Grid Middleware & Interoperability

dCache, Storage Interoperability beyond WLCG

WLCG Data Grid meets reality ….

patrick FUHRMANN

And with many thanks to
Jillavisia Lin
for her patience.

WITH CONTRIBUTIONS BY

dcache TEAM

And in particular

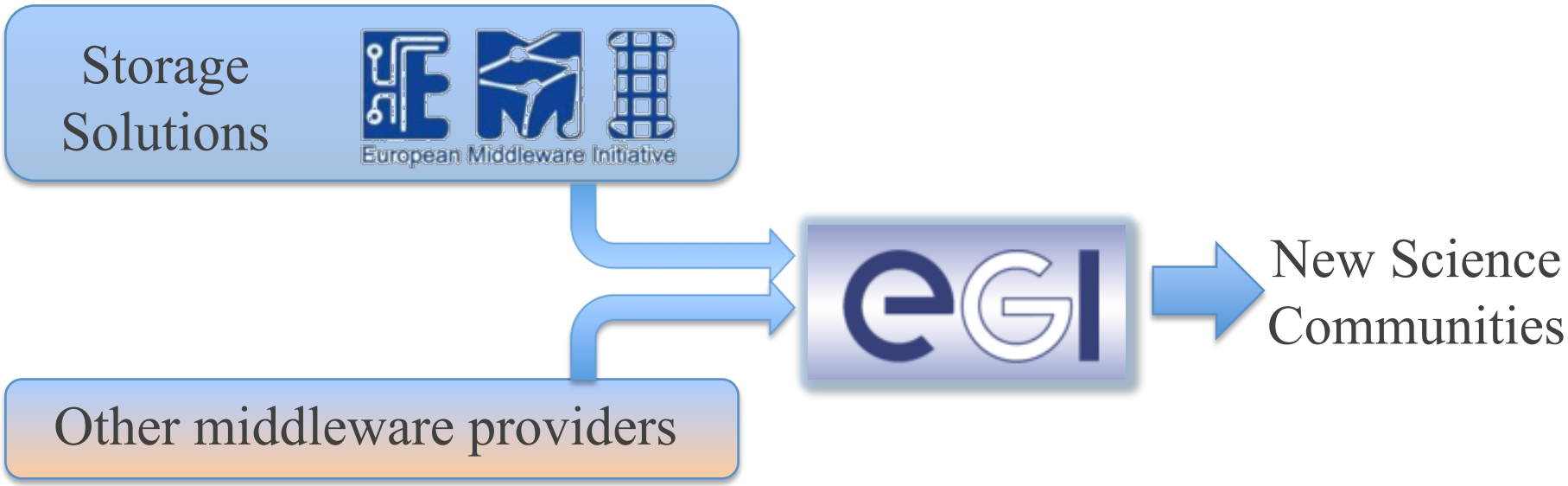
gerd BEHRMANN, NDGF
tigran MKRTCHYAN, DESY
hanno HOLTIES, LOFAR
tom LANGBORG, SNIC
anton BARTY, CFEL

Content

- Some examples of new, data intensive communities.
- Collecting their mass storage requirements.
- Can EMI provide a solution ?

The question is :

Will the WLCG/EGEE storage middleware stack, as provided to EGI through the European Middleware Initiative (EMI), be able to satisfy the needs of new data intensive communities ?



Using three examples, I tried to find out what modern science groups need in terms of storage and data-access.

All three communities have in common that they

- ✓ Intend to utilize existing storage facilities, most of which are serving WLCG storage already. (Tier I and II)
- ✓ Are not paid for using the Grid.
- ✓ And not to forget : they are all using dCache.

Examples for new data-intensive communities/groups

Would like to use the SARA storage facility, which is currently serving as WLCG Tier.

Would like to utilize DESY storage facilities currently being used as HERA Tier-0, Atlas, CMS and LHCb Tier-IIs and for many more groups and experiments.

Would like to utilize the Swedish dCache Tier II facility.

