dCache
in the
NDGF Distributed Tier 1

Gerd Behrmann

Second dCache Workshop

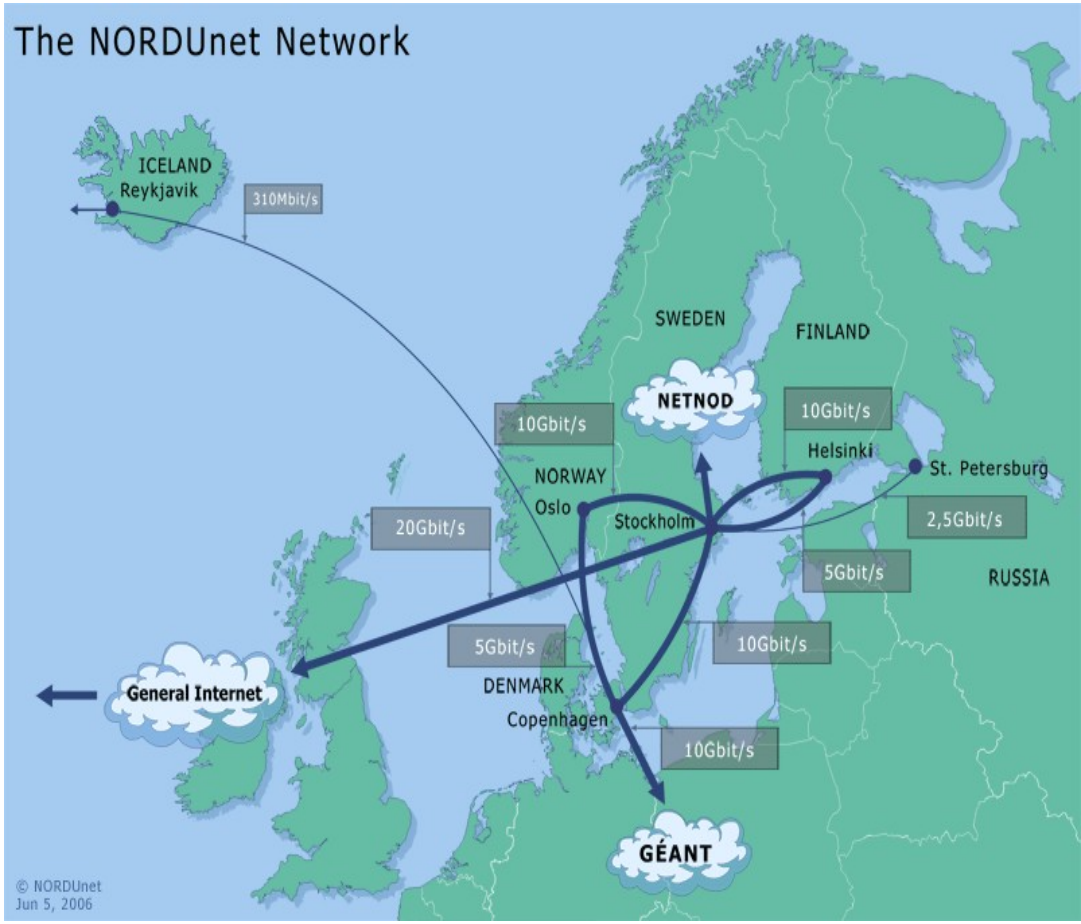
Hamburg, 18th of January 2007

NDGF

Nordic DataGrid Facility

NORDUnet A/S

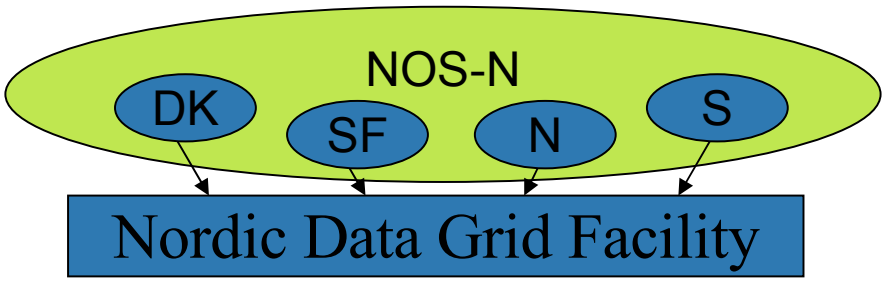
- The Nordic Regional Research and Educational Network (RREN)
- Owned by the 5 Nordic National RENs
- 25 Years of Nordic network collaboration
- Leverage National Initiatives
- Participates in major international efforts
- Represents Nordic NRENs internationally, gateway to the Nordic area



