

dCache, the agile storage technology

NEC 2013, Varna

Patrick Fuhrmann



Content

- The dCache Cheat Sheet
- The project structure
 - Partners and people
 - Our funding
 - Sustainability/Networking
- Deployments
 - WLCG overall
 - Big
 - Wide
 - Super Mini
- Customers Relation
 - Deployment Channels
 - User Support channels
- Technology (Design -> consequence)
 - Metadata , Data separation (Multi Tier 3)
 - gPlazma (IdP)
- Work in progress
 - Cloud software and service

Cheat Sheet

Cheat Sheet

- dCache.org is an international collaboration, developing and distributing storage software (dCache)
- dCache is in production in about 60 places around the world and stores (roughly) about 120 Pbytes in total for WLCG.
- dCache supports different storage media, like disk, SSD and tape and provides mechanisms for manual and automated internal and external replication and transitions.
- dCache storage can be accessed via standard protocols like WebDAV, NFS, and gridFTP and proprietary protocols like dCap and xrootd.
- dCache supports a variety of authentication and mapping mechanisms, e.g. Kerberos, X509, User/Password, LDAP, NIS, NSSWITCH.

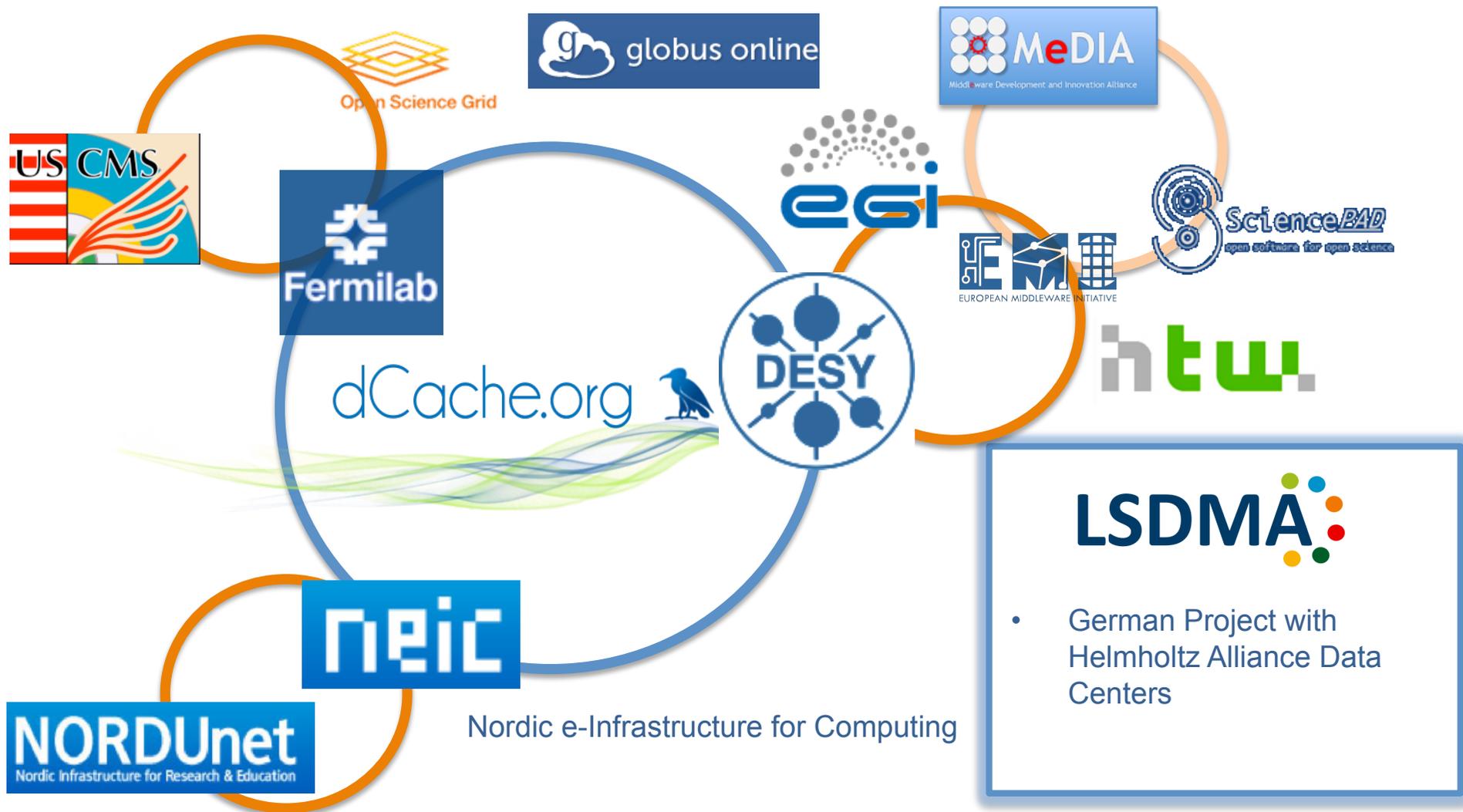
Project Structure

The dCache partners and team



dCache.org@DESY will start hiring in a couple of weeks.

dCache partners bridging national projects and activities.



Data Lifecycle Labs (Customers)

- Energy
 - smart grids, battery research, fusion research
- Earth and Environment
- Health
- Key Technologies
 - synchrotron radiation, nanoscopy, high throughput microscopes, electron-microscope imaging techniques
- Structure of Matter