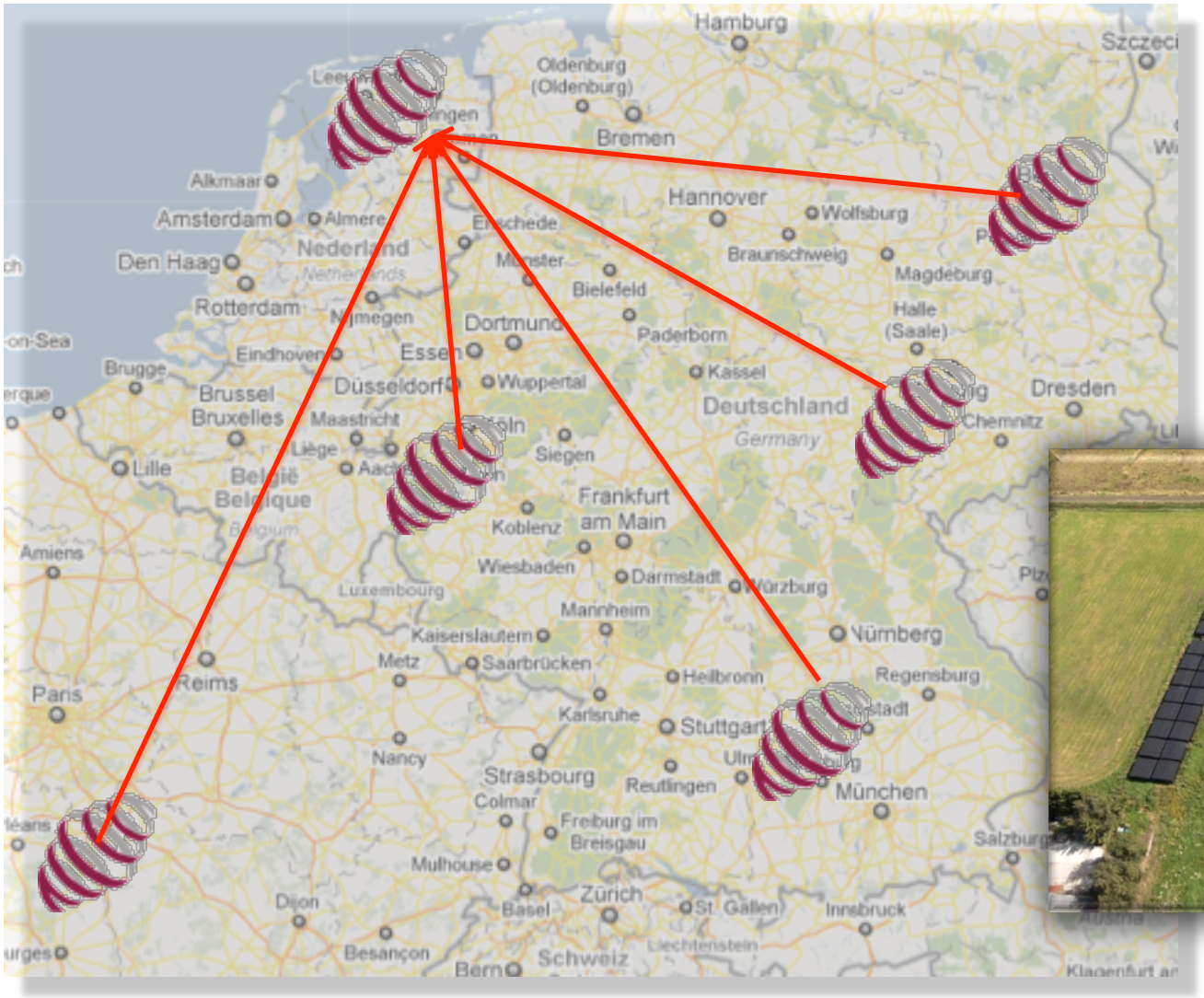
The International LOFAR Radio Telescope

(The first software telescope)

Information provided by

hanno HOLTIES, LOFAR

The International LOFAR Radio Telescope

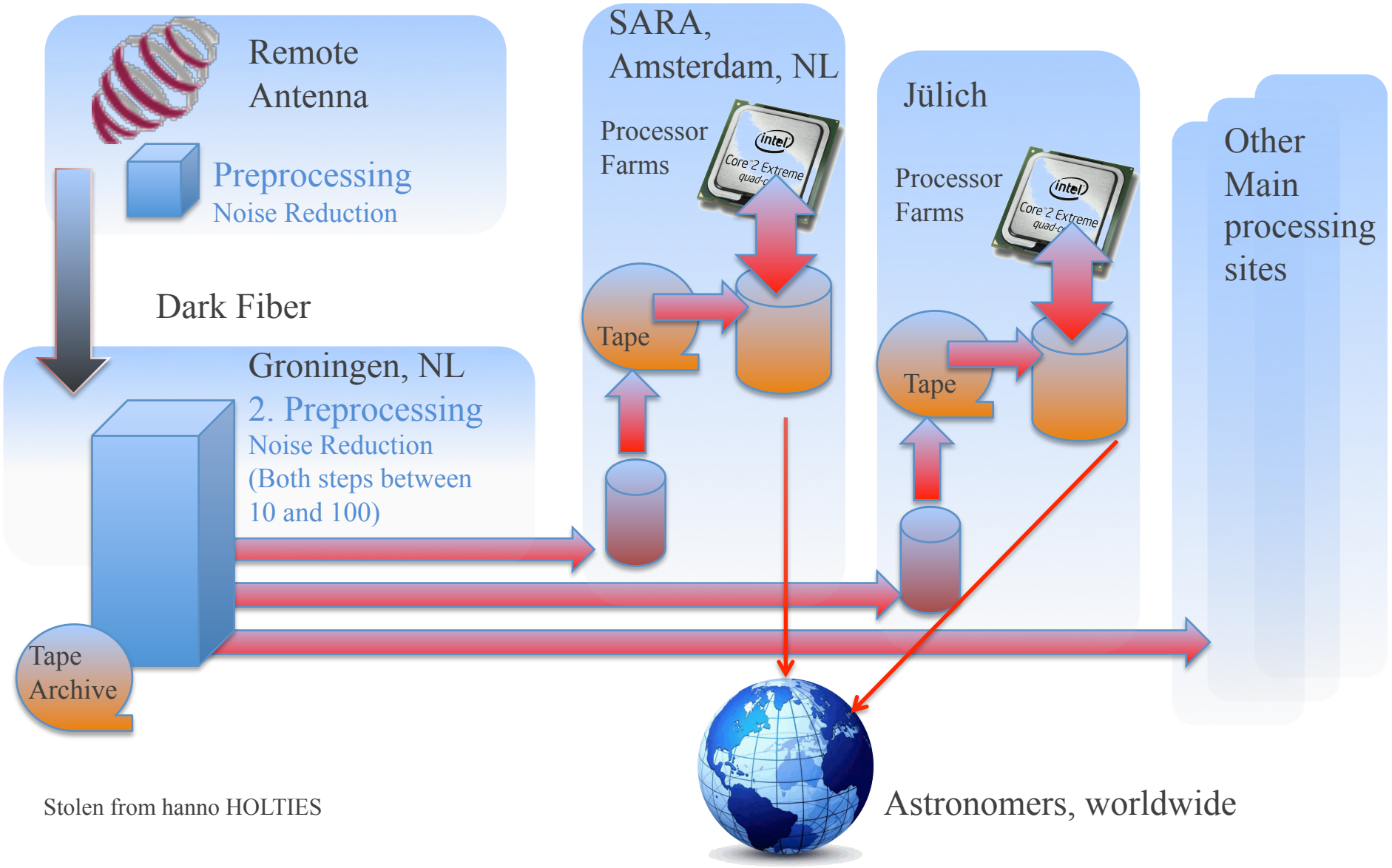


As of Feb 24, 2010 :

- ✧ 21 Complete Stations
- ✧ 10 In Progress
- ✧ 13 Planned
- ✧ NL, DE, UK, FR, SE

