

# dCache Hands-on

ISGC 2014, Taipei

Christian Bernardt  
(Slides: Patrick Fuhrmann)



# Agenda



Time	Topic	Speaker
09:00 - 09:20	Introduction to dCache	Christian Bernardt
09:30 - 10:30	Installation + Reading/Writing in dCache	Hands-on
10:30 - 11:00	Coffee Break	
11:00 - 11:15	Reading/Writing dCache (continued)	Hands-on
11:15 - 12:00	gPlazma Talk	Paul Millar, Tigran Mkrtchyan
12:00 - 12:30	Certificate-based WebDAV	Hands-on
12:30 - 14:00	Lunch	
14:00 - 14:40	Pool Management/Pool Selection	Oleg Tsigenov
14:40 - 15:30	Hardware Lifecycle	Hands-on
15:30 - 16:00	Coffee Break	
16:00 - 16:30	Hardware Lifecycle (continued)	Hands-on
16:30 - 16:50	CDMI	Paul Millar
16:50 - 17:10	Feature/ Future Outlook (ownCloud)	Patrick Fuhrmann

- The dCache Cheat Sheet
- The project structure
  - Partners and people
  - Our funding
  - Sustainability/Networking
- Deployments
  - WLCG overall
  - Big
  - Wide
  - Super Mini

