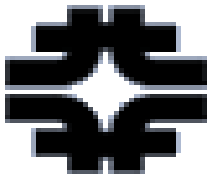
dCache.ORG

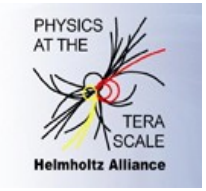
dCache seminar at FERMIlab



Patrick Fuhrmann et al.

and slides stolen from nearly everywhere

additional funding, support or contributions by



d-grid
DGI II

Roadmap for today

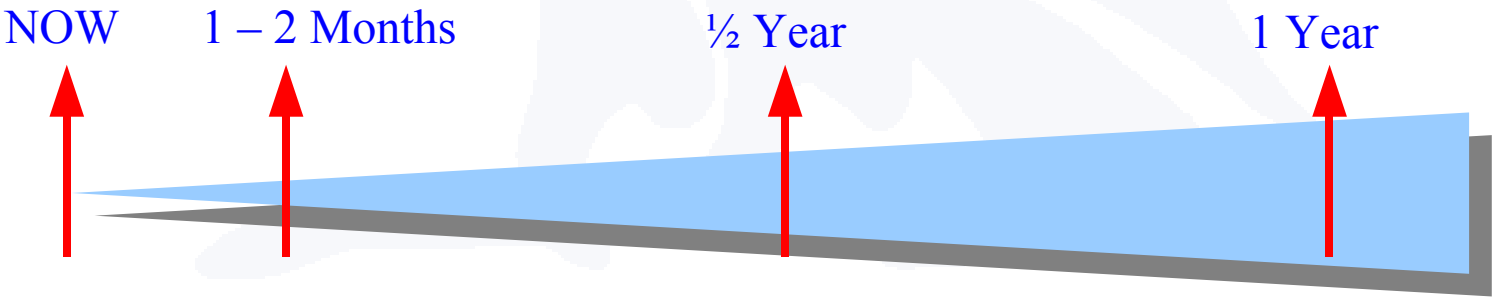
Preliminaries

- Who is behind dCache ?
- What is dCache.org ?
- Market share and support model.
- Open Science Grid -> Tanya

Selected Topics

- Chimera
- NFS 4.1
- NDGF Approach -> Gerd

What's the plan ?



dCache.ORG

Who is behind dCache ?

No

Not yet

But ....

The Team

dCache.ORG

dCache.ORG

Head of dCache.ORG

Patrick Fuhrmann

Core Team (Desy, Fermi, NDGF)

Andrew Baranovski
Gerd Behrmann
Bjoern Boettscher
Ted Hesselroth
Alex Kulyavtsev
Iryna Koslova
Tanya Levshina
Dmitri Litvintsev
David Melkumyan
Paul Millar
~~Martin Radicke~~
Owen Synge
Neha Sharma
Vladimir Podstavkov

Head of Development FNAL :

Timur Perelmutov

Head of Development DESY :

Tigran Mkrtchyan

Head of Development NDGF:

Gerd Behrmann

External Development

Abhishek Singh Rana, SDSC
Jonathan Schaeffer, IN2P3

Support and Help

German HGF Support Team
Greig Cowan, gridPP
~~Stijn De Weirdt (Quattor)~~
Flavia Donno, CERN

Need a job ?

dCache.ORG

2 job positions offered at DESY

(Europe,Germany, Hamburg),

starting end of October.