Stolen from hanno HOLTIES

LOFAR (simplified) data flow model



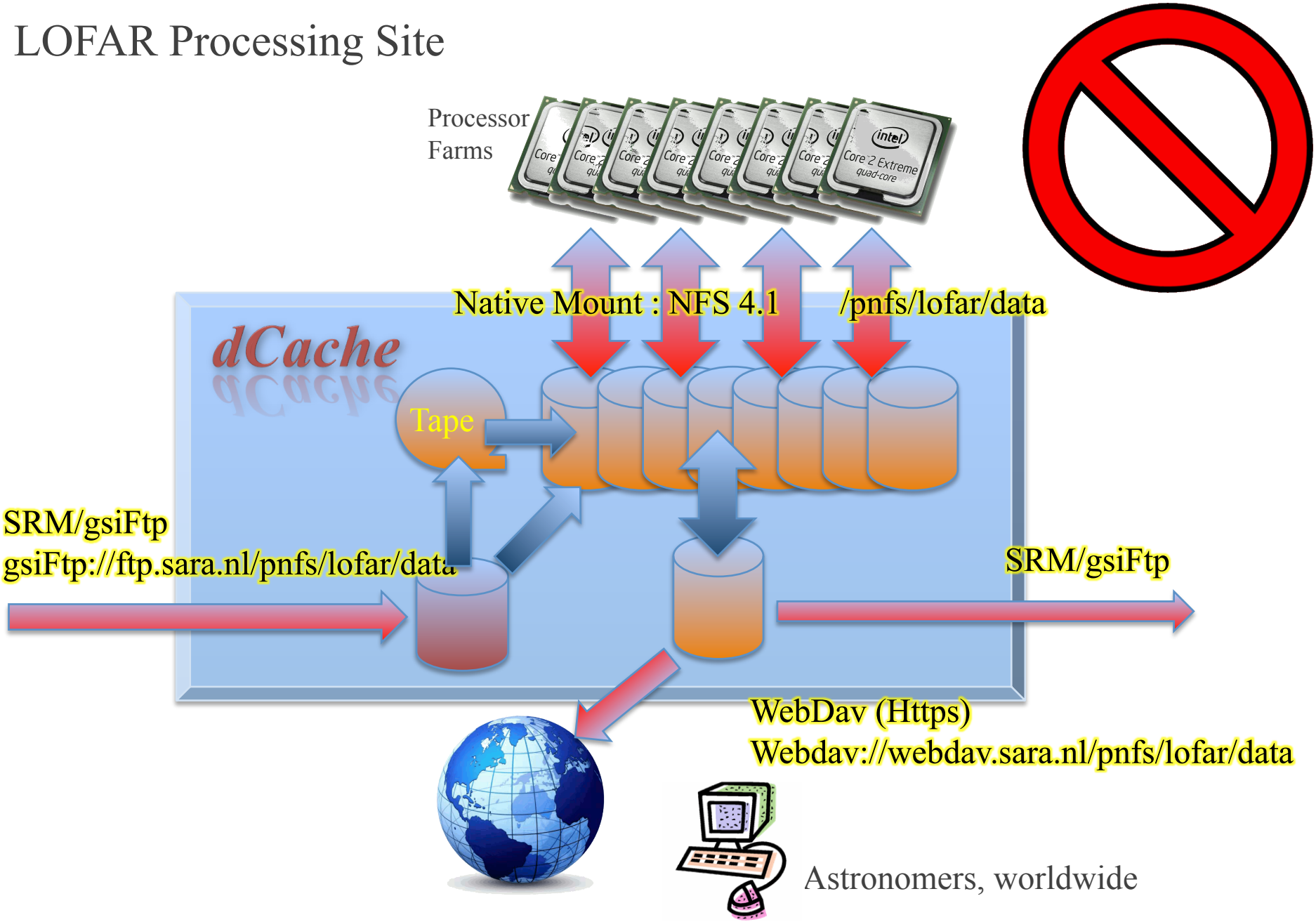
Stolen from hanno HOLTIES

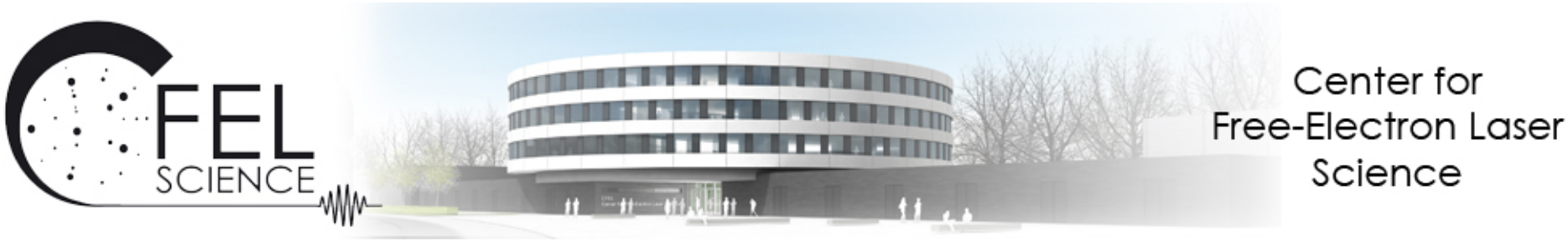
LOFAR Requirements

- ✓Low threshold data retrieval
  - ➢Access only by registered LOFAR members.
  - ➢CERTS are not desirable for all members.
  - ➢Owner of data needs to disable directory browsing.
  - ➢Common protocols : Mounted file system, http/WebDav
- ✓Roles
  - ➢OPERATIONS can put data into permanent storage.
  - ➢USER may retrieve data from permanent storage.
  - ➢Quotas on 'tape backend usage'.
  - ➢Groups storage areas for read/write
- ✓Integration with external (non-EGEE) identity management system.
- ✓Accounting
  - ➢Per VO, user, directory
  - ➢Quotas
- ✓Data integrity
- ✓Fixed URLs (to support external catalogues)

