



**3<sup>rd</sup> TF-Storage meeting**  
**Thursday-Friday, 12-13 February, 2009**  
Dublin, Ireland

**Table of contents**

<b>1. Welcome and apologies</b> .....	<b>1</b>
<b>2. Approval of agenda</b> .....	<b>1</b>
<b>3. Minutes of last meeting and update of action list</b> .....	<b>2</b>
<b>4. Participants' presentations</b> .....	<b>2</b>
• IBM cloud technologies, <i>Pol MacAonghusa (IBM)</i> .....	2
• Sun storage direction, <i>Phil Lawrence (Sun)</i> .....	3
• iPODS: Intelligent Performance Optimisation of Virtualised Data Storage Systems, <i>Nicholas John Dingle (Imperial College London)</i> .....	3
• dCache, <i>Paul Millar (Desy)</i> .....	4
• CineGrid project, <i>Jeroen Roodhart (Univ. of Amsterdam)</i> .....	4
• GRnet simple storage service, <i>Kostas Koumantaros (GRNet)</i> .....	5
• High speed storage transfer, <i>Sajid Qureshi (Attoware)</i> .....	5
<b>5. TF-Storage Work Item related talks</b> .....	<b>6</b>
• NREN disaster recovery services, <i>Jan Meijer (UNINETT)</i> .....	6
• Poste Restante service, <i>Jan Meijer (UNINETT)</i> .....	7
• Federating SSH access, <i>Cándido Rodríguez Montes (RedIRIS)</i> .....	7
• Shibbolized iRODS, <i>David Corney (Rutherford Appleton Lab.)</i> .....	7
• Overview on TF-Storage Work Items .....	8
<b>6. Date of next meeting, aob and close</b> .....	<b>9</b>

**1. Welcome and apologies**

The third TERENA Storage Task Force meeting was held on 12-13 February, 2009, in Dublin, Ireland hosted by HEAnet, the Irish NREN. Jan Meijer (UNINETT), as the chair of TF-Storage, welcomed the participants and asked for a roll call. Peter Szegedi (TERENA) introduced the European NREN community and the role of TERENA for the large number of participants attending first time to the TF-Storage meeting. The history of the storage activities at TERENA and the main objectives of the Storage Task Force were also presented.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d1-0-Intro.pdf>>

**2. Approval of agenda**

The proposed agenda was agreed with the participants without any changes. The presentations are available on the TF-Storage website: <http://www.terena.org/activities/tf-storage/ws5/agenda.html>

### 3. Minutes of last meeting and update of action list

There were some updates on the action list defined during the last TF-Storage meeting in Riga. The comments are shown in the table below:

Reference	Who	Action	Status
Tsec(08)068-1	Jan Meijer (UNINETT)	<i>Investigate coordinated action towards IBM on GPFS academic licensing issue.</i>	<b>On-going.</b> IBM had been invited to give a presentation.
Tsec(08)068-2	Kaspars Krampis (SigmaNet)	<i>SigmaNet to investigate availability of trial version of commercial CleverSafe product.</i>	<b>Open.</b> SigmaNet was not represented at the meeting.
Tsec(08)068-3	Christoph Witzig (SWITCH)	<i>Create test accounts for those interested in a joint Poste Restante software project and investigate the possibility to open the software for use by others with SWITCH as godfather of the project.</i>	<b>Activities are on hold, waiting for a decision at University of Basel. This decision won't be taken before April 2009, earliest.</b> There is a new direction: a Flash-based open source software development project is being established.
Tsec(08)068-4	Peter Szegedi (TERENA)	<i>Function as editor of a document listing the features of the available Poste Restante implementations, based on input from those involved in the implementations.</i>	<b>Changing.</b> Focus point might be shifted towards the documentation of an open source, Poste Restante service development project.

### 4. Participants' presentations

The first day of the TF-Storage meeting was a seminar day. Vendors (e.g., IBM, SUN, Attoware) had been invited to present their views on the major directions of cloud technologies and storage developments.

- **IBM cloud technologies, Pol MacAonghusa (IBM)**

Pol (CTO Emerging Technologies, IBM) gave a talk about the Europe's first Cloud Computing Centre and its facilities. It was established by IBM and the Industrial Development Agency of Ireland in March 2008 in Dublin. The cloud computing centre has been designed especially for proof of concepts, novel cloud developments and tests. The main aim is to run try outs on clouds (no free hosting). The centre does work for the likes of Google, Amazon and Ebay. Virtualization and cloud computing with autonomic management is a key part of the solution. Except some older components everything can be virtualized in a data centre, hence scheduling and provisioning of virtualized resources are very important. Regarding the autonomic

management, e.g., in the Google solution, a new machine is automatically identified, the proper software is installed and the resource immediately becomes part of the cloud. Of course, there are some interoperability issues among the vendors' solutions, so standardization is important. However, there are some open source agreements e.g., between IBM and Google. In general, the proper SLA handling is still an open question; it is sometimes overestimated, sometimes not even considered. One solution could be to offer a mix of SLAs as a menu for the generic users. The key is to force the users to select the optimal SLA, because higher SLA means higher costs.

