

dCache, managed Cloud Storage

2nd International Scientific Conference Science of the Future

Patrick Fuhrmann

On behave of the project team



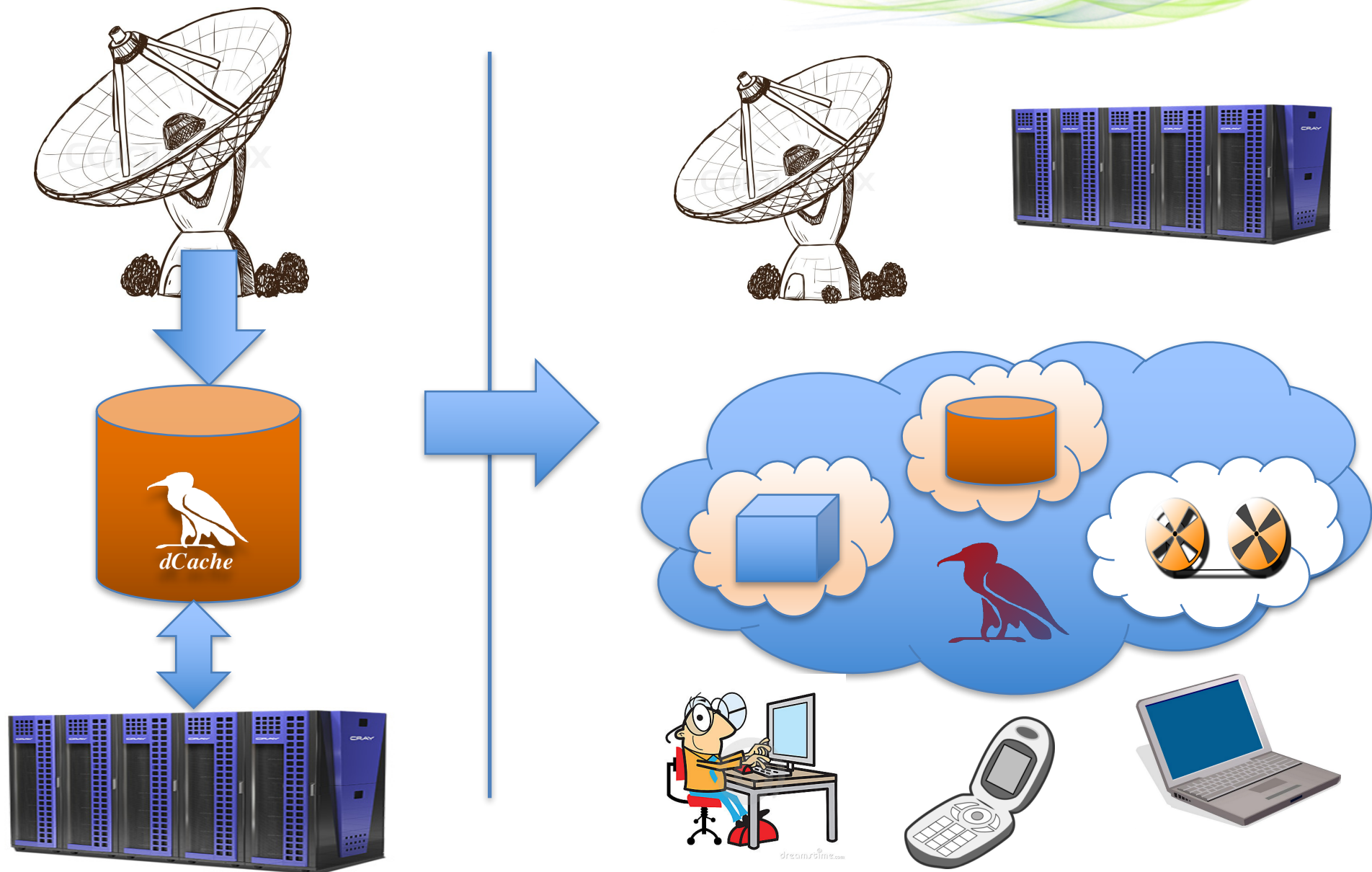
That's this about



- The Technology
- The Deployments
- The Collaboration
- The Funding influence on the development
- Design Principle
- Consequences of the design
- Improvements in Operations
 - Unbreakable
 - Adopt object stores
- Improvements for the customer
 - Quality of Service in Storage
 - Sync'n Share
- The ultimate scientific life cycle engine

Or in other words ...

In other words



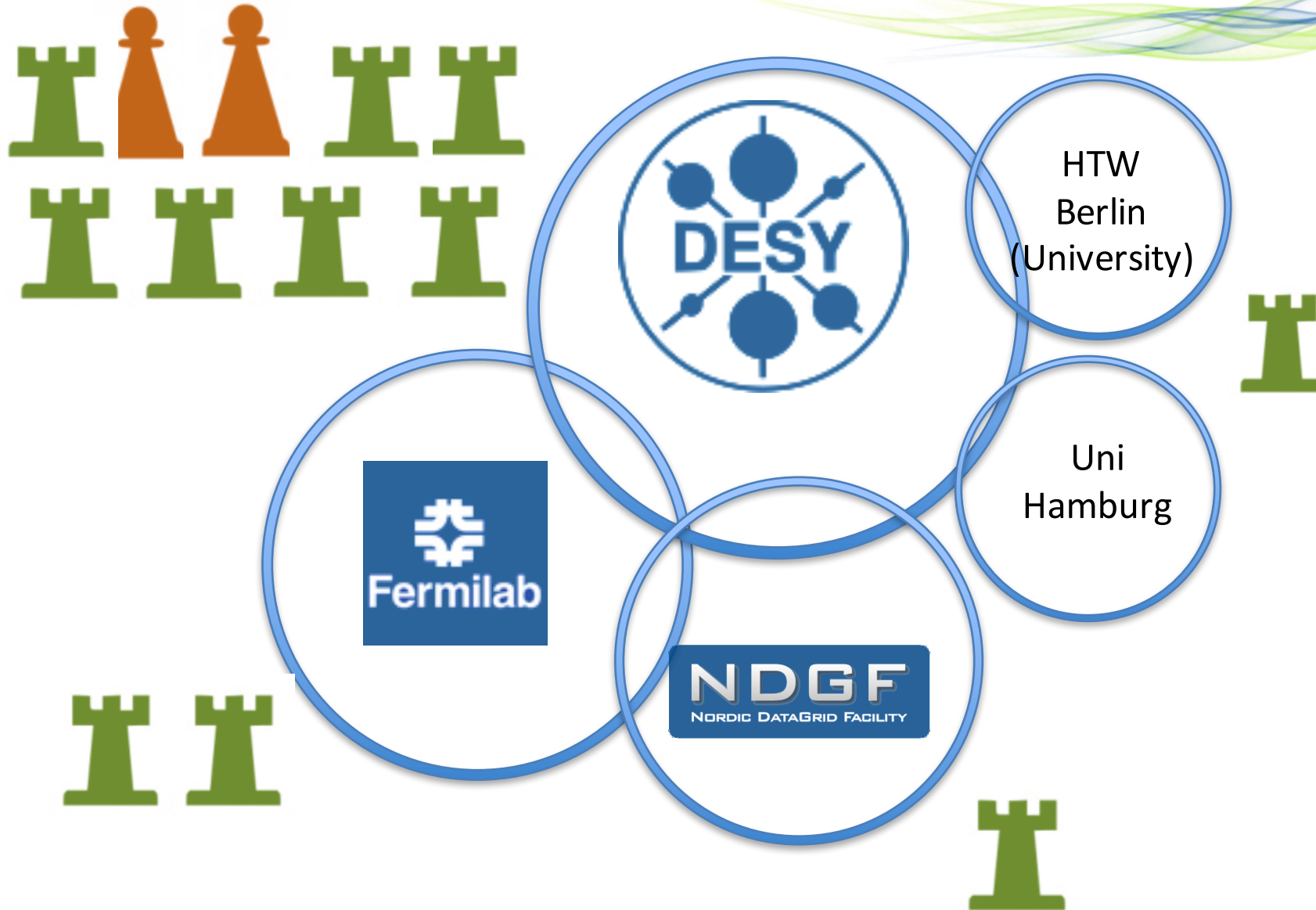
The Technology Cheat Sheet



- Combines heterogeneous storage nodes under a common virtual file system tree and scales into 100PB region.
- Provides access to data via a variety of protocol, e.g. NFS4.1, WebDAV, GridFTP, etc.
- Provides a variety of authentication mechanisms, like User/Pass, X509 Certificates, Kerberos, in preparation SAML and OpenID Connect, Macarons.
- Multi Tier support: moves data around between different media types, like Tape, Spinning Disks and SSDs.
 - By user request.
 - Automatically based on the access profile, hot spot.
- Provides resiliency, e.g. through multiple copies.