Stolen from hanno HOLTIES

LOFAR Processing Site

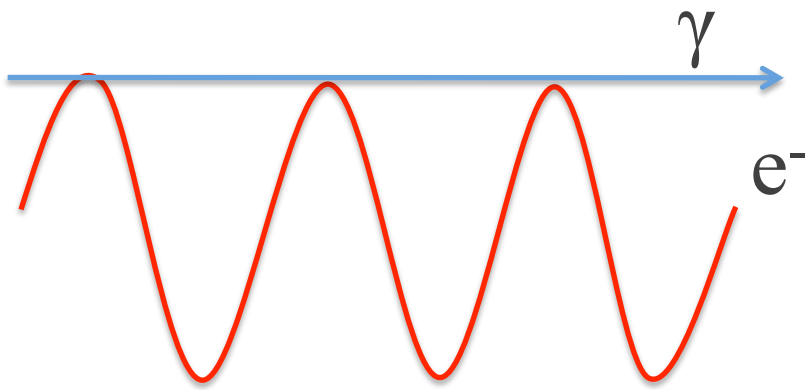




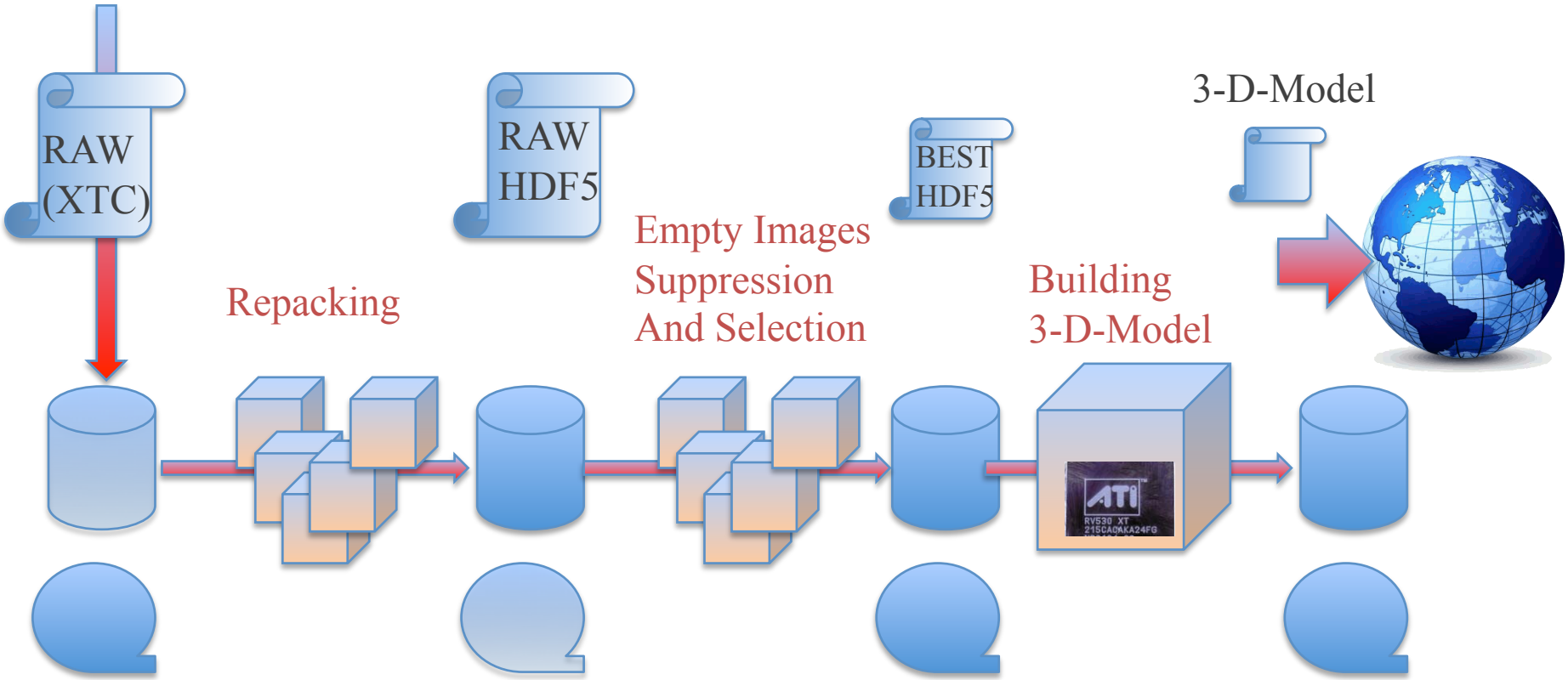
The Center of Free-Electron Laser Science, CFEL

Information provided by

anton BARTY, CFEL



Free Electron Light Sources



Stolen from anton BARTY

CFEL Requirements

- ✓Authorization Authentication
  - ➢Different Authentication Mechanisms must point to the identity
    - •Kerberos
    - •Certificates
    - •User/Password
  - ➢Fine grained access control. Protect data till publication.
- ✓Access
  - ➢Fast access from worker-nodes for coordinated processing.
  - ➢As not all applications can be re-linked: standard POSIX access is required.
  - ➢Scientists need access from outside the laboratory.
    - •Either browser or
    - •OS integrated mechanisms (WebDav)
- ✓Data integrity
- ✓Storage Policy / Attributes
  - ➢Data location disk/tape must be defined by experiment manager role.
  - ➢Some data but be 'retrievable' by all group members.

Stolen from anton BARTY

Swedish National Infrastructure for Computing

Information provided by

tom LANGBORG, SNIC

Uppmax	Uppsala Multidisciplinary Center for Advanced Computational Science
Lunarc	scientific and technical computing for research at Lund University
HPC2N	High Performance Computing Center North
C3SE	center for scientific and technical computing at Chalmers University of Technology in Gothenburg
NSC	National Supercomputer Center in Linköping
PDC	Center for high performance computing

SNIC

SNIC National storage is an infrastructure for archiving data.

Swestore Project Jan 20, 2010

Create an infrastructure for storage for Swedish Research and Swedish Universities.

