

## dCache, the Project

People, Funding, Plans and stuff

by patrick FUHRMANN

and

*Antje, Dmitry, Gerd, Irina, Jan, Paul, Tanja, Tigran,  
Timur, Thomas, Owen, Paul, Kai.*

# Content

## ➤ dCache, the project

- ✧ The Partners
- ✧ Then Funding
- ✧ European Middleware Initiative
- ✧ The Process

## ➤ dCache, the customers

- ✧ At DESY
- ✧ dCache at WLCG
- ✧ dCache in the future
- ✧ Collecting requirements

## ➤ dCache, the system

- ✧ Design
- ✧ Managed Storage
- ✧ Protocols

dCache, the project  
Partners

DESY [8]



Management: Quality, PR, Release

dCache.org infrastructure: Communication, Repository, Release

Testing: Functional, Stability

Development: NFS 4.1, Web-Interface, everything else

FERMIlab [2]



Development : SRM, gPlazma, Resilient Manager

Contact : Open Science Grid

Nordic Data Grid Facility (NDGF) [2]



Development : Whatever NDGF needs (essentially everything)

Contact : Nordic Countries

# dCache, the project Funding

Labs

**DESY**  
**FermiLab**

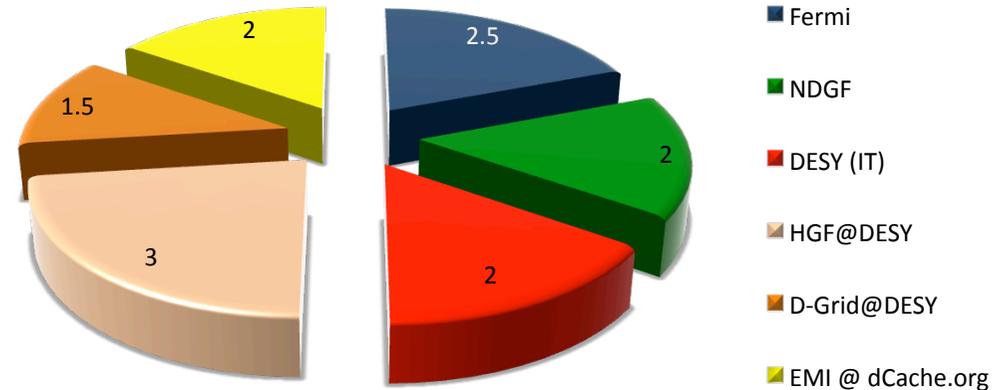
Organizations

**NDGF**  
**European Middle Initiative, EMI**

Open Science Grid (US) [no funding, only first level support]

German Government

**Helmholtz Alliance, "Physics at the Terra Scale"**  
**German D-Grid, "Integration Project II"**



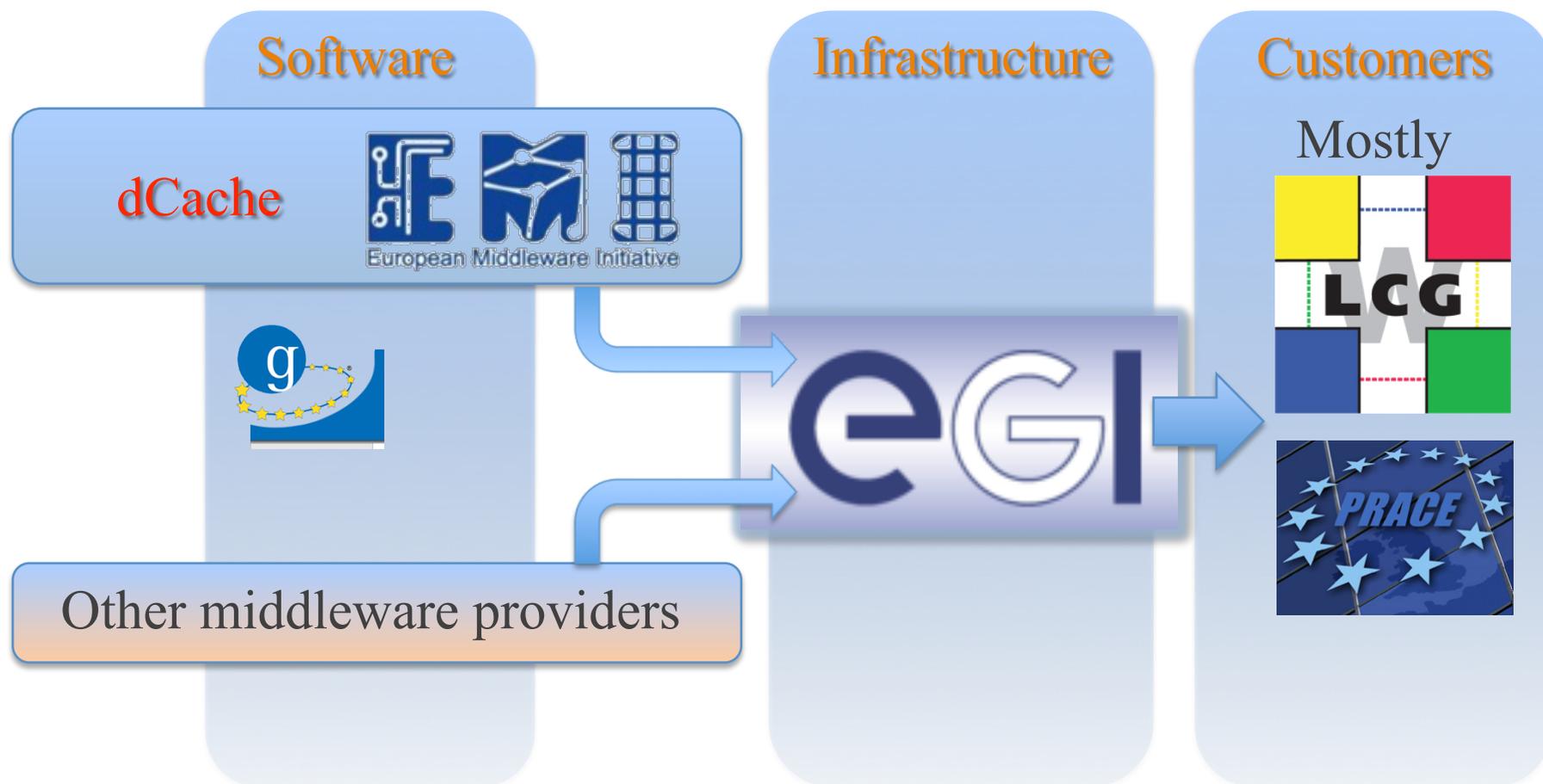
For at least 6 team members, funding is save for another 3 year.