Some of the recent research topics were highlighted in the talk:

- Storage is the typical resource in a cloud. Despite of the virtualizations data is stored physically on the disks hence defragmentation is still an issue.
- Mobile devices can also be part of the cloud. The large number of mobile devices and peer-to-peer communications has to be managed and administrated.
- Some research topics exploit the extreme scalability of clouds; like video surveillance systems, grid processing, streaming technologies, Software as a Service (SaaS) applications on virtual networks, etc.

At the end of the presentation it was noted by the audience that the federated NREN environment is different from a single vendor (IBM) based cloud (various vendors, HW and SW architectures exist). It has to be managed differently. As a general conclusion; standardization is needed in that area.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d1-1-IBM.pdf>>

- **Sun storage direction, *Phil Lawrence (Sun)***

Phil (Senior Solution Architect, Sun Microsystems) presented about the Sun S7000 series and why it was disruptive, and a number of other Sun storage technologies. The Sun strategy appears to be based on heterogeneous industry standard hardware components, with Sun open source software on top, allowing disruptiveness. Sun will make money on shipping and servicing appliances, like the S7000 series, allowing you to go from zero to storage in 5 minutes.

The S7000 series uses hybrid storage pools where SSDs (solid-state drives) are included. ZFS (file system designed by Sun Microsystems) is used and part of the ZFS design is to make intelligent use of SSDs, impacting performance in interesting ways.

Features like active-active clustering, snapshots, clones, etc. are available free of extra charge, no extra licenses are needed.

There is a big box with some 154 disk slots that goes for about 150K GBP. Encryption will be built in next year. The appliances are planned to do inline deduplication as of Q4 this year. Phil also talked about the dTrace utility, used to trace performance issues, and the Sun storage configuration tool, a point/click storage configuration tool.

When asked who will use the Sun Unified 7000 series, Phil responded "Customers who are struggling to keep pace with rapid storage growth, looking for a radically easier and faster way to manage storage at a substantially better ROI".

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d1-2-SUN.pdf>>

- **iPODS: Intelligent Performance Optimisation of Virtualised Data Storage Systems, *Nicholas John Dingle (Imperial College London)***

iPODS (Intelligent Performance Optimisation of Virtualised Data Storage Systems) is a three

year project (2007-2010) funded by the UK Engineering and Physical Sciences Research Council (EPSRC), supported by industrial partners (Reuters and IBM).

The main objective of iPODS is "to develop more sophisticated fabric intelligence that is able to autonomously and transparently migrate data across tiers and organise data within tiers to deliver the required QoS in terms of factors such as response time, availability, reliability, resilience, storage cost and power utilisation."

Recently the project partners are seeking to develop: intelligent data placement and migration strategies and performance evaluation tools to assess the benefits of such strategies quantitatively. The presented performance evaluation tool is based on mathematical representation of virtualised storage systems. Basic queuing theories and heuristic data placement strategy are used to obtain the analytical results. It was demonstrated by measurements that the RAID model matches well with observed reality.

Current limitations of the model are:

- caching is not represented in the model.
- FIFO is good assumption but sometime to minimize the seek time the jobs are chosen in a different order.
- distribution of the arrival times are not so close to the reality (real life traces are needed).

Details about the future work can be found here: <http://aesop.doc.ic.ac.uk/>

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d1-3-ipods.pdf>>

- **dCache, Paul Millar (Desy)**

dCache is both the name of the project and the name of the file based, write once read many times (grid) storage system.

dCache can combine lots of heterogeneous servers (the only requirement: Java). It provides a rooted name-space, support for HSM storage and many file access protocols.

The supported industry standards are: FTP, HTTP, NFSv2&v3 (name-space only), NFSv4.1 (coming with dCache v1.9.4). The overall aim is to move towards more standard protocols.

However, dCache have had strong involvement in community standards (e.g., syncat, GLUEv2.0 and SRM (v1.1, v2.2)). dCache was the first implementation of GridFTPv2. It was noted that performance evaluation/benchmark is not available yet between the native NFS and dCache NFS systems.