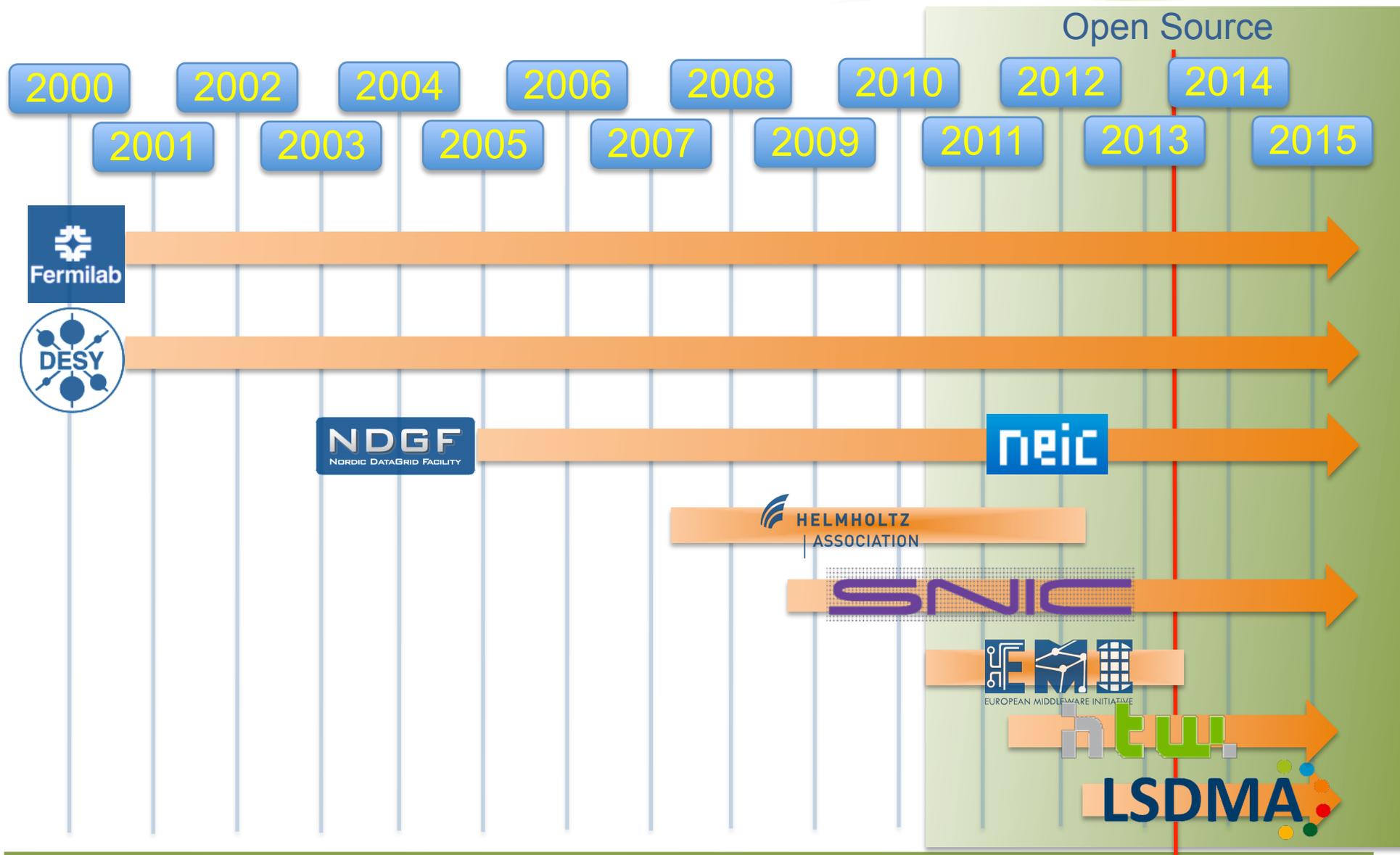
Data Service Integration Team

dCache.org 

- **Federated Identity**
- Federated Data Access
- Metadata Management
- Archiving

Funding

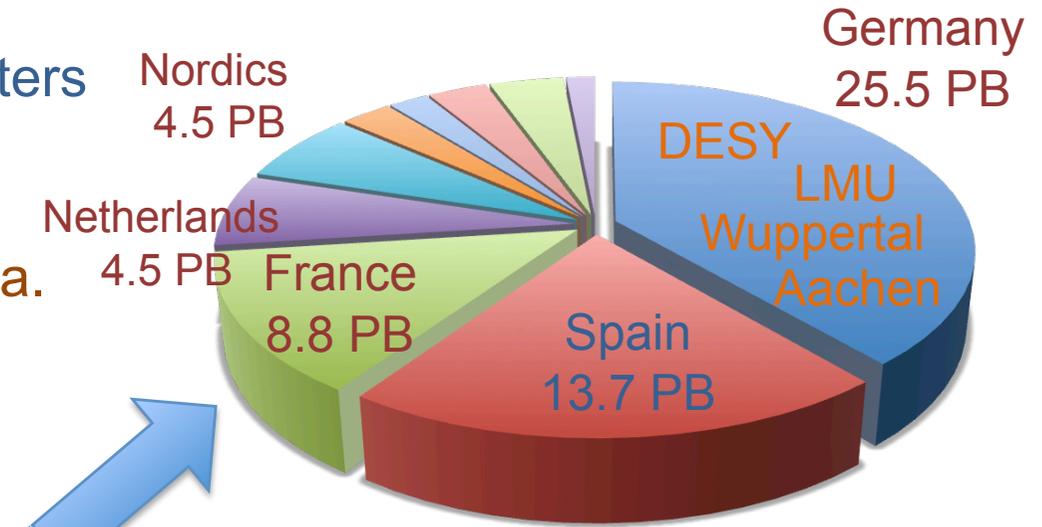
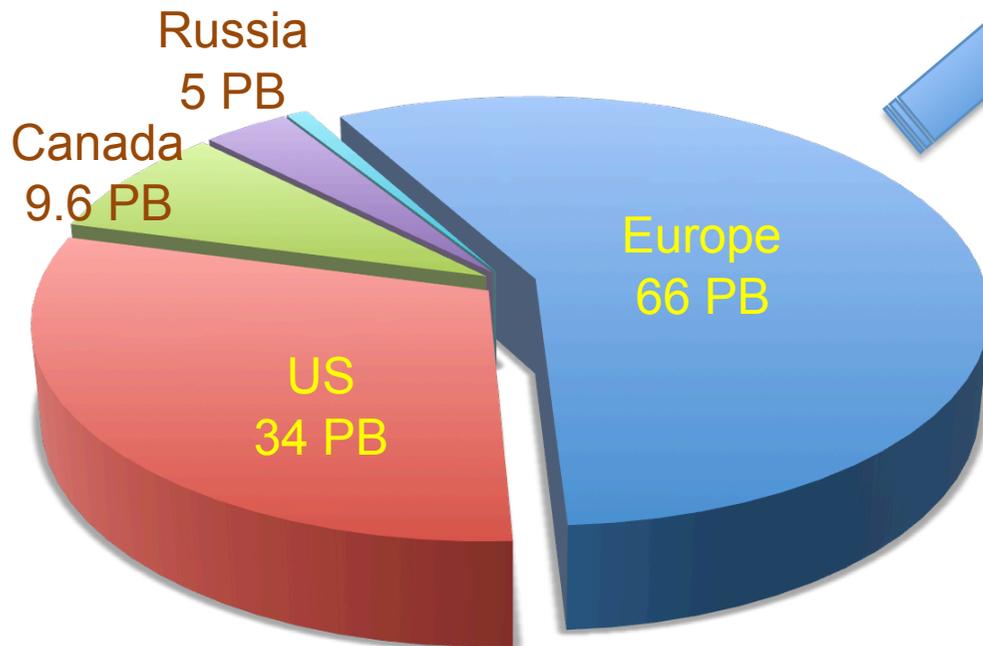
dCache project timeline