Planned Data Access

"SRM, WebDav and **gsiFtp** are examples of protocols for communicating with the National Storage. Authentication method are **X509 Certificates. Kerberos** could be used in some special cases" , Tom Langborg, SNIC

Internal	External
SRM	SRM
gsiFtp	gsiFtp
WebDAV	WebDAV
NFS 4.1	Web Portal/Gateway

Stolen from tom LANGBORG

Translating the collected requirements into our language

Collected requirements

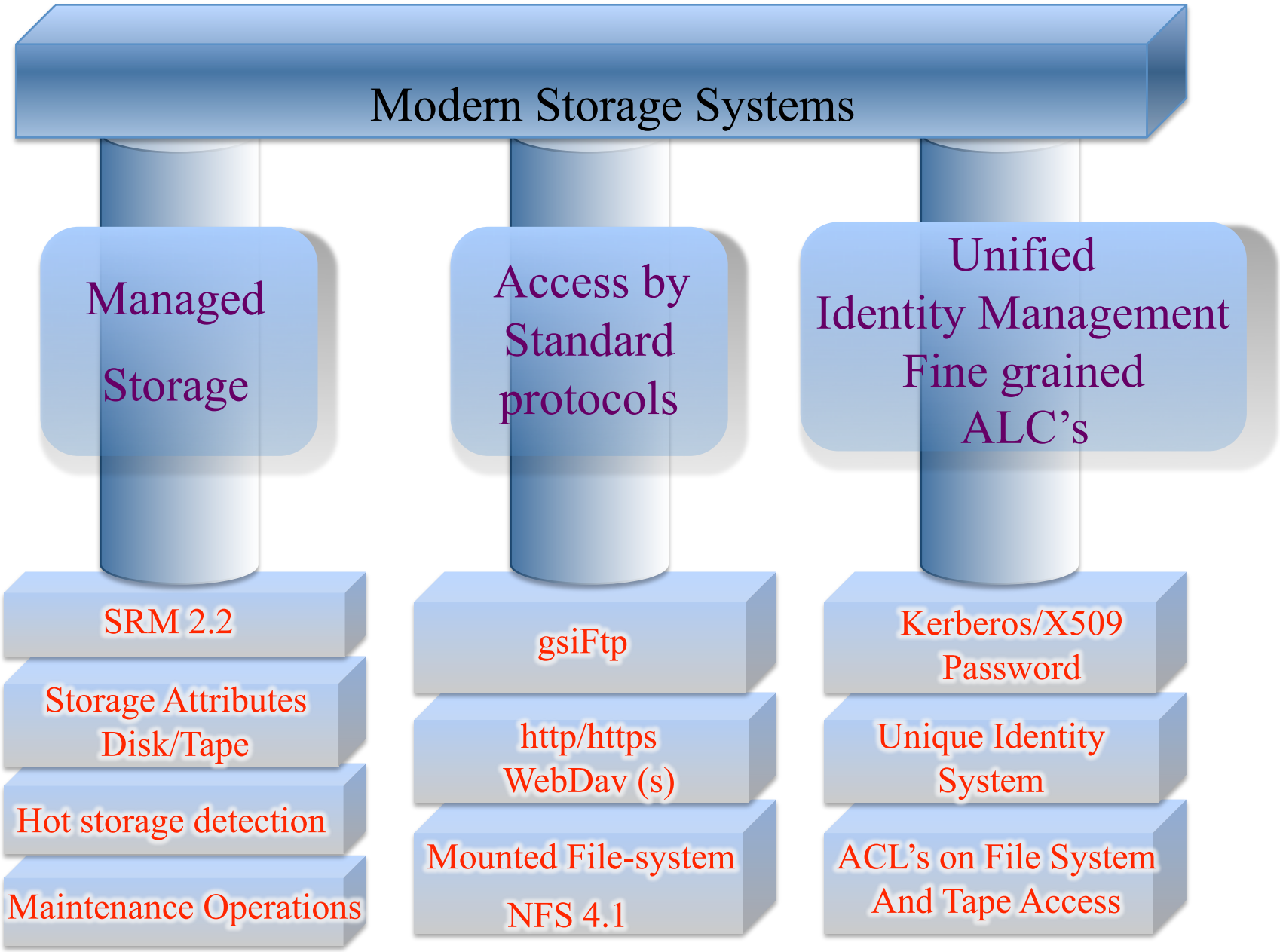
- ✓Data access
  - ✓Standard POSIX access (by mounting a file system space)
  - ✓Remote access via a standard client (browser, curl, OS mechanisms)
- ✓Storage management
  - ✓Definition of storage location e.g. Tape, Disk per directory or file.
  - ✓Manual or automatic data location management/transition
    - ✓Pinning
    - ✓Bring online (by authenticated User)
  - ✓Quotas on storage.
  - ✓Quotas on data transitions.

Collected requirements

- ✓Authentication
  - ✓Different authentication mechanisms must point to the same identity
  - ✓Support required for
    - •User/password (https)
    - •Certificates
    - •Kerberos
  - ✓Connectivity to external identity management.
- ✓Authorization
  - ✓Fine grained access control (ACLs) on file system.
  - ✓Access control on tertiary storage (tape) access.