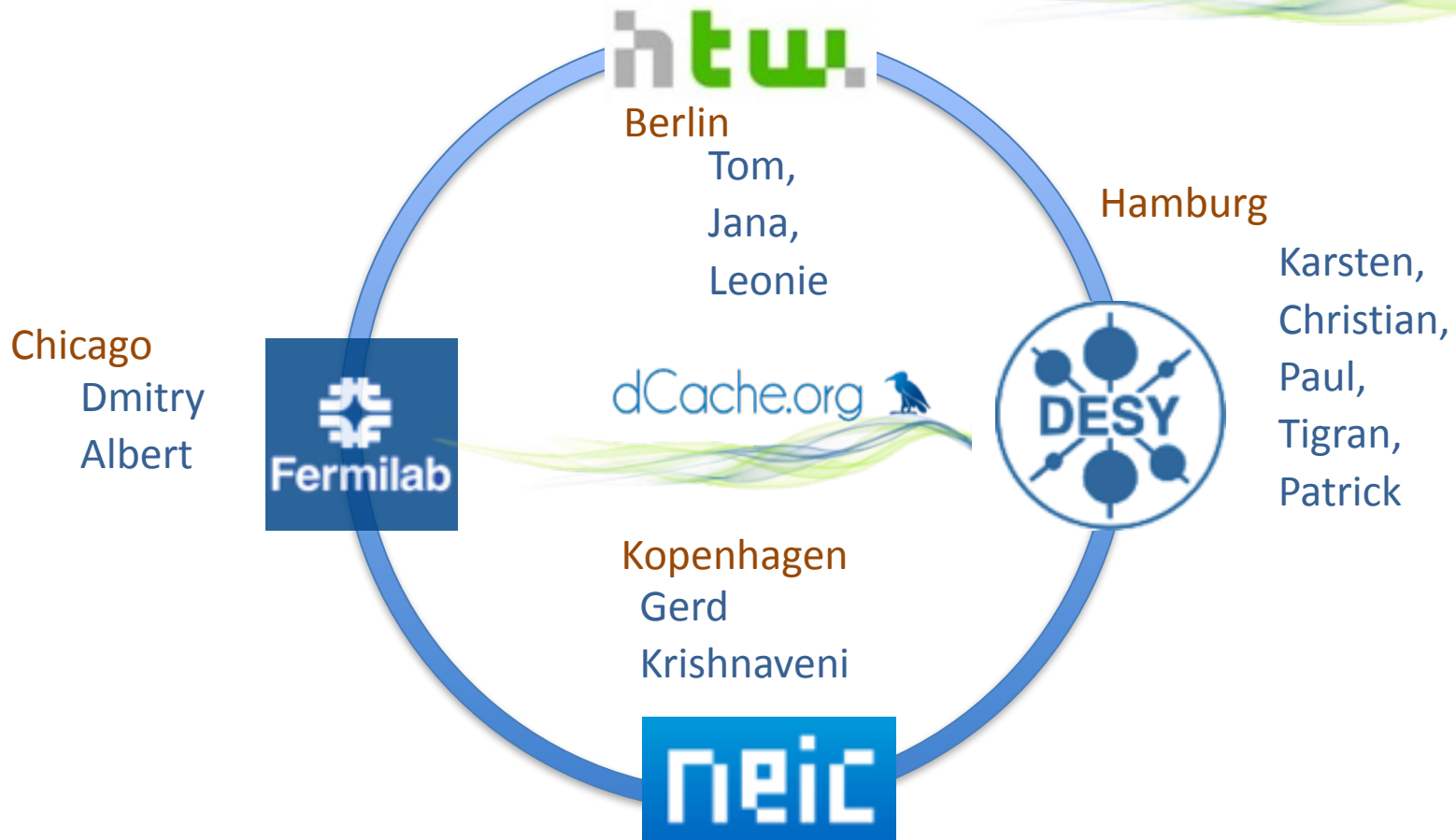
# Cheat Sheet

## Cheat Sheet

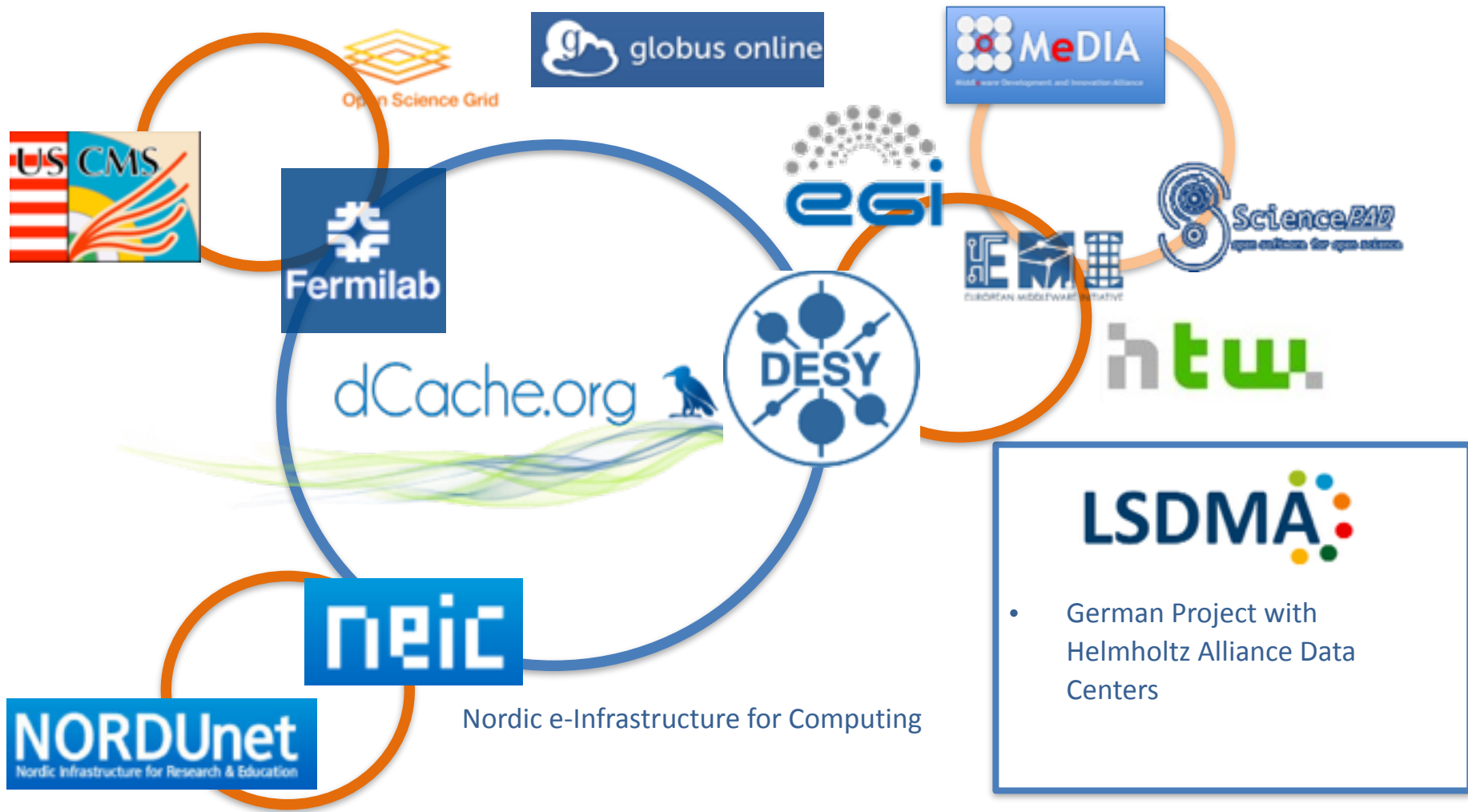
- dCache.org is an international collaboration, developing and distributing storage software (dCache)
- dCache is in production in about 60 places around the world and stores (roughly) about 120 Pbytes in total for WLCG.
- dCache supports different storage media, like disk, SSD and tape and provides 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



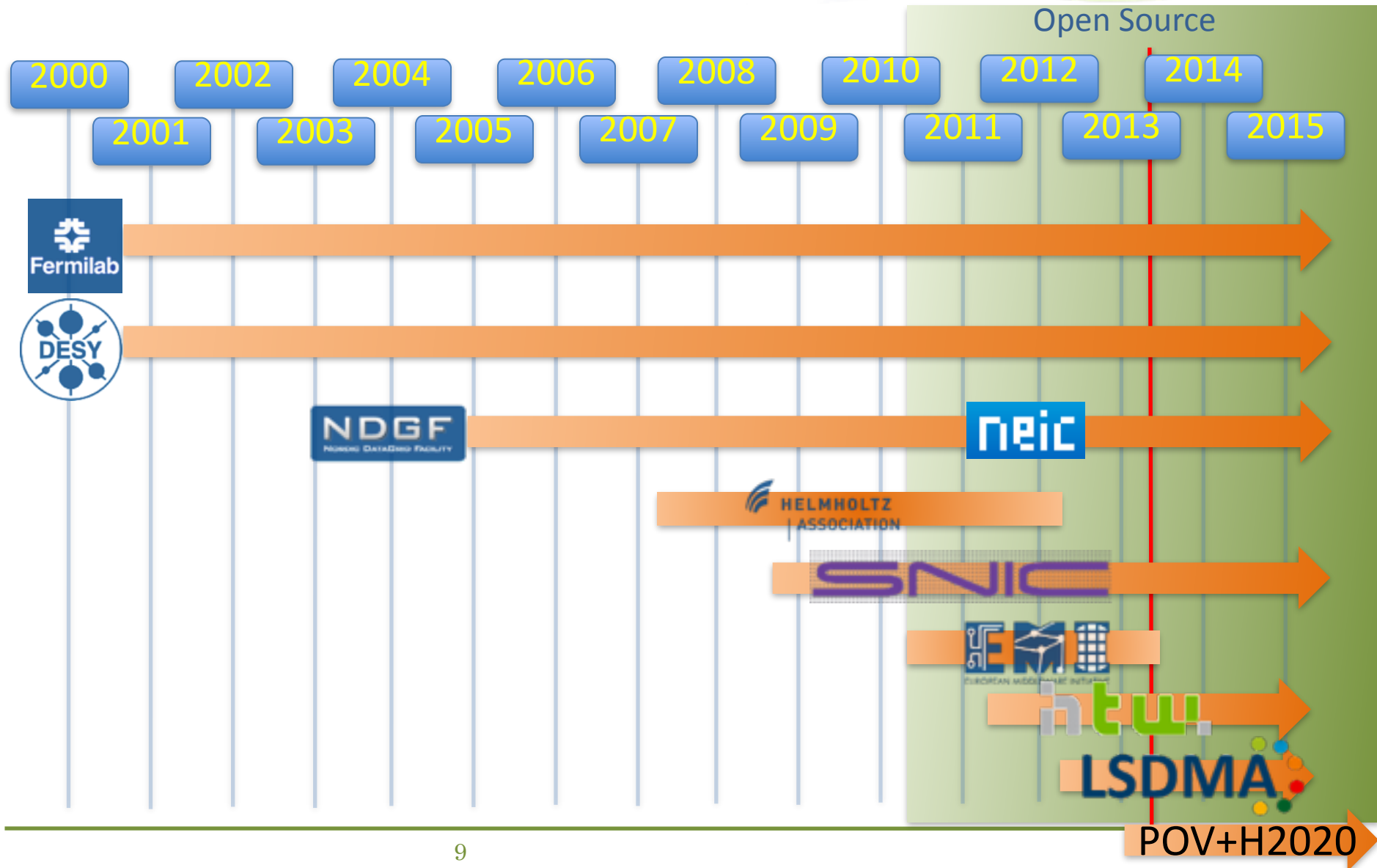
# dCache partners bridging national projects and activities.