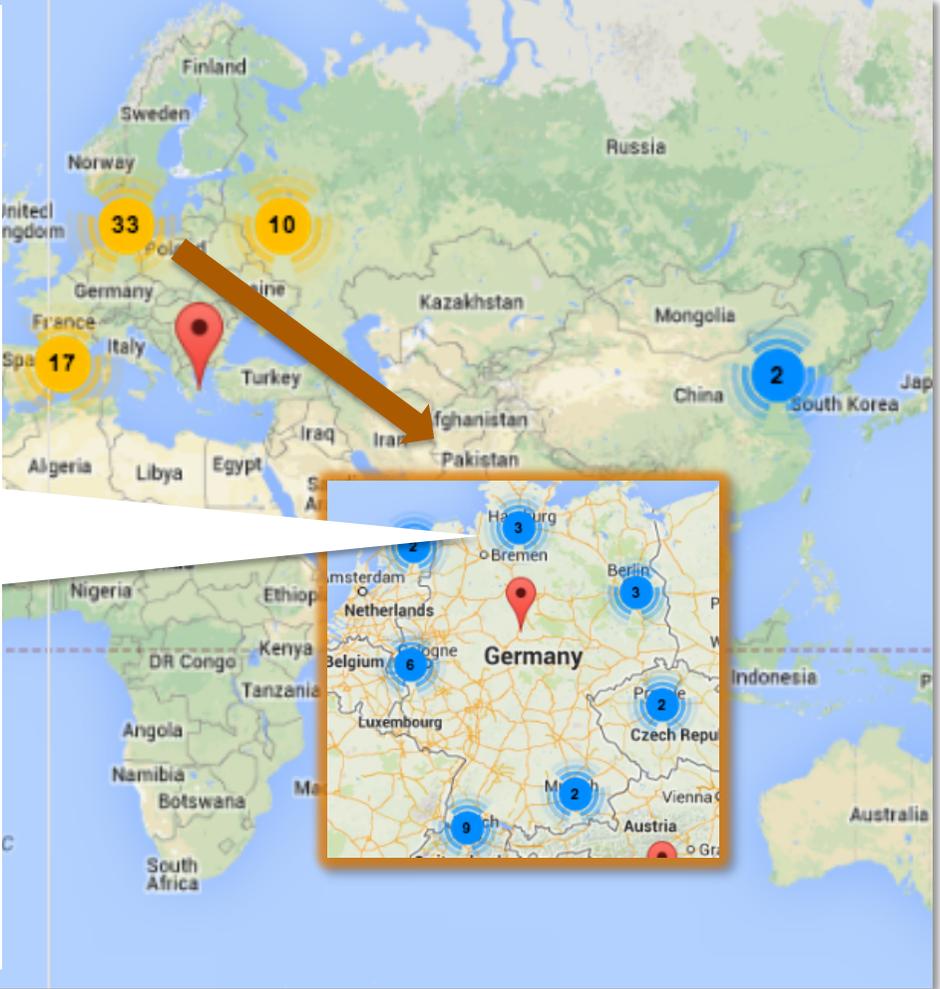
Deployments

dCache storage for WLCG

- About 115 PBytes just for WLCG
- In 8(+2) out of 11(+3) Tier 1 centers
- And about 60 Tier 2's, which is
- about 1/2 of the entire WLCG data.



Tigrans new dCache world map dCache.org



The image displays a world map with several dCache sites marked by colored circles and numbers. A red pin is placed on Hamburg, Germany. A brown arrow points from this pin to a zoomed-in map of Germany. The zoomed-in map shows several sites in Germany: Hamburg (3), Bremen (3), Berlin (3), Cologne (6), Luxembourg (2), and Munich (2). Other sites are visible in Europe: Poland (33), Italy (17), and South Korea (2).

DESY Hamburg
Location: Hamburg, Germany
Site URL: http://grid.desy.de/
End Point: dcache-se-desy.desy.de
Version: 2.6.5 (ns=Chimera)
Total Size: 718.2 TiB
Used Size: 246.6 TiB

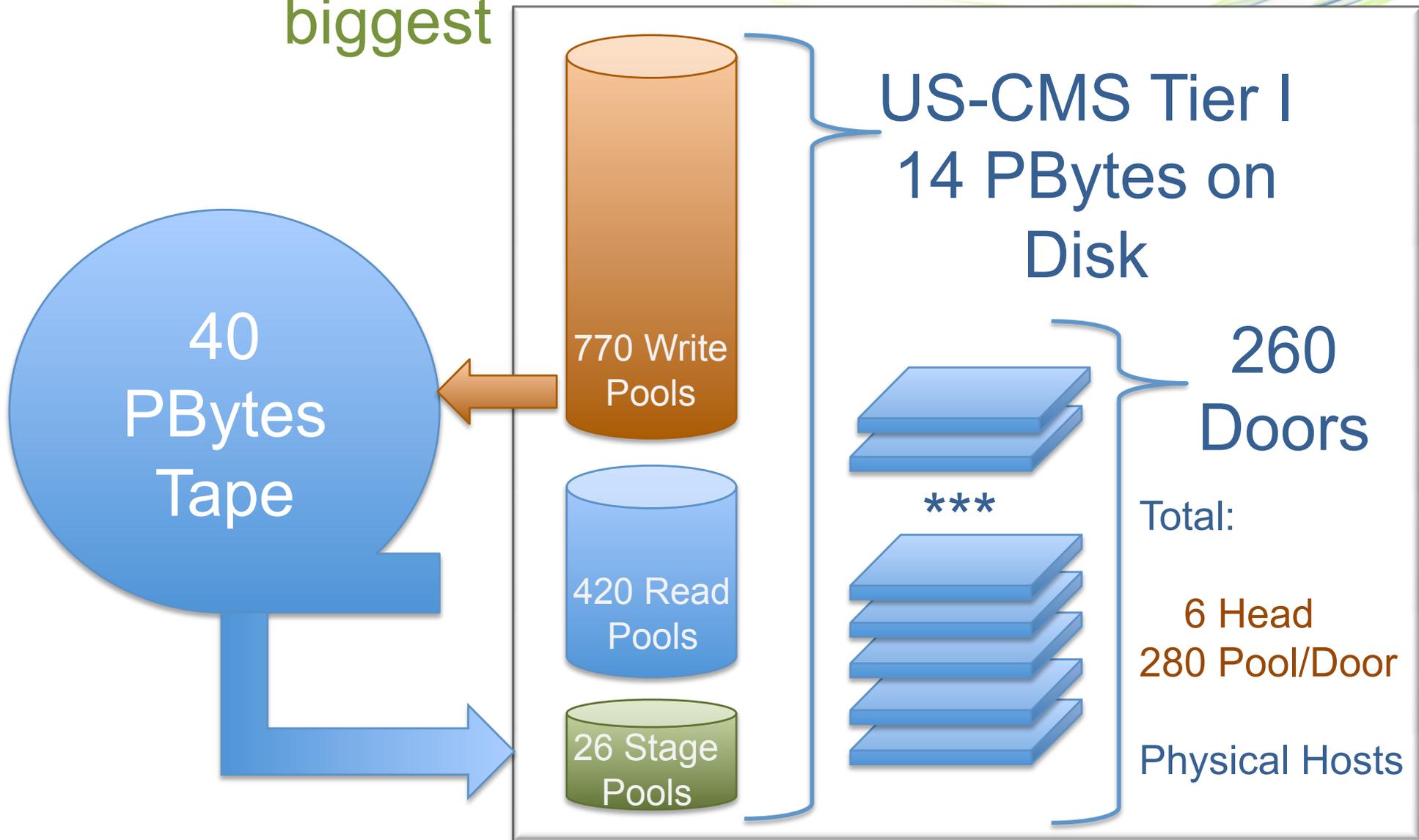
DESY Hamburg
Location: Hamburg, Germany
Site URL: http://grid.desy.de/
End Point: dcache-se-cms.desy.de
Version: 2.6.6 (ns=Chimera)
Total Size: 3.9 PiB
Used Size: 3.6 PiB