dCache, the project  
Funding

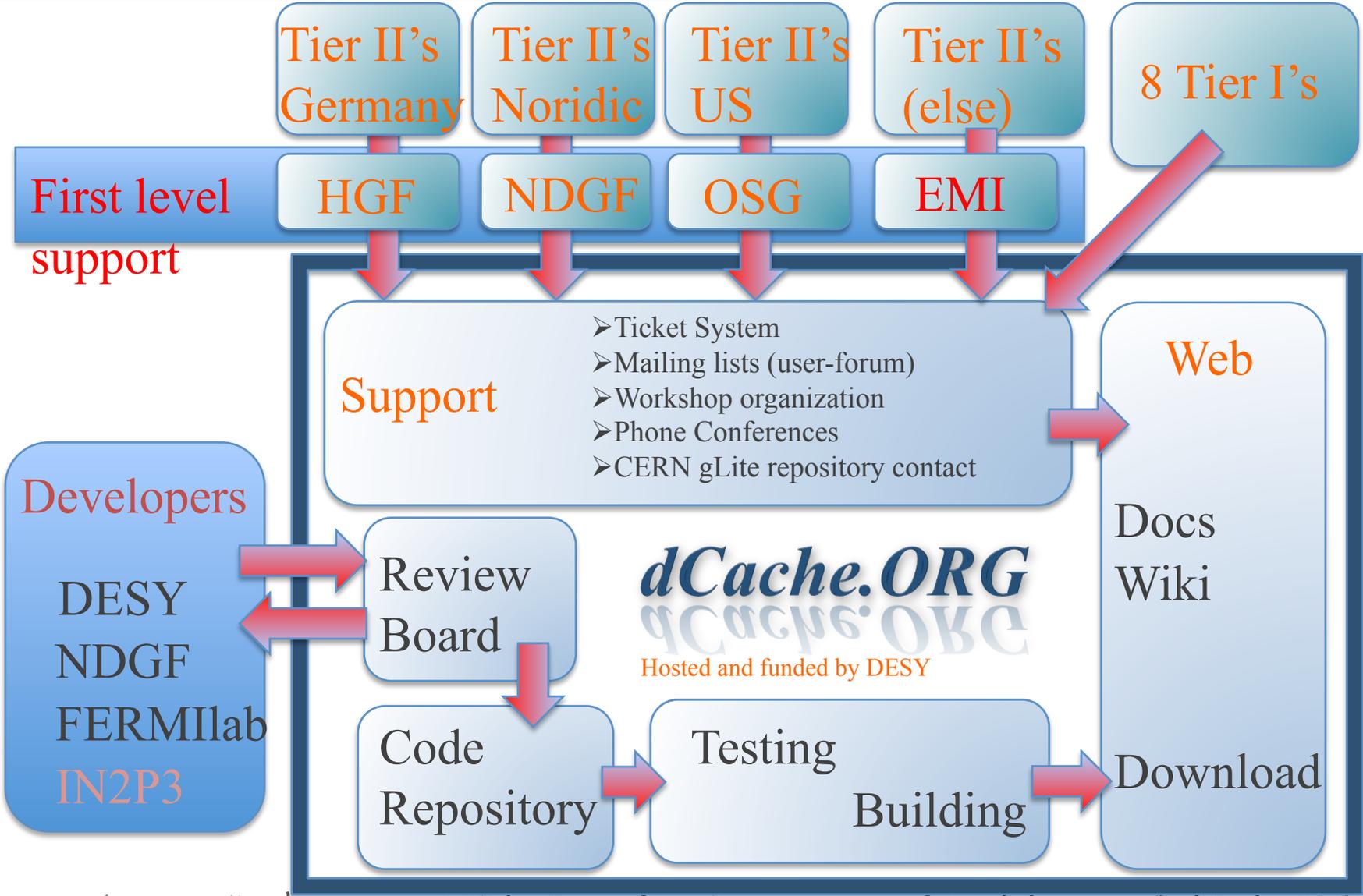
DESY invests 2 FTE for dCache development  
and gets 10 in return.

dCache, the project  
The European Middleware Initiative.