The Collaboration

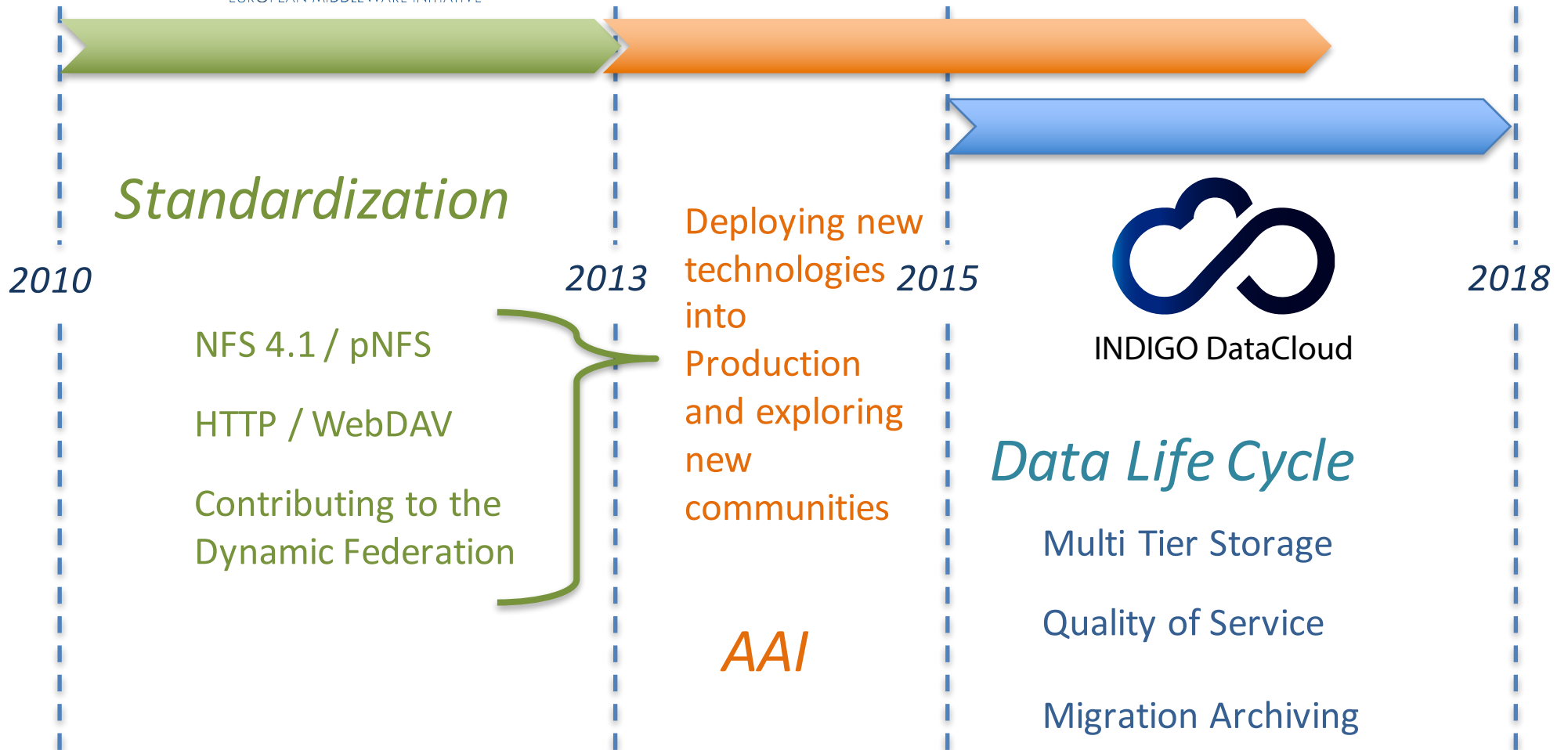
The dCache.org collaboration



On funding and technical directions



Funding influences dCache development topics



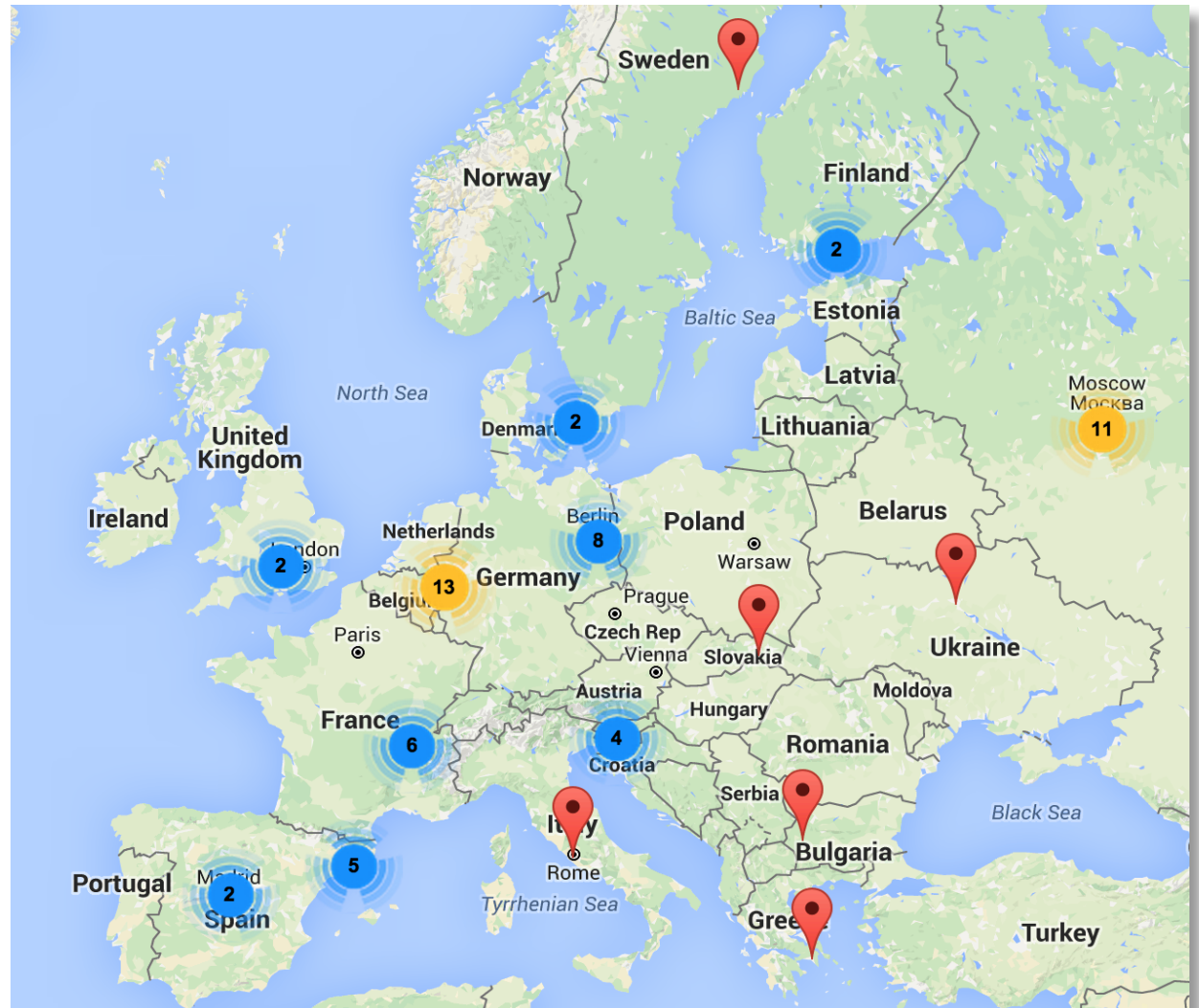
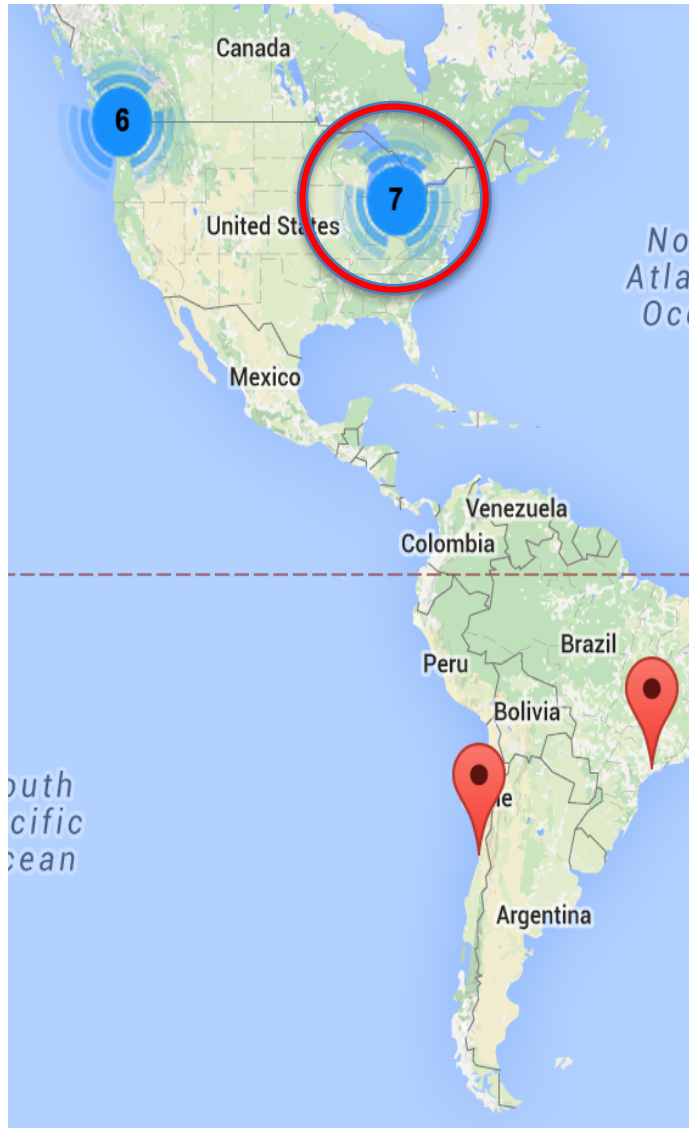
dCache Deployments