DESY Hamburg
Location: Hamburg, Germany
Site URL: http://grid.desy.de/
End Point: dcache-se-atlas.desy.de
Version: 1.9.12-12 (ns=Chimera)
Total Size: 2.6 PiB
Used Size: 2.0 PiB

Available at dCache.org

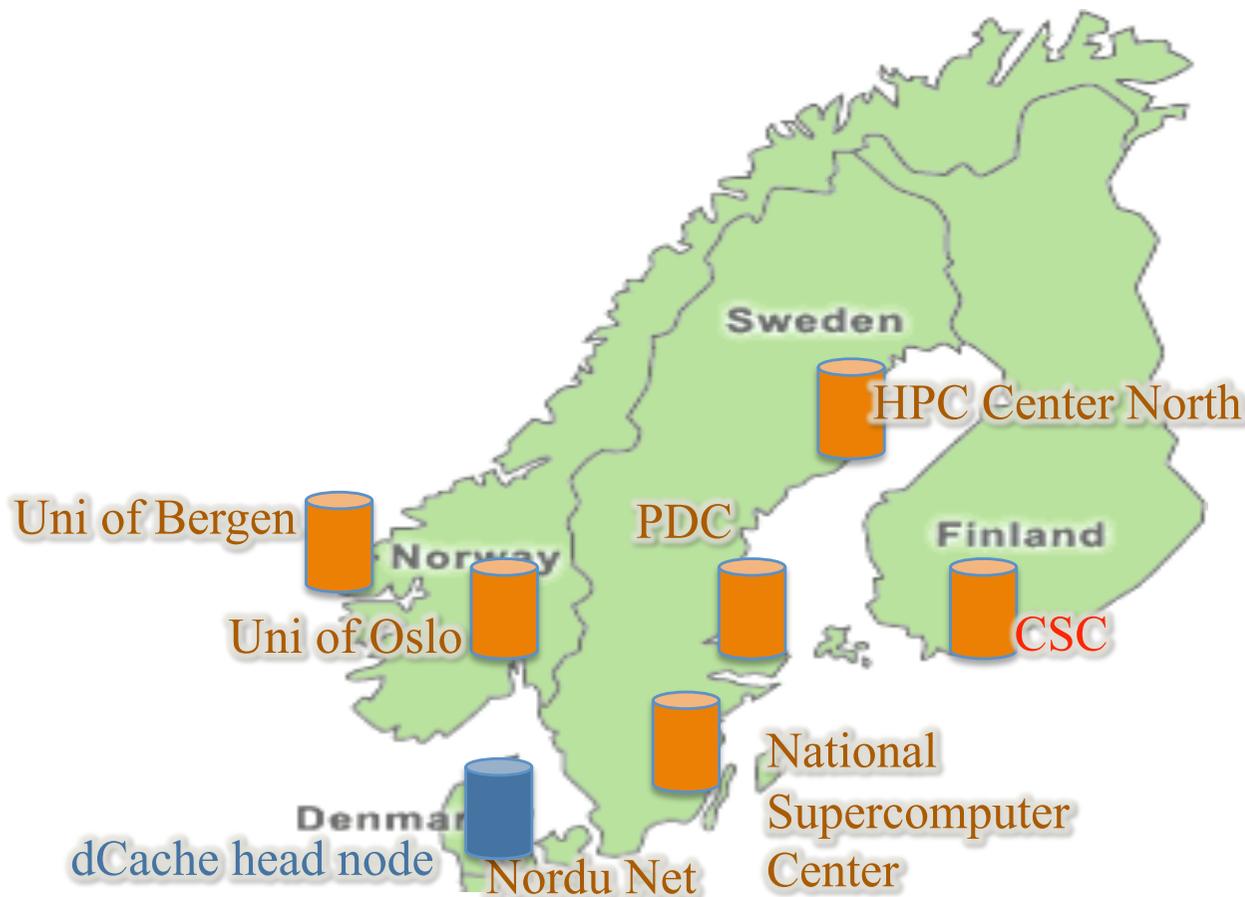
Interesting installations

Starting with possibly the
biggest



Information provided by Catalin Dumitrescu and Dmitry Litvintsev

To certainly the
most widespread



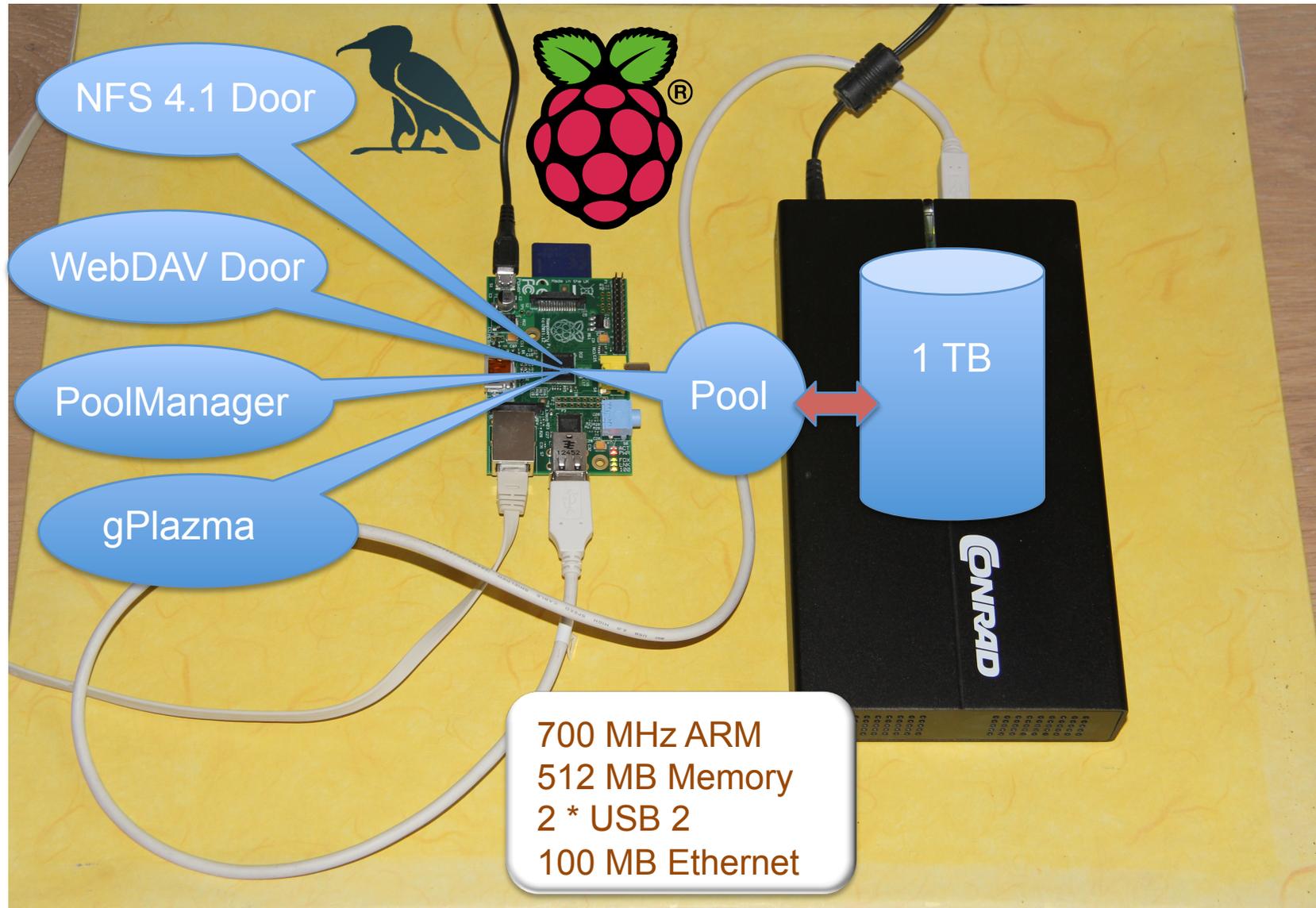
4 Countries

One dCache

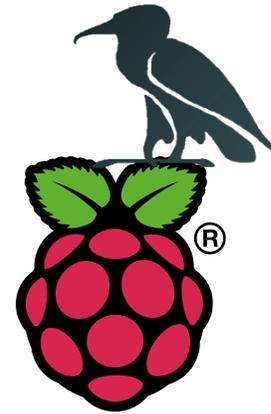
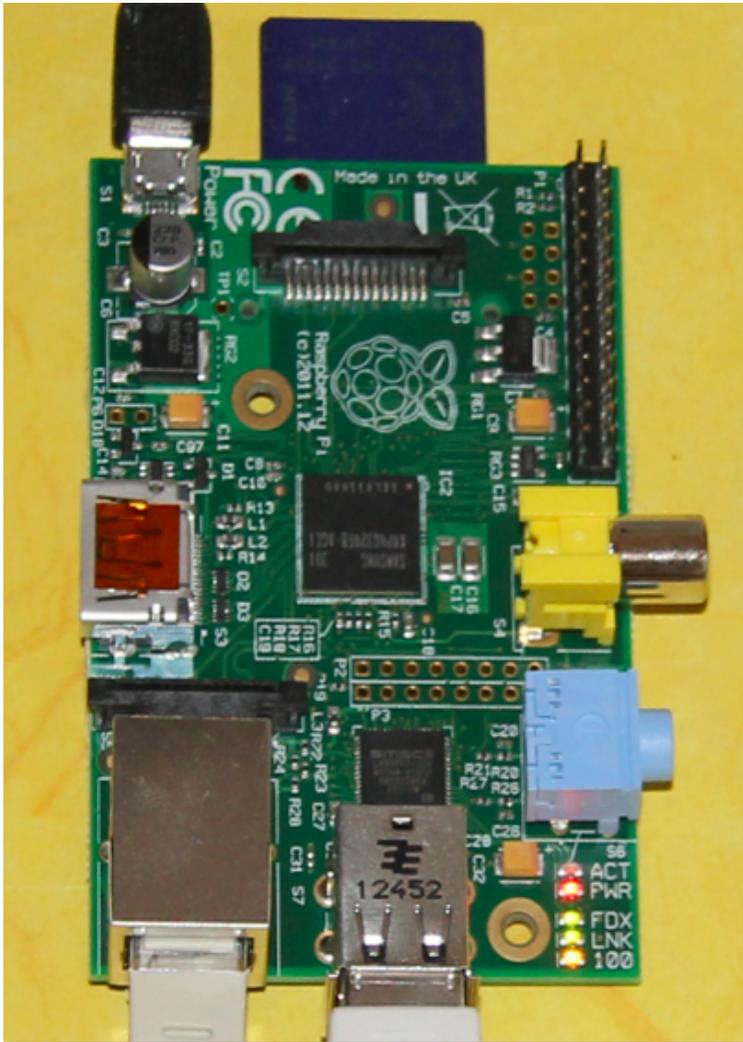
Slide stolen from Mattias Wadenstein, NDGF

To very likely the smallest

One Machine – One Process



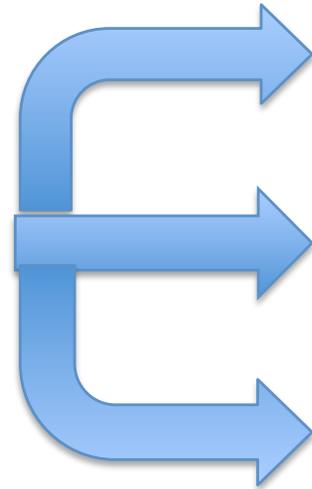
The raspberry dCache



700 MHz ARM
512 MB Memory
2 * USB 2
100 MB Ethernet

Customer Relations

Deployment Channels



- dCache.ORG / Web Pages
-  UMD
- Targeting: EPEL

Support Channels

- support@dCache.org (security@dCache.org) for all bug reports, feature requests and requests for help. Tickets are distributed to all dCache partners.
- **German Support Group:** Group composed of German dCache sites, helping each other with monitoring and daily operational work and organizing the dCache tutorial of the annual GridKA school of computing
- **EGI.eu:** First level support for dCache packages taken from UMD.
- Weekly customer **phone meetings**
- 2 dCache **workshops/year**