EMI is more than funding.  
It provides a deployment chain to the customer (distribution)



dCache, the project  
The process



dCache, the customers  
Customers

## The customers



dCache, the customers

## Customers at DESY

HERA

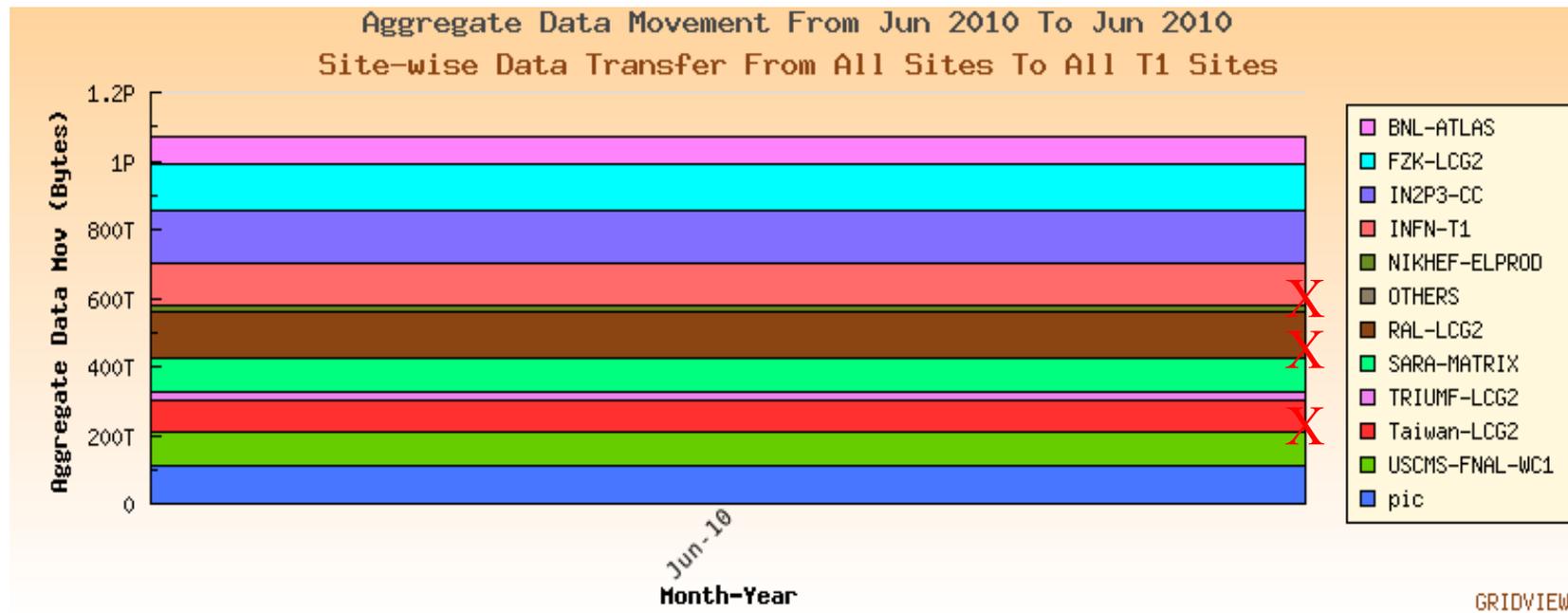
ILC

See Birgit's presentation in a bit

dCache, the customers  
dCache in WLCG

Certainly most impressive

- ✓ dCache stores more than ½ of all WLCG data
- ✓ in about 40 Tier II's and
- ✓ 8 out of 11 Tier I's
- ✓ Data stored in dCache (worldwide for WLCG) approaches 20 PBytes



dCache, the customers  
dCache, new customers



new customers



new requirements

So we tried to find-out what our new customers need  
(Not so much what they want)

dCache, the customers  
dCache, recent, future customers



Planned to utilize DESY storage facilities.



Is using dCache at the Netherland Tier I  
in Amsterdam and in Jülich.



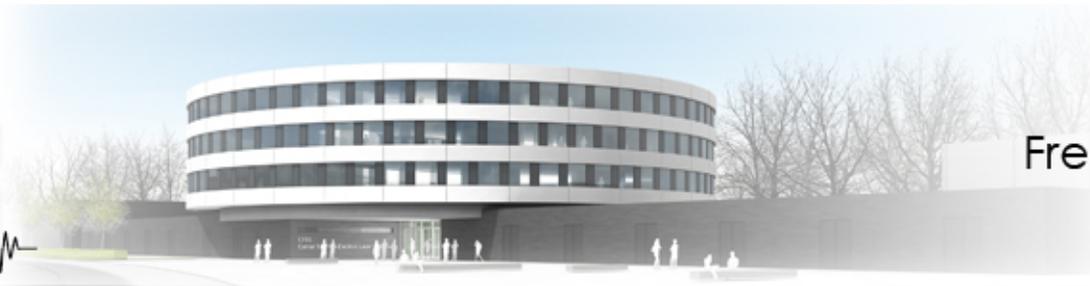
Would like to utilize the Swedish  
dCache Tier II facility.

dCache, the customers at DESY  
European XFEL and CFEL



## THE COMPANY

The European XFEL is being constructed by the European XFEL GmbH. For the construction and operation, the company collaborates closely with DESY and many other institutions.