# Funding

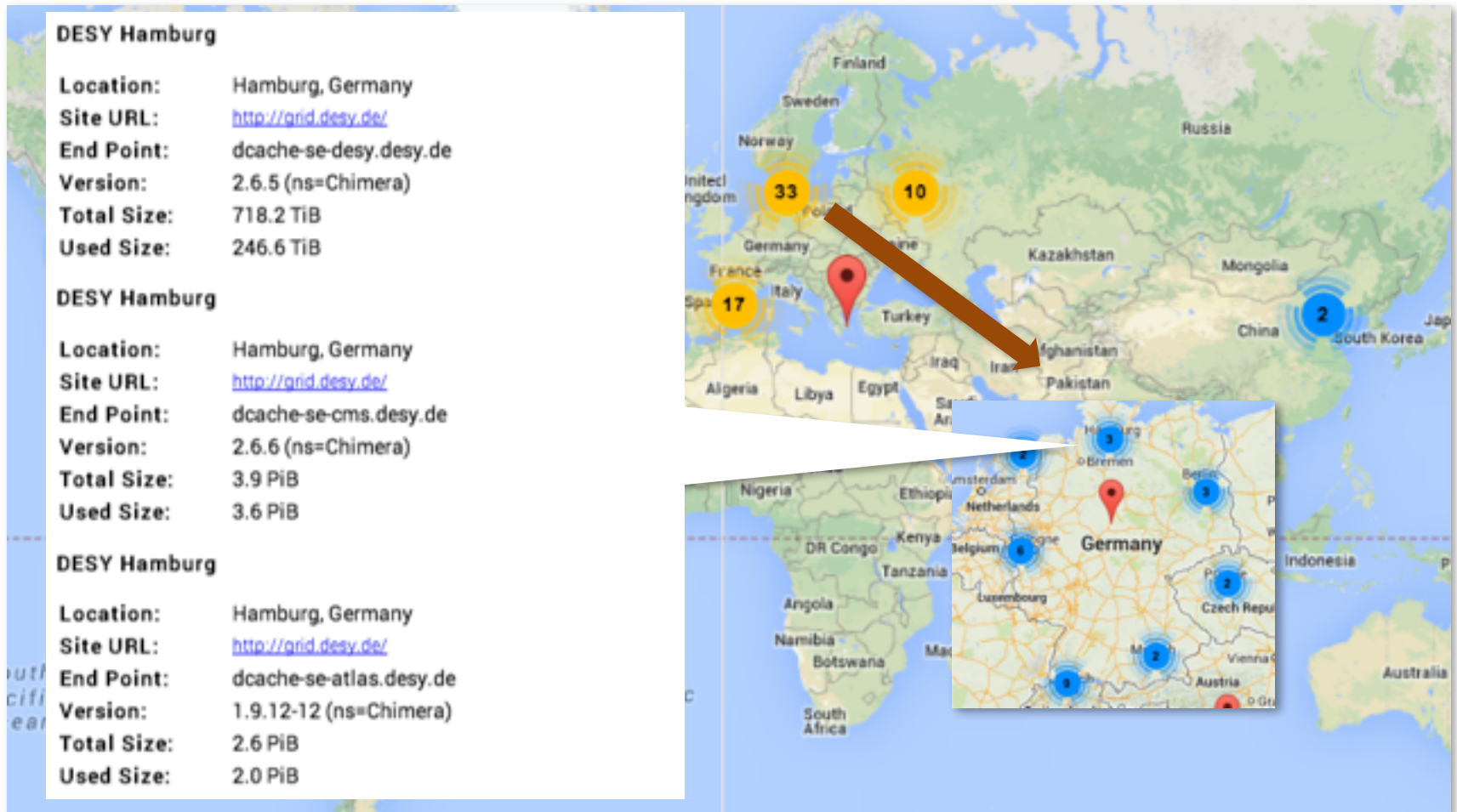
dCache project timeline

dCache.org 



# Deployments

# Tigrans new dCache world map

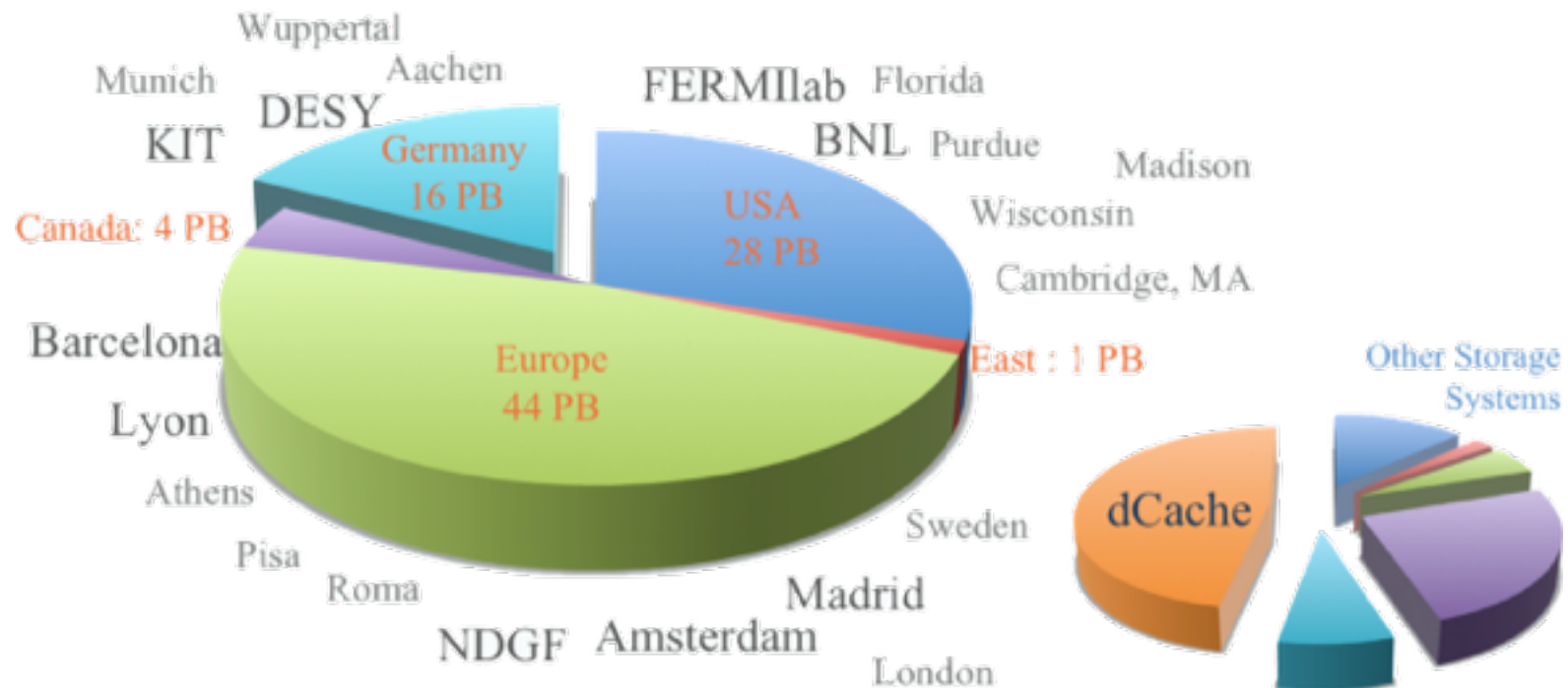


Available at dCache.org

# dCache in WLCG (Worldwide LHC Computing Grid)

## Worldwide dCache Managed Storage Space in WLCG

94 PB in total

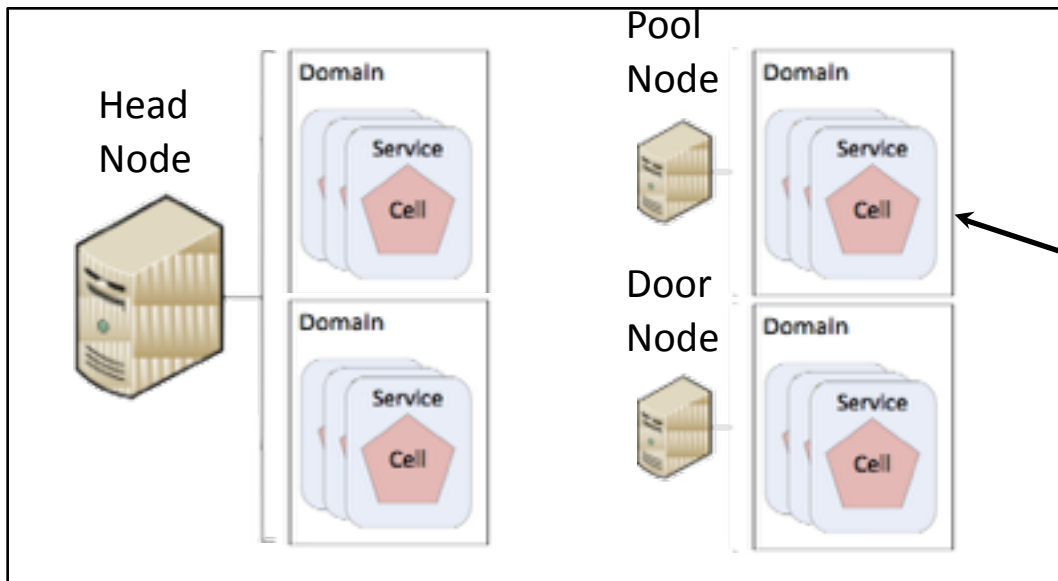


# Technicalities

# dCache Technical Design

- Java based

## Storage Element

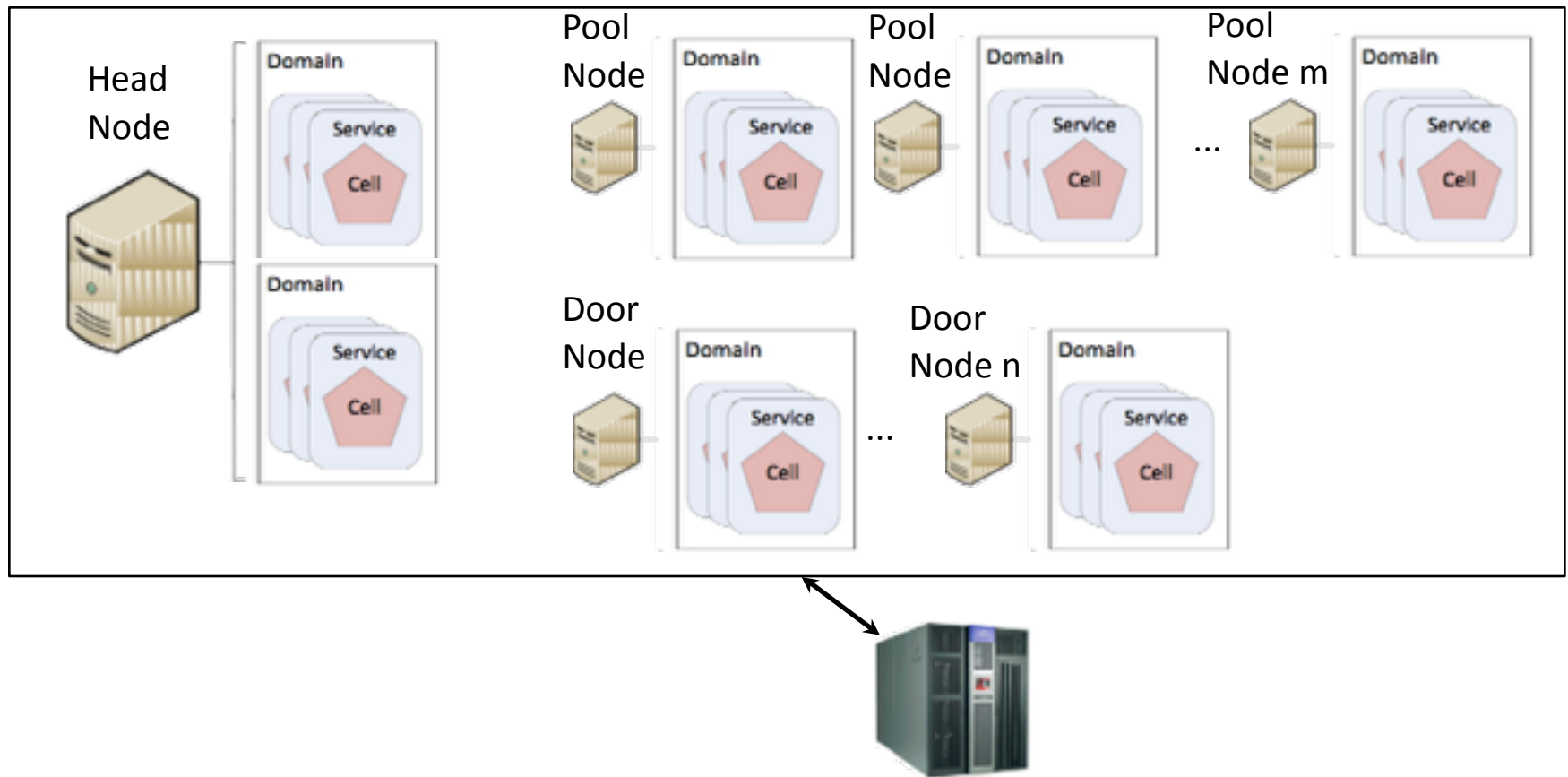


## Tape Robot



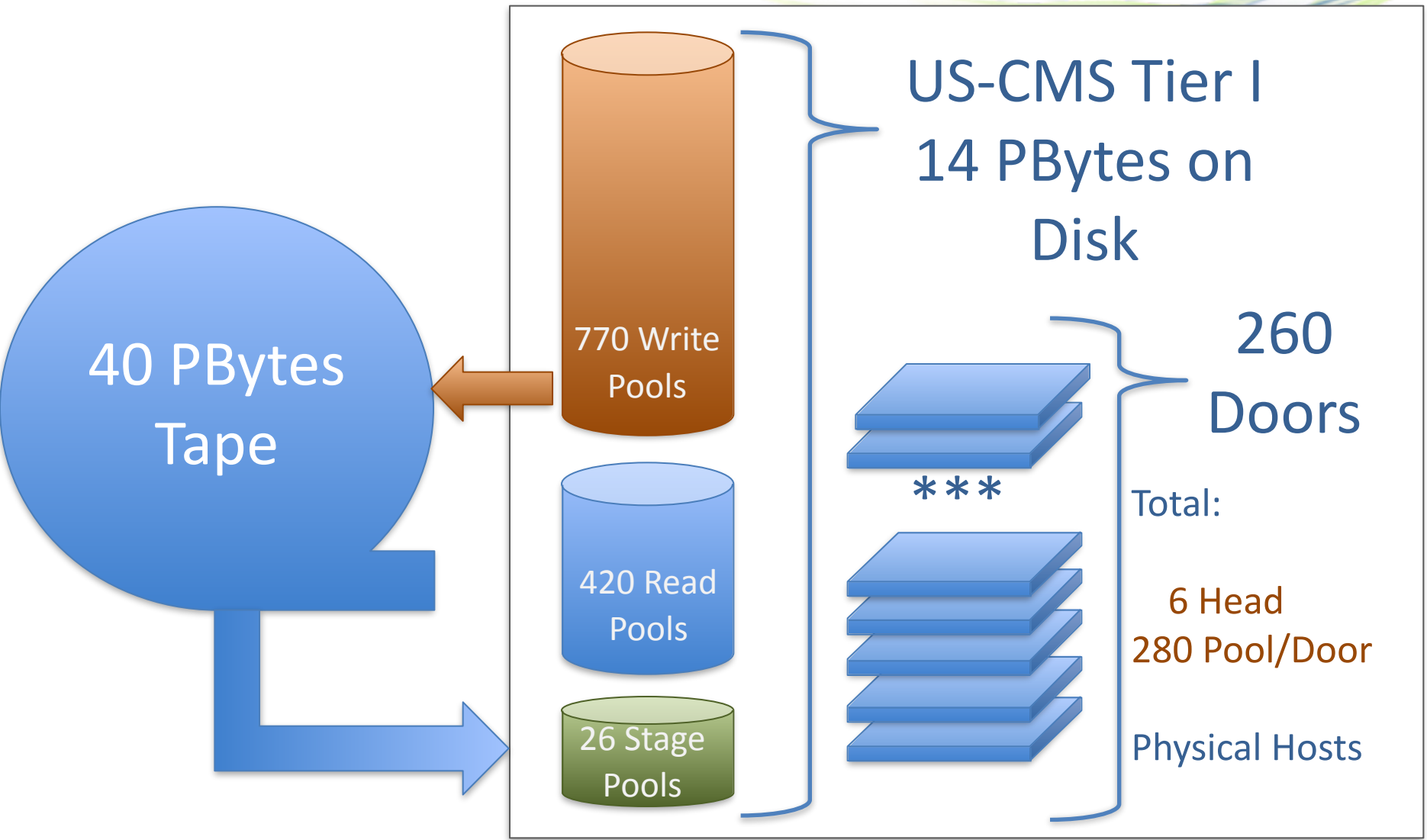
# dCache Überblick - Skalierbarkeit

## Storage Element



## Interesting installations

Starting with possibly the biggest

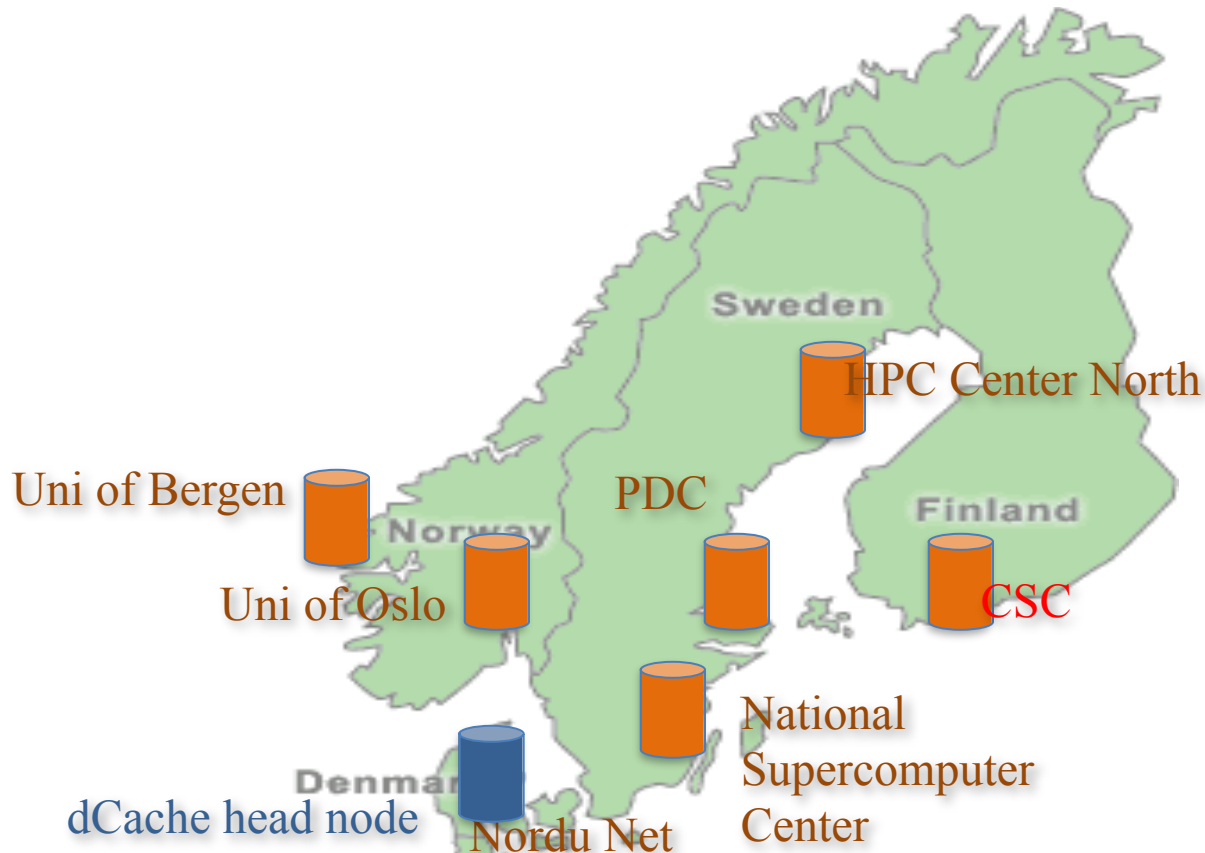


Information provided by Catalin Dumitrescu and Dmitry Litvintsev

