorduGrid ARC
Middleware Components
Mandatory elements
Optional elements
Computing Servers 1..*
Logging Server
Data Indexing Server
Brokering Clients 1..*
Resource Index Servers 1..*
Monitoring Clients 1..*
User Database Server
Storage Servers 1..*

Example: ARC



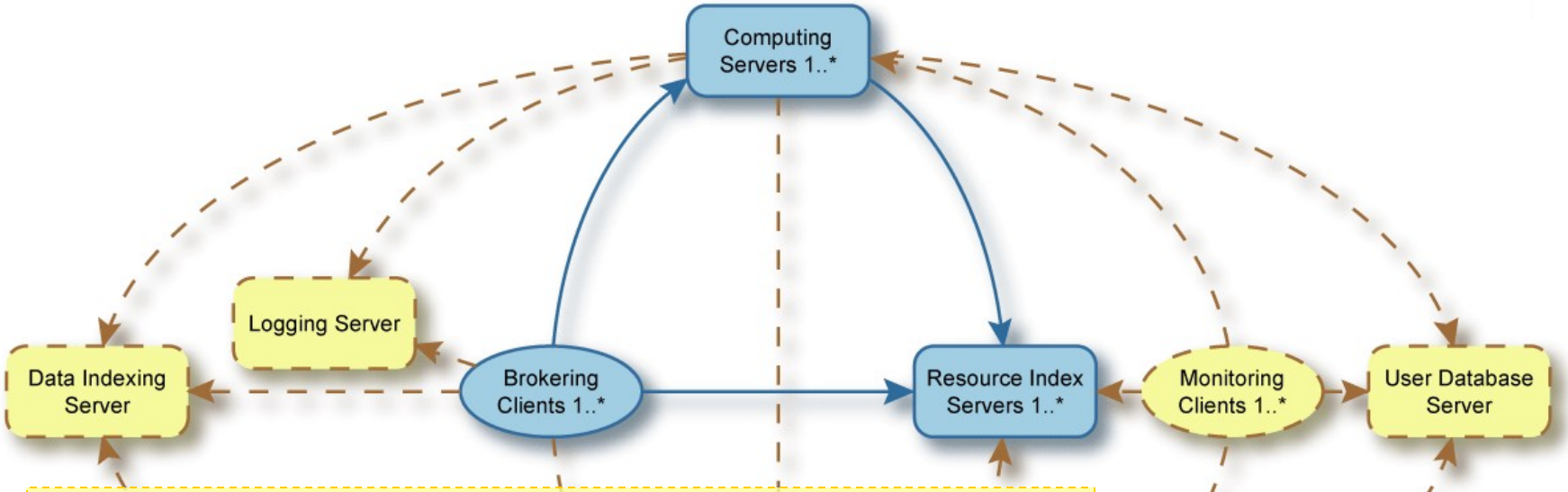
- Computing resources: Linux clusters/pools or workstations
  - Addition of non-Linux resources is possible via Linux front-ends
- Front-end:
  - Runs custom pluggable GridFTP server for all the communications
    - Accepts job requests and formulates jobs for LRMS/fork
    - **Performs most data movement** (stage in and out), **cache** management, interacts with data indexing systems
    - Manages user work area
  - Performs all kinds of job management upon client request
  - Publishes system and job information

Example: ARC



- Client: a lightweight *User Interface* with the built-in Resource Broker
  - A set of command line utilities
  - Minimal and simple
  - Under the hood: resource discovery, matchmaking, optimization, job submission
  - Complete support for single job management
  - Basic functionality for multiple job management
  - Support for single file manipulations
- Portals and GUI clients are being developed

Example: ARC



- Information System: based on Globus-patched OpenLDAP: it uses GRIS and GIIS back-ends
  - Keeps strict registration hierarchy
  - Multi-rooted
  - Effectively provides a pseudo-mesh architecture, similar to file sharing networks
  - Information is only kept on the resource; never older than 30 seconds
  - Own schema and providers

Example: ARC

- ARC is the NDGF job submission system
- Fully redundant services
- Highly decoupled
- Well defined interfaces

Example: ARC

- Services:
  - CE: Compute element – a gateway between grid and resource
    - starts jobs on LRMS
    - stages data in
    - caches data
    - stages data out
    - announce information
  - GIIS: top level info
    - redundant
    - like top level DNSes
  - UI: User interface – not really a service