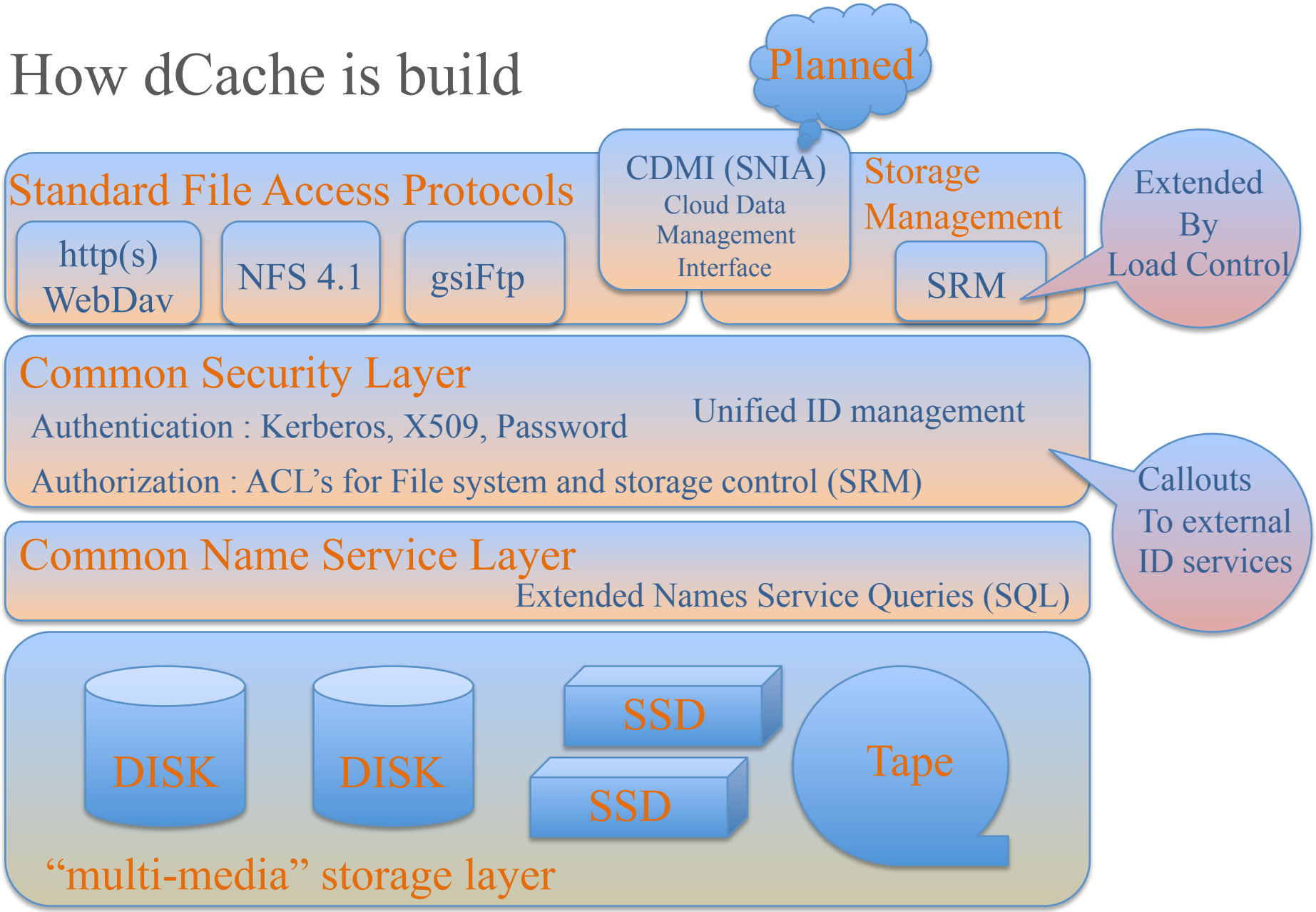
Collected requirements

- ✓Data integrity
  - ✓Check sum checking with all data location changes
    - ✓Arrival
    - ✓Tape → Disk
    - ✓Disk → Disk
  - ✓Bad checksum detection on sleeping data.

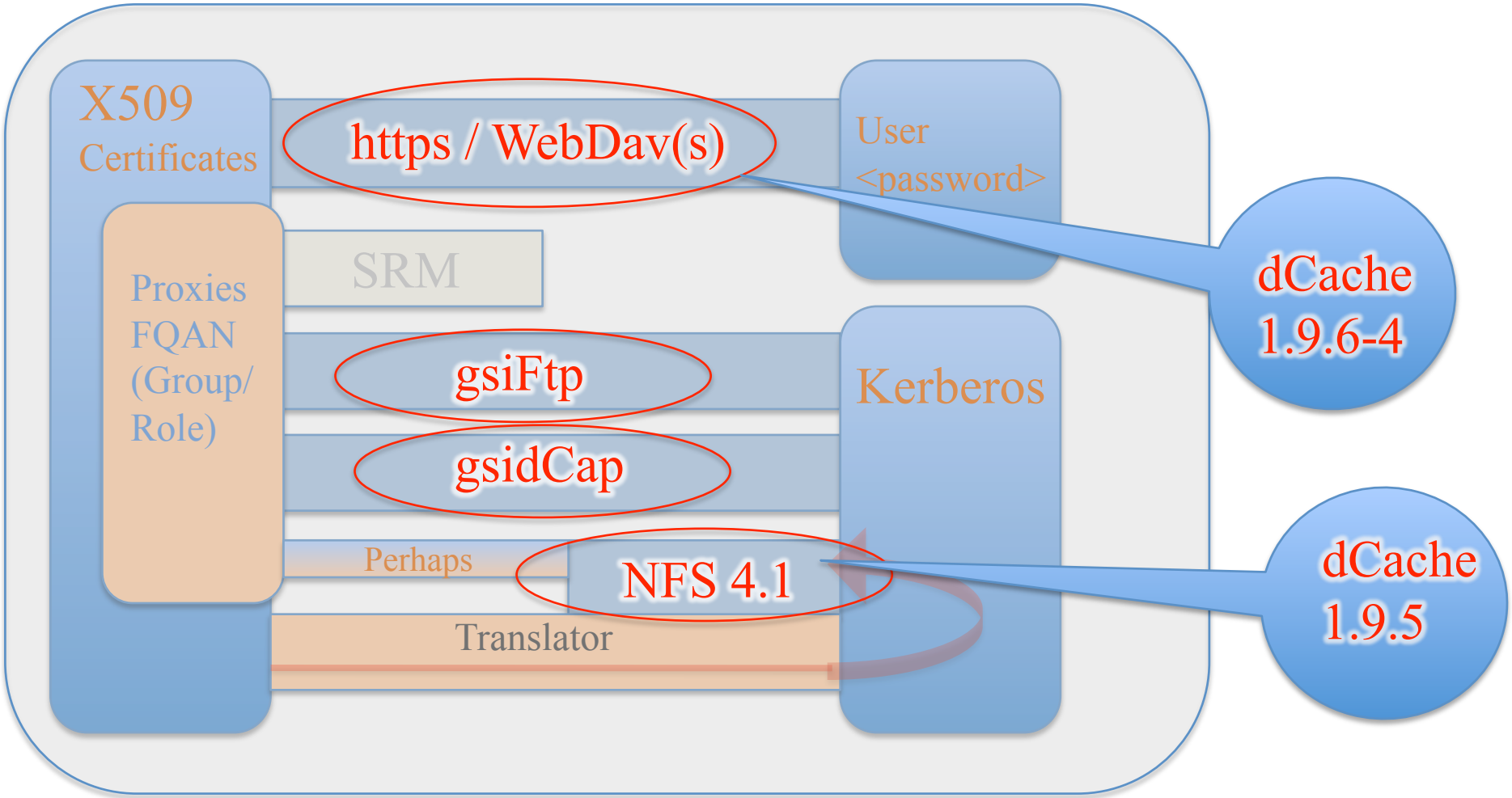

Modern Storage Systems
Managed Storage
Access by Standard protocols
Unified Identity Management Fine grained ALC's
SRM 2.2
Storage Attributes Disk/Tape
Hot storage detection
Maintenance Operations
gsiFtp
http/https WebDav (s)
Mounted File-system NFS 4.1
Kerberos/X509 Password
Unique Identity System
ACL's on File System And Tape Access

Can we solve this with dCache ?