Technology and design

Design Patterns and consequences

(stolen from a dCache tutorial)



Design #1 Service Modules & Message Passing

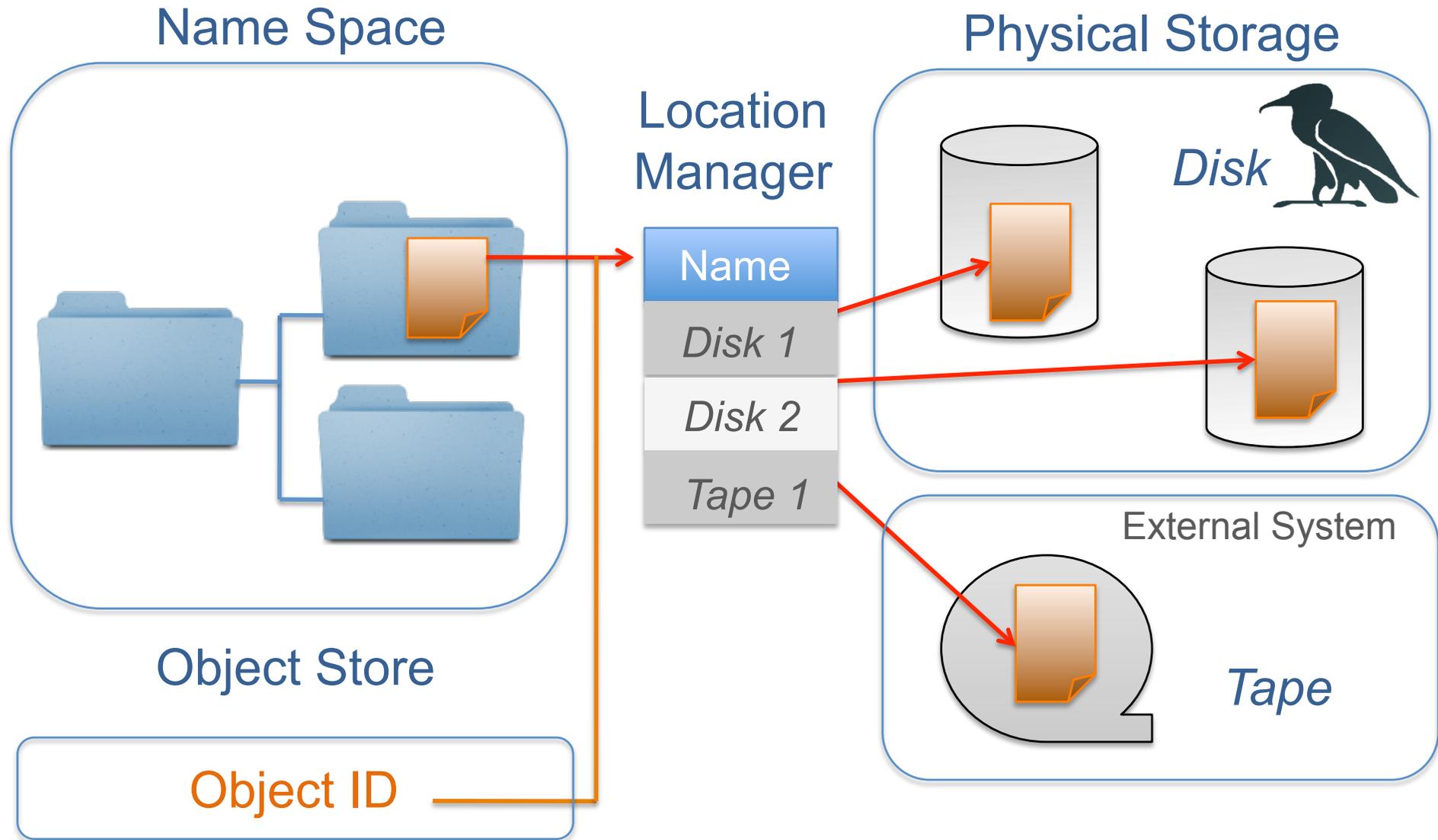
Design #2 Namespace – Physical Storage separation