To certainly the most widespread

4 Countries

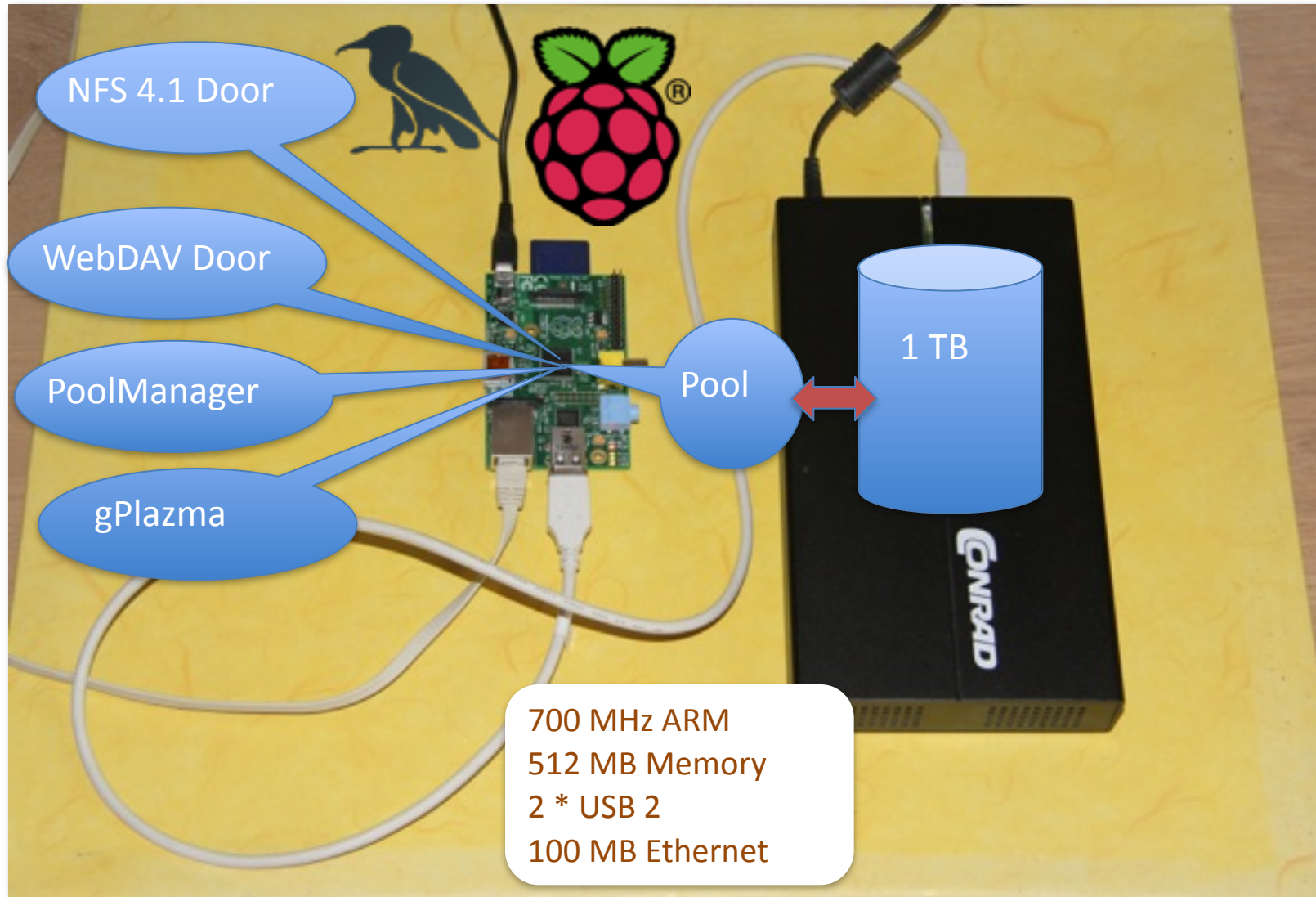
One dCache



Slide stolen from Mattias Wadenstein, NDGF

# To very likely the smallest

One Machine – One Process



# Hands-on

# Introduction

- Some things you need:
  - Your
    - login:root
    - password:dcacheWS@tw2014
  - WLAN: ISGC-?
- Wiki Page for this hands-on:  
<http://trac.dcache.org/wiki/Taipei2014>

# Content

- Installation of dCache server
- Writing to / Reading from dCache
- WebDAV using certificate-based authentication
- Pool Management and Pool Selection
- Hardware Lifecycle (adding storage, migrating)

# Installation of dCache server

- Installation from RPM
  - Configure NFSv4.1 and mount locally
  - Read/ Write File using mounted NFS