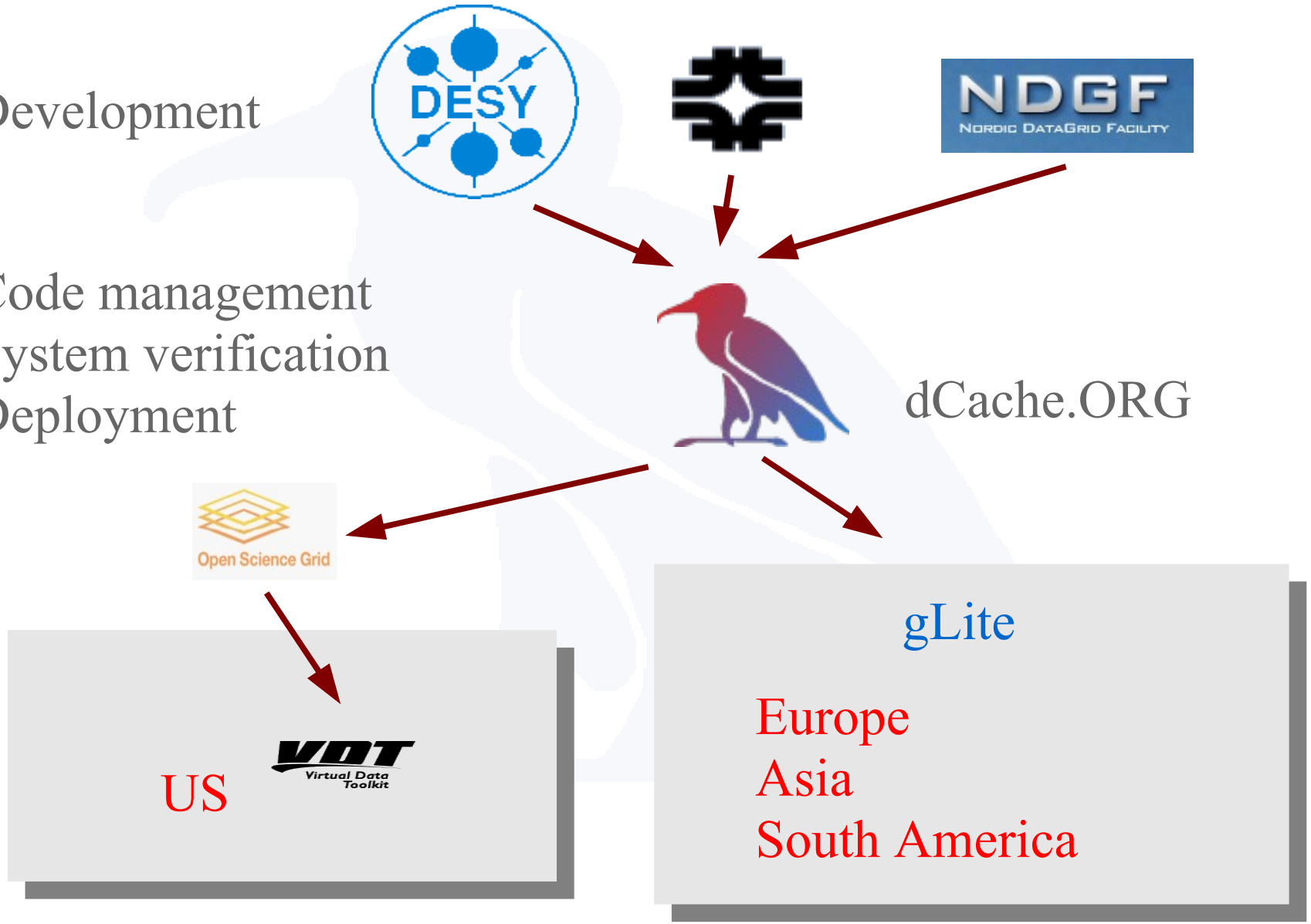
dCache.ORG

What is

dCache.org ?