Huge, Wide and small

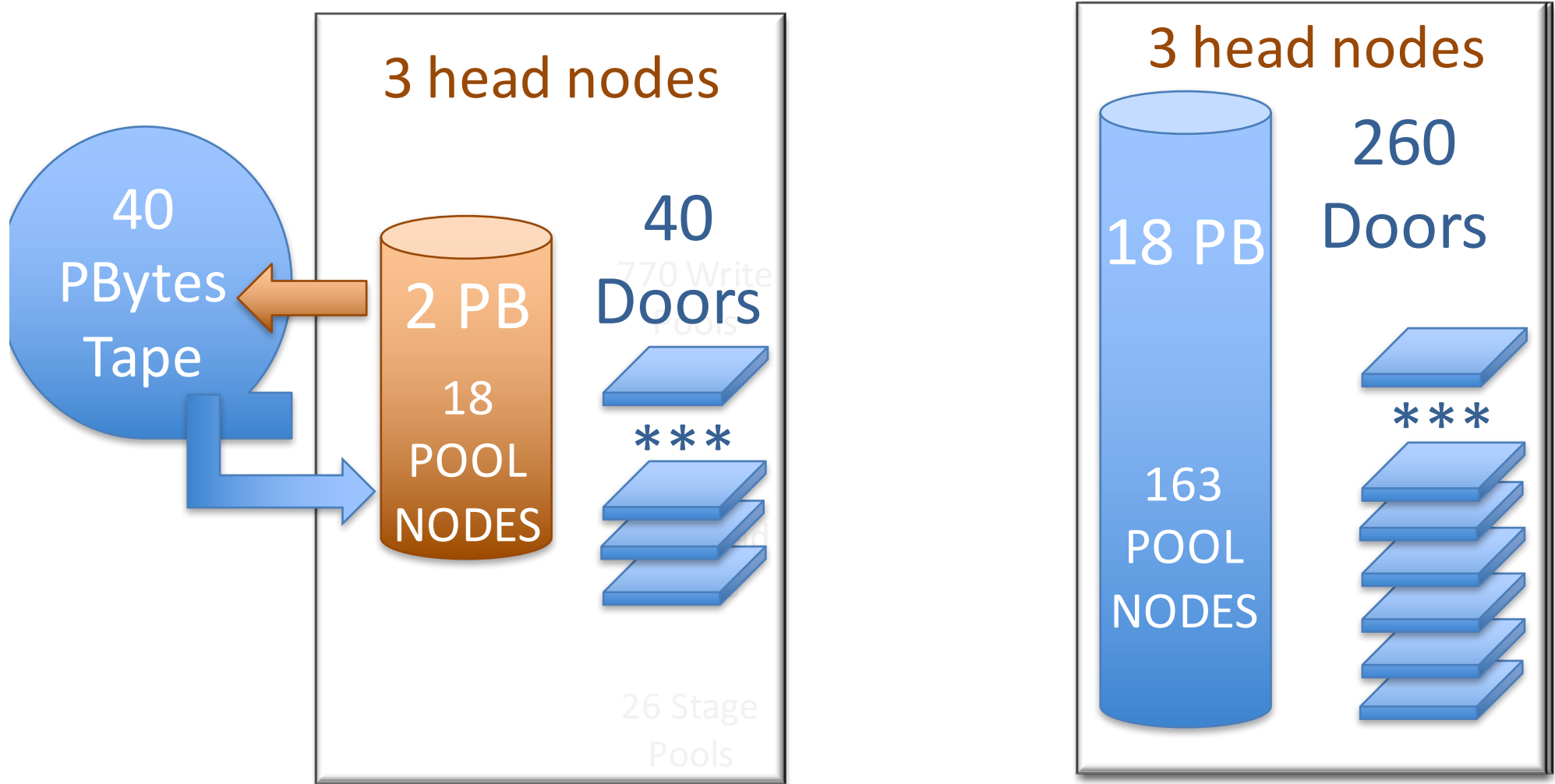
Worldwide distribution



Worldwide distribution

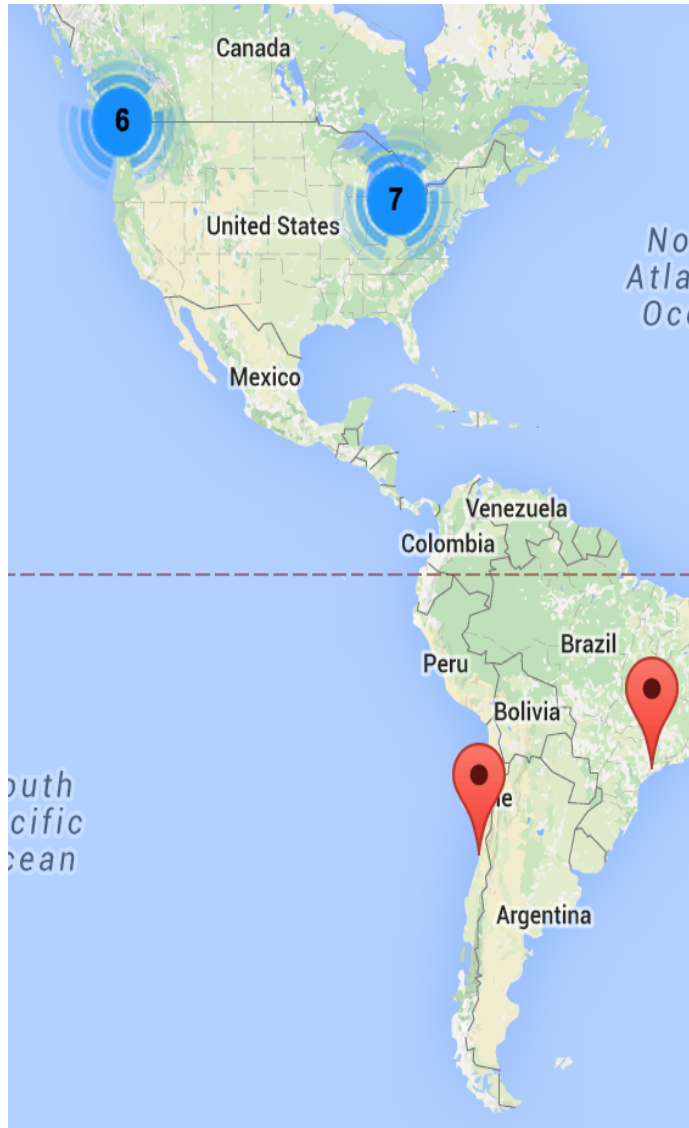


Starting with possibly the biggest US-CMS Tier I 18 PBytes on Disk