- put [twgrid.org](http://twgrid.org) behind the machine name,

# Writing / Reading

using webDAV door

- Configure webDAV access
- Read / Write file using user/password based authentication

# Certificates and dCache

using webDAV door + user certificate

- Configure webDAV access
- Establishing host certificates and CA-certificates in dCache
- Read / Write file using certificate based authentication (on client and server side)

# Pool Management

- Map pools to directories (pool selection)
- Familiarisation with admin interface
- Finding files in dCache

# Hardware Lifecycle

- Add pool
- Balance data between pools
- Migrate data to new pool
- Switch off old pool
- Access data that was there before
- Without service interruption

# Wrap Up

# What have you done:

- Installation of dCache server
- Writing to / Reading from dCache
- WebDAV using certificate-based authentication
- Pool Management and Pool Selection
- Hardware Lifecycle (adding storage, migrating)

Reproduction of workshop machine:

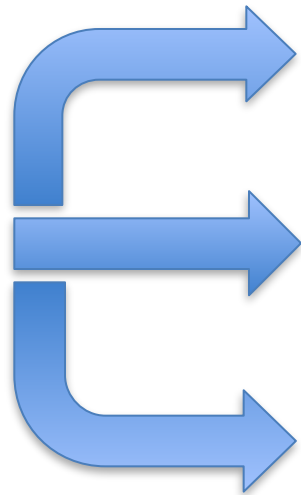
Github: <https://github.com/chrber/hands-ons.git>

## Customer Relations

# Deployment Channels

dCache.org 

dCache.org 



dCache.ORG / Web Pages



UMD

Targeting: EPEL

# Support Channels

- [support@dCache.org](mailto:support@dCache.org) ([security@dCache.org](mailto: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
- **EGL.eu**: First level support for dCache packages taken from UMD.
- Weekly customer **phone meetings**
- 2 dCache **workshops/year**