Center for  
Free-Electron Laser  
Science

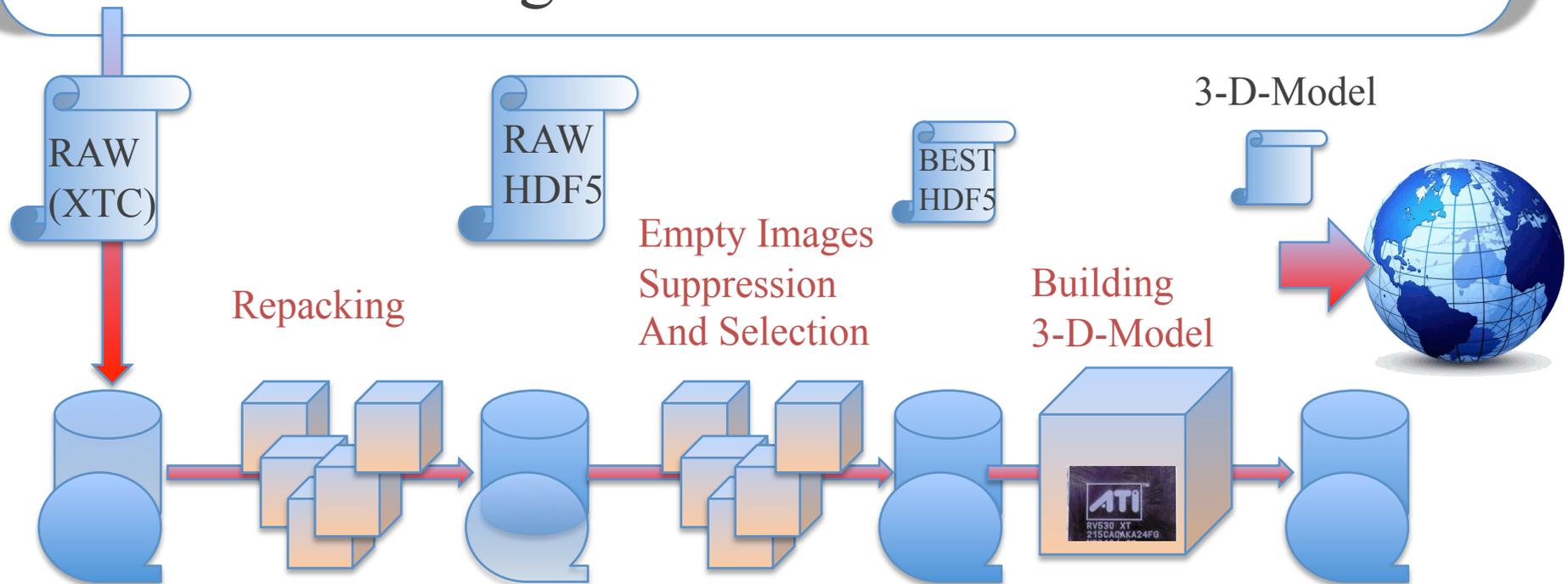
# The Center of Free-Electron Laser Science, CFEL

dCache, the customers  
CFEL as an example

Stolen from anton BARTY



## Free Electron Light Sources



dCache, the customers  
LOFAR

# The International LOFAR Radio Telescope

(The first software telescope)



Information provided by

hanno HOLTIES, LOFAR

dCache, the customers  
LOFAR

As of Feb 24, 2010 :

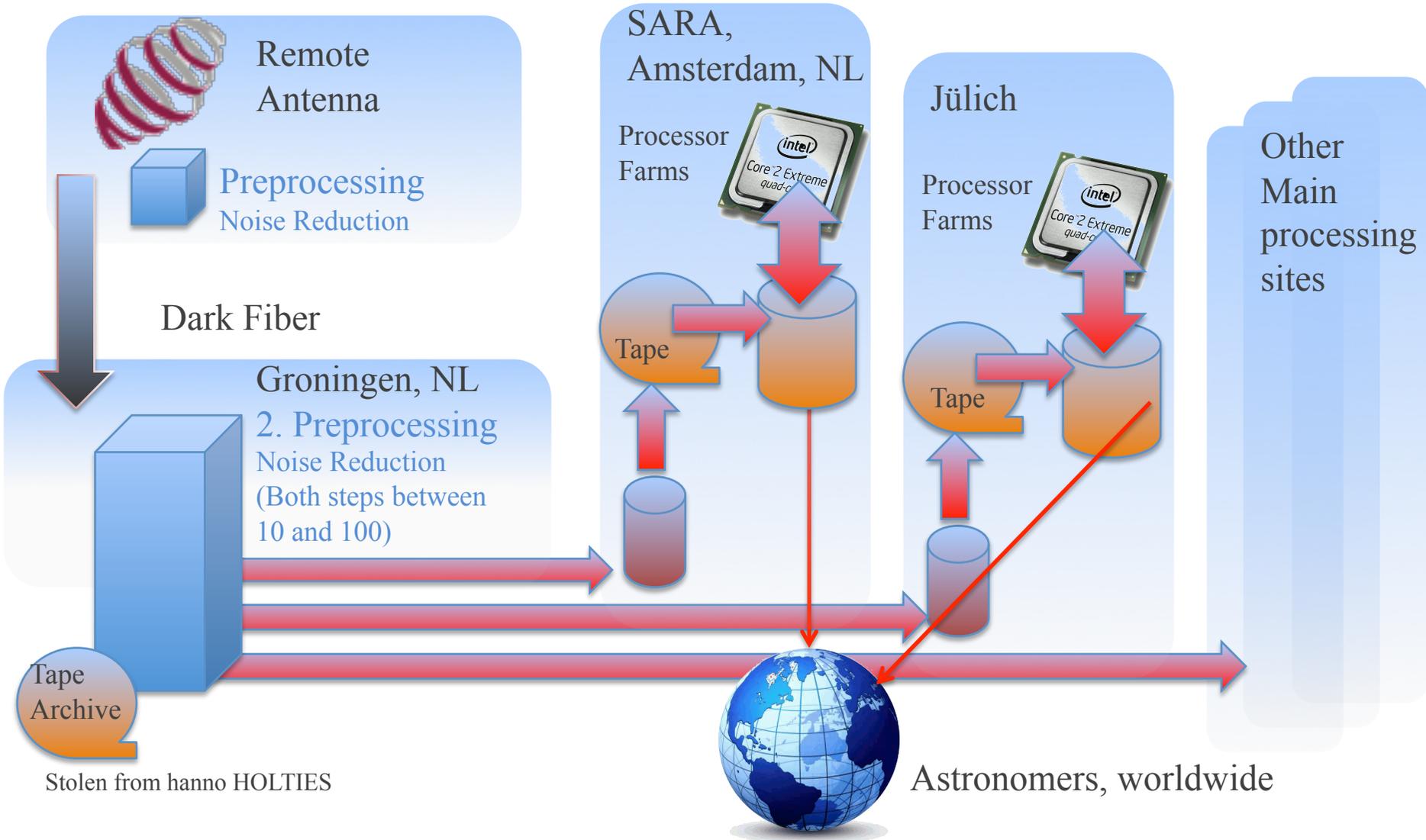
- ✧ 21 Complete Stations
- ✧ 10 In Progress
- ✧ 13 Planned
- ✧ NL, DE, UK, FR, SE