Design #3 Services allow plug-ins

Design #2 Namespace – Physical Storage separation

Design

Namespace – Storage separation

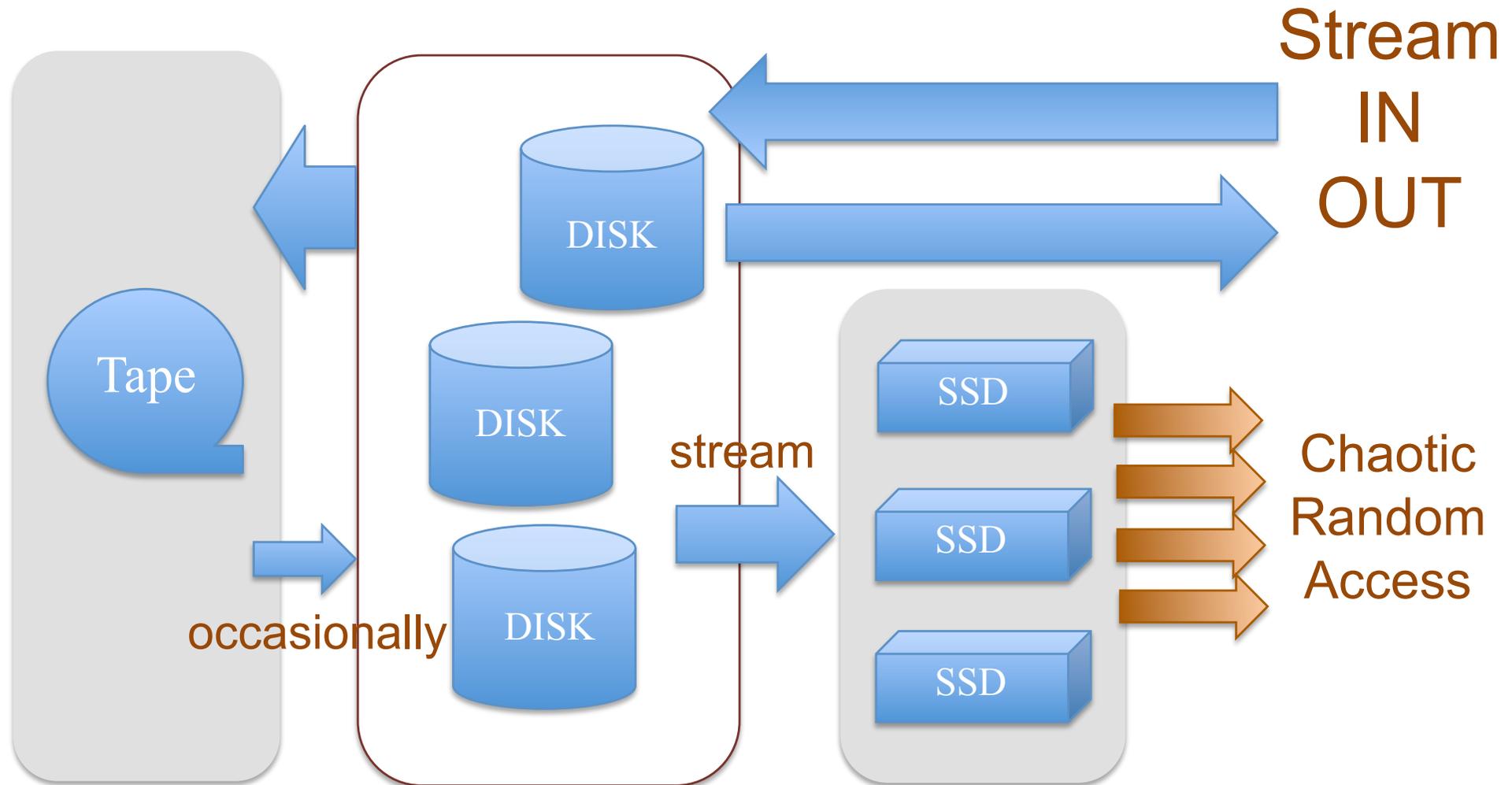


Resulting in Replica Management

- Hot Spot detection
 - Files are copied from ‘hot’ to ‘cold’ pools
- Multi Media Support
 - File location is based on access profile and storage media type/properties
 - Fast streaming from spinning disks
 - Fast random I/O from SSD’s
- Migration Module(s)
 - Files can be manually/automatically moved or copied between pools.
 - Rebalancing of data after adding new (empty) pools.
 - Decommission pools.
- Resilient Manager
 - Keeps max ‘n’ min ‘m’ copies of a file on different machines.
 - System resilient against pool failures.
- Tertiary System connectivity (Tape systems)
 - Data is automatically migrating to tape.
 - Data is restored from tape if no longer on disk

In preparation : Multi Tier Storage dCache.org

Analysis



Design and consequence

(stolen from a dCache tutorial)

Design #3

Services allow plug-ins

Plug-in Facility

Standard File Access Protocols

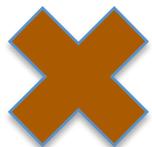
http(s)
WebDav

NFS 4.1

gsiFtp

Storage Management

SRM



Common Security Layer

Authentication : Kerberos, X509, Password

Unified ID management

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

Common Name Service Layer

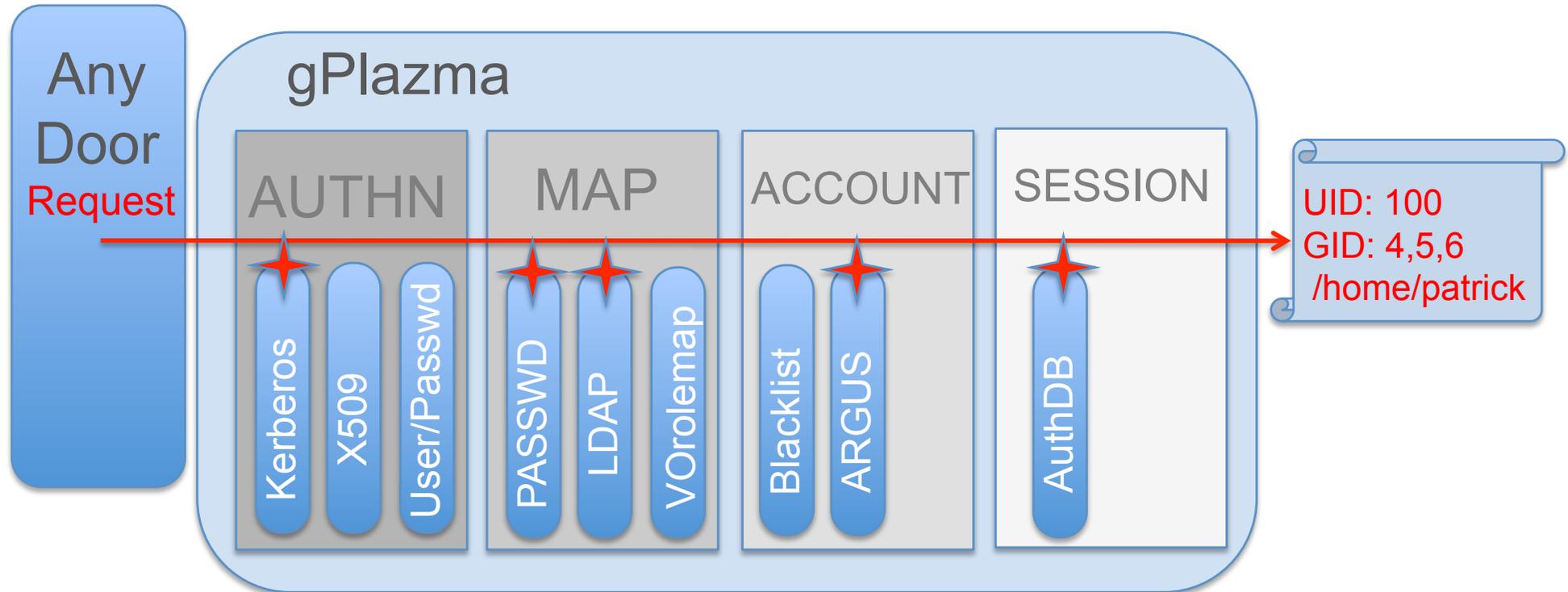
Extended Names Service Queries (SQL)

“multi-media” storage layer



gPlazma (AAI) pluggin

Design stolen from Paul



Out of the box plug-ins

- Kerberos
- X509 Certs and proxies
- User / Password
- NIS/LDAP
- NSSWITCH
- GridMapFile
- VO Rolemap
- Argus
- Local Blacklist