## Technology and design

## Design #1 Service Modules & Message Passing

## Design #2 Namespace – Physical Storage separation

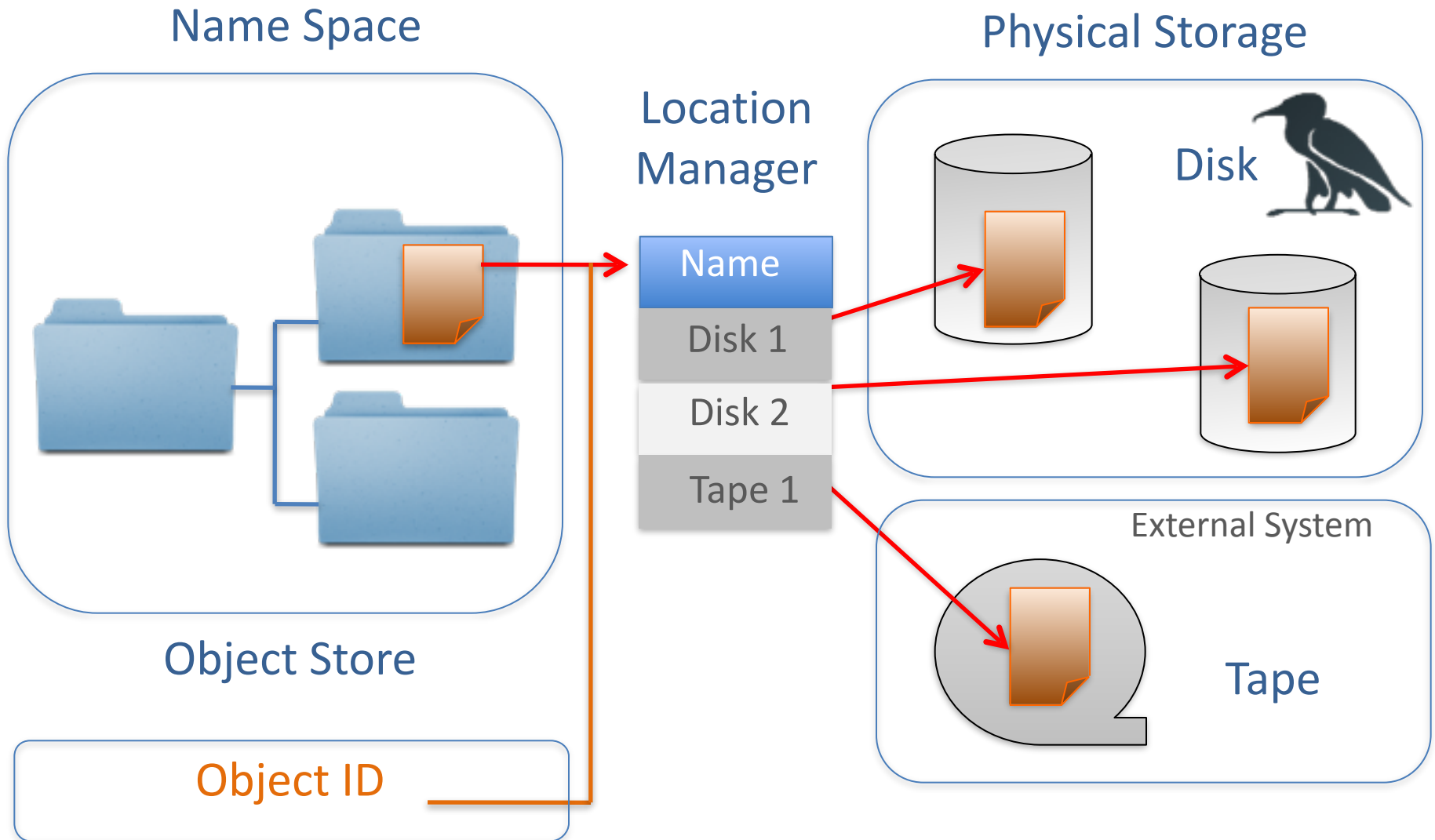
## Design #3 Services allow plug-ins

## Design #2

### Namespace – Physical Storage separation

# Design

## Namespace – Storage separation

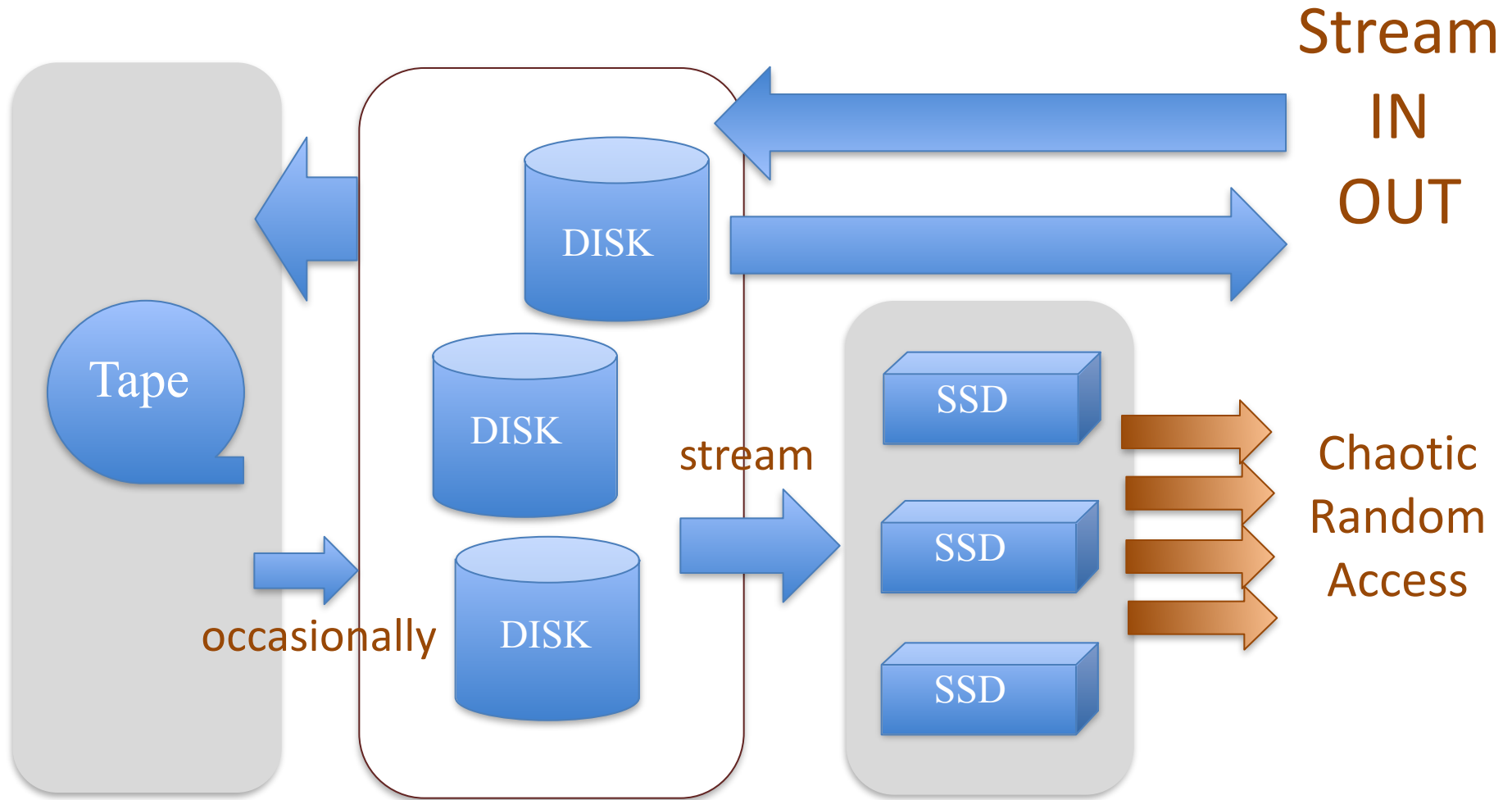


Resulting in .... Replica Management

- 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

## 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



DISK



DISK



SSD



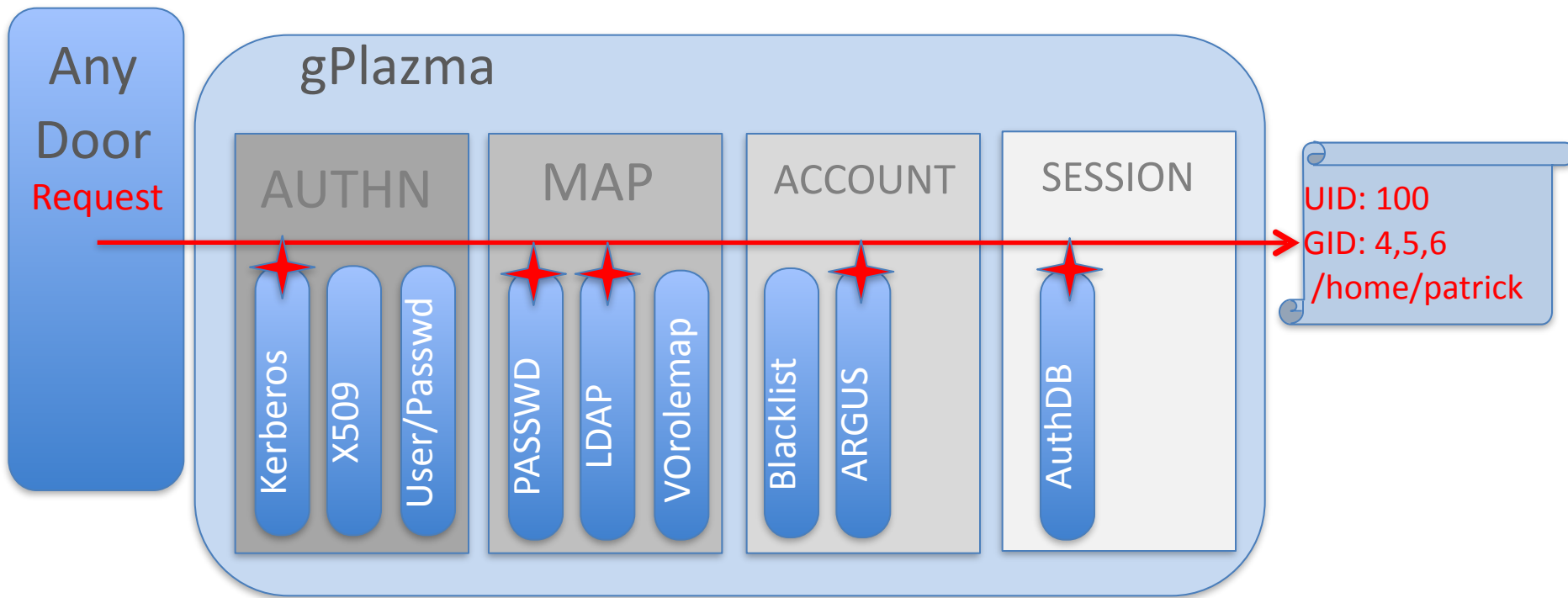
SSD



Tape

# 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

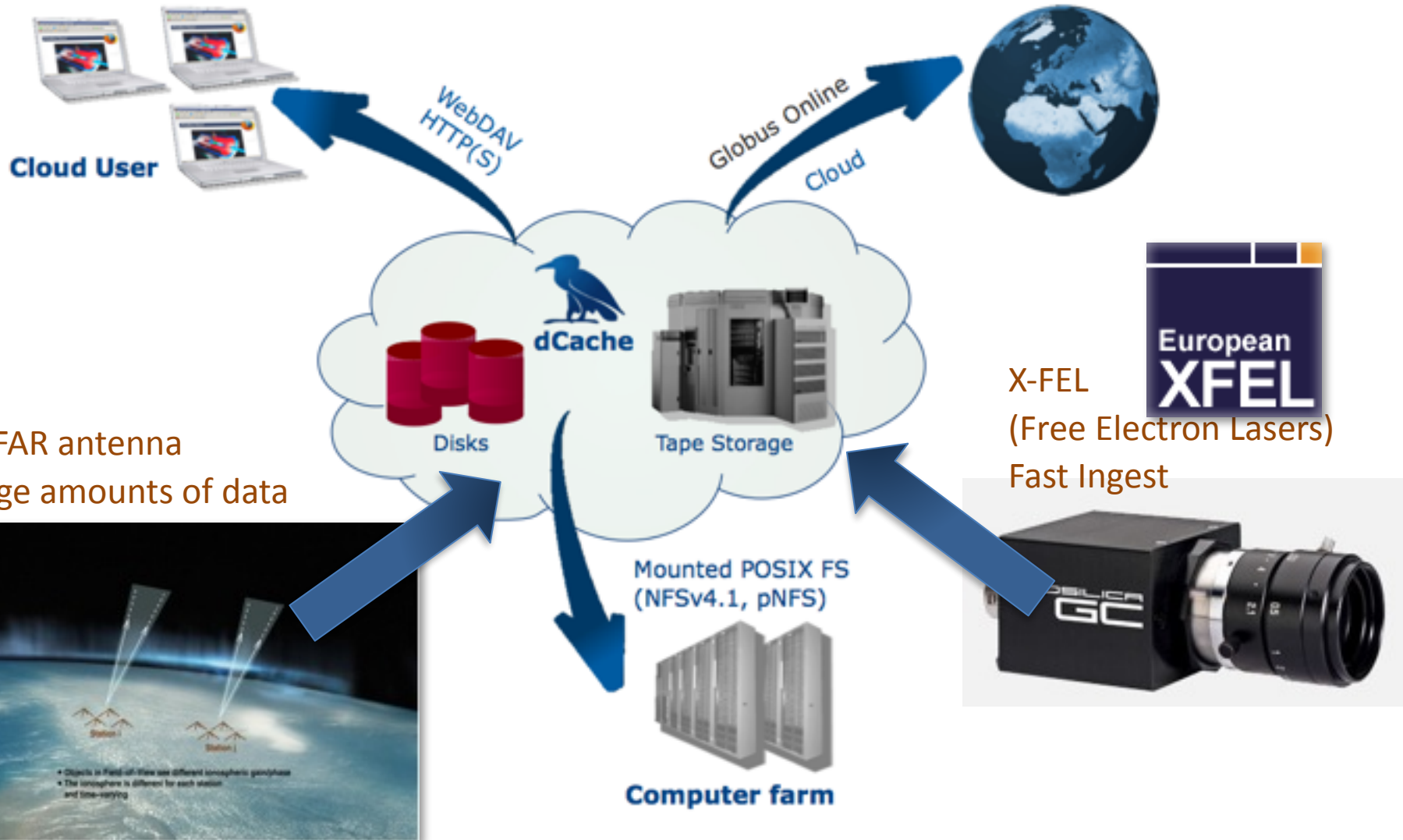
# Consequence of #3

- 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.

## 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



# dCache.org Cloud as a service

- 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:

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.

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

- Storage is about persistency
- So, a key component of a successful Storage Technology is a sustained project structure.
- How did dCache achieved that for the past 10 years and how will we proceed ?

# Who did we achieve that ?

- Three large institutions contribute to dCache with developers, development infrastructure and massive test systems.
  - DESY, FERMILab and NDGF
- For all three institutions, dCache is a strategic component, running in production.
- A continuous success of dCache is therefore in their interest.
- They build the bridge to national projects and activities.

- dCache has a long history in strong involvements in European and national activities.
  - Contributed to EGEE x.
  - Provided the data area leader of the European Middleware Initiative, EMI.