What is dCache.org ? High level overview

dCache.ORG



dCache.ORG

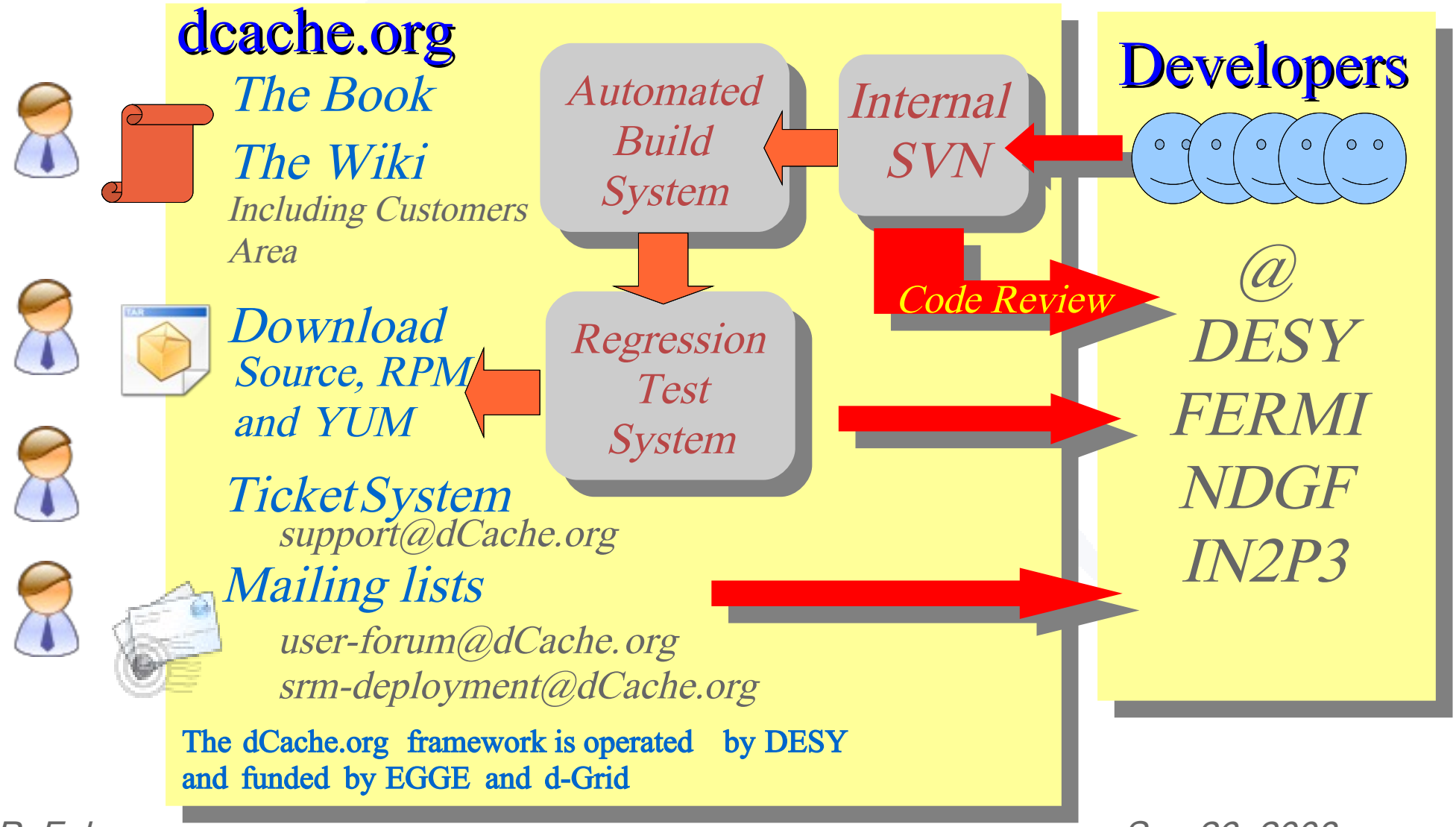
Customer interactions

- User helping user : user-forum@dCache.org
- SRM 2.2 deployment : srm-deployment@dCache.org
- Ticket system : support@dCache.org
- Regular phone conference with some big sites (on request)
- Wiki/SVN area for customers feedback and contributions
- Weekly phone conferences with dCache Tier I's
- Organisation of dCache workshops and tutorials
  - e.g. gridKa school two weeks ago.

What is dCache.org ? The infrastructure

dCache.ORG

- *dCache.ORG is an infrastructure*
- *dCache.ORG is the door into the dCache team*



dCache.ORG

dCache distribution

dCache is distributed (YUM at DESY and CERN) and configured through YAIM for the Tier II's in Europa.

dCache is distributed and configured through VDT for OSG supported Tier II's

dCache is distributed through dCache.org for the Tier I's and configured manually due to the complexity of the Tier I setups.