NDGF
Nordic DataGrid Facility

What is the 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 e-Science projects (i.e., Nordic WLCG Tier-1)
  - Develop grid middleware and services
- NDGF 2006-2010
  - Funded (2 M.EUR/year) by National Research Councils of the Nordic countries
  - Builds on a history of Nordic grid collaboration
  - Strategic planning ongoing.



NDGF
Nordic DataGrid Facility

NDGF Vision

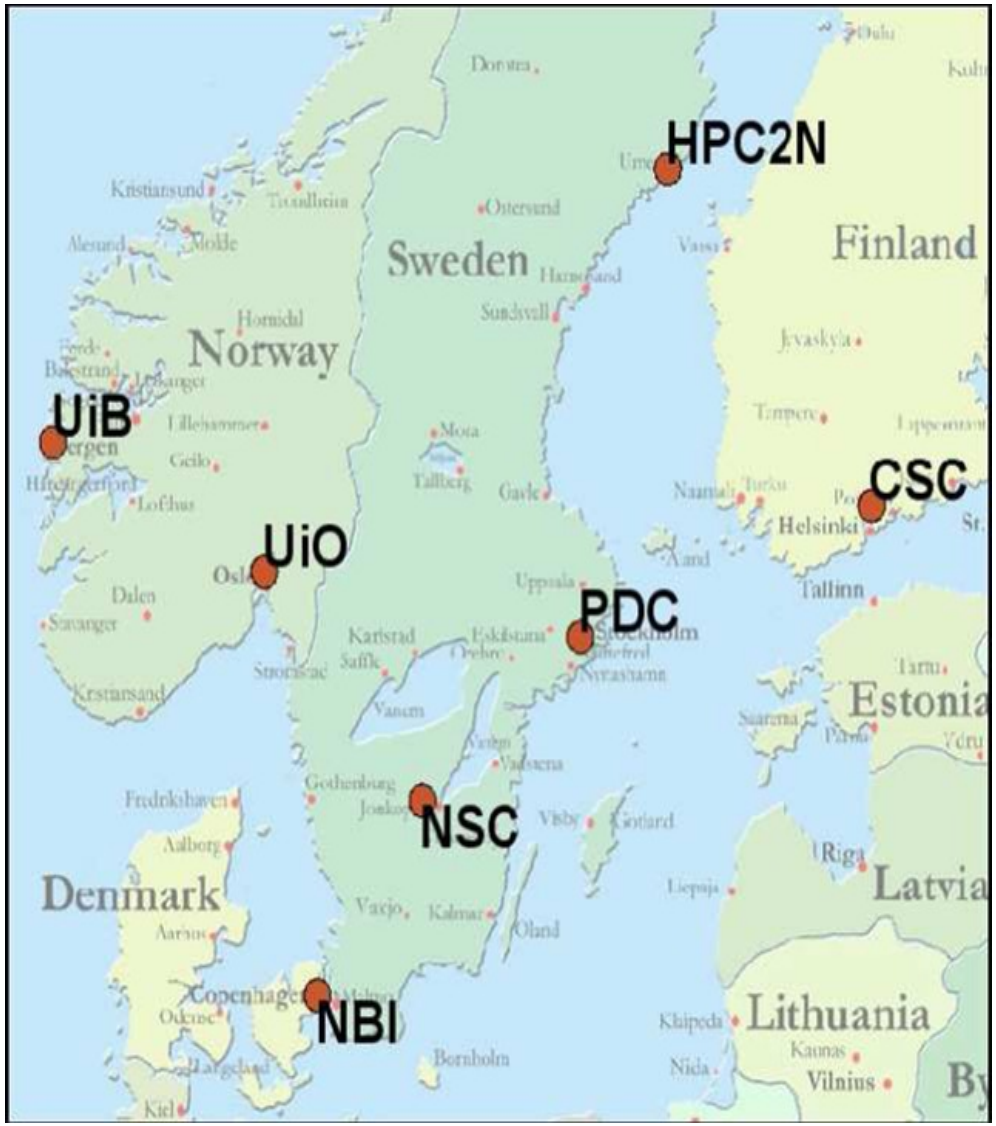
- "...to establish a Nordic Data Grid Facility and to involve Nordic countries in European and global co-operation in data sharing in a variety of fields."
- To *coordinate* and *facilitate* the creation of a Nordic e-Infrastructure sharing platform
- To enable Nordic researchers to participate in major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