In conclusion can be said that the dCache (Grid) Storage Software:

- supports HSM backends.
- fault-tolerant.
- namespace and data-storage are separated.
- supports HTTP and (with v1.9.4) NFSv4.1
- majority of LHC data (outside of CERN) is stored on dCache servers.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d1-4-dCache.pdf>>

- **CineGrid project, Jeroen Roodhart (Univ. of Amsterdam)**

CineGrid's mission is to build an interdisciplinary community focused on the research, development and demonstration of networked collaborative tools, enabling the production, use and exchange of very high-quality digital media over high-speed photonic networks.

CineGrid experiments with 4K video streams that require a significant amount of storage for the media files, as well as some serious performance.

Lessons learnt during the project:

- The disks are not big enough and the traditional storage solutions are not fast enough for HD digital media. Compression is usually pre-processed but decompression has to be done on the fly. Some hardware implementations are needed because of the speed.
- Raw storage system speed may be enough for 1 uncompressed 4K stream, but does not scale to concurrent streaming. Typically, large data sets have different storage requirements.
- Considering the CineGrid application, there is no standardized workflow from storage to display moreover the conventional streaming tools use "file system paradigm" not suitable for proper storage.

CineGrid will be using iRODS. The proposed contribution to iRODS is to place NDL and semantic information (e.g., storage/content delivery nodes, transcoding) within iRODS.

Novel approach is needed where "streaming nodes" can access data using cluster technology including; fast interconnect (RDMA/QDR InfiniBand), more than one storage server and new technologies that may lead to more elegant designs (e.g., SSD, ZFS, Lustre).

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d1-5-Cinegrid.pdf>>

- **GRnet simple storage service, *Kostas Koumantaros (GRNet)***

Kostas (GRNet) introduced the GRNet Simple Storage System (GSS) to provide free storage for research and academic community (10G/person). Users will be able to upload, share, and index their files. The idea has been inspired by Amazon S3, but going beyond that.

GSS currently is running in BETA test version. It offers users a file system abstraction with file/folder hierarchical structures and usual file system operations. Users are able to share their files with selected other users, or defined user groups. GSS enables users to version their files automatically and full text search is also provided. Initially, all users are equals, policy issues are not solved yet (improvements are needed).

The final GSS version will use Shibboleth for AA. GRNet has prepared a Shibboleth infrastructure for all institutions in Greece. The service will be in production mode soon.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d1-6-GRNET.pdf>>

- **High speed storage transfer, *Sajid Qureshi (Attoware)***

Sajid (Attoware) introduced the challenges of high speed data transfer. During the transfer of the huge data you may have many copies in the systems. Some of those copies (e.g., in the NIC buffer) cannot be touched. In case of system crashes physical storage can be restored but the data in memory may go.

Attoware provides a solution to do the data transfer stress free, installation free and flexible. They have developed a microkernel running over Windows. It does not modify the operating system just sends requests and calls. The main achievement: 1 TB+ data can be transferred per hour between two PCs, each with 4 Gbits/s NICs.

The current code version is Intel specific and very small (only 100K: 10K Assembly + 90K C-based). It is feature-complete on Windows but it can be put on a mobile device too.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d2-2-Attoware.pdf>>

## 5. TF-Storage Work Item related talks

The second day of the meeting was focused on the task force related presentations and discussions with the active contribution of the task force participants.

- **NREN disaster recovery services, Jan Meijer (UNINETT)**

Jan (UNINETT) presented an initiative to provide backup service for the Norwegian University Collages. The business case behind that could be to provide a cheaper, better (in terms of reliability, scalability, functions, etc.) and future proof solution. In stead of each college establishing their own off site backup facility with server hardware to run virtual machines on, you'd have one or more centralized platforms to do this. The chance that any two data centres burn down on the same day is rather small.

The service could be the first step towards realization of the longer term vision; cloudify the Norwegian higher education sector (i.e., create a nation-wide virtual data centre, so especially smaller sites don't need to operate physical servers and storage).

From the network connections' perspective, Norway can use any services via Nordunet and GEANT. In addition, with potential direct peering contracts, bandwidth cost could be cut when connecting to commercial storage/compute cloud providers.

Based on the preliminary user consultations in Norway, the general impression is that the customers do not really know what kind of service parameters (RPO (Recovery Point Objective), RTO (Recovery Time Objective)) are suitable for them making it slightly difficult to assess in which direction the service development should be taken. We have to identify and prioritize the major events affecting the IT services (i.e., human failure, equipment failure, site failure, etc.).

It was commented that to formulate a service offer (Business Case for backup services), the basic business processes of the university have to be well understood and political backing needs to be established. Jan commented that university colleges, the primary target for this service, tend to