Market share and support model

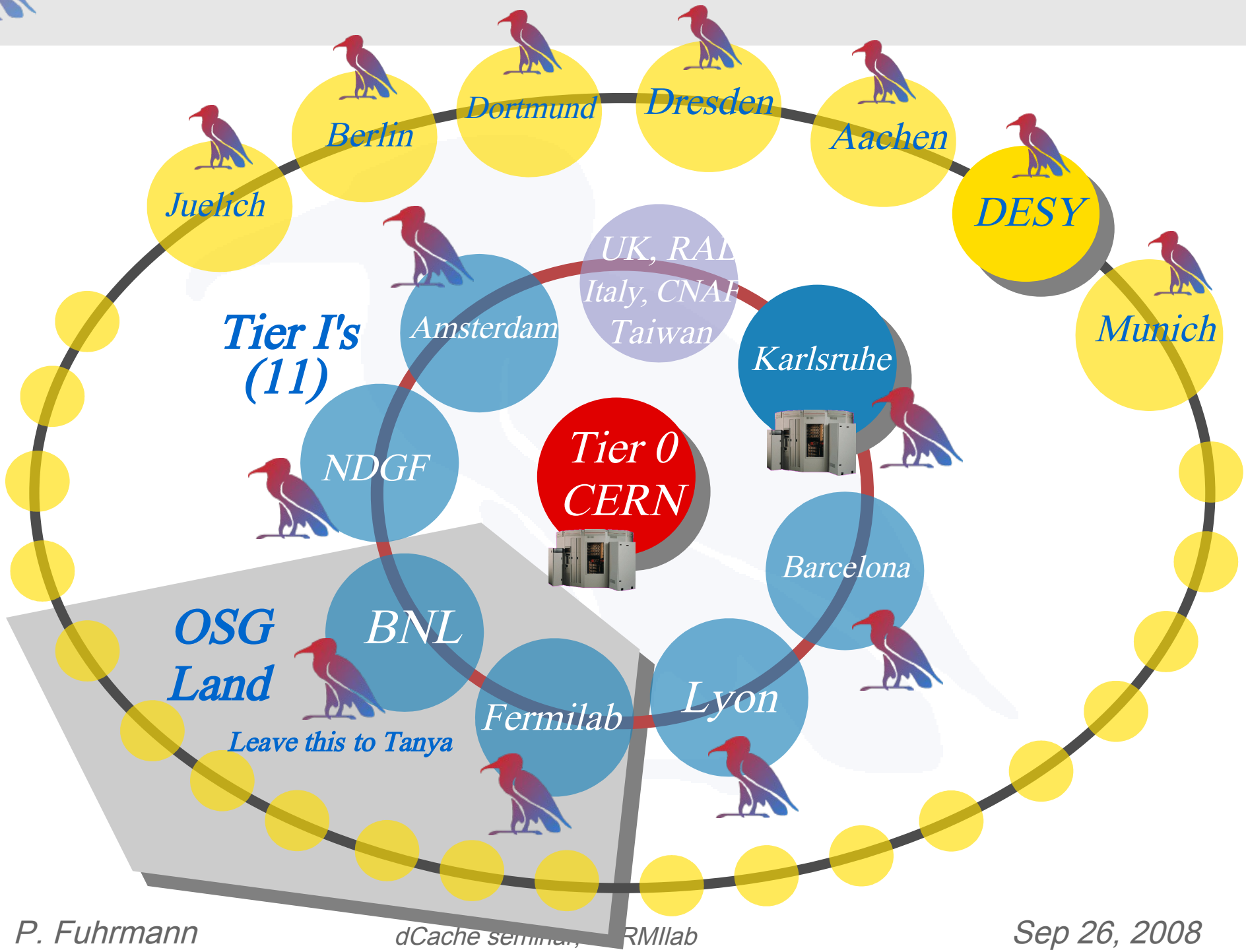
dCache.ORG

Market Share

Support Model

Open Science Grid : covered by Tanya

7 out of 11 Tier I's and more than 70 Tier II's using dCache



dCache.ORG

dCache.ORG

Current support model

dCache.ORG

Tier I's

FERMI(Timur), BNL(Tigran), NDGF(Gerd), gridKa(Patrick),SARA,Pic,In2P3

Open Science Grid

Tanya

US Tier II's

support@dCache.org

GGUS

PHYSICS AT THE TERA SCALE Helmholtz Alliance

German Support

C. Jung et al.

Distributed Storage Support

German Tier II's

Flavia, Giancinto, Greig

Other Tier II's

dCache.ORG

Selected Topics