June 14, 2010:

Stolen from hanno HOLTIES

dCache, the customers  
LOFAR, data flow model



Stolen from hanno HOLTIES

dCache, the customers  
SNIC

# SNIC

## Swedish National Infrastructure for Computing

Information provided by

tom LANGBORG, SNIC

<b>Uppmax</b>	Uppsala Multidisciplinary Center for Advanced Computational Science
<b>Lunarc</b>	scientific and technical computing for research at Lund University
<b>HPC2N</b>	High Performance Computing Center North
<b>C3SE</b>	center for scientific and technical computing at Chalmers University of Technology in Gothenburg
<b>NSC</b>	National Supercomputer Center in Linköping
<b>PDC</b>	Center for high performance computing

SNIC National storage is an infrastructure for archiving data.

## Swestore Project Jan 20, 2010

Create an infrastructure for storage for Swedish Research and Swedish Universities.

### Planned Data Access

“**SRM, WebDav** and **gsiFtp** are examples of protocols for communicating with the National Storage. Authentication method are **X509 Certificates. Kerberos** could be used in some special cases” , Tom Langborg, SNIC

Internal	External
SRM	SRM
gsiFtp	gsiFtp
WebDAV	WebDAV
NFS 4.1	Web Portal/Gateway

Stolen from tom LANGBORG

dCache, the customers  
Some work to do

## Translating the collected requirements into our language



dCache, the system  
dCache, requested features

## Sys-admin

### Managed Storage

Manual Storage Control  
(Retention Policy,  
Access Latency)

SRM

Automatic Storage Control

### Unified

Identity  
Management

ACL's

## User

### Access Protocol

Standard Protocols e.g.

NFS 4.1 (native mount)

http(s)/Web-Dav

gridFtp

dCache, the system  
dCache, high level design

Planned

Standard File Access Protocols

http(s)  
WebDav

NFS 4.1

gsiFtp

CDMI (SNIA)  
Cloud Data  
Management  
Interface

Storage  
Management

SRM

Extended  
By  
Load Control

Common Security Layer

Authentication : Kerberos, X509, Password

Unified ID management

Authorization : ACL's for File system and storage control (SRM)

Callouts  
To external  
ID services

Common Name Service Layer

Extended Names Service Queries (SQL)

DISK

DISK

SSD  
SSD

Tape

“multi-media” storage layer

## GRID protocols, the boring bit

- **gsiFTP** : Wide area transfers
- **dCap (gsidCap)** : Posix Like File Access (client library required)
- **Xrootd** (Alice security) : Posix Like (client library required)
- **SRM** : Storage Resource Manager (Storage Control)

dCache, the system  
dCache, standard protocols

## Standard protocols : The cool bit.



Message :

New communities require access to their data by standard protocols or clients which are coming with their different OS'es.

dCache, the system  
dCache, WebDav

## ➤ WebDav

- Linux (Gnome, KDE)
- Windows
- Mac OSX
- Browser

Available in recent dCache releases

dCache, the system  
dCache, NFS 4.1

## ➤ NFS 4.1

Available in recent dCache releases

## Compared to WLCG protocols (dCap/rfio/xroot)

- NFSv4.1 is an industry standard (will be supported by many vendors)
- NFSv4.1 clients are provided and maintained by other people
- Client caching is coming for free (regular file system cache)
  - ✓ Caching algorithms are designed by file system experts.
  - ✓ Security of files in cache is consistent.

## Compared to previous NFS protocol versions

- pNFS makes use of highly distributed data (client redirect, layout)
- Compound RPC calls (multiple ops, one rpc call)
- Security gss api defined in spec and not added later. (Secure)

## Using NFS 4.1 with dCache

dCache can be mounted on your WN as any other network file system and dCache data can be directly accessed through this protocol.

## NFS 4.1

## Contributors

Coordinated by the Center of Information Technology Integration (U. Michigan)

Slide is stolen from “Lisa Weeks” presentation :

pNFS: Blending Performance and Manageability

### Blue Arc

CITI

CMU

EMC

IBM

LSI

OSU

Net App

Ohio SuperComputer

Panasas

Seagate

StorSpeed

Sun Microsystems

Desy

### Clients

- › Sun (Files)
- › Linux (Files / Blocks / Objects)
- › Desy / dCache (Java-based / Files)