Information provided by Catalin Dumitrescu and Dmitry Litvintsev

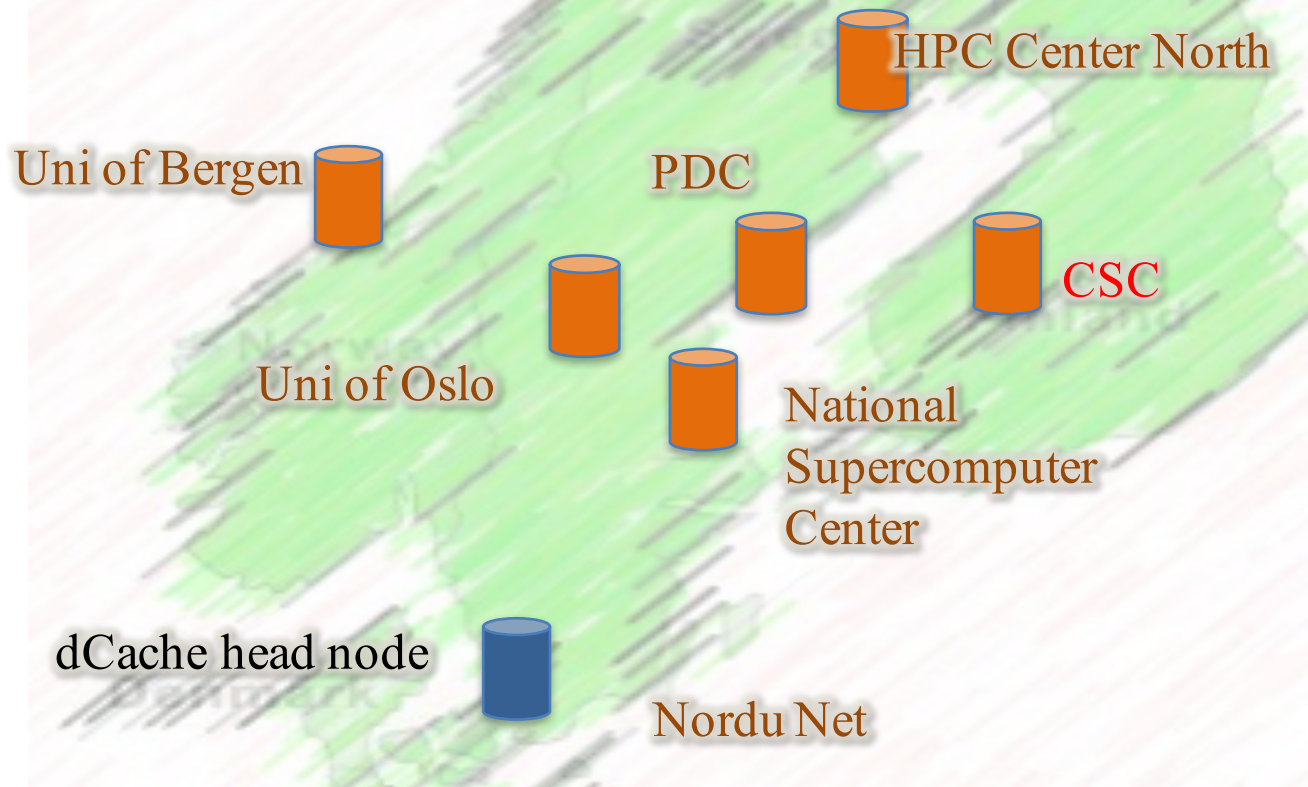
Worldwide distribution



To certainly the
most widespread

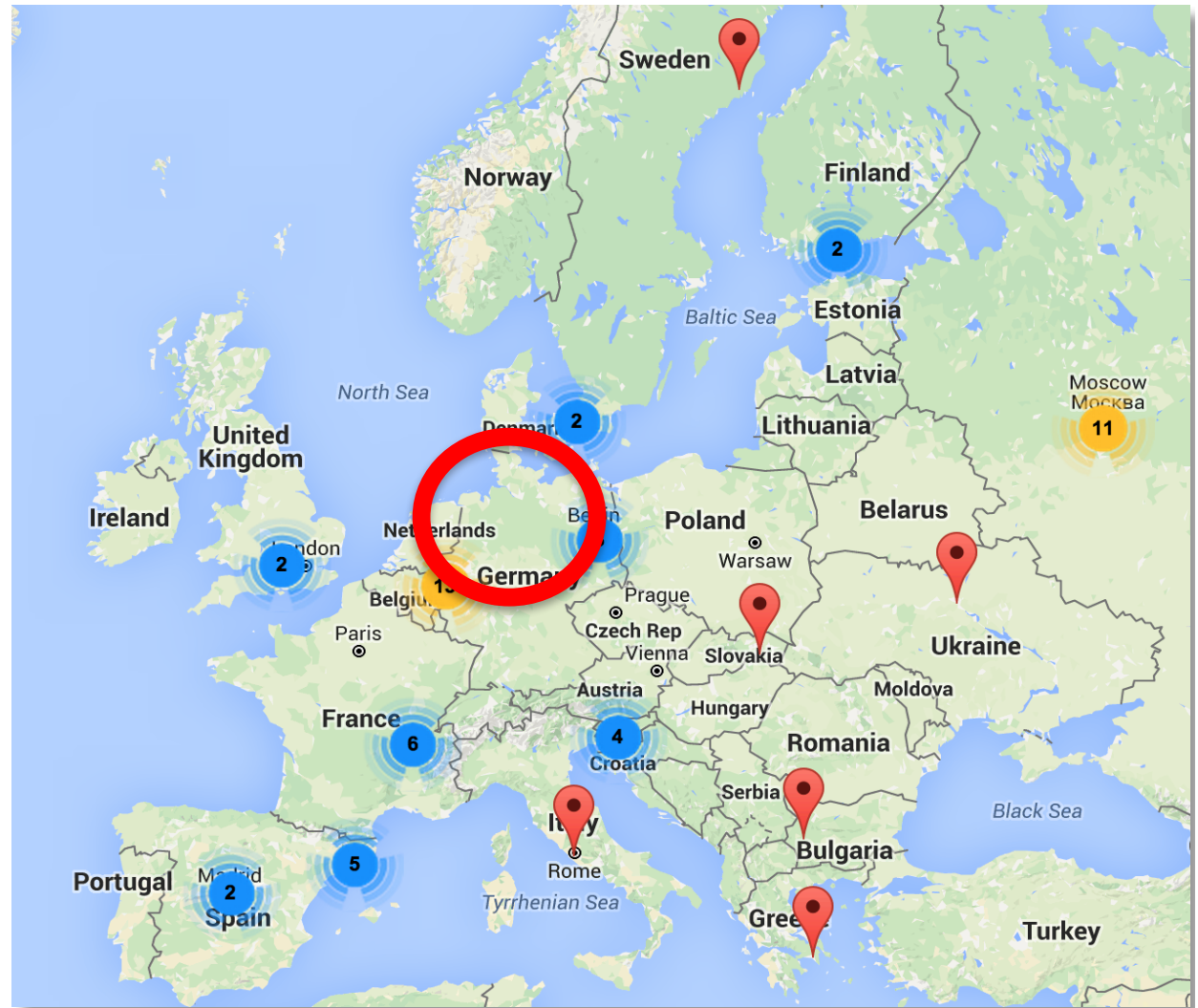
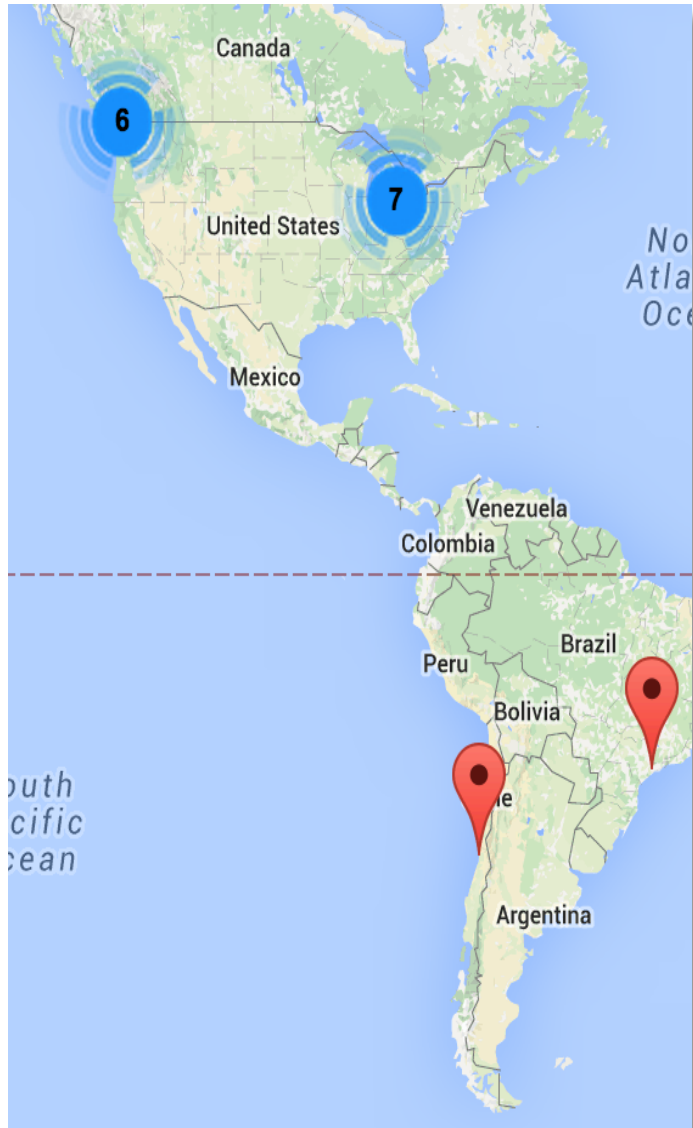
5 Countries