Chimera

~~ACL's : will be covered by a dCache seminar~~

NFS 4.1

The NDGF approach : covered by Gerd

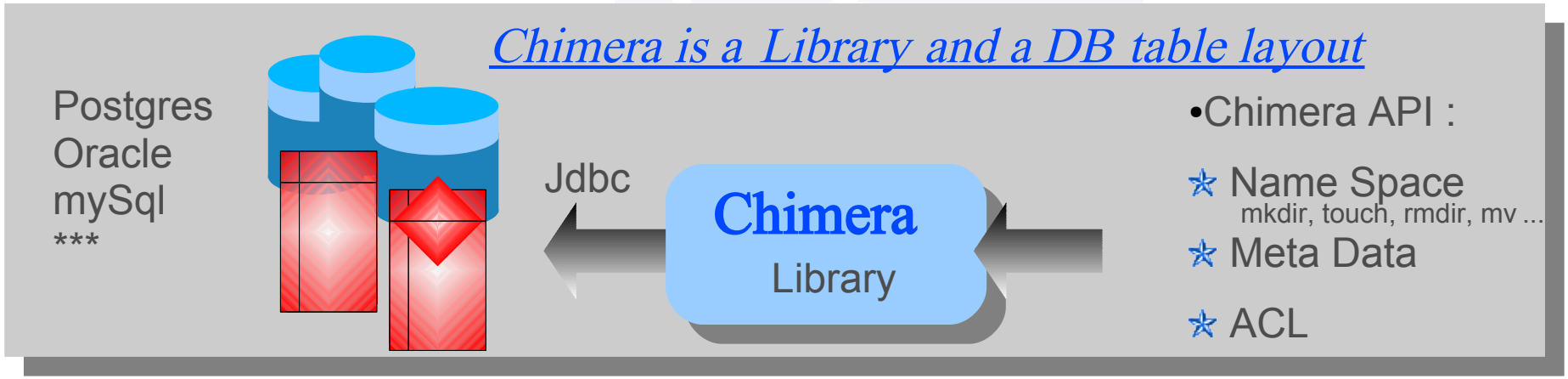
dCache.ORG

Chimera

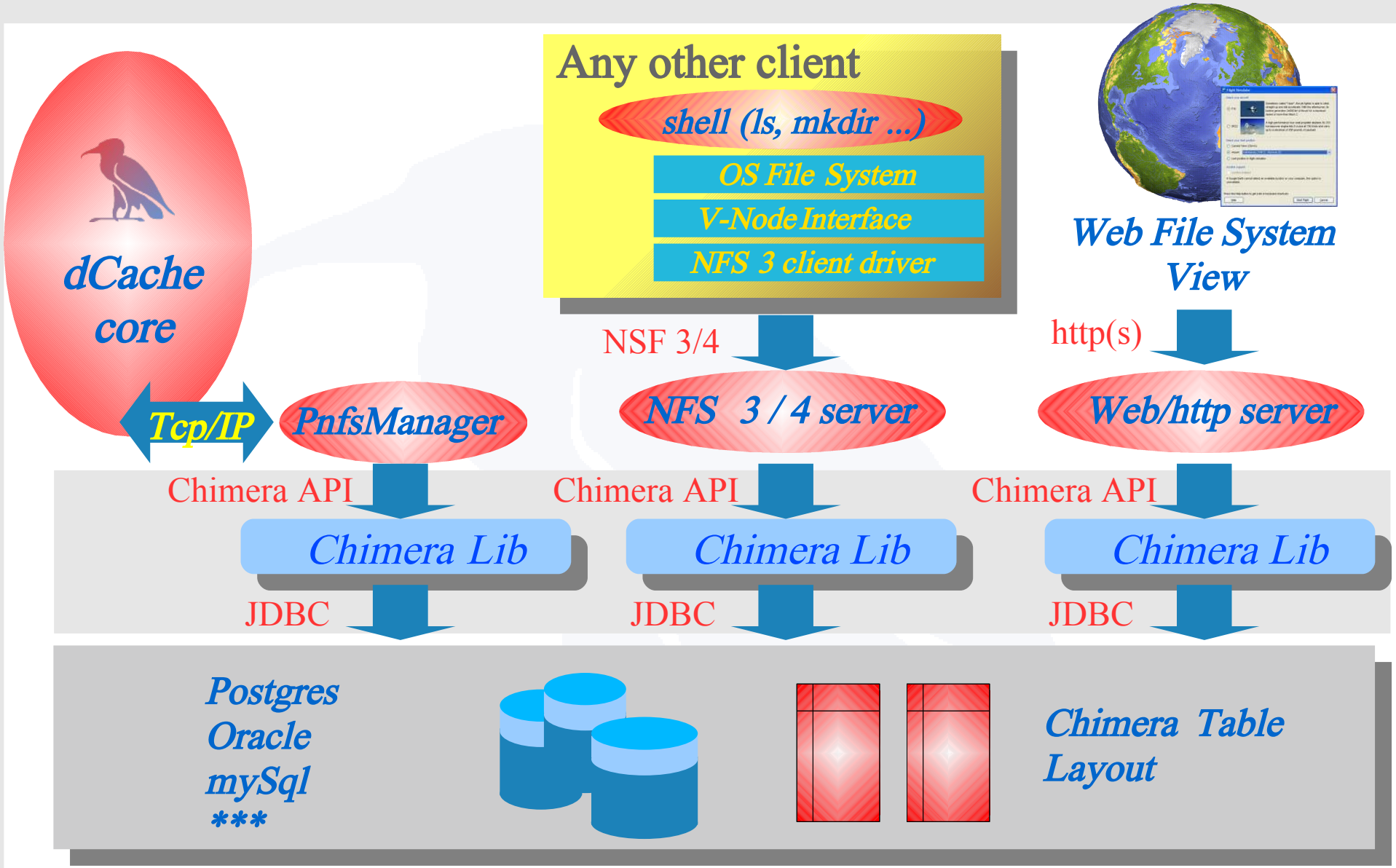
What is Chimera (Basics) ?

dCache.ORG

- Chimera provides the same functionality to dCache as Pnfs does.
- Only the pnfs manager driver within the PnfsManager has to be adjusted.
- Chimera is a Java API, a library and a database table layout.
- There is nothing like a Chimera server.
- Consequently it can make use of any DB performance improvements.



How does dCache interact with Chimera ?