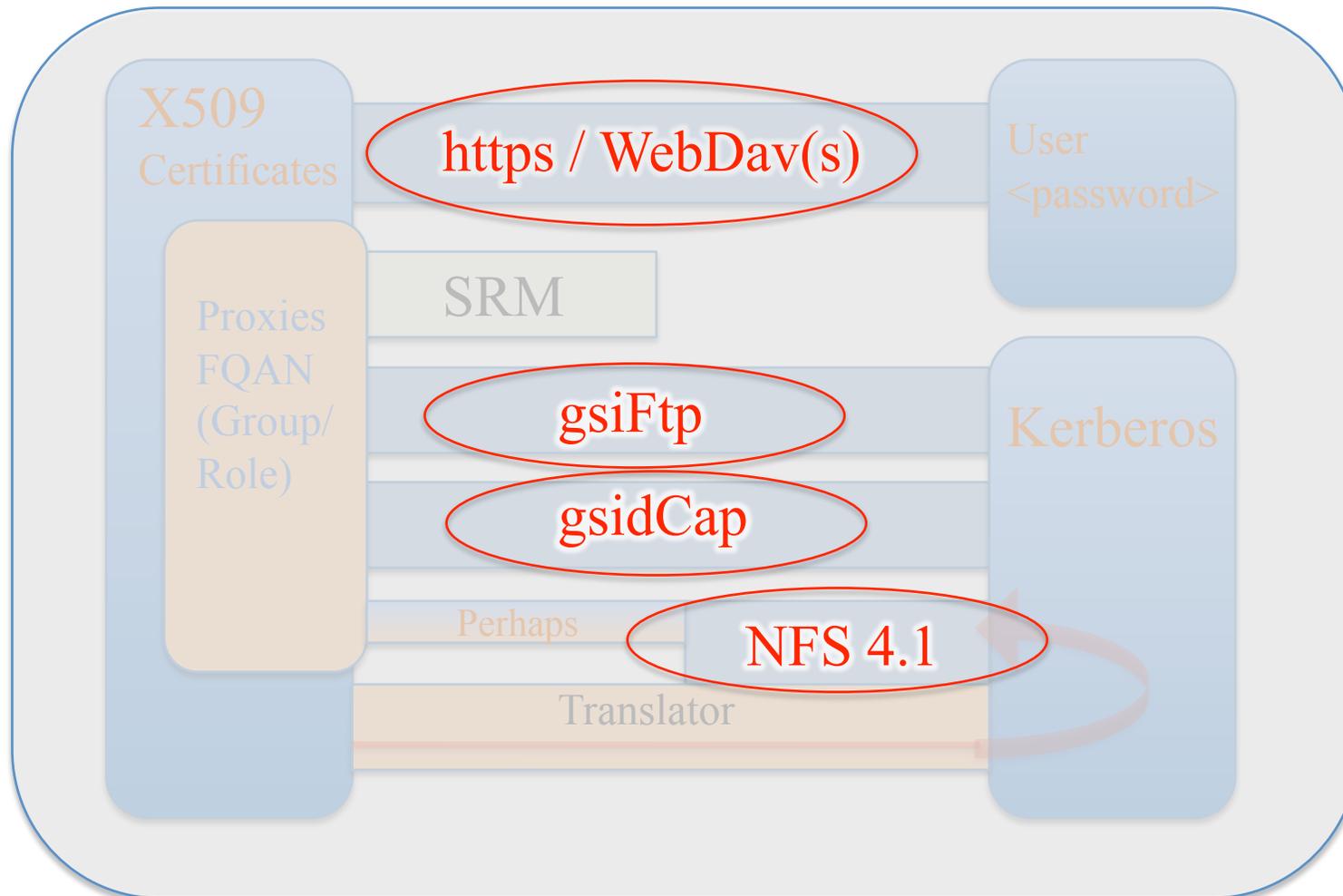
### Servers

- › Sun (Files)
- › Linux (Files)
- › NetApp (Files)
- › EMC (Blocks)
- › LSI (Blocks)
- › Panasas (Objects)
- › Desy / dCache (Java-based / Files)

dCache, the system  
dCache, NFS 4.1 clients

- NFS 4.1 and the linux kernel
  - NFS 4 already in SL5
  - NFS 4.1 in 2.6.32
  - NFS 4.1 plus pNFS in 2.6.33/34
- Kernel 2.6.34 will be in Fedora 13 and RH6 Enterprise (summer)
- **NFS 4.1 (pNFS) Kernel available in Fedora 12 (NOW)**
- Windows Client expected 4Q10.
- DESY grid-lab is testing with :
  - SL5 and 2.6.33 kernel plus some special RPM. (mount tools)
  - **See our wiki for further information**

dCache, the system  
dCache, protocols and security



## dCache, the system Storage Control

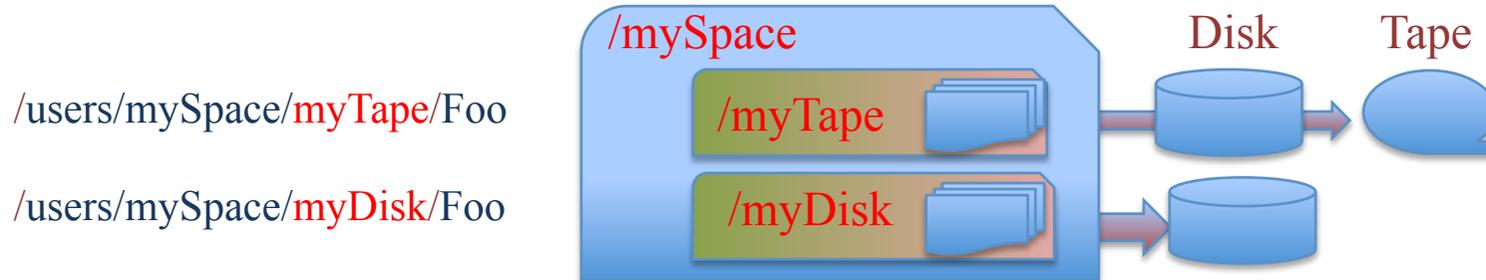
Some works on Storage Control in dCache.

Mostly important for dCache sys admins.



