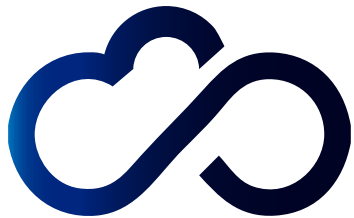


# Sync and Share, Quality of Service and dCache

Patrick Fuhrmann

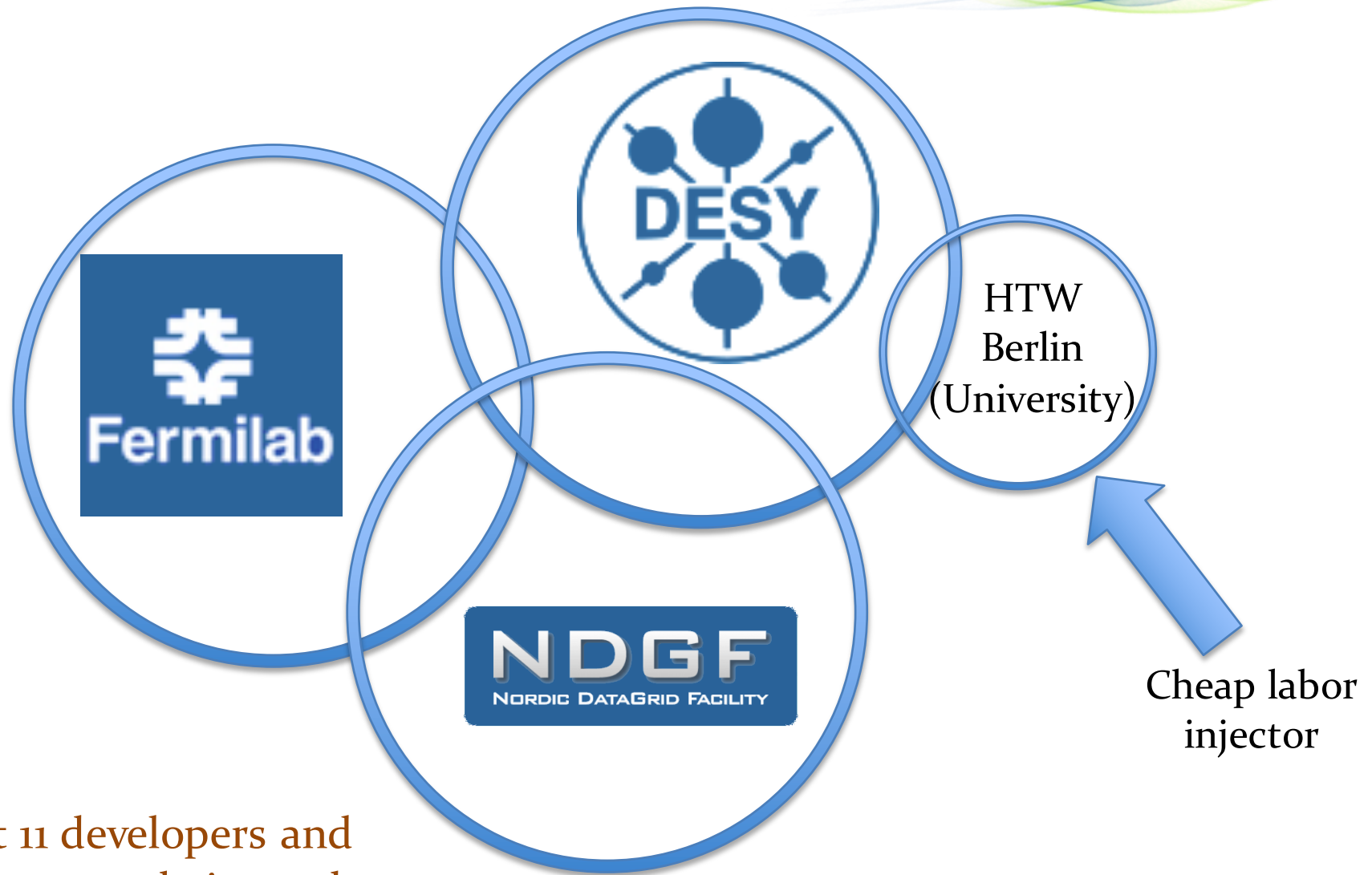
On behalf of the dCache team  
but especially  
Tigran, Lusine and Quirin