What does Chimera provide ?

dCache.ORG

- *The dCache PnfsManager talks directly to Chimera. (no intermediate layers).*
- *Performance independent of number of files per directory.*
- *Chimera can distinguish between dCache core and the various client interfaces.*
- *Chimera allows ACLs to be plugged in (Posix implementation already av.)*
- Chimera takes advantage of the performance of the chosen database back-end.
  - *If the database back-end can span various hosts, Chimera can do as well.*
  - *No central database table locks.*
- *Customised SQL queries can easily be applied. (e.g. Space Usage by uid, group time....)*
- *Chimera allows at least 10 times more dCache file access operations per second than pnfs (using postgres and similar hardware)*

Status of Chimera

dCache.ORG

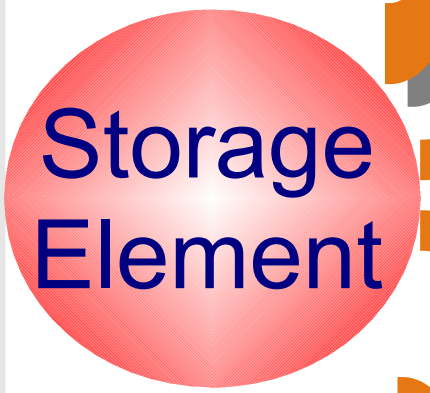
- Chimera is ready (in 1.8.0-15 and higher
- dcache.org now provides a fast conversion mechanism.
  - 100 – 400 records per second
  - Goal : largest site should be able to migrate within one day.
- Weare investigating in a 'way back' as well.
- Based on the production installations in DESYHamburg and Zeuthen, minor issues could be resolved.
- BNL is preparing to move the Phenix dCache to chimera (using Oracle)
- The 6 month delay of LHC possibly allows NDGF to move to chimera as well.

dCache.ORG

NFS 4.1

Motivation

dCache.ORG



Space/Protocol Management

SRM Storage Resource Management

OGF

Wide Area Transport Protocol

In use : gsiFtp

Discussed : http(s)

IETF

Information Service Protocol

Transport : LDAP

Content : GLUE Schema

IETF
OGF

Local Access Protocol

(gsi)dCap or rfio and xRoot

These is not at all a standard

And another project : NFS 4 within CITI

dCache.ORG

University of Michigan

"We are developing an implementation of NFSv4 and NFSv4.1 for Linux."

Introduction of RFC 3530

The Network File System (NFS) version 4 is a distributed filesystem protocol which owes heritage to NFS protocol version 2, RFC 1094, and version 3, RFC 1813. Unlike earlier versions, the NFS version 4 protocol supports traditional file access while integrating support for file locking and the mount protocol. In addition, support for strong security (and its negotiation), compound operations, client caching, and internationalization have been added. Of course, attention has been applied to making NFS version 4 operate well in an Internet environment.

Why is NFS 4.1 : technical perspective

dCache.ORG

- NFS 4.1 is aware of distributed data
- Faster (optimized) e.g.:
  - Compound RPC calls
  - e.g. : 'Stat' produces 3 RPC calls in v3 but only one in v4
- GSS authentication
  - Built-in mandatory security on file system level
- ACL's
- dCache can keep track on client operations
  - OPEN / CLOSE semantic (so system can keep track on open files)
  - 'DEAD' client discovery (by client to server pings)
- smart client caching.

Why is NFS 4.1 : project perspective

dCache.ORG

- POSIX Clients are coming for free (provided by all major OS vendors).
- NFS 4.1 is aware of distributed data.
- Will make dCache attractive to other (non-hep) communities.
- (W)LCG could consider to drop LAN protocol zoo (dcap,rfio,xroot) and goes for standards.
- First step to make WLCG independent of self written storage software.
- Which means : give industry a chance to provide full solutions instead of just selling cheap boxes.

Breaking News

dCache.ORG

Bakeathons last week :