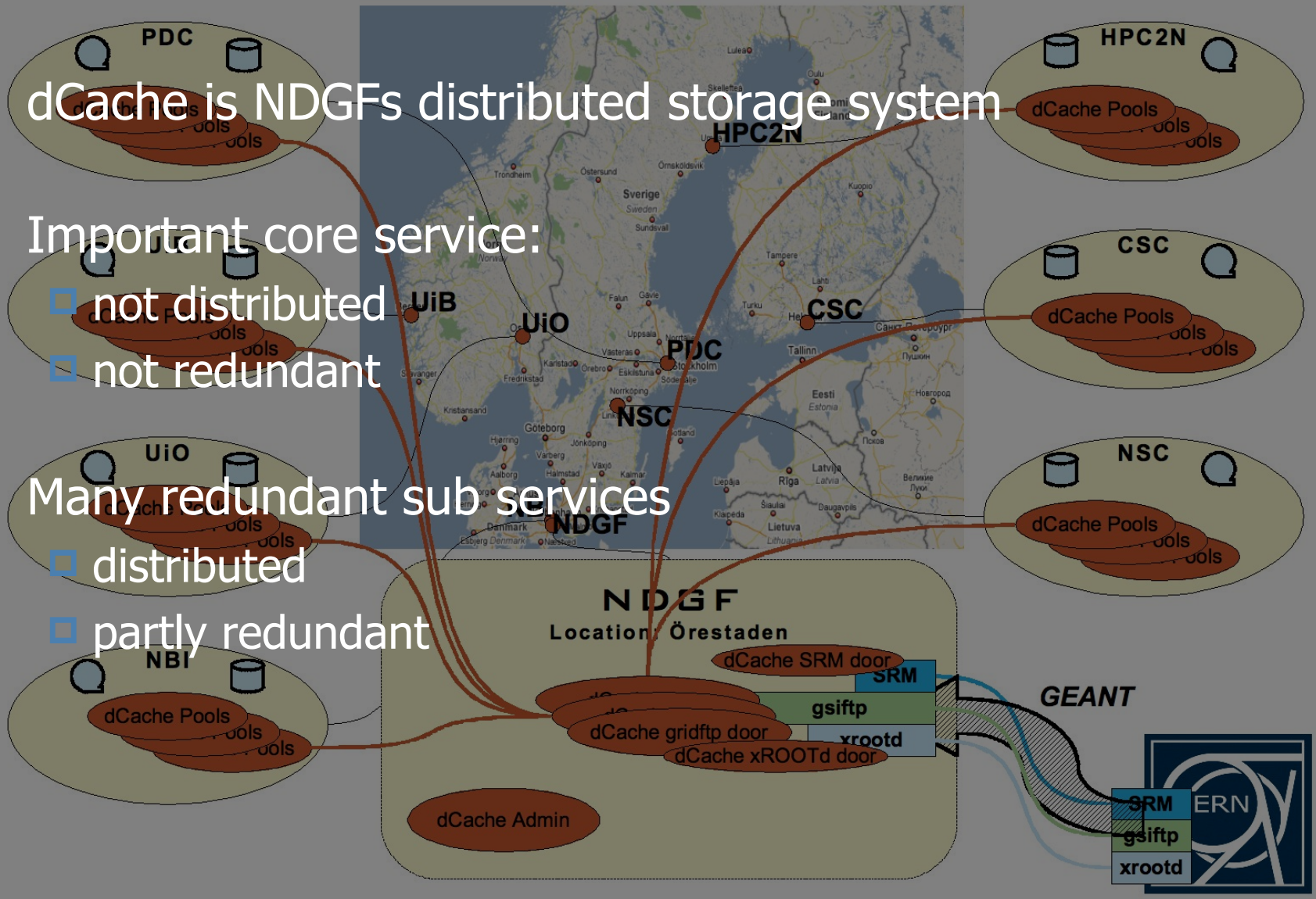
Example: ARC

- Feature additions:
  - trivial – CEs / UIs can be upgraded asynchronously
- Feature changes:
  - requires backwards compatibility
  - requires forward compatibility

NDGF
Nordic DataGrid Facility

Example: dCache

- dCache is NDGFs distributed storage system
- Important core service:
  - not distributed
  - not redundant
- Many redundant sub services
  - distributed
  - partly redundant



Example: dCache

- dCache services:
  - admin database
    - statefull
    - single point of failure
  - doornodes
    - stateless
    - can be heartbeat monitored
  - pool nodes
    - statefull for data reading
    - stateless for data writing

Example: dCache

- Feature additions and changes:
  - tightly coupled
  - organize an upgrade window
  - important to test

Summery

- Non-redundant systems *will* have downtime
- Watch the interfaces !
- Consider the need for a testbed
- and: test, test, test

Thanks!