One dCache



Slide stolen from Mattias Wadenstein, NDGF

Worldwide distribution

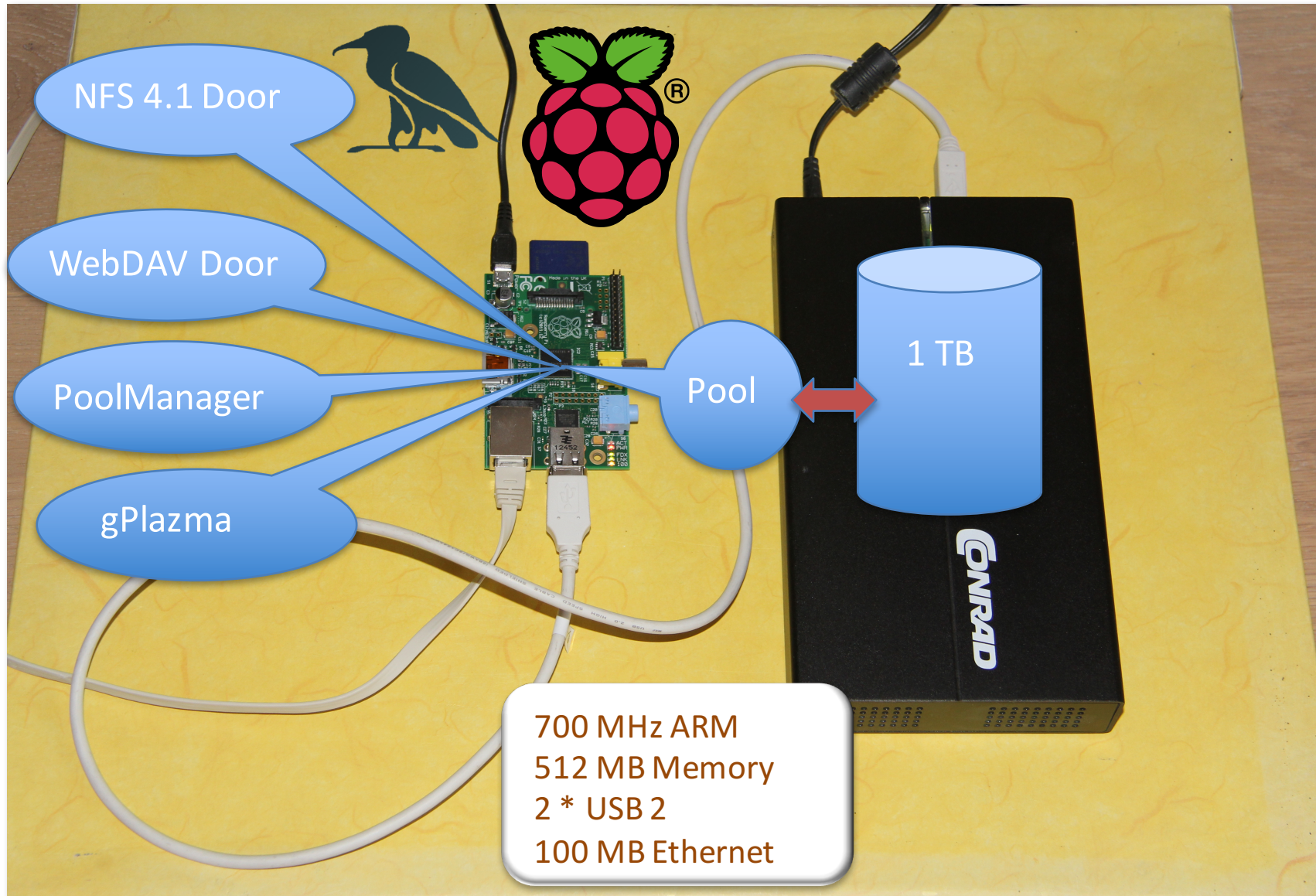


Hamburg, Eimsbuettel



To very likely the smallest

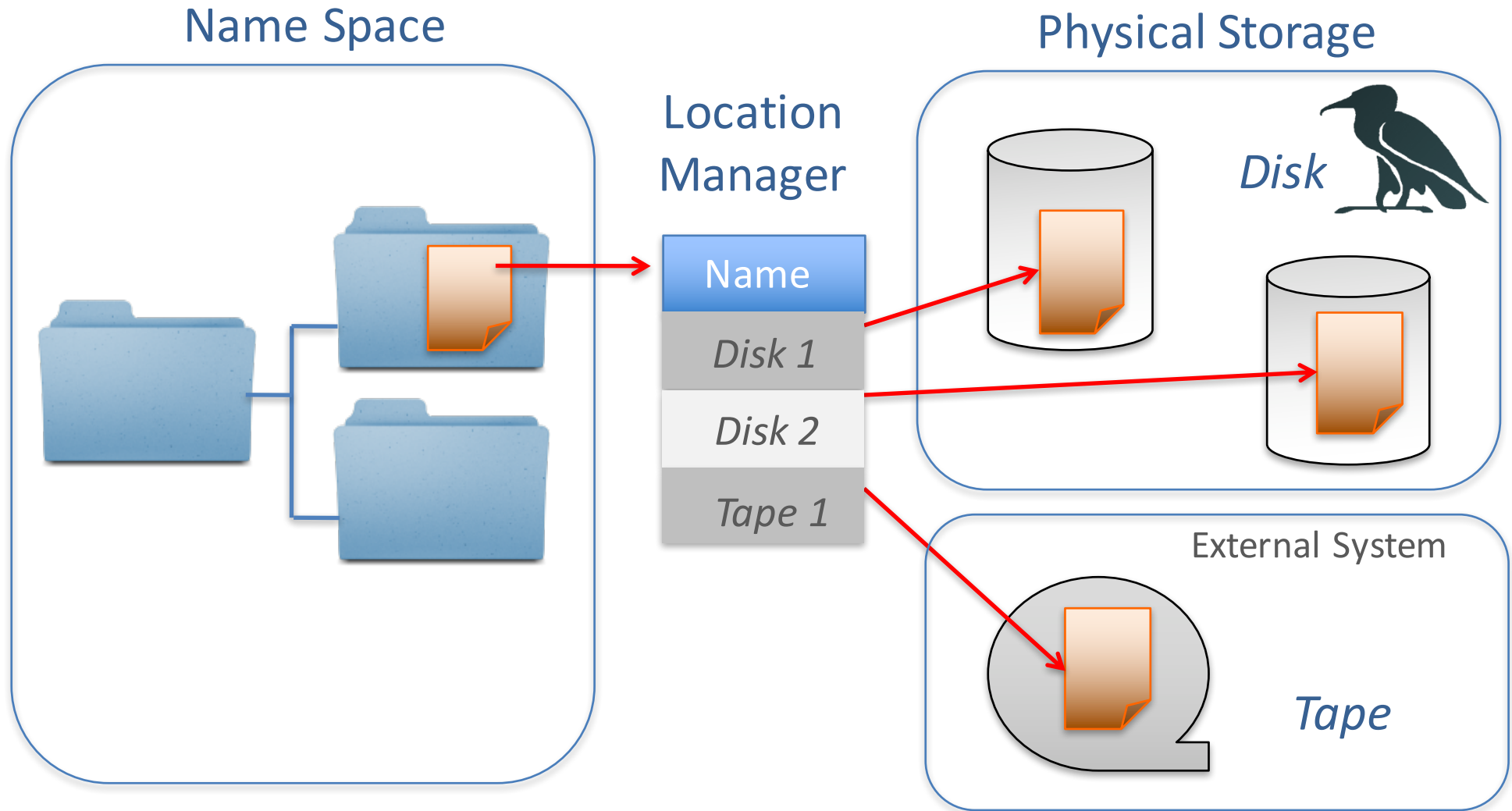
One Machine – One Process



Design Principles

Design

Namespace – Storage separation

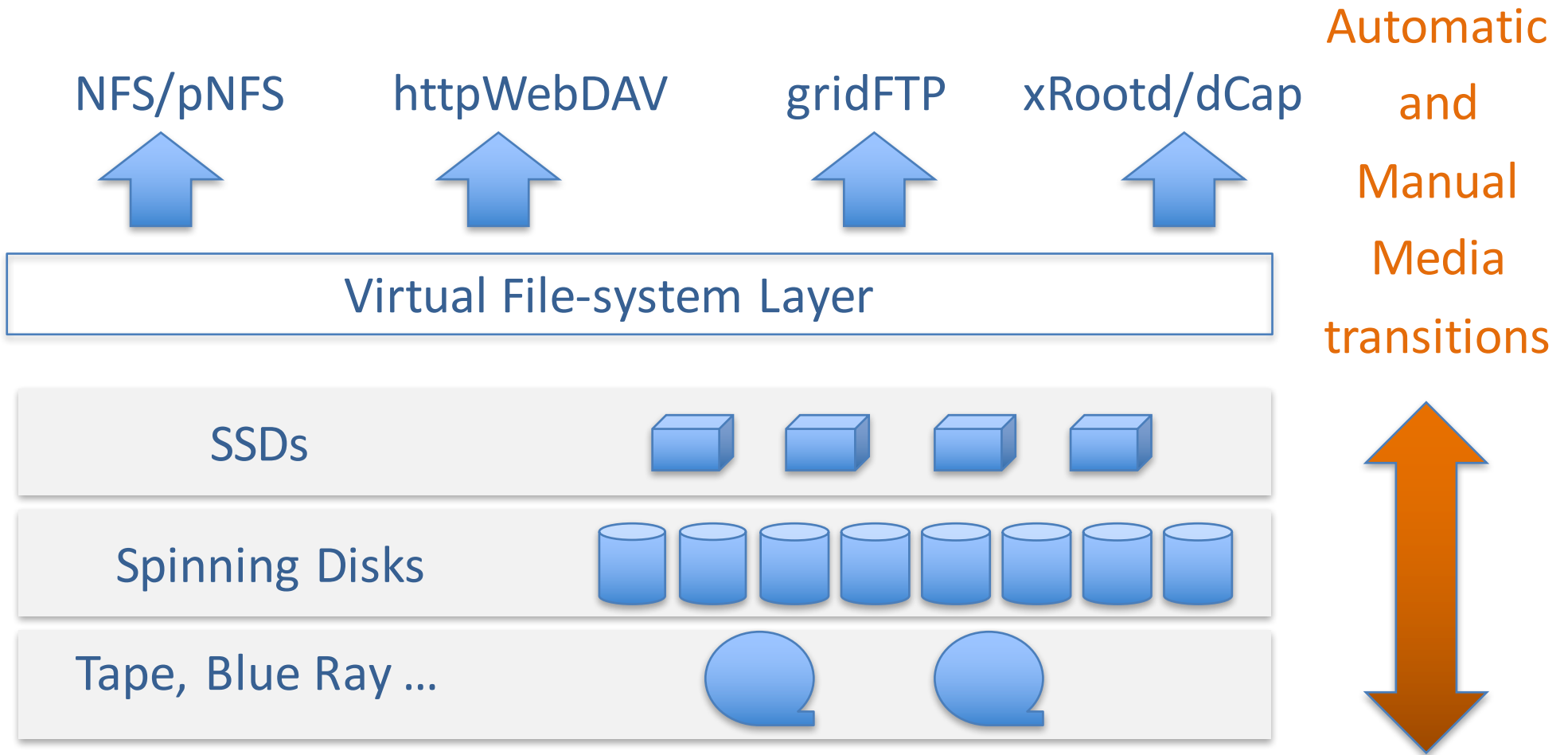


Consequences of this design pattern

Consequence : Multi Tier support

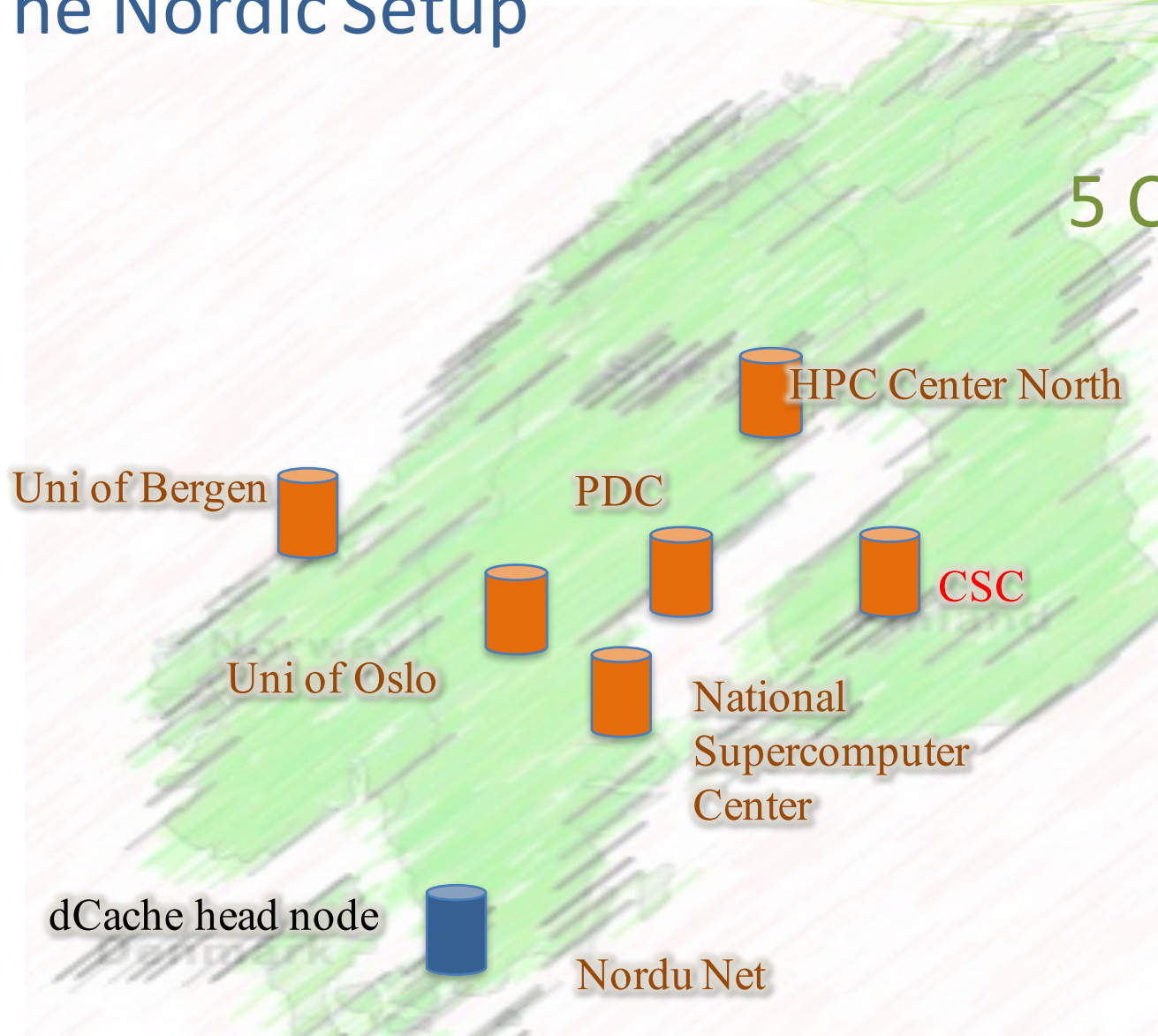
SSD, Spinning, Disk, Tape, multiple file copies.

dCache supports multi-tier storage and transitions.