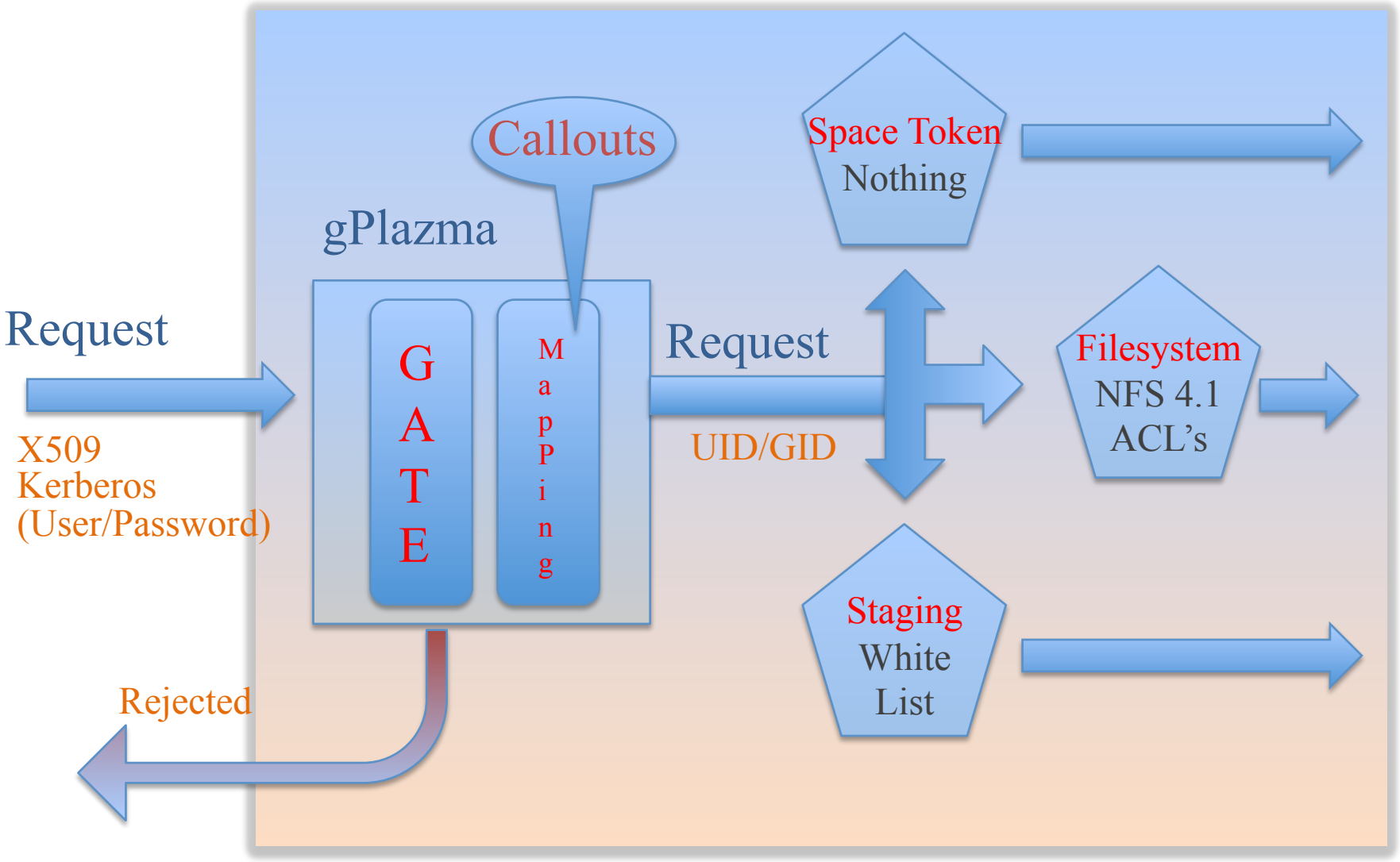
How dCache is build



dCache supported data access protocol suite.



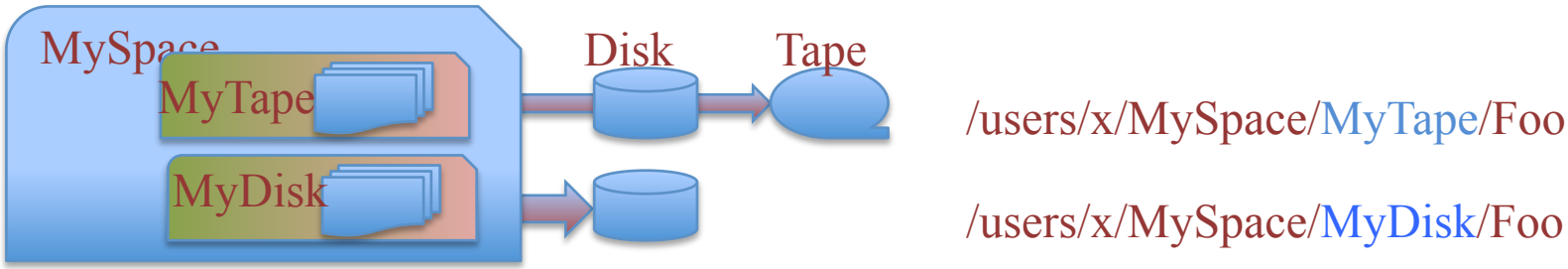
Authentication / Authorization Flow



dCache storage control (Spec)

Manual storage control (aka Managed Storage)

- SRM 2.2 (WLCG & Addendum & Addendum) compatible.
  - ✓ Define storage media (Disk/Tape) per file or "Space".
  - ✓ Pin / Unpin files
  - ✓ Bring Online file(s)
- Storage Media can be assigned to directory (sub) structure.