## Manual Storage Control

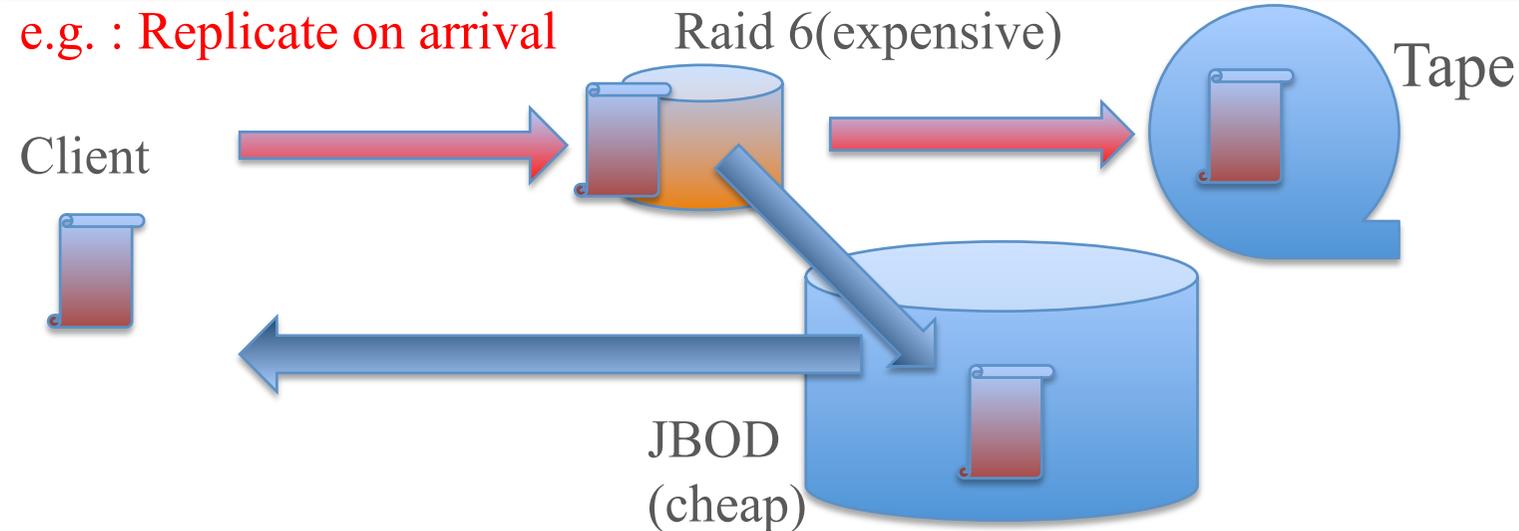
- SRM 2.2 (WLCG & Addendum & Addendum) compatible.
  - ✓ Define storage media (Disk/Tape) per file or “Space”.
  - ✓ Pin / Unpin files
  - ✓ Bring Online file(s)
- Storage Media can be assigned to directory (sub) structure.



- Data can be scheduled for replication for maintenance or performance reasons. (Migration Module)
  - ✓ Scheduled server downtimes
  - ✓ Server decommissioning
  - ✓ Multiple copies to increase throughput

## dCache, the system Automatic Storage Control

- Data stored to tape and retrieved when needed.
- Files are **automatically** replicated to cope with **high server load**.
- Files **replicated “on arrival”** to ensure second copy while not yet on tape.
- Configuration can enforce a **permanent second or n<sup>th</sup> copy** of each file.
- File hopping from tape to temporary disk to **optimize tape access**.



## In summary

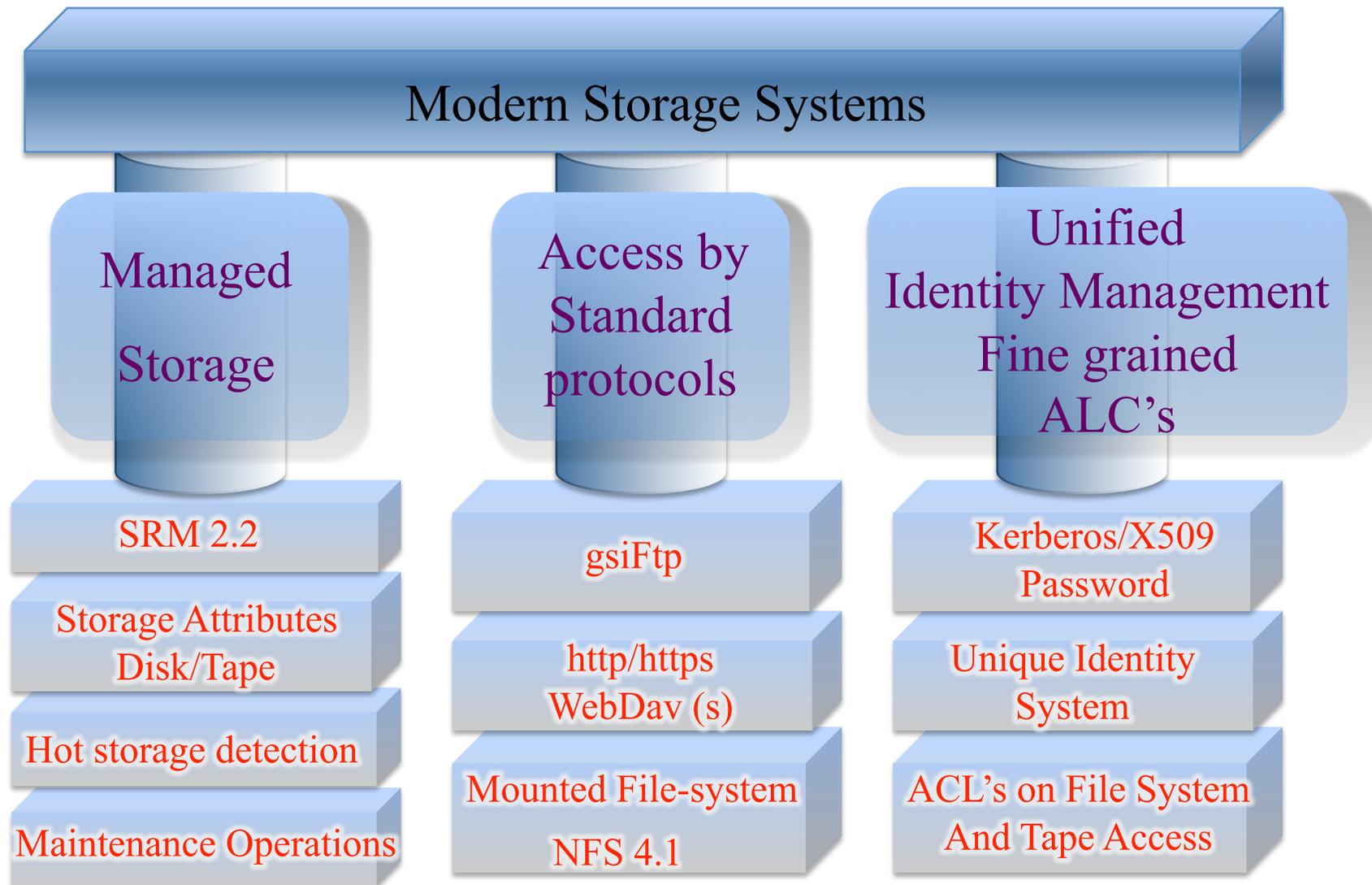
dCache combines well known and  
standardized data access  
mechanisms, e.g. mounted file-system, web access,  
browser/WebDav, with a broad  
automatic and manual storage control functionality,  
under a  
common file name space and security umbrella.