INDIGO DataCloud

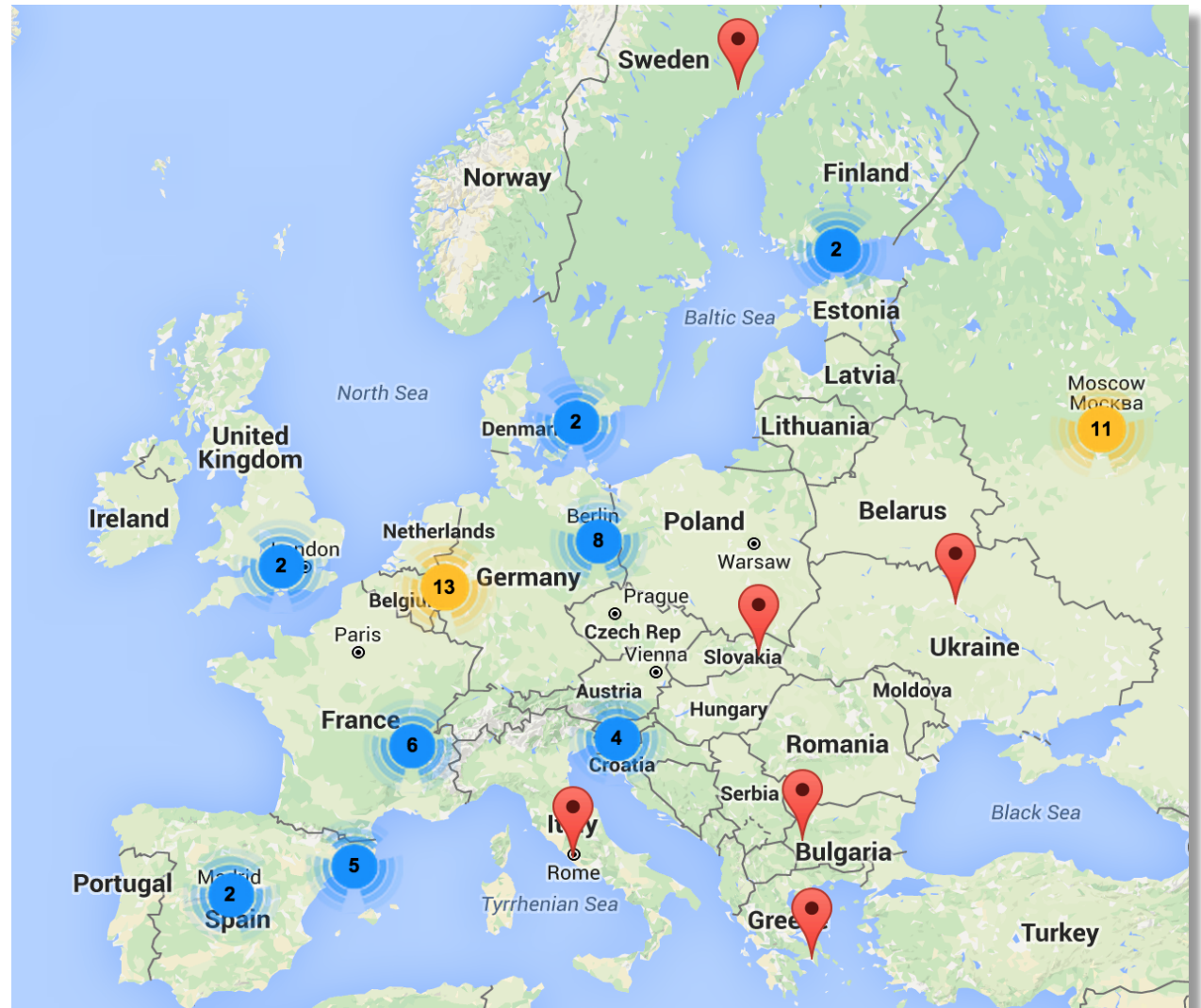
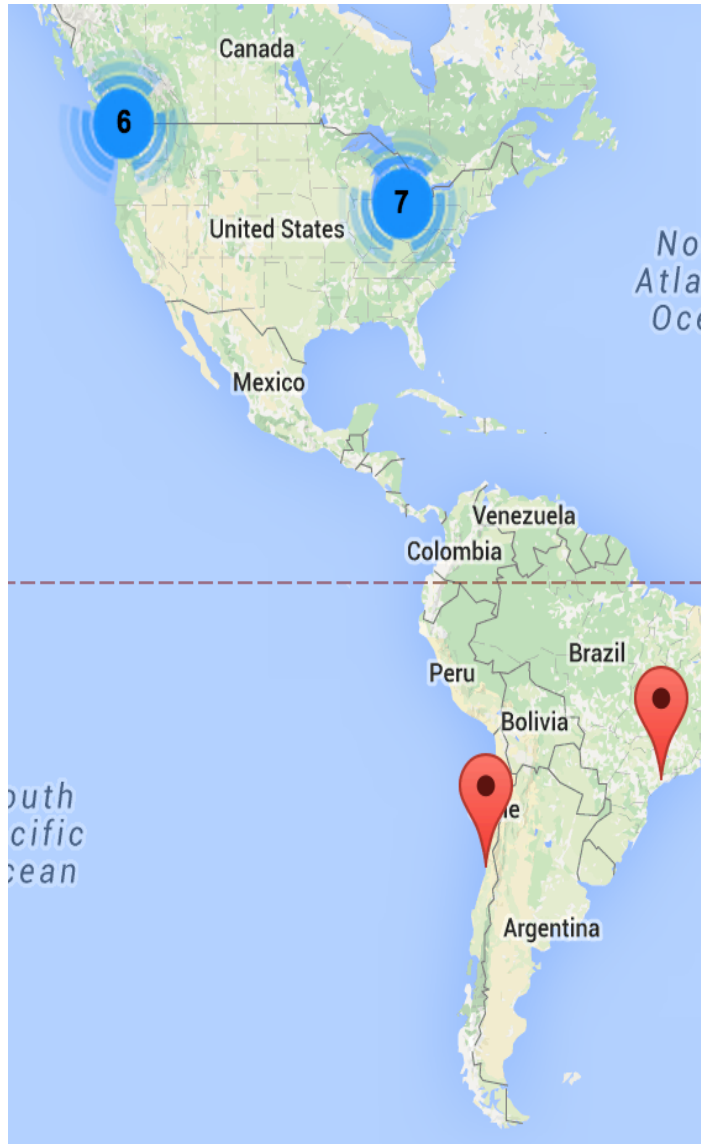


# The dCache.org collaboration



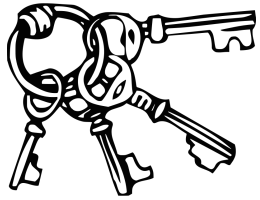
About 11 developers and support people in total

# Usage



To proceed we need to learn a bit about dCache

# Features needed for this presentation



## Access via variety of Credentials Can be all mapped to the same individual

Kerberos



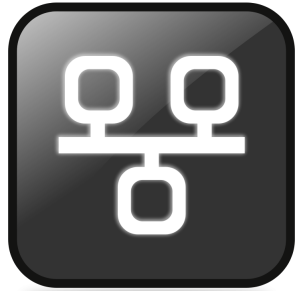
Username  
Password



X509



# Features needed for this presentation



## Access via a variety of Protocols All to the same file

http/WebDAV



NFS/pNFS



GridFTP

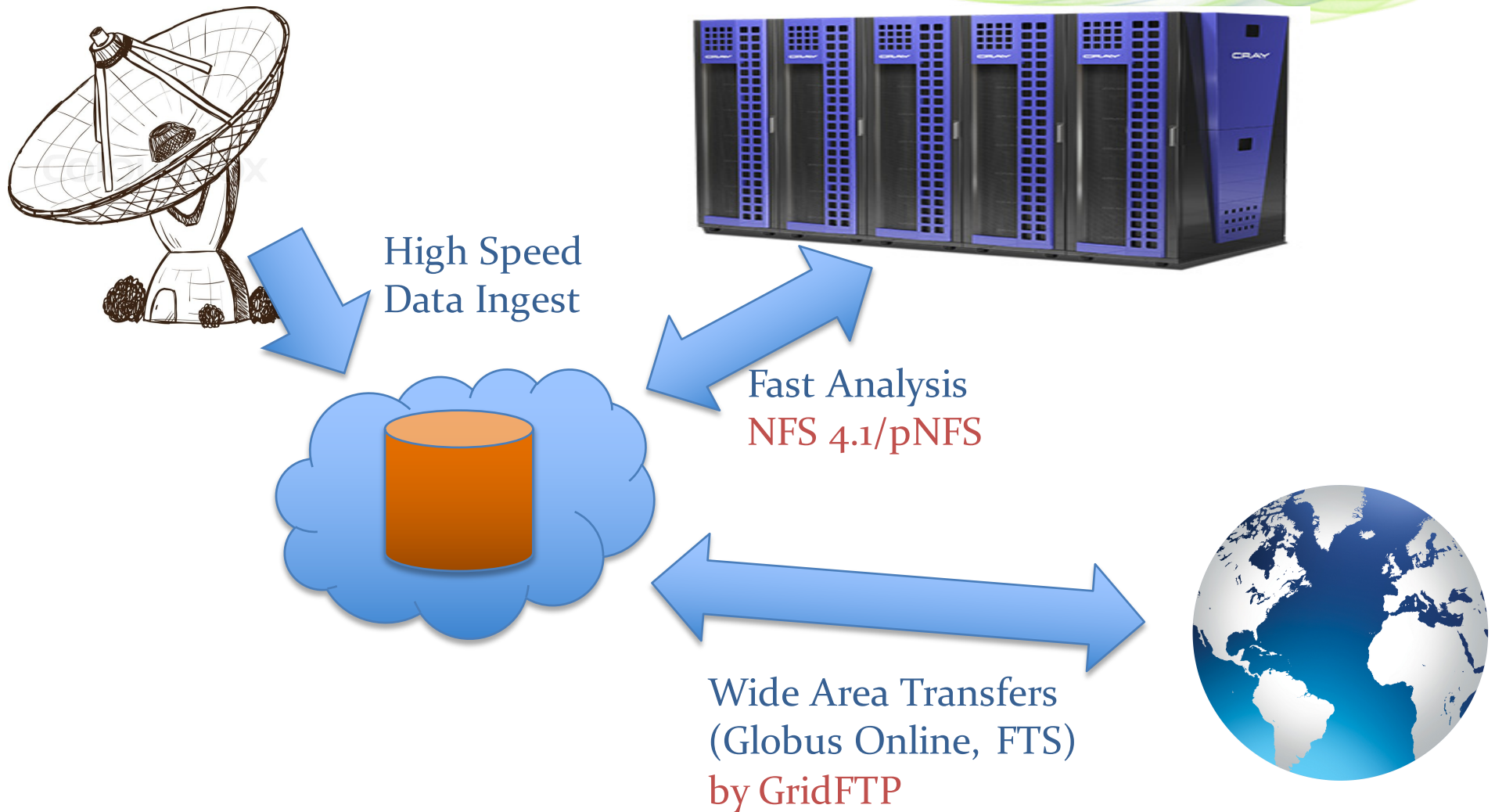


## Consequence

We support a typical scientific data life cycle



# Scientific Data Cloud



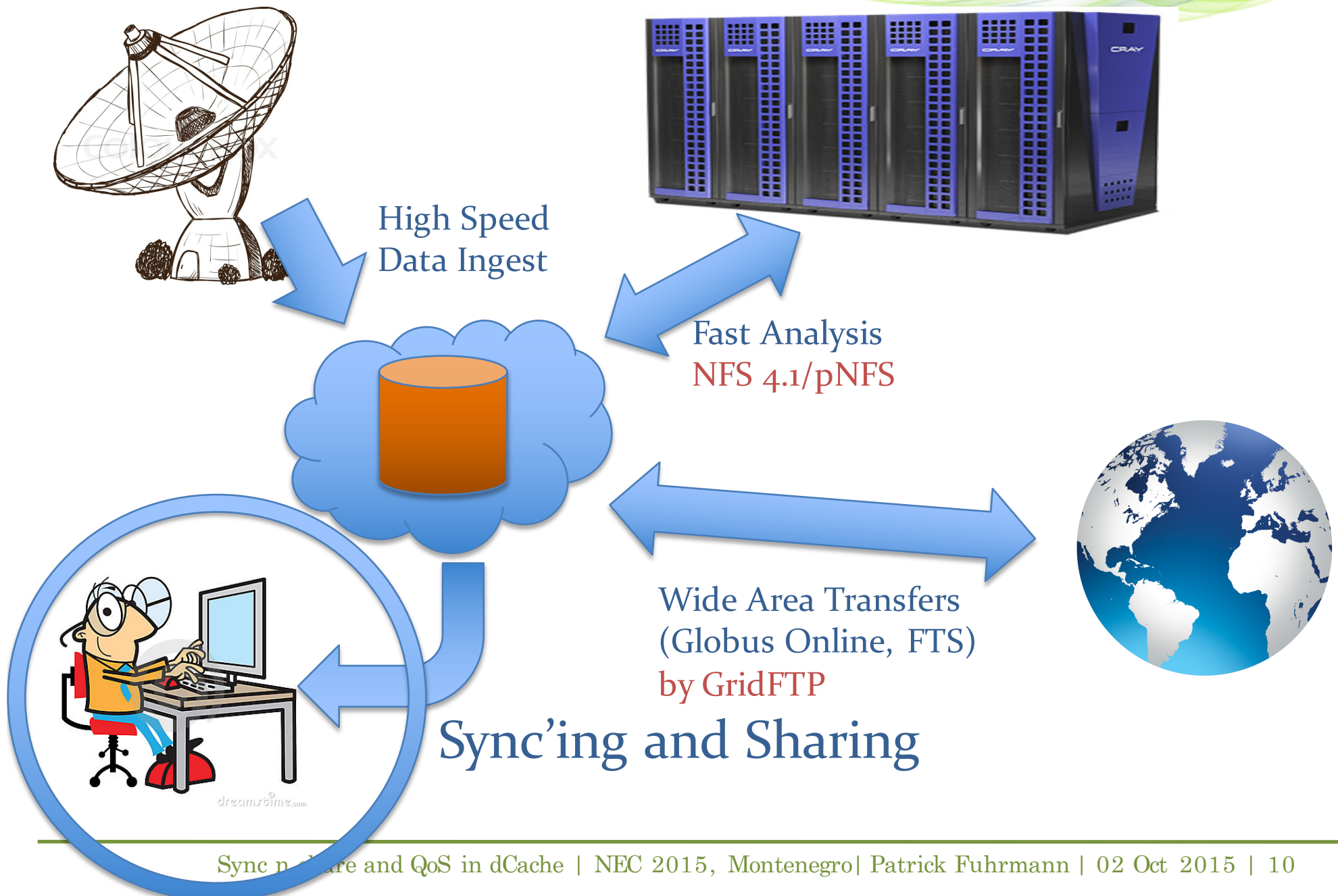


Except, something is missing !

The final scientist needs to :

- **Sync with his/her devices**
- **Share data with colleagues**

# Scientific Data Cloud

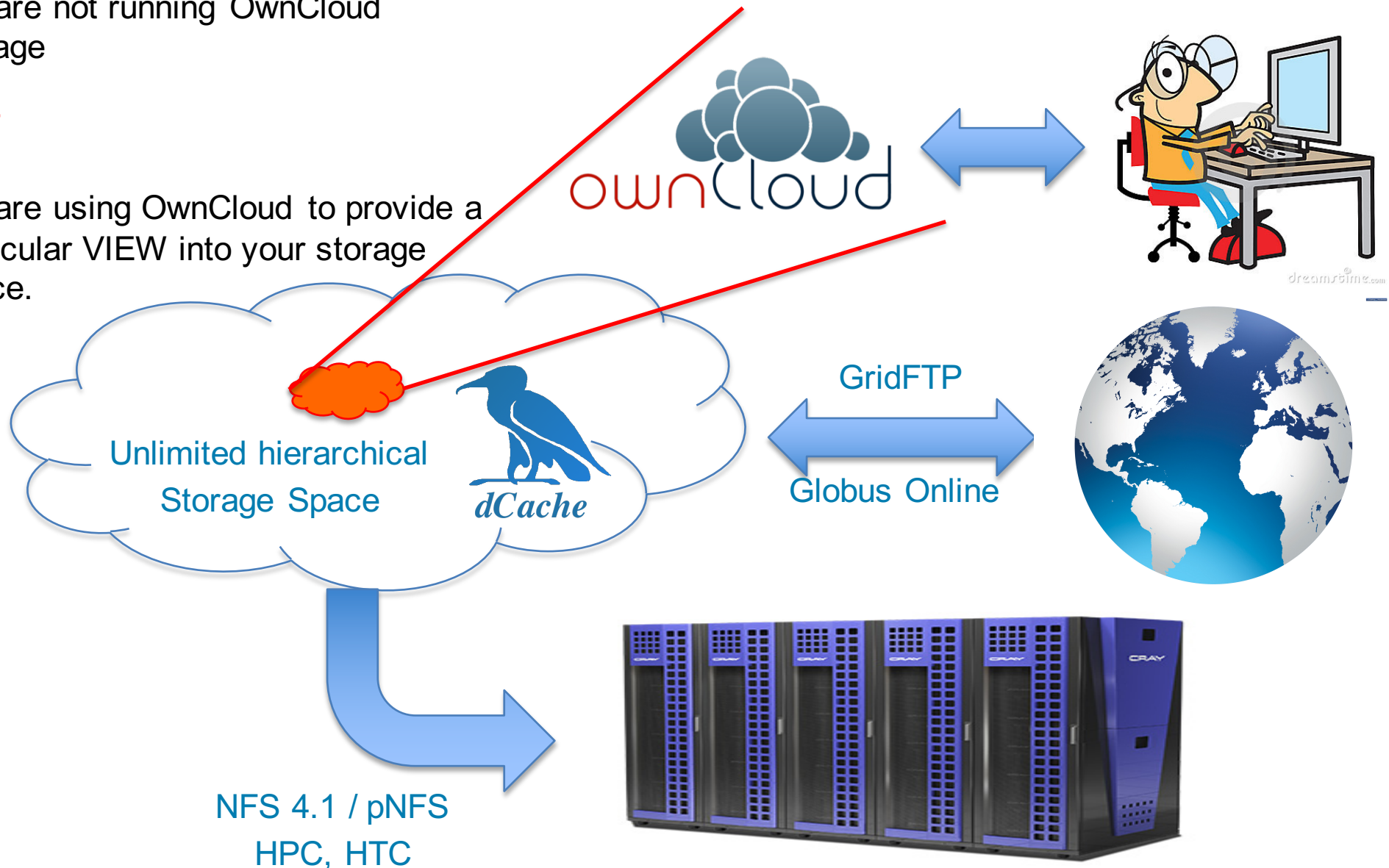


# Why not using ownCloud

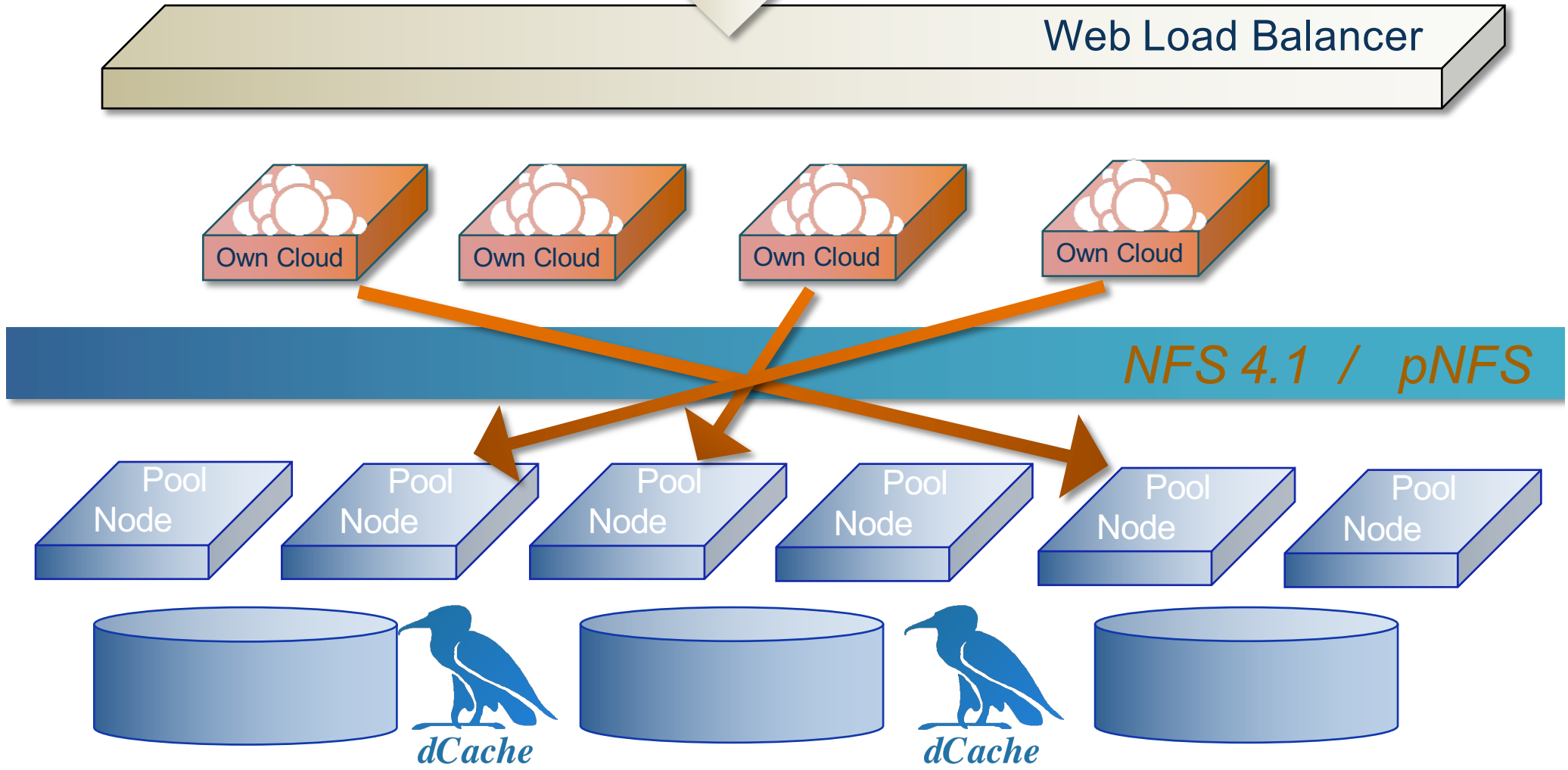
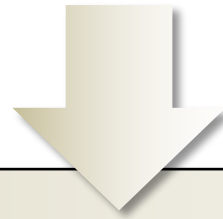
We are not running OwnCloud storage

**BUT**

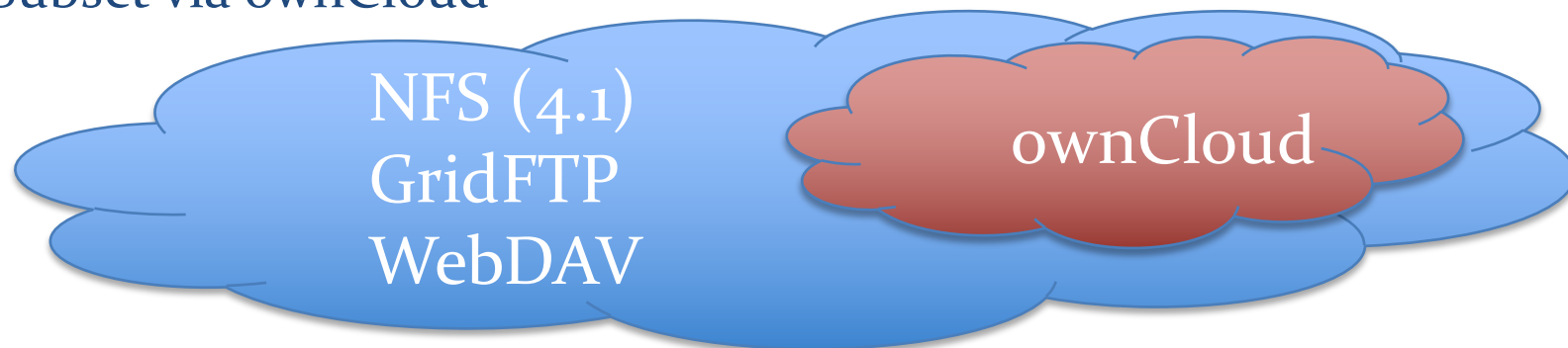
We are using OwnCloud to provide a particular VIEW into your storage space.



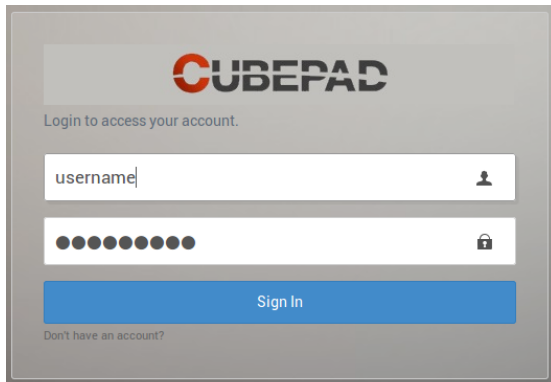
# How to scale out !



- Fully integrated into DESY infrastructure
  - Monitoring
  - Kerberos
  - LDAP
- Groups are added one by one to check scalability.
- Currently
  - 650 Users
  - 7 Tbytes (2 replicas)
  - Some power users up to 200 Gbytes / each
- Idea: Unlimited space (XXL)
  - Subset via ownCloud



But ownCloud is not the only  
and possibly not the best  
solution.



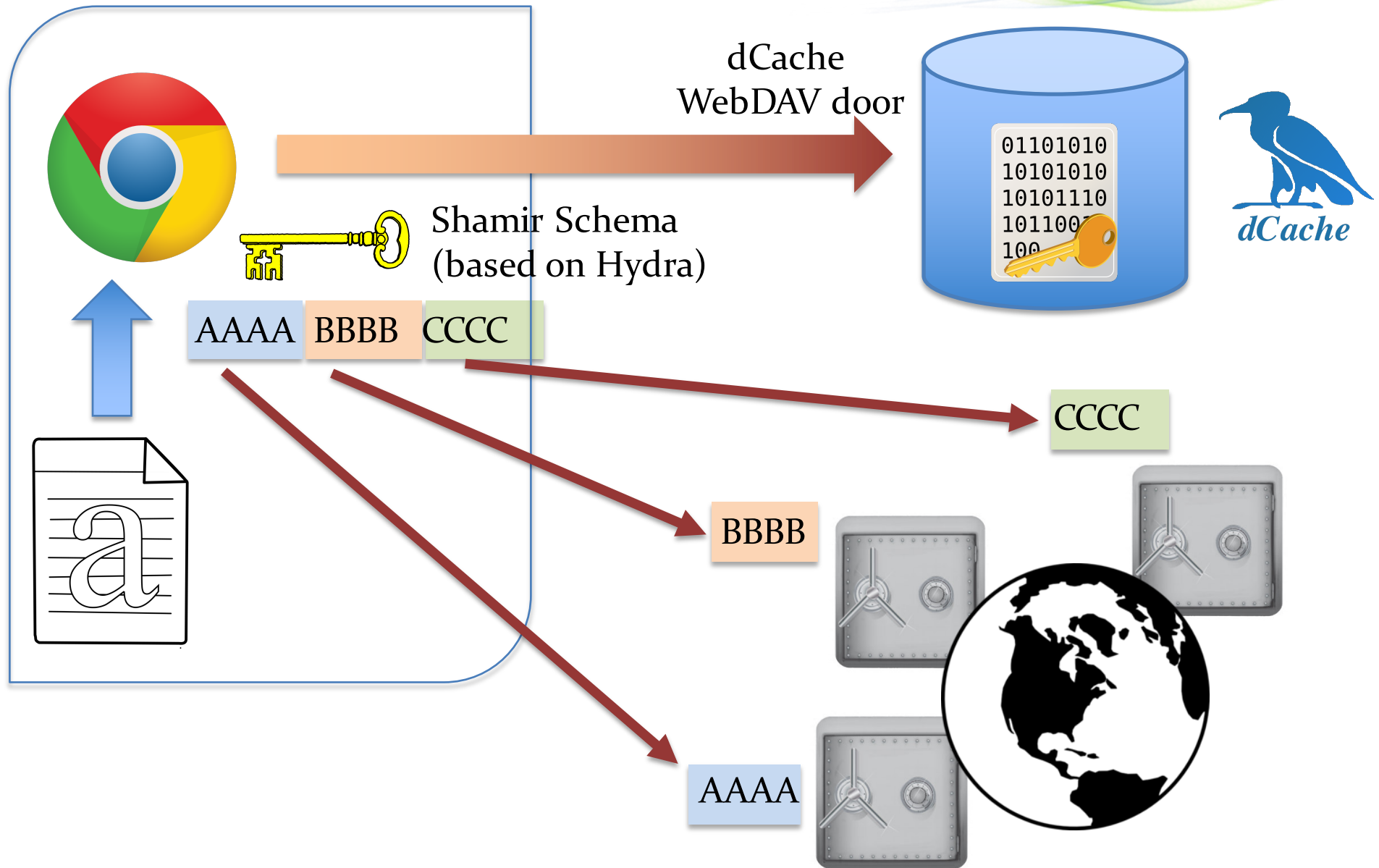
## Cube PAD

- We are investigating further
- dCache collaborates with DCORE
- DCORE provides CubePAD
- Besides other advantages: focus on strong privacy plus sharing
- Tighter integration with dCache
- Final goal : dCache namespace holds CubePAD metadata.



# Cube Pad File Manager

# Encrypting and sharing



- File is encrypted within the browser on the fly to the server (dCache WebDAV).
- Each file gets its own secret symmetric key.
- Symmetric key is split into 'n' pieces and stored at 'n' different geographical and political Locations. (Shamir Schema).
- One needs to break into 'm' < 'n' servers to get the entire key.
- Sharing works by sharing the keys.

Now we have :

- Scalable storage
- Access via scientific mechanisms concerning
  - credentials and
  - protocols
- Sync'n Share for easy access from
  - Laptop
  - Mobile devices
  - Browser

Still bits and pieces missing :

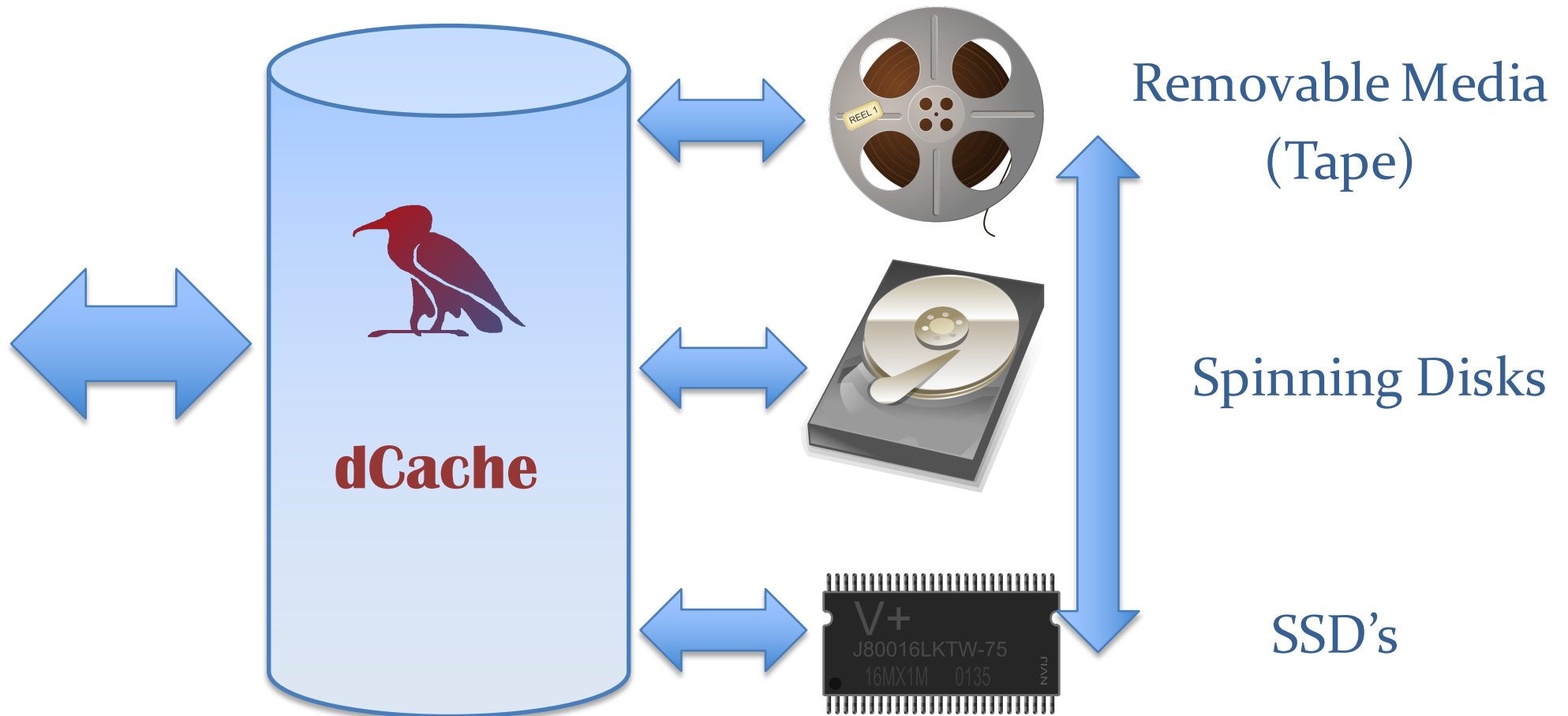
## Selection of Quality of Service for your storage.

- QoS : SSD, Tape, Spinning disk, # of copies
- Or in other words :
  - Access latency : low <-> high
  - Probability of data loss : low <-> high
- Considerations :
  - High Quality of Storage is expensive
  - Not all data is equally important
- So the user or experiment framework should be enabled to pick the right compromise based on his/her
  - Requirements
  - Size of your wallet

# Storage Quality

- Amazon
  - S3 : online
  - Glacier : nearline
- Google
  - Standard
  - Durable Reduces Availability (DRA)
  - Nearline
- IBM (HPSS, GPFS)
  - Storage classes (user defined)
- dCache
  - Storage groups (user defined)
  - Tape
  - Disk (spinning or SSD)
  - Resilient Management ('n' copies)

## Multi Tier / Quality of Service



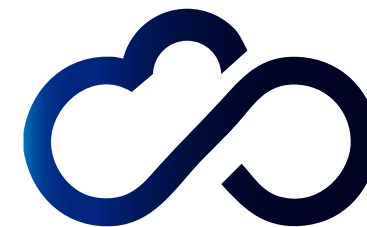


In order to get this sorted out consistently,

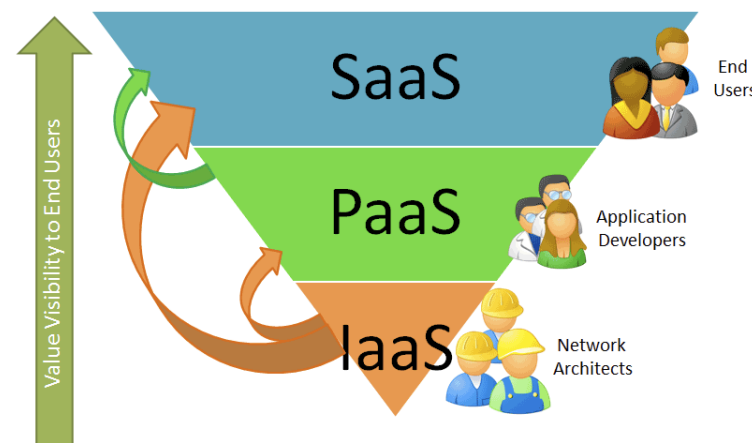
dCache is following two strategies.

- Providing API and GUI for customers to specify personal QoS setup.
- Agreeing on standard vocabulary to enable PaaS to consistently describe QoS
  - Trying this with RDA and OGF
  - Hope is to agree on a http/REST based protocol to negotiate QoS with arbitrary endpoints.  
(CDMI good candidate)
  - dCache is part of this activity within INDIGO DataCloud

- 11 ++ Million Euros
- 30 months duration
- 26 partners
- *The project aims for an Open Source Data and Computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over private and public e-infrastructures.*
- About 800.000 Euro for dCache.
- ~ 2 more FTEs
- Major objectives for dCache is :
- “Data LifeCycle Support” and
- “Software Defined Storage”



INDIGO DataCloud



# Summary

- dCache extends its multi protocol, multi credential Mantra by typical Cloud Access Mechanisms.
- Successfully production system with ownCloud but evaluating other systems (CubePad) especially for ‘high privacy’ mechanism.
- Making already established QoS mechanisms in dCache available
  - via GUI for individuals and
  - trying to agree on a standard vocabulary and management protocol with European and International standardization organizations to support the use of QoS by platform services (experiment frameworks)

# The END

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