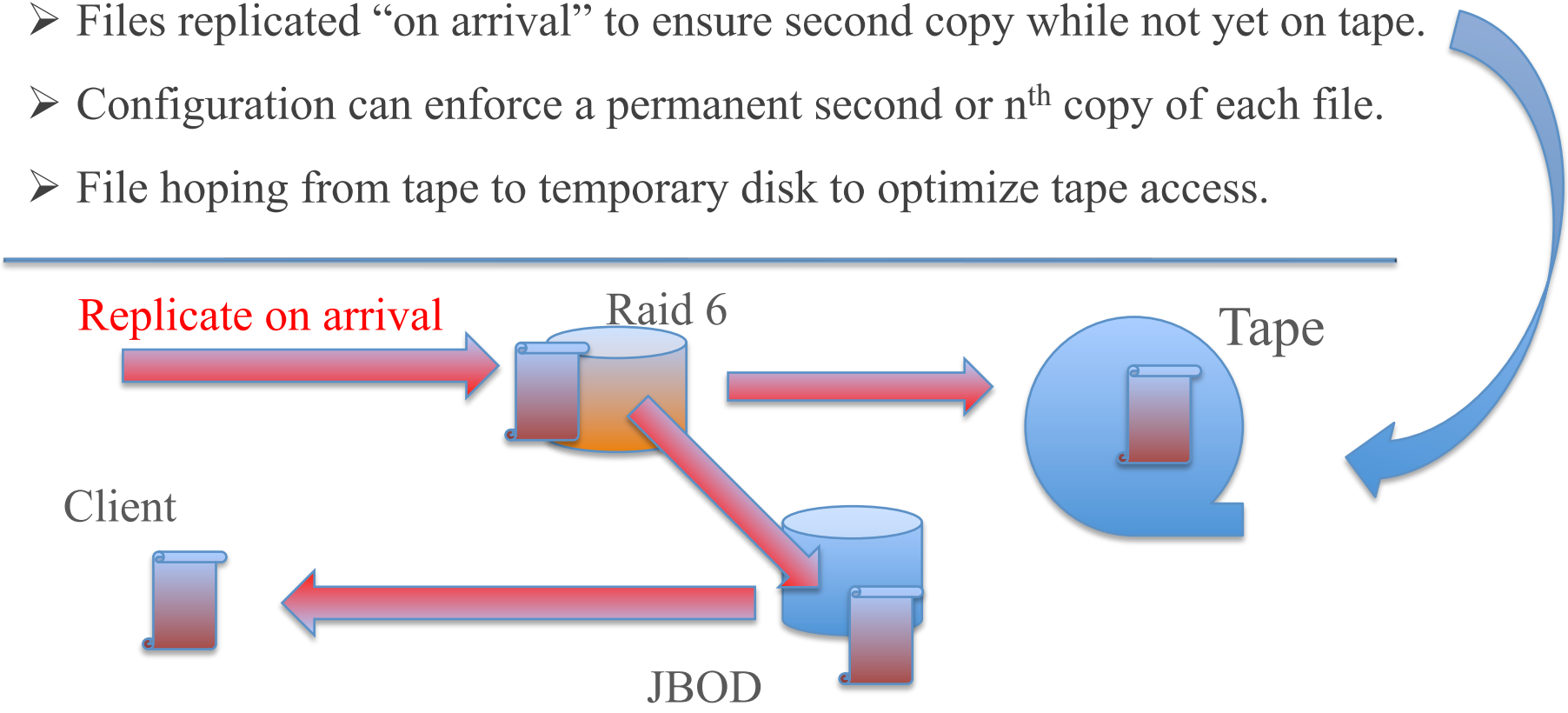


- Data can be scheduled for replication for maintenance or performance reasons.
  - ✓ Scheduled server downtimes
  - ✓ Server decommissioning
  - ✓ Multiple copies to increase throughput

dCache storage control (Spec)

Automatic storage control (aka dCache file hopping)

- Data stored to tape and retrieved when needed.
- Files are automatically replicated to cope with high server load.
- Files replicated "on arrival" to ensure second copy while not yet on tape.
- Configuration can enforce a permanent second or n^{th} copy of each file.
- File hoping from tape to temporary disk to optimize tape access.



In summary

dCache combines well known and
standardized data access
mechanisms, e.g. mounted file-system, web access,
browser/WebDav, with a broad
automatic and manual storage control functionality,
under a
common file name space and security umbrella.

With dCache, EMI and with that EGI is well prepared to serve new data intensive communities.

About supporting NFS 4.1

Or

Why is NFS 4.1 more than just file://…

NFS 4.1 in a mini nutshell

- NFS 4.1 (pNFS) is a IETF standard
- NFS 4 defines security standards (gss e.g. Kerberos)
- NFS 4.1 pNFS honors distributed data.
- All important storage vendors (IBM, PANASAS, EMC, NETAPP, dCache) are part of the NFS 4.1 working group under the roof of CITI (University of Michigan) and have an implementation ready.
- NFS 4.1 is available for Solaris and Linux (kernel 2.6.34)
- It will be in RH6 enterprise editions till end of the year.
- Back-ports for SL5 are in discussion.
- No vendor locking (e.g. GPFS, Lustre)
- dCache supports NFS 4.1 since 1.9.5 (Golden Release)

Storage Developers Conference (St. Clara, 2009)

NFS 4.1 Contributors

Coordinated by the Center of Information Technology Integration (U. Michigan)
Slide is stolen from "Lisa Weeks" presentation :
pNFS: Blending Performance and Manageability

Blue Arc
CITI
CMU
EMC
IBM
LSI
OSU
Net App
Ohio SuperComputer
Panasas
Seagate
StorSpeed
Sun Microsystems
Desy

Clients

- Sun (Files)
- Linux (Files / Blocks / Objects)
- Desy / dCache (Java-based / Files)

Servers

- Sun (Files)
- Linux (Files)
- NetApp (Files)
- EMC (Blocks)
- LSI (Blocks)
- Panasas (Objects)
- Desy / dCache (Java-based / Files)

About supporting NFS 4.1

As industry is preparing to provide a powerful remote file access protocol for distributed data, ***replacing proprietary vendor locking protocols*** *like gpfs, Lustre, Panasas and Netapp, it is time for us to* ***get rid of the HEP data access protocol zoo****.*

Why not jumping on the train and using NFS 4.1.

The client would come for free and

for the application software that would just be a file://...
Instead of loading/linking weird libraries.

Further Reading

www.dCache.org