  - Provided the Storage Leader of the HEPCG division of the German D-Grid project.
- Currently:
  - EGI and LSDMA (but see later)

## Current project activities and next steps

- In order to get young people involved in our development and in using dCache, we started a close collaboration with the “University of Applied Sciences, HTW” in Berlin.
- Right now, 3 students are part of the dCache team on their way to earn a Bachelor or Masters degree.

- With “Photon Science” at DESY (X-FEL and CFEL) and “Intensity Frontier” at FERMILab, we are extending our feature set to support non-WLCG and non-HEP communities. (See later, e.g. storage management through nfs, high through put single stream input and support of modifiable files, AAI).
- They benefit from dCache’s strong move towards standards in data access, data transfer and identity management.

- EMI funding, which ended May this year, was taken over by the German “Large Scale Data Management and Access” project. (See presentation by Andreas Heiss, SCC).
- dCache is leading the LSDMA “Federated Identity” and the “Federated Data Access” work packages and it contributing to the “Meta Data” work package.

- People like “Storage Management” features in dCache:
  - Flush to tape
  - Recall from tape
  - Etc
- People like mounting dCache with NFS 4.1/pNFS

# LSDMA Data Management

## (More examples)

- Implementing the CDMI Cloud protocol in dCache, possibly followed by S3. UNICORE will implement the client part as a proof of concept.
- Supporting the NEXUS and HDF5 file format in dCache, which essentially means supporting non-immutable data.
- Building an NFS 4 German data federation, possibly using FedFS.

## 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

# 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 ½ of the entire WLCG data.