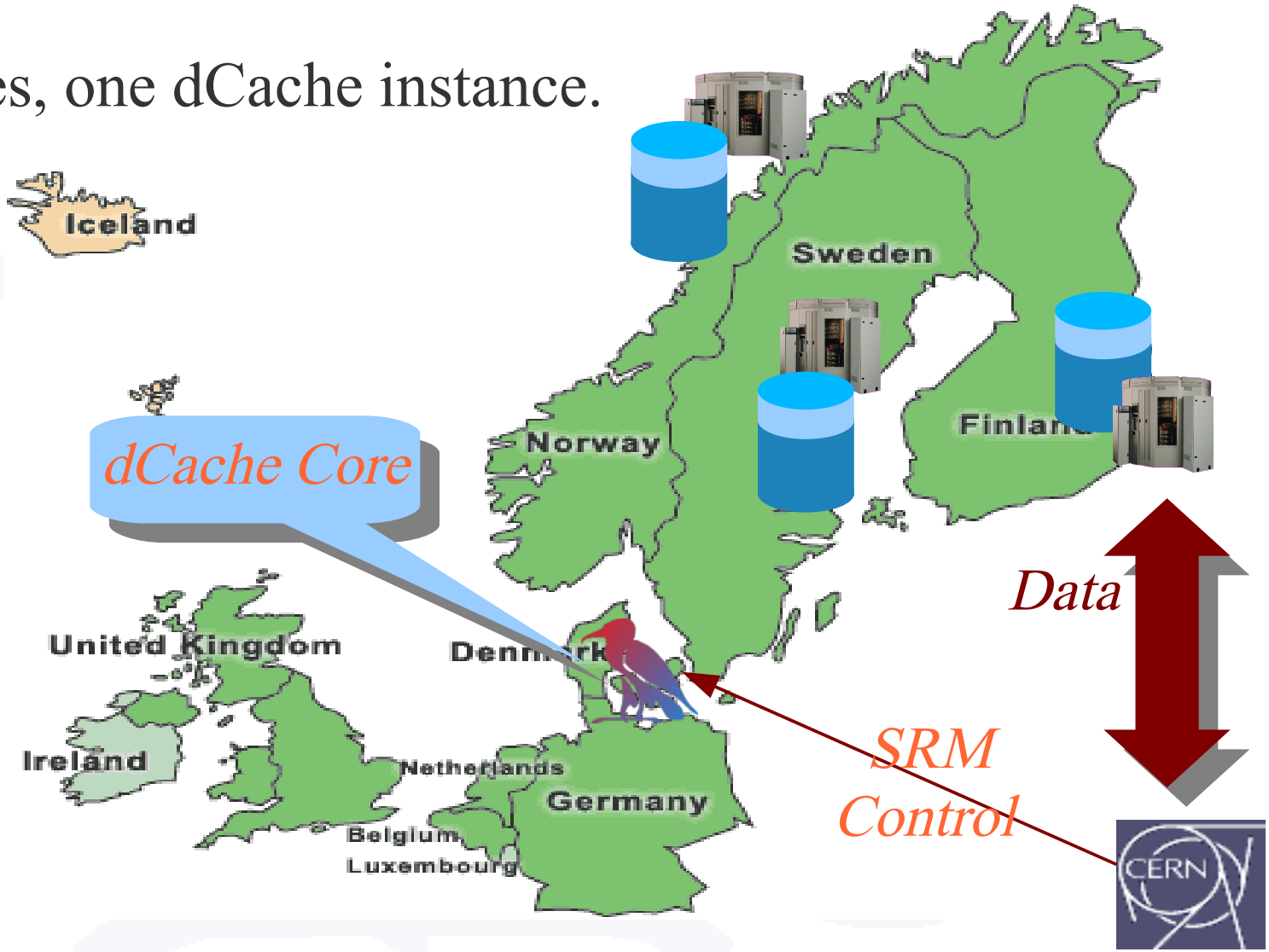
- dCache server can talk to all known NFS4.1 clients
- Some limitations : no modify, no striping but not a problem for clients
- NFS4.1 will be in official standard linux kernel Q1 2009
- full IETF approval till end of 2008
- Client versions
  - Linux
  - SUN (Solaris)
  - CITI will work on Mircosoft client very soon
- Server vendors : IBM, SUN, Panasas, netApp, LSI, EMC, dCache

NDGF : Certainly the most challenging approach

Leave this to Gerd

dCache.ORG

4 Countries, one dCache instance.

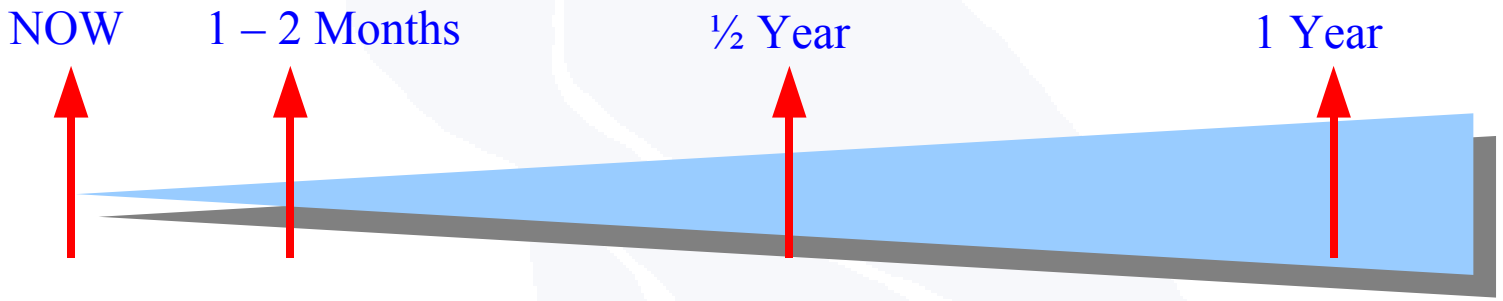


Idea :

At any time a country may 'go down' though raw data storage proceeds.

dCache.ORG

What's the plan ?



What did we achieve yet ?

10 Sep 2008

Jamie Shiers back in 2005

When the LHC starts operating in 2007, it will be the most data-intensive physics instrument on the planet, producing more than *1500* megabytes of data every second for over a decade.