With dCache, EMI and with that EGI is well prepared to serve  
new data intensive communities.

# Further Reading

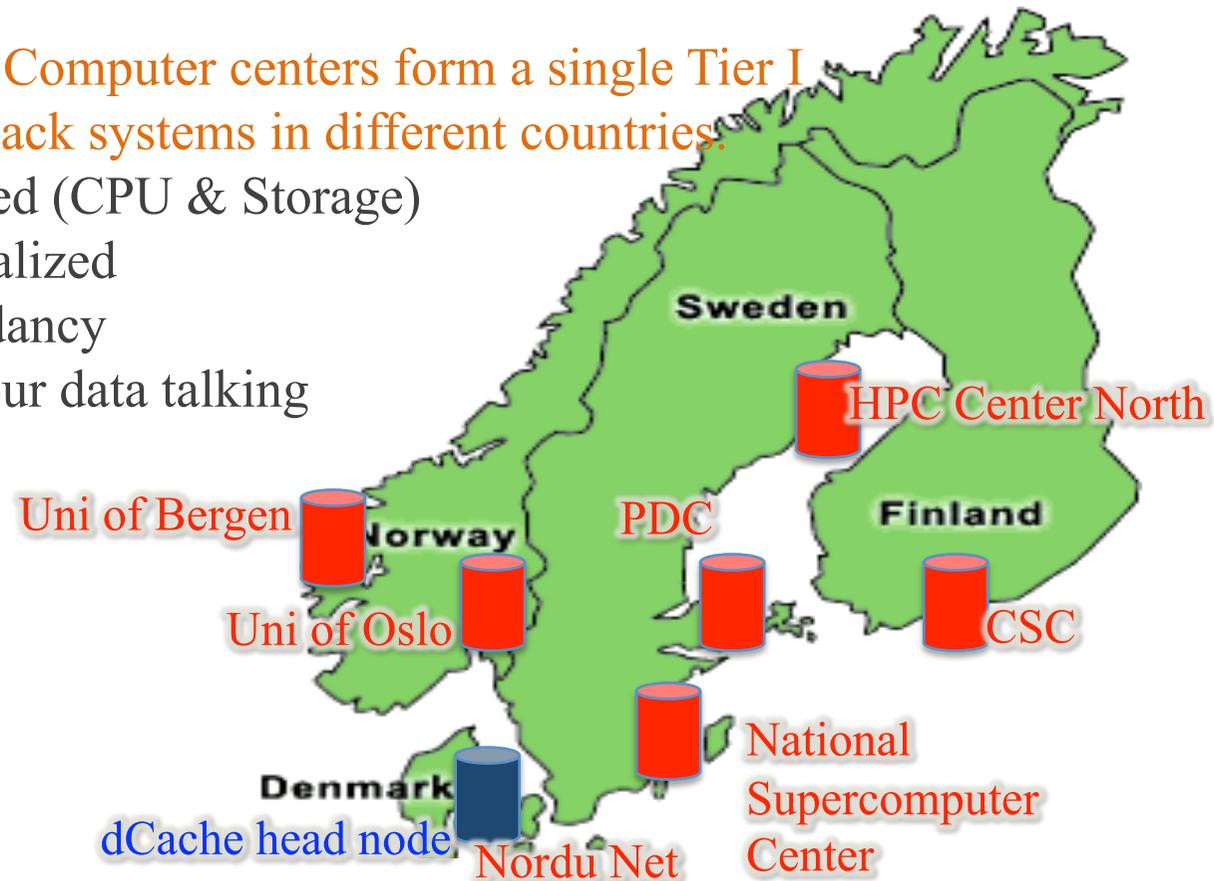
[www.dCache.org](http://www.dCache.org)

dCache, the system  
Thee pillars of modern storage



## The advanced dCache installation (NDGF)

- ✓ The 7 biggest Nordic Computer centers form a single Tier I
- ✓ Many different tape back systems in different countries.
- ✓ Resources are scattered (CPU & Storage)
- ✓ Services can be centralized
- ✓ Advantages in redundancy
- ✓ Especially in 7\*24 hour data talking



Slide stolen from Mattias Wadenstein, NDGF