Consequence of #3

- Authentication, mapping and user attributes are handled separately and in independent pluggins. So any reasonable combination is possible.
- Data access protocols and gPlazma are orthogonal. So the same mapping and user handling can be applied to any protocol.
- Easy to add a new auth/mapping pluggin to handle local infrastructure. (e.g. SARA added their own LDAP pluggin)

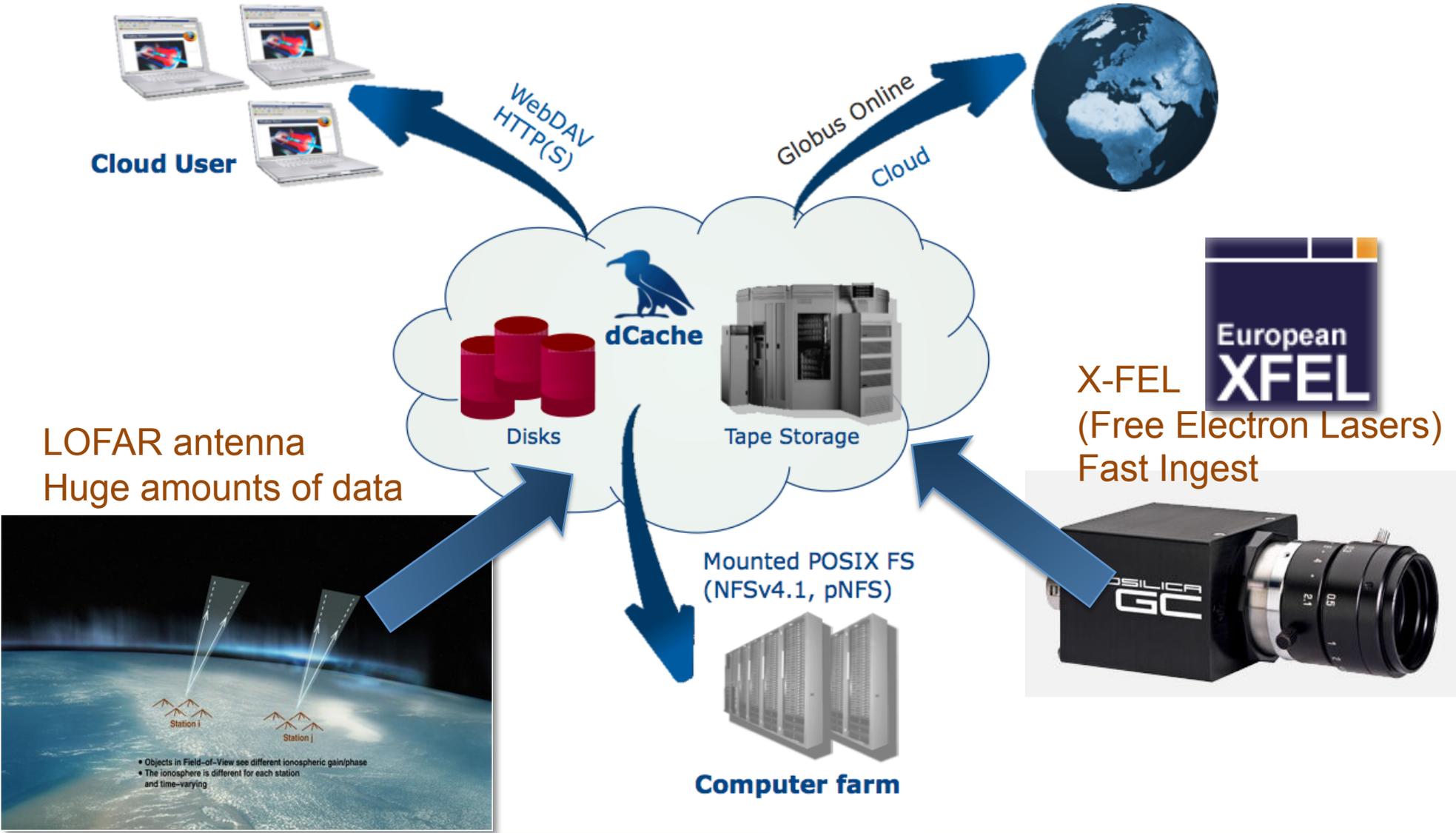
Work in progress



Cloud software and service

- Data can be accessed by a variety of protocols
 - Globus-online transfers via **gridFTP**
 - FTS Transfers for WLCG via gridFTP or **WebDAV**
 - Private upload and download via **WebDAV**
 - Public anonymous access via plain **http(s)**
 - Direct fast access from worker-nodes via **NFS4.1/pNFS** (just a mount like GPFS or Lustre but with standards)
- Individuals are authenticated by different mechanisms
 - X509 certificates or proxies
 - Username/password
 - SAML assertions (from IdP)
 - Kerberos tokens

Scientific Storage Cloud



- In collaboration with the HTW Berlin, dCache.org makes a storage cloud available for students.
- They get unlimited storage for free and their masters or bachelor degree.
- We get:
 - Cloud CDMI Implementation
 - Knowledge what young people need from cloud.
 - Client software on mobile devices.
- Currently 3 students are working with us.

Work in progress



Migrate HEP (including WLCG) local storage access from proprietary protocols to industry standards



- Quick reminder:
 - pNFS allows GRID storage elements (e.g. dCache and DPM) to be mounted like regular disk systems.
 - But provides scaling by letting the client directly exchanging data with the individual storage node.
 - Photon Science and BELLE (1&2) are already accessing their data via NFS at DESYs dCaches.
- As SL6 is now ready for WLCG, NFS 4.1/pNFS clients are available on work group servers and worker nodes.
- CMS and ATLAS dCache at DESY have been upgraded, supporting latest NFS4.1/pNFS server.
- DESY is evaluating NFS for CMS (thanks to Christoph Wissing and DOT Team), starting with the “National Analysis Facility”, followed by GRID worker nodes.
- Next step will be evaluations at FERMILab.

LSDMA Federated Identity

(Just one example)



- Organizing German sites/universities to provide IdP.
- Integrating those IdPs into the German DFN Online CA. (Hopefully including Umbrella/PanData).
- Goal: Easy access for all scientists, registered at any IdP, to access scientific resources (e.g. dCache) w/o handling X509 Certificates.

In summary

- Due to the broad developers base across international institutions and projects, dCache.org doesn't see any issues in continuous future funding.
- For the same reason, dCache.org is well integrated into the existing infrastructures and communities and keeps on track on upcoming requirements in storage management and access.
- By involving universities and students in the design and development process, dCache is keeping up with the latest developments in computers science and on the requirements of young people in data access and data sharing.

The End

further reading
www.dCache.org