Consequence : Federated Storage Structures

Federating Storage The Nordic Setup



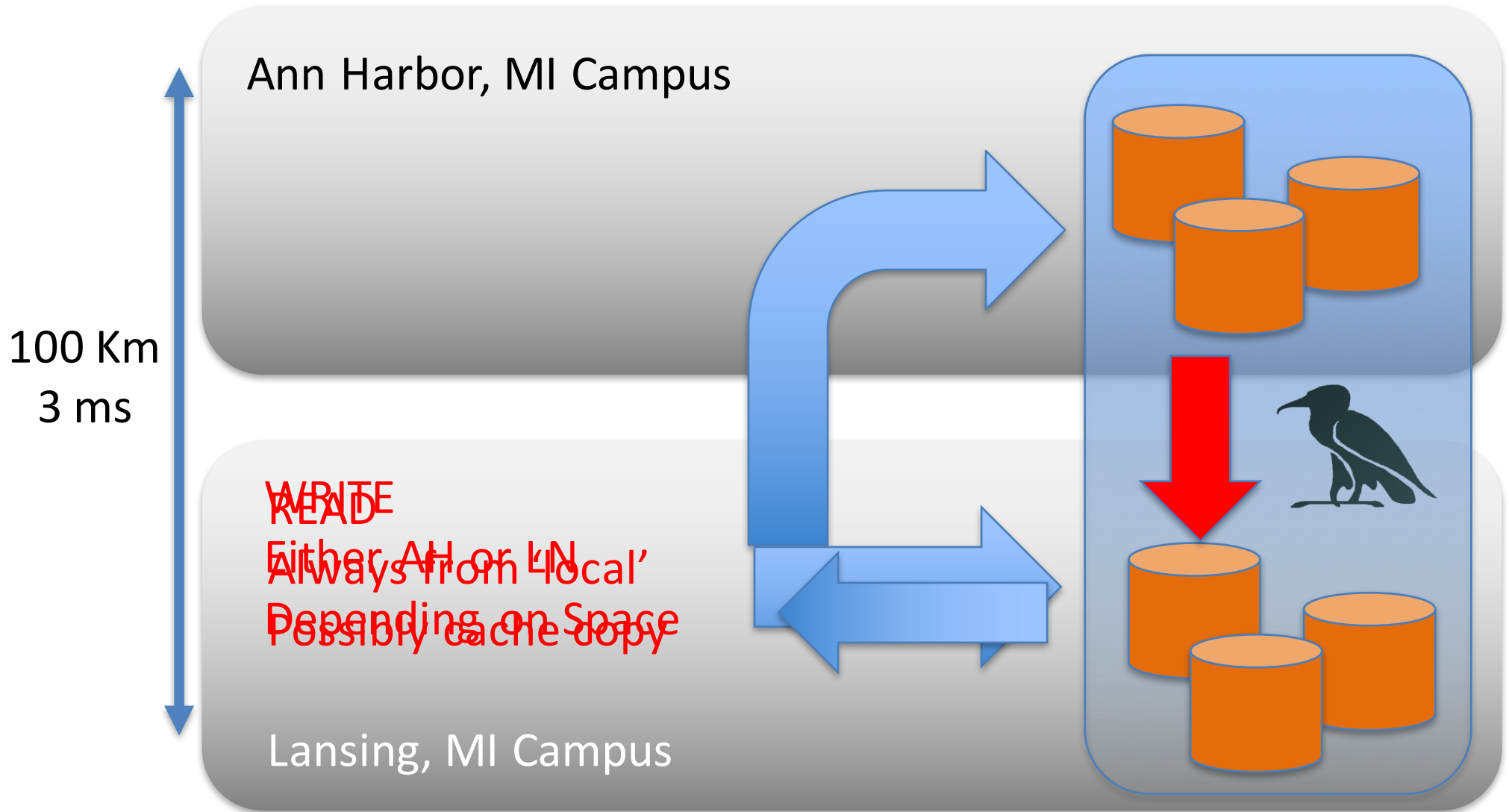
5 Countries

One dCache

Slide stolen from Mattias Wadenstein, NDGF

Federating Storage

The Michigan Setup



More consequences

- Hot spot detection and mitigation.
- Data migration to add or decommission hardware
- Resilient Manager : creating 'n' copies on different pools nodes to allow 'n-1' pools to fail before the system degrades.
- And many you can think of

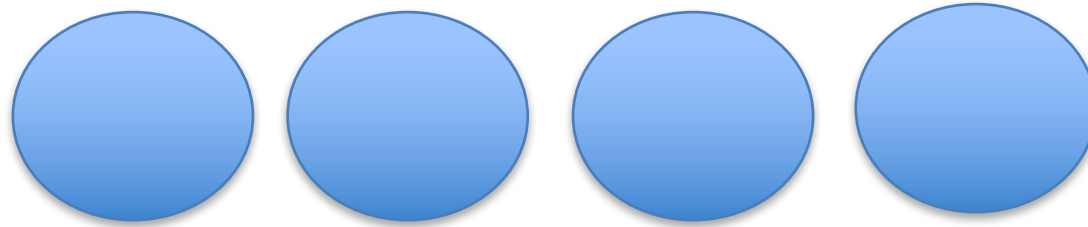
What do we need to make this even better
suited for cloud applications ?

Improved operations (I)

Make it unbreakable

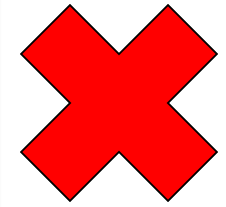
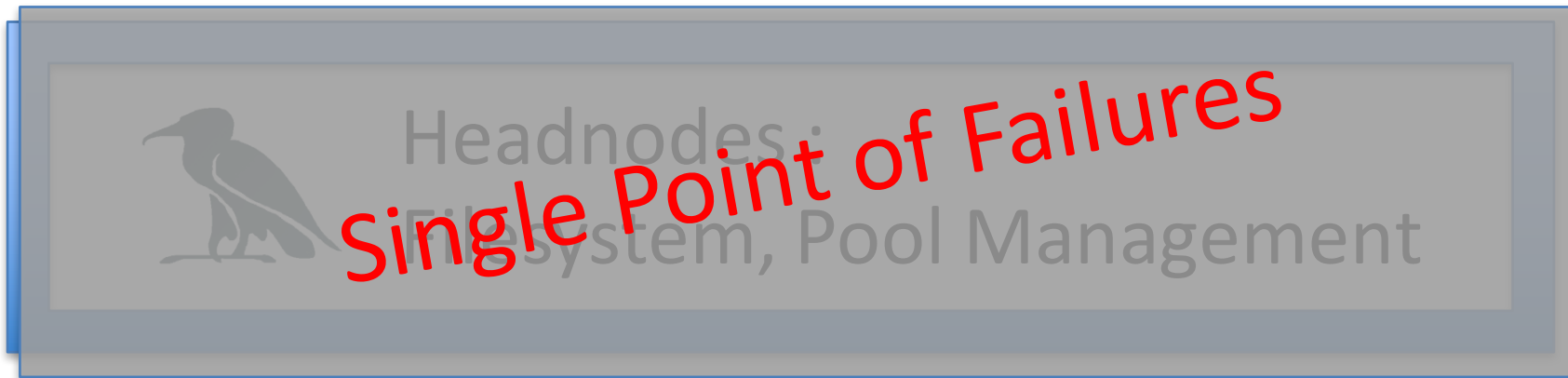
High Availability

Redundant
Doors
(Protocol Engines)

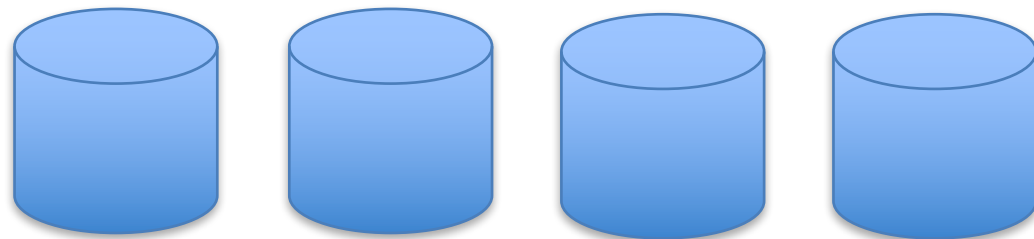


Headnodes:
Filesystem, Pool Management

Single Point of Failures



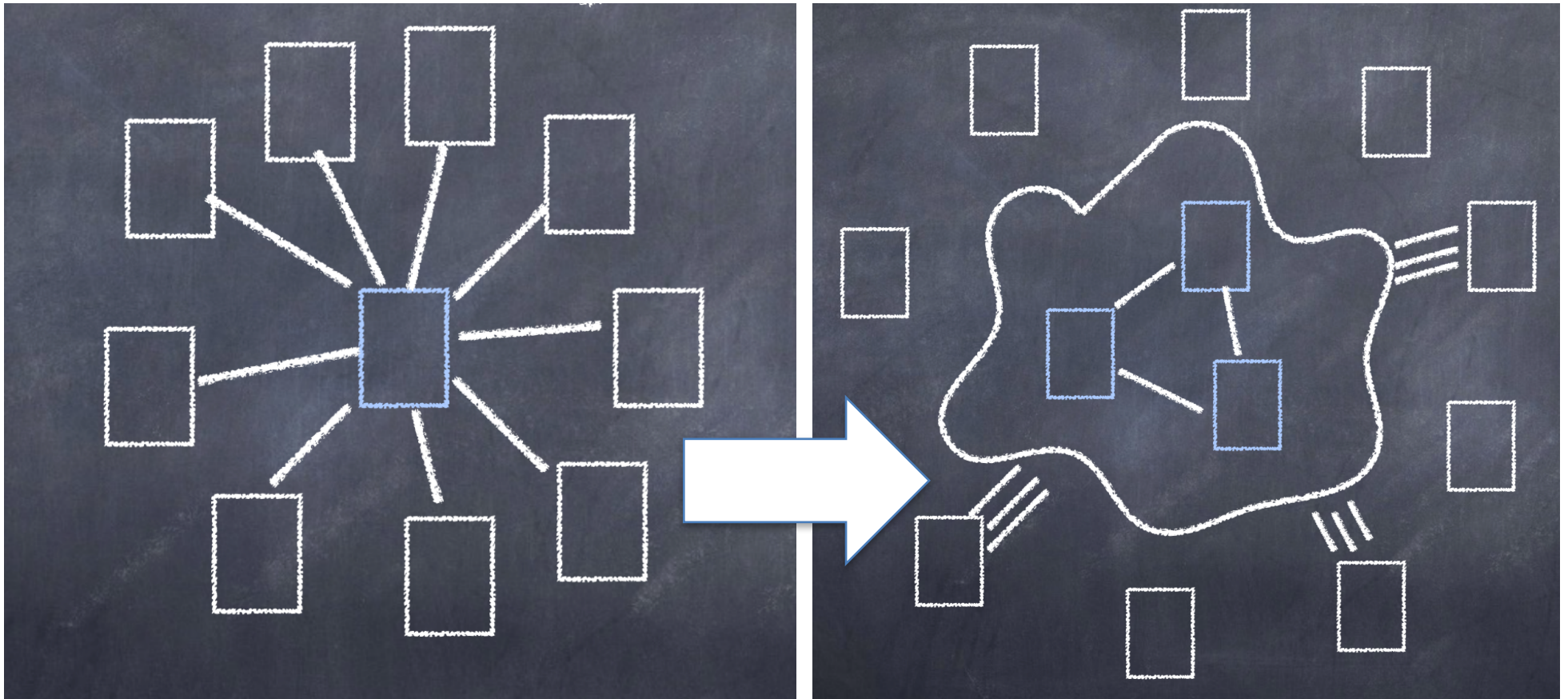
Redundant
Pools
(Replica Manager)
Multiple File Copies



Some remaining issues to fix

- The message passing has be fixed to overcome failures of an essential path segment. 'rerouting'
- How do we make services redundant which are singletons ?
- How do we manage failures of state-less services ?

High Availability



Any single component can fail, w/o breaking the service

Stolen from Gerd Behrman, NDGF

High Availability

- Singletons (build quorums, e.g. using Zookeeper)



- Stateless services : use publish subscribe



High Availability



Result : at any point in time, one internal service (node) can fail without consequences on the overall service.

Essential for a huge 24/7 installation.