Les Robertson end of 2007

With the start of LHC the largest share of LHC data outside CERN will be stored in dCache.

dCache.ORG

Short Term Roadmap, or the 1.9.x story

dCache.ORG

Leaving details to Gerd

new features

1.9.2

1.9.1

Today, Sep 26

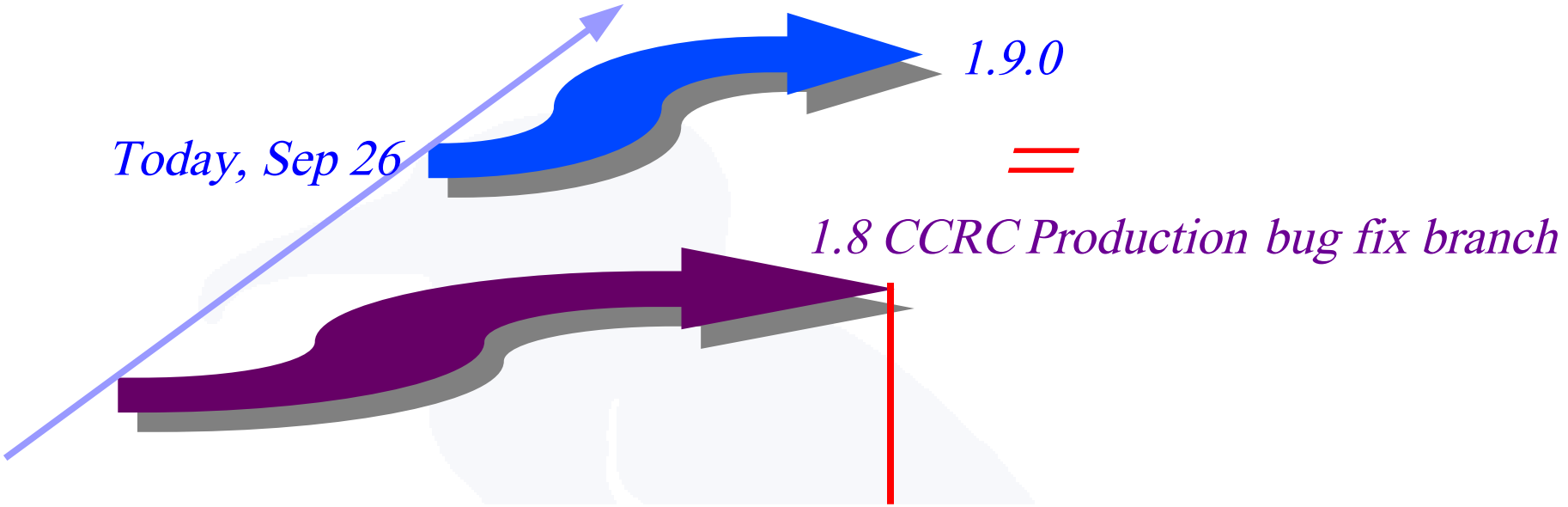
1.9.0 feature branch

=

1.8 CCRC Production bug fix branch

BTW : Support of 'phasing out' releases

dCache.ORG



- *1.9.0 is a low risk upgrade to 1.8*
- *(Changed our mind after last GDB)*
- *We'll encourage sites to follow this path as soon as they can.*
- *We'll apply critical fixes to 1.8 for some short time.*

As soon as we have proven that 1.9.0 is stable and behaves well, support for 1.8 will be terminated.

The 6 months plan

dCache.ORG

Components in the queue :

- Code is ready :
  - Chimera
  - ACL's
  - Improved Information System
  - Improved Pool Code
- Code is nearly ready :
  - modern gPlazma (Ted,Tanya) modified DN/FQAN -> UID/GID(s) mapping
  - unified log file format
- Code is on the way :
  - improved PinManager (Timur may report on this)

The 6 months plan (Cont.)

dCache.ORG

The following 6 months will be dominated by feeding those new components into the 1.9.x series (as Gerd will explain) and to test, deploy and stabilise them.

Improvements will cover the short term MoU agreement with WLCG.

We will try to deploy Chimera at as many sites as possible. There is no need to rush. Chimera is a significant change. So people should be convinced by success of other sites using it.

We'll professionalise the system release process.

We are preparing for online video tutorials (as already started with) which should give our users some confidence in upcoming releases and features.

The 1 Year plan is of course confidential

dCache.ORG

Getting rid of legacy local access protocol and moving towards NFS4.1

Make dCache attractive for the non HEP community.

Collaborating with the CASTOR III team on new storage control protocols and optimised transfer mechanisms.

Prepare for dCache 2.0 (3 Year plan)

dCache.ORG

Further reading

www.dCache.ORG