- *Think of NDGF as one big Super Computer Center – spanning the entire Nordic area*

NDGF
Nordic DataGrid Facility

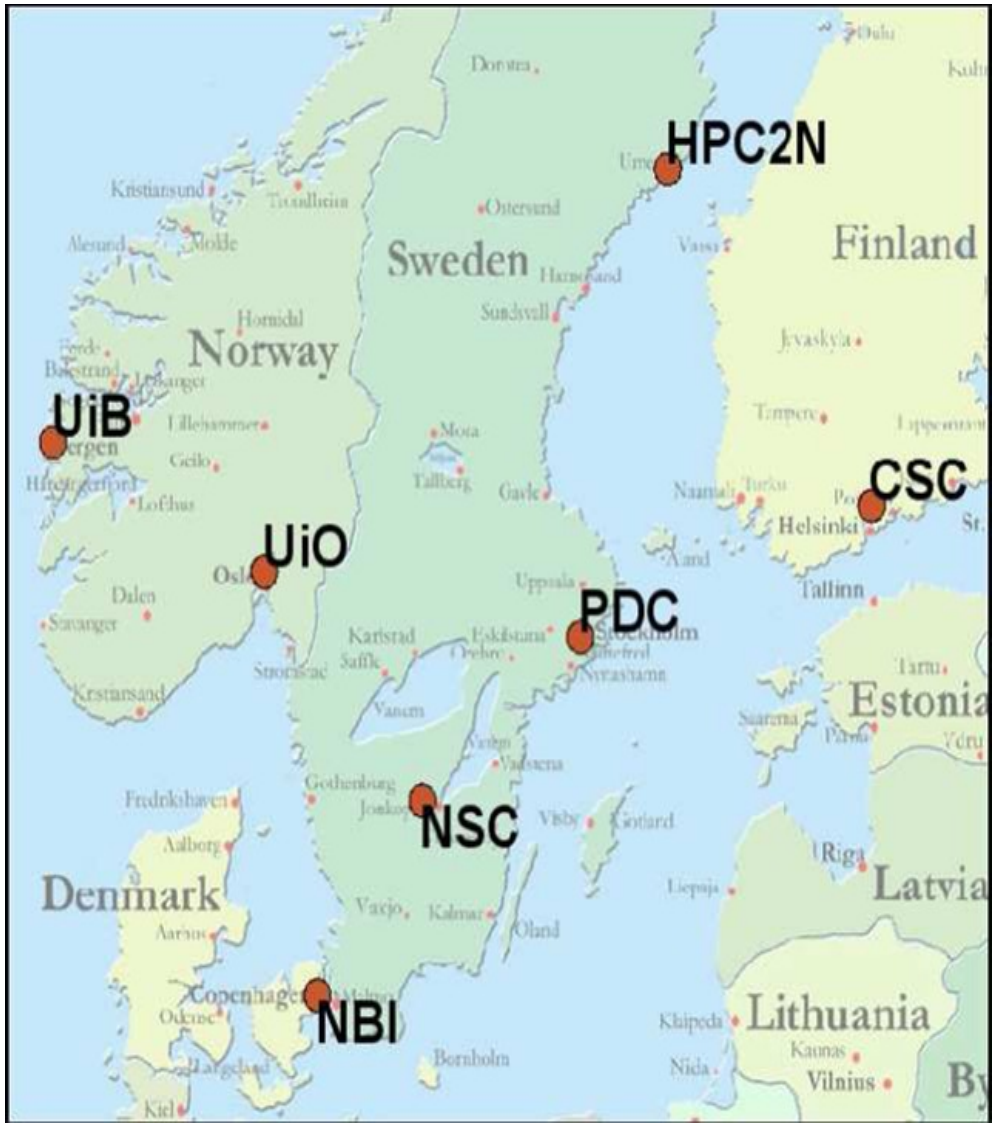
dCache at NDGF

- NDGF Tier 1 sites connected by dedicated 10Gbit fiber (end of 2007/ beginning of 2008)
- Storage resources not located at NDGF and not under direct control of NDGF
- Storage ressources not necessarily dedicated to dCache
- Still, we are required to expose all sites as a single Tier1 with a single entry point.
- Two installations are established, remaining sites have received funding and will be established during 2007.
- No “worker nodes”



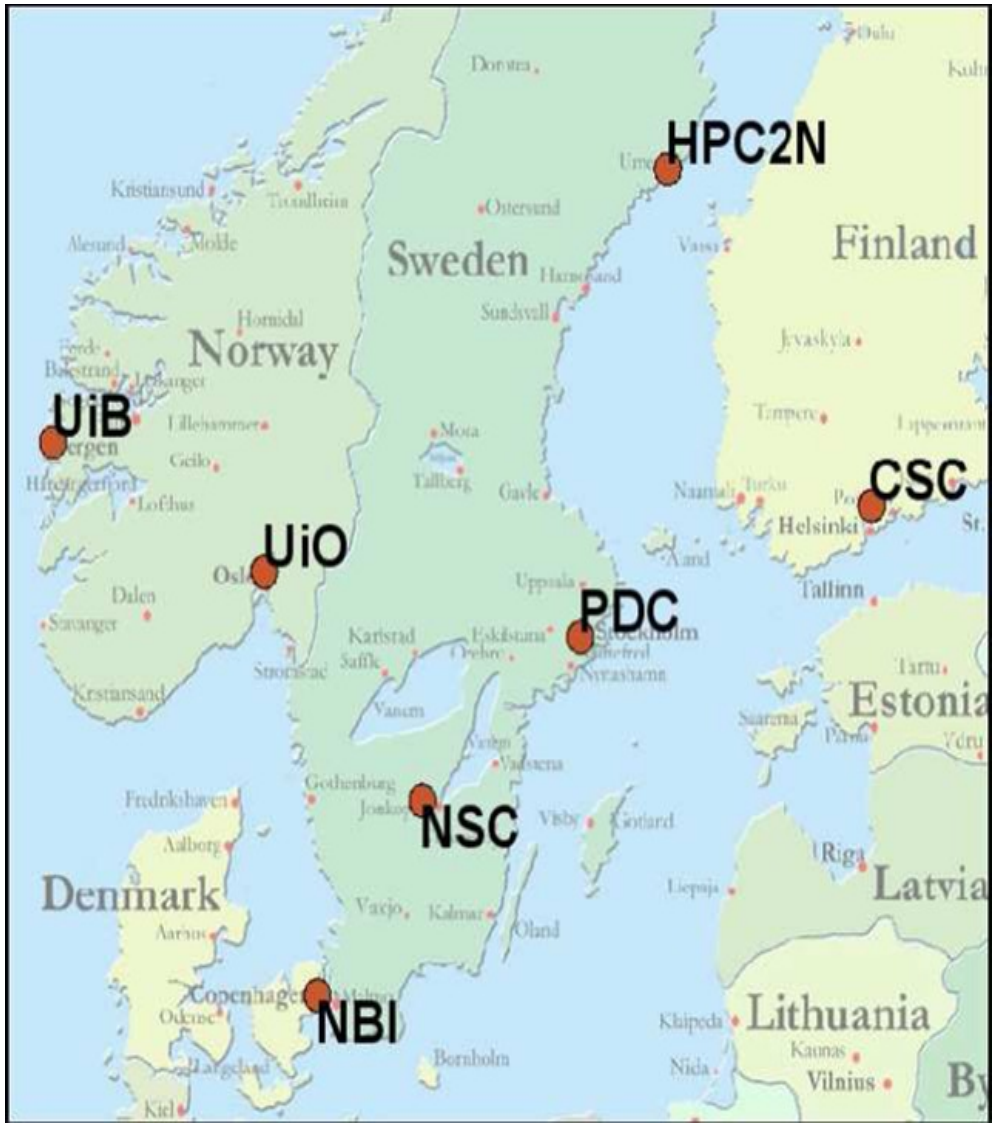
Scenario A

- Central dCache installation with minimal storage (first-level dCache)
- Standalone dCache installation at each site (second-level dCache)
- Use HSM interface to stage data between first-level and second-level



Scenario A

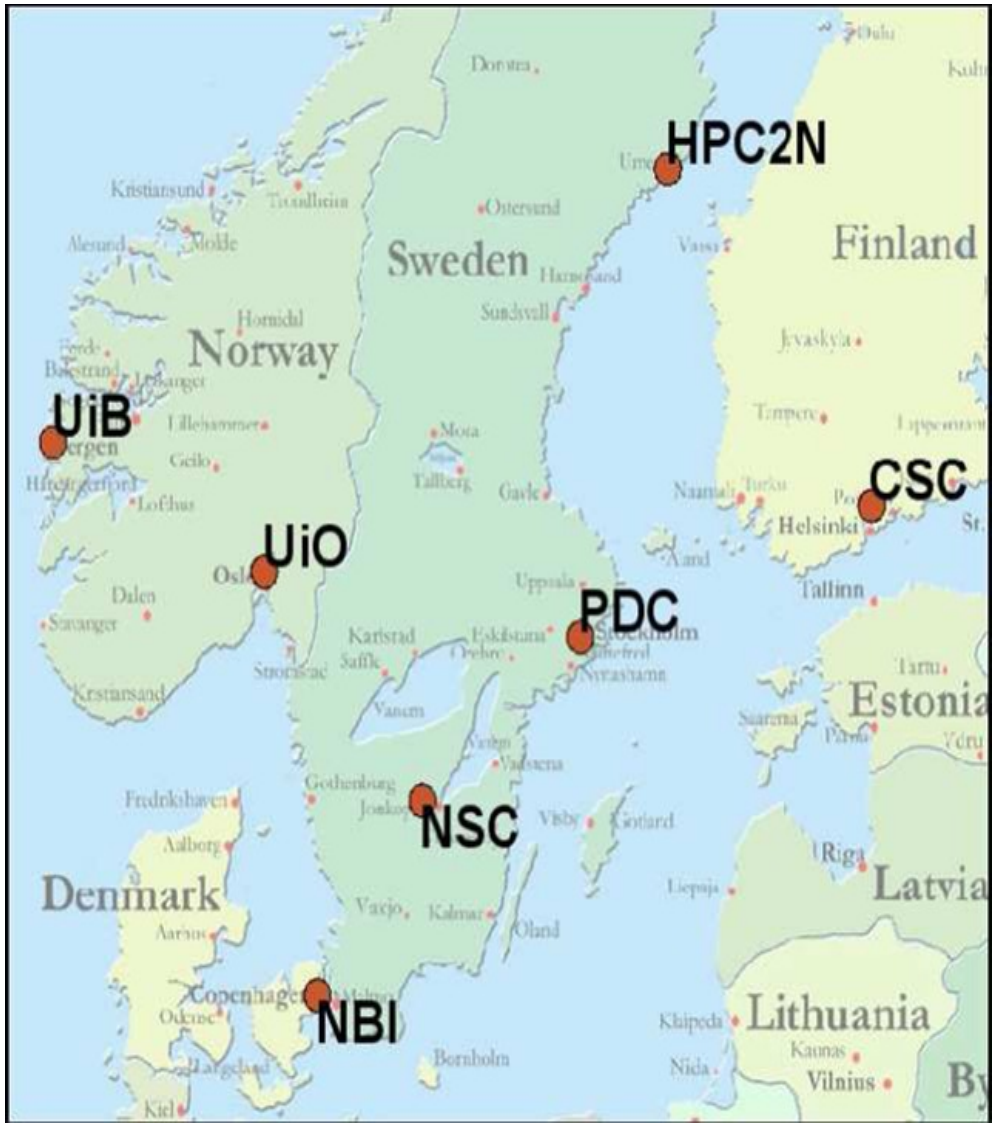
- Problems:
  - Latency
  - Inherently centralized
  - Central buffering required
  - No standalone operation
  - Improvement possible by new type of pool
    - Avoid central storage by storing directly on remote systems



Scenario B

- One uniform dCache spanning all sites
- dCache head nodes operated by NDGF and placed in Copenhagen
  (7 x Dell 1950, Dual Intel Core Xeon, 4GB RAM, 2 x External RAID boxes)
- dCache pools operated by site owners
- Still centralized

- We currently deploy scenario B



Challenges

- Security
  - Many administrative domains
  - Local and national rules
  - Internal node communication over WAN
  - Mounting NFS over WAN is out of the question
- Administration
  - Site administrators are worried about loosing control
  - Mechanisms for delegating control over local ressources

Challenges

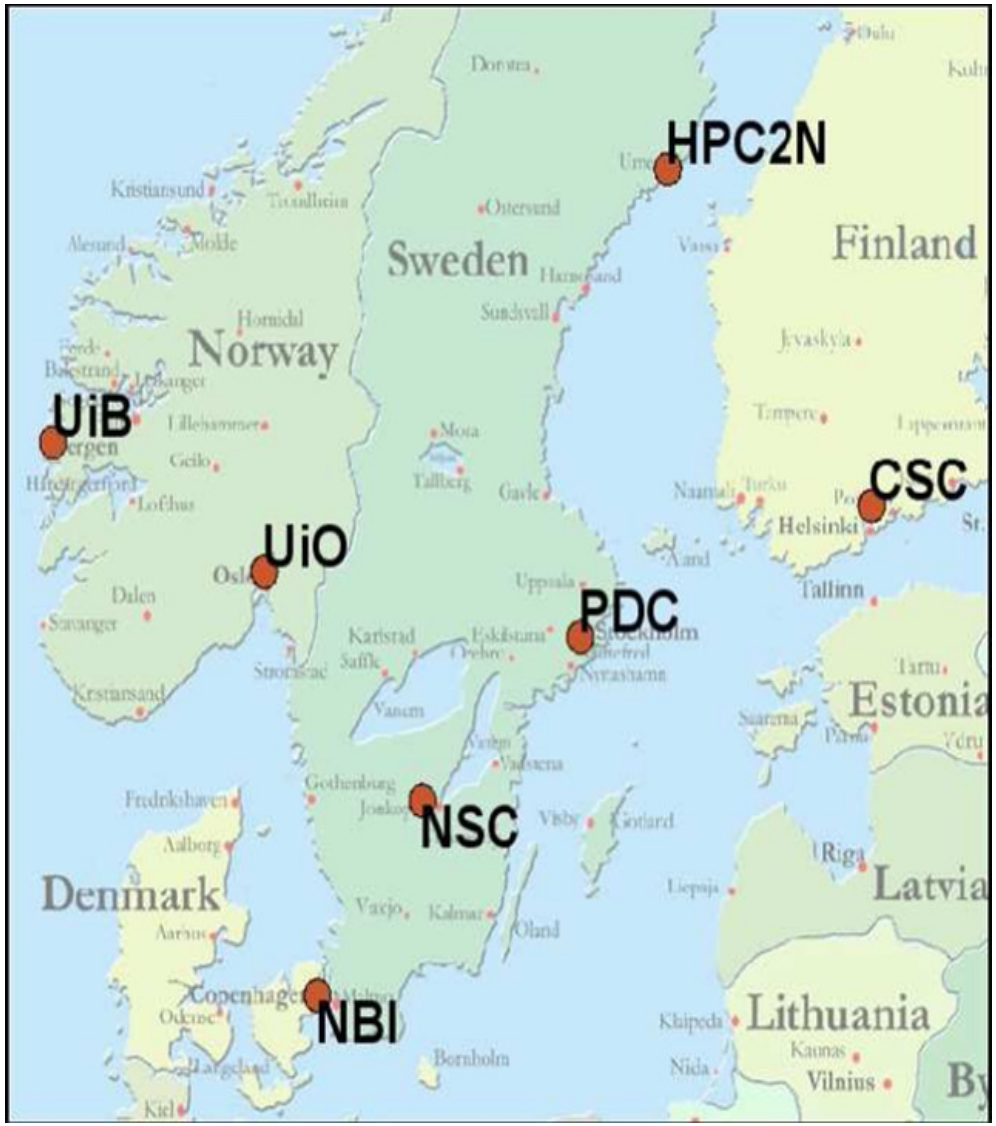
- Maintenance
  - Platform (SL is not widely used in NorduGrid)
  - Upgradability
  - Autonomeous operation
- Reliability
  - dCache is fairly resilient against pool failures
  - Head nodes provide central point of failure
  - Network saparation in WAN
  - Disconnected operation (at least read-only)
    - Brain dump ideas: Replicated name space provider, DHT
  - Long term hope that dCache becomes less centralised

Challenges

- Performance
  - No network model
    - e.g. SRM door assummes all GridFTP doors are equal (except for current load)
  - Proxy operation of GridFTP
- Functionality
  - HSM without PNFS (done in head)
  - Heterogenous access to HSM
    - Stage-in must happen to connected pool
    - Tape0Disk1 -> Tape1Disk1 may require file migration to another pool
  - Tivoli (TSM) integration
  - User friendly view of logical name space without PNFS (beyond FTP access and beyond admin shell)

Problems with (Grid)FTP

- PASV/PORT before STOR/RETR
  - dCache in PASV mode will always use door as proxy
- In mode E (extended block mode), the sender is always active.
  - Uploads to dCache always use door as proxy



Solution A

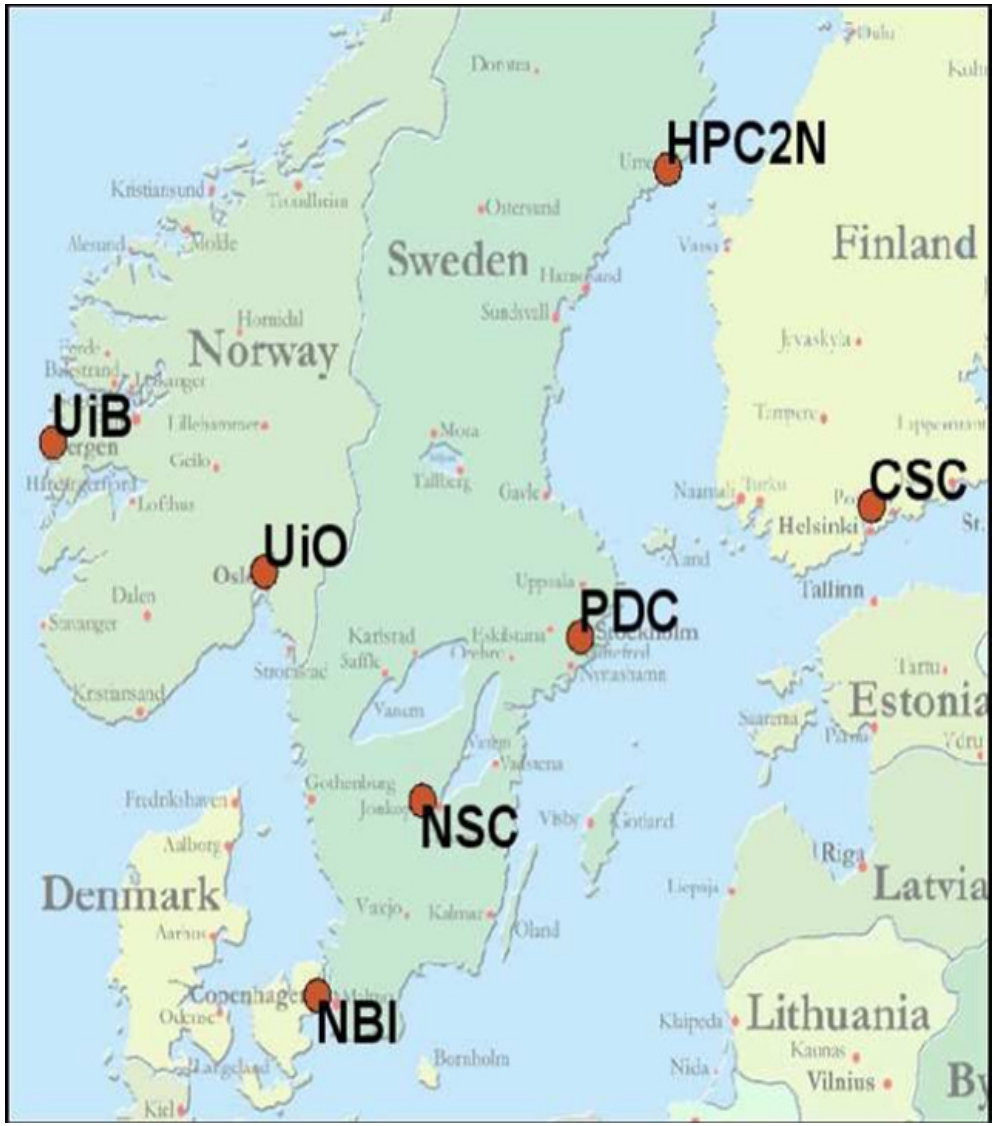
- SRM performs SURL to TURL translation
  - Currently TURL points to some GridFTP using LFN.
- Instead, let TURL be:

  gsiftp://pool:port/pfn

  i.e. let TURL point directly to where the data is stored/is supposed to be stored.
- Consequences:
  - SRM door gets more work to do, and
  - we need to embed an FTP implementation in a mover
  - could make it possible to queue requests in SRM rather than in pools

Solution B

- Use a proxy, but do not place it at the door
- Problems
  - Requires a network model
  - Assumes that control channel and data channel is established from the same location



Solution C

- GridFTP version 2
  - Draft, GFD-R-P.047
    Mandrichenko, Allcock, Perelmutov
- Relevant highlights
  - GET and PUT commands replace PASV/PORT and STOR/RETR:

    GET file=/foo/bar;pasv;mode=e;

    127 PORT=(130,225,33,7,156,7)
  - New eXtended block mode: Mode X
    - No restriction between direction of connection and direction of transfer, e.g. sender can be passive.
    - Number of concurrent connections can adapt; check sum on blocks; concurrent transfers on shared data channels...

Solution C

- Problems:
  - Draft status. Some clarifications needed, IMO.
  - No signs of progress since June 2005.
  - No implementations, except
    - dCache head has GETPUT feature and mode X is on my laptop
    - Started developing a patch for GLOBUS.

Commitment

- 1-2 FTE for dCache development and initial deployment.
- As long as needed and as long as dCache moves in “the right direction”.
- We avoid long term promisses. Concrete development plans are only made for one month at a time.
- We currently focus on GridFTP2 (solution C). Next in line will likely be solution A and/or HSM issues.

Summary

- WAN deployment of dCache at the NDGF distributed Tier 1
- Provides unique and interesting problems
- NDGF is committed to contribute to dCache to resolve these problems ... and continue to do so as long as we have the need and dCache continues to move in “the right direction”.
- Current focus is on immediate problems with data flow, HSM, security and administration.