Improved operations (II)

Integrate scalable easy to maintain solutions e.g.
CEPH

Delegating work to external Storage Layers

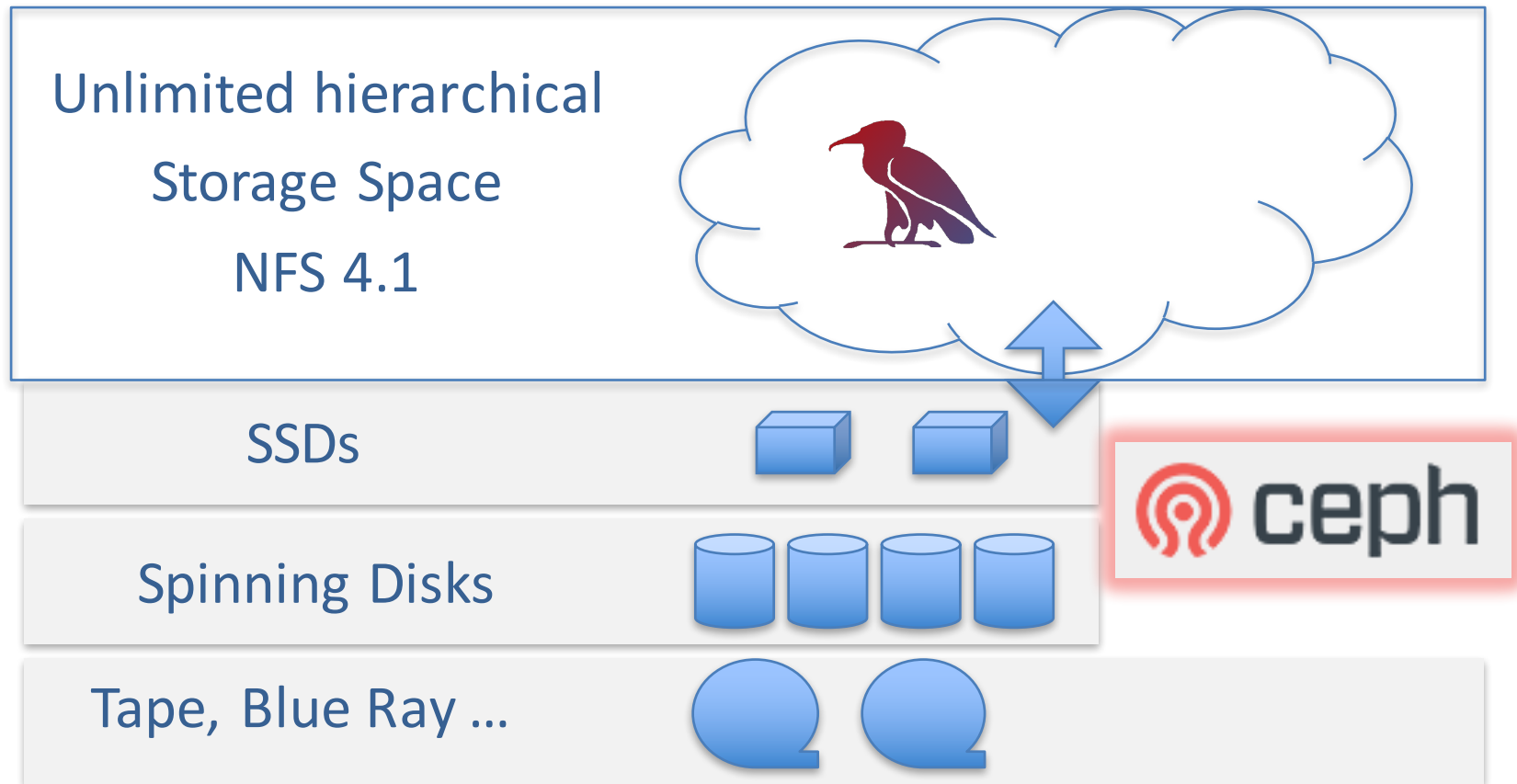
- Provides a single-rooted namespace.
- Metadata (namespace) and data locations are independent.
- Uniquely handles different Authentication mechanisms, like x509, Kerberos, login+password, auth tokens.

Can be delegated

- Provides access to protocols (WebDAV, NFS, DCAP).
- Provides data migration between multiple tiers of storage (DISK, SSD, TAPE).
- Aggregates multiple storage nodes into a single storage system.
- Manages data movement, replication, integrity.

Slide stolen from Tigran Mkrtchyan

New Technologies in dCache

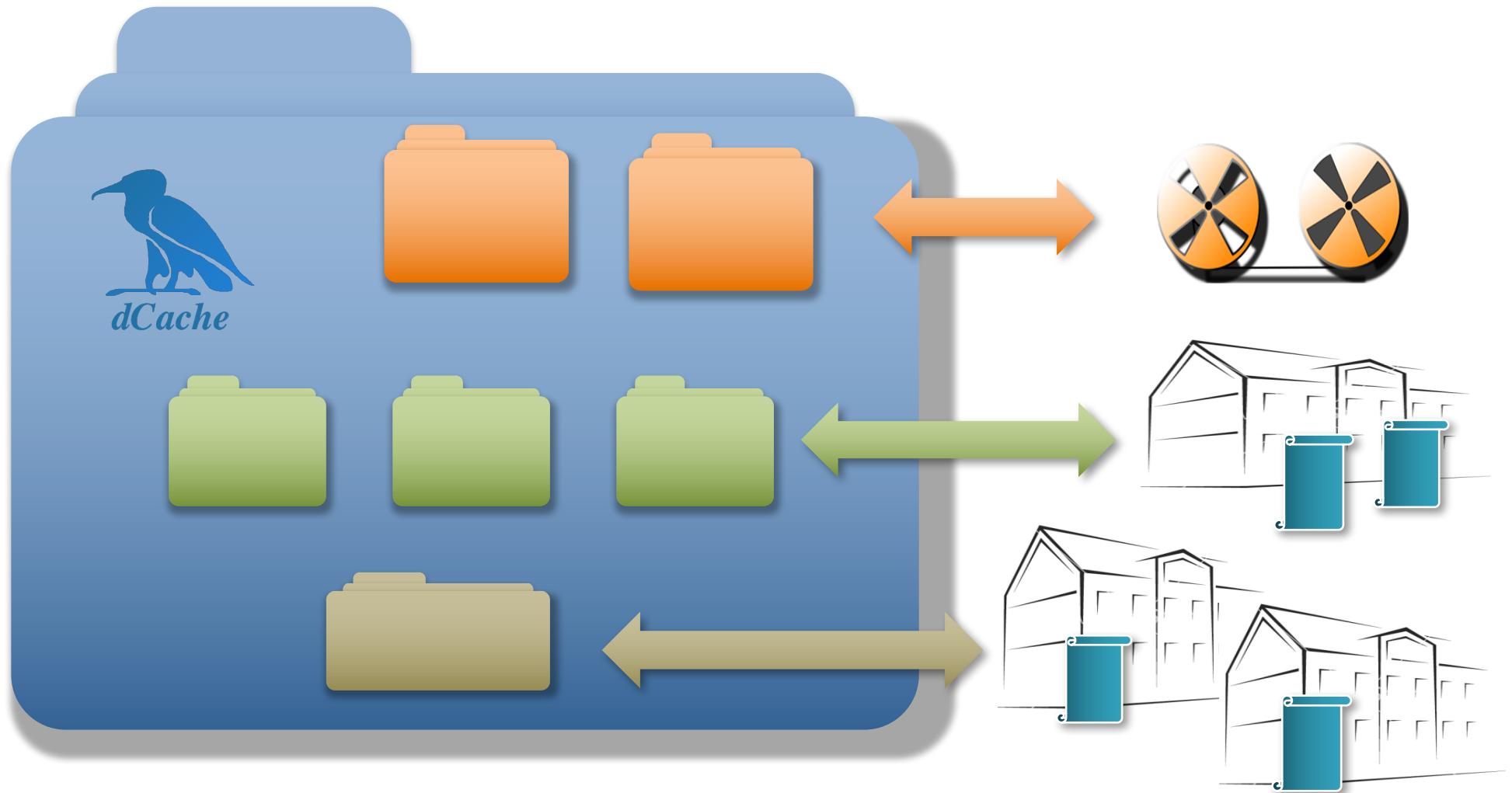


Storage has different qualities

- Example from home
 1. A ripped DVD is not really valuable but
 2. Pictures of your kids are irreproducible.
- There are similar examples in science e.g. collected earth climate data.
- Google and Amazon already provide different service levels in storage (e.g. S3 and Glacier)

We can do the same, even on directory basis.

Storage Quality selection



That was an easy one with dCache,

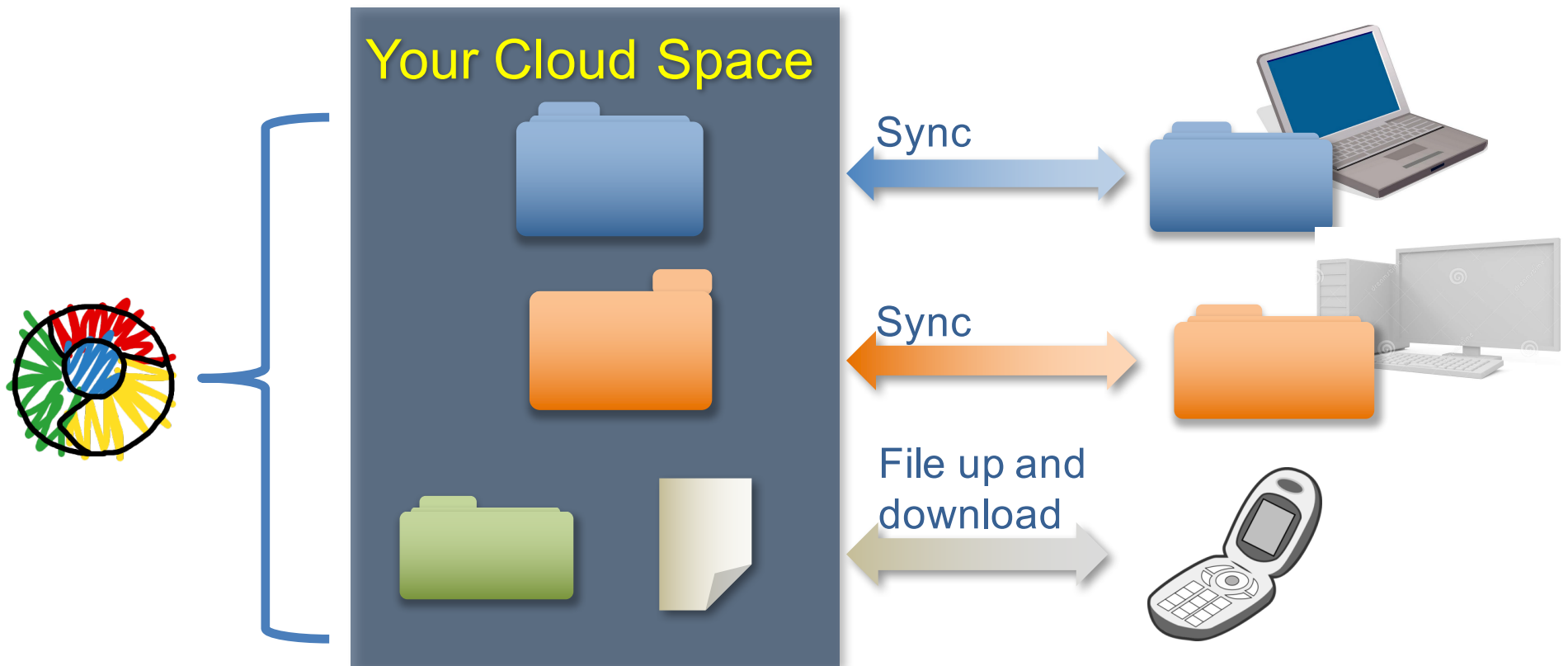
however

Make sure to store and analyse the billing information to charge customers for high quality storage, like multiple copies and tape usage.

Now, how about the web 2.0 feeling

Sync'n Share ?

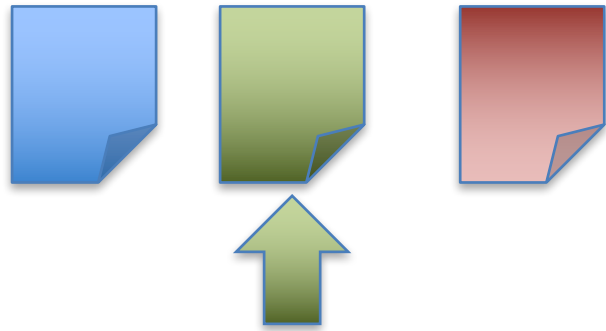
The cloud feeling, Sync'n Share



The cloud feeling, Sync'n Share

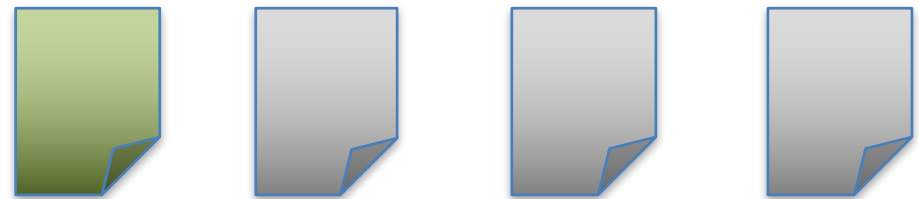


My Home



Share this one with you

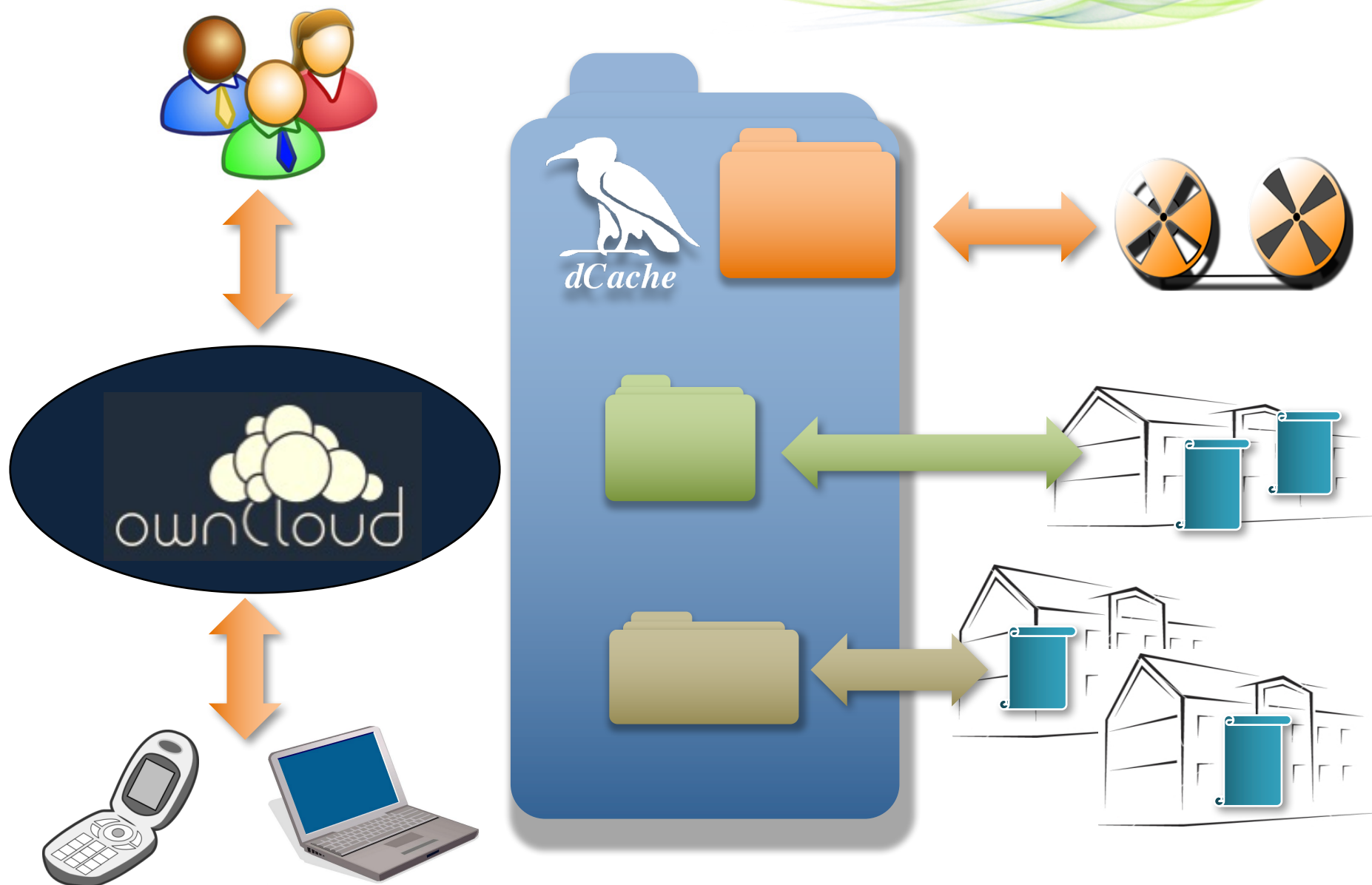
Your Home



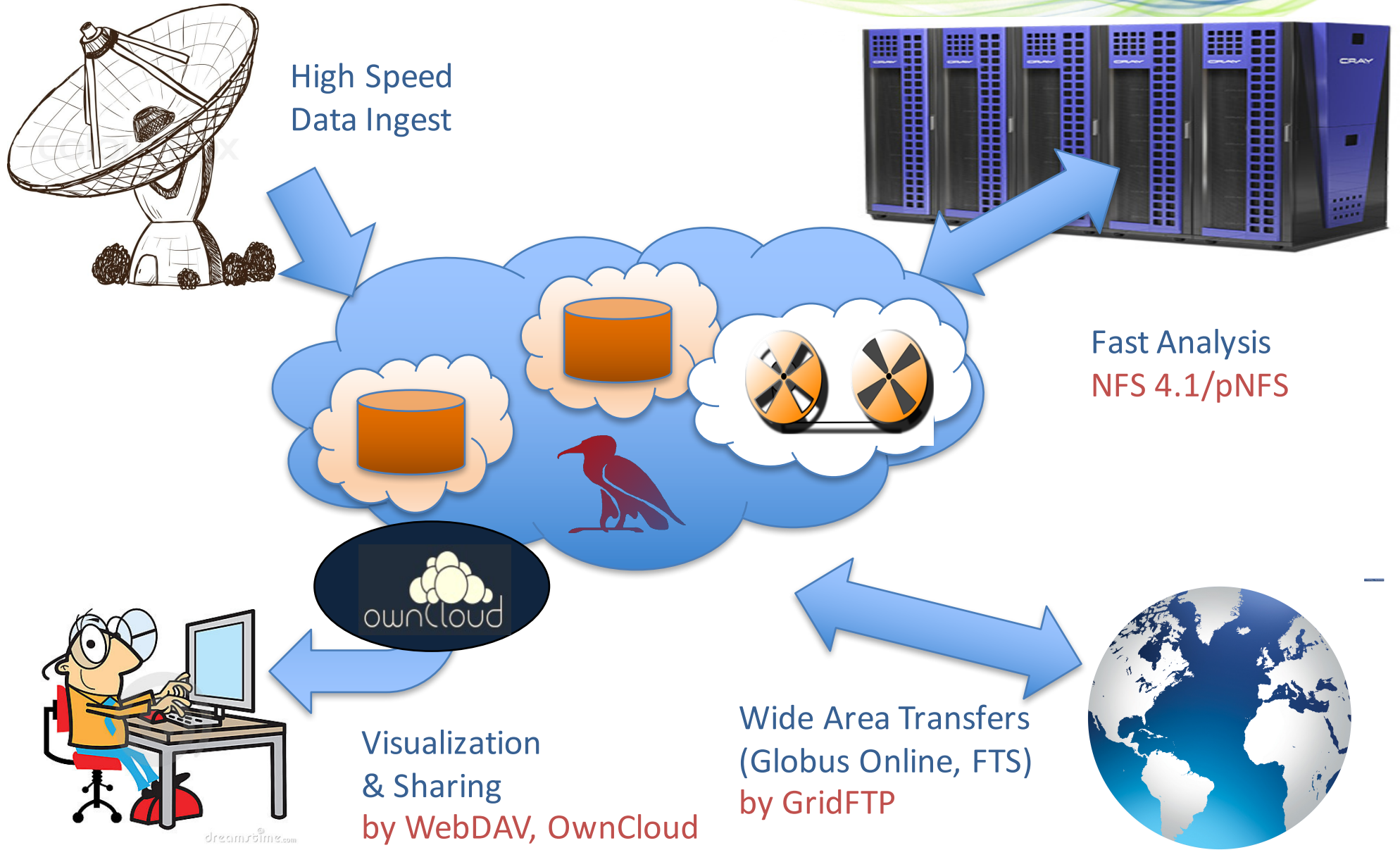
And as we didn't want to invent the wheel again, we picked the ownCloud (nextCloud) software do to the trick for us.

So in summary :

The Hybrid System



Scientific Data Lifecycle



The END

further reading
www.dCache.org