



# dCache, sharing data, federating storage

EGI User Community Board Meeting

Patrick Fuhrmann



# Content

- What is this about ?
- Motivation.
- Status of ‘Scientific Storage Cloud’ system
- “Modern” sharing
- Authentication
- Access Protocols
- User defined retention policies
- Deployment
- Federated Identities
- Federated Storage
- Summary

## What is this about ?



- Providing unlimited storage space for all interested scientific communities:
  - Via standard protocols
  - Commonly used authentication mechanisms
  - Allowing modern sharing mechanisms
  - And user defined data retention policies.
- dCache.org and it's partners are providing this as a “service in Europe”, not as a software distribution.



- Within the German LSDMA project, dCache is supposed to provide storage solutions for German scientific communities.
- In order to learn more about ‘modern storage and sharing’ requirements, we would like to involve Universities and students (= scientists of tomorrow)
- Due to some disagreements between European and other countries on ‘data privacy’ Germany would prefer to store data in Europe and not in ‘Dropbox’

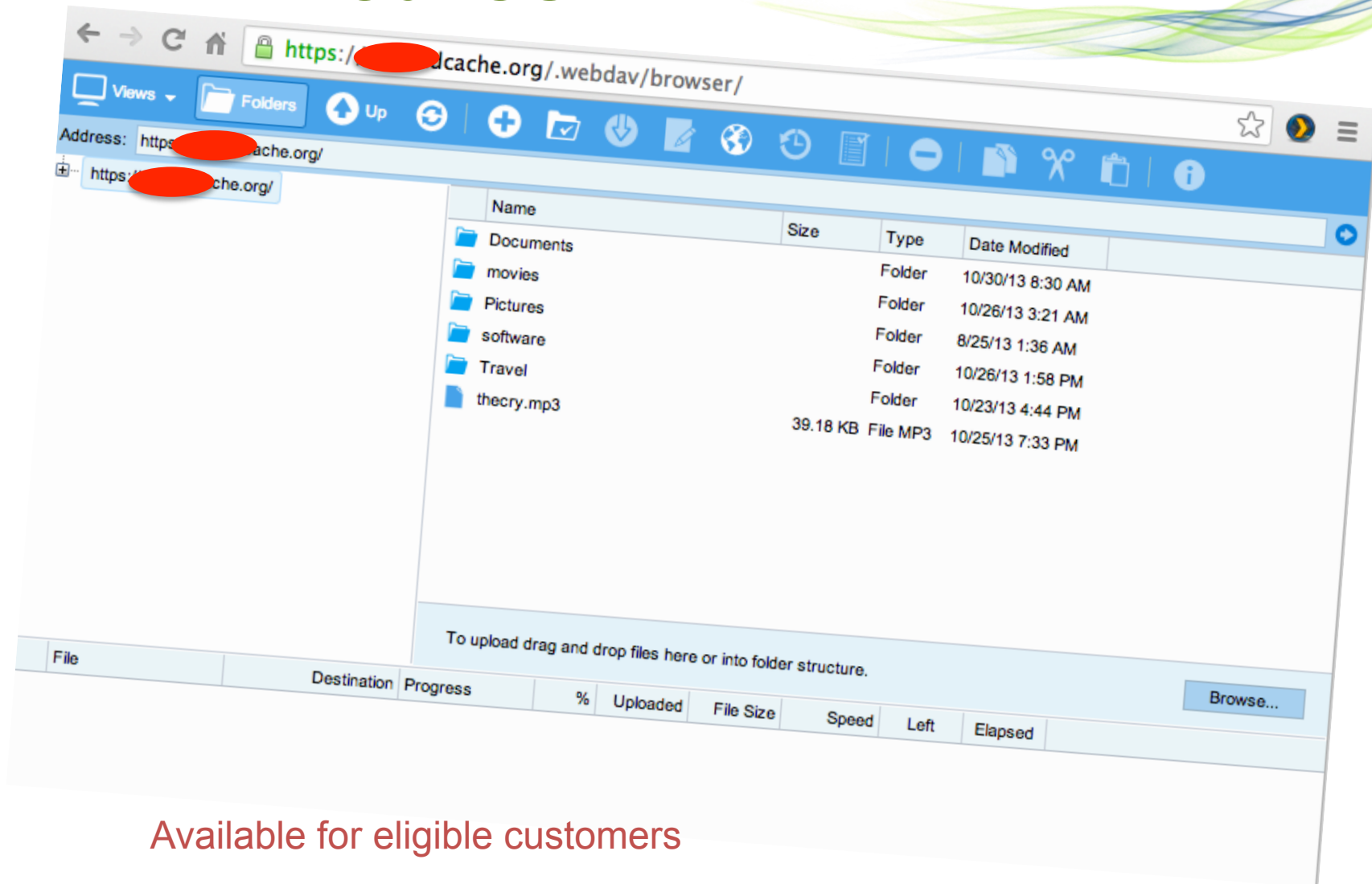
# Status of the Scientific Storage Cloud

dCache.org



- Service is installed and in use.
- Current access via username/password and WebDAV
- Registration easier than getting a Google account. (w/o human interaction on our side)
- For convenience, we provide a browser GUI page for up/download and namespace operations.
- You may as well use your OS WebDAV interface.
- Drivers/Apps for ANDROID already available from other Open Source projects. (Tested by students from HTW Berlin)
- Absolutely no sharing for now !!!
  - <https://☺.dCache.org/> is always only your home.

# Our GUI



Available for eligible customers

# Next steps: “Modern” Sharing

dCache.org



- Right now : no sharing
- Next is: public sharing
  - `Xxx.dcache.org/docs/first.pdf` -> `xxx.dcache.org/public/564-465-765`
  - One needs to know the Public URL to fetch the data
  - Public URL can't be guessed
  - Public URL doesn't allow to guess the owner nor the original file name
  - Not clear yet, if multiple public links should be allowed.
- Next is: sharing with registered users
  - `Xxx.dcache.org/docs/first.pdf` in 'patricks' account becomes
  - `Xxx.dcache.org/shared/patrick-first.pdf` in share destination account.
- Are there any other (more advanced) sharing techniques ?

# Next Steps: Authentication

- Right now: registration with X509, but cloud storage access with selected username and password only.
- Next: Adding more authentication methods:
  - X509 certificates or proxies in browser or through CURL or DavIX.
  - Accepting XAML assertions from different IdPs
  - Using Social Network account credentials
  - Kerberos for site-local access
- Allowing to point different authentication credentials to one account. (Credential Mapping)
- User authentication and mapping is done through the infrastructure (in this case DESY), not by dCache.
  - Right now OpenLDAP (Self registration though av. Certificate)
  - Later DESY registry including self registration system





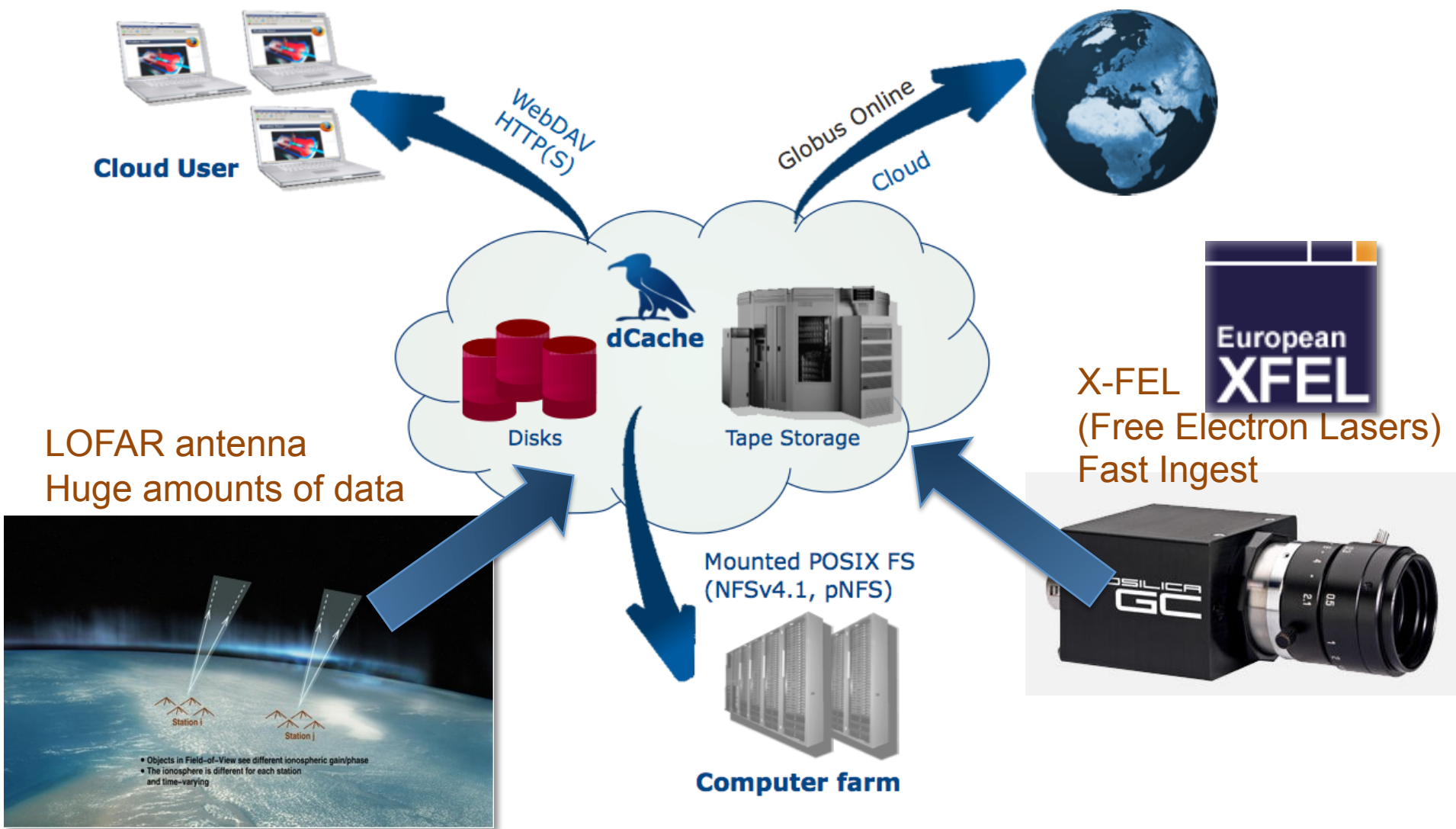
- Right now: Only https/WebDAVs
  - Server automatically provides Javascript GUI for browser access.
  - Most OSes support WebDAV services for read/write and namespace operations.
  - HTW Berlin students are evaluating mobile device clients.
- Next up: GridFTP for transferring data from/to other systems via GlobusOnline or FTS3
- Next up: Cloud Data Management Interface
  - Implemented by HTW Students
  - In addition to data access: Data Management.
  - Plus: Meta Data management
- NFS 4.1 for direct analysis access from local worker nodes.



- Right now: two copy of all files on different servers.
- Next up: User defined number of copies.
- Next up: Requesting tape copies of data
- Next up: Automatic building of containers on the server, for tertiary storage or wide area transfers.

- Right now: in use by dCache.org partners, some DESY IT colleagues and ‘randomly’ selected people at CHEP and Hepix (from Beijing to Chicago)
- Starting Monday next week: Next ‘computing’ semester at HTW Berlin.
- Next: DESY scientists
- Next: Advertised through the DFN ‘data program’

# Scientific Storage Cloud





- Federated Identity is a ‘nice to have’ by all sciences (however not necessarily x-science)
- Federated Identity was mentioned by Kostas during CHEP as an EU goal which points to H2020 funding.
- Success in Federated Identity is slow and only partial. See the nice summary presentation by David Kelsey from this Hepix.
- EU wide success within next 3 years is questionable.

# Federated Storage



- Federated Storage not yet clearly defined.
- ATLAS and CMS have started with a private proprietarily federation.
- CERN DM and dCache are building up an http ECO system including federation abilities.
- At this point we would need concrete use cases from non-WLCG communities to move forward.
- LSDMA in Germany is trying to build federated storage on top of a variety of different storage technologies as a ‘show case’ for its “Data Lifecycle Labs”.

# Summary



- dCache.org within LSDMA installed a concrete Storage Cloud System for scientific communities.
- Sharing, protocols, authentication and other storage properties will be added step by step based on feedback from our users.
- European Federated Identity would make the sharing of data much easier for storage providers and users. Slow progress due to legal issues.
- Federated storage still needs better defined use cases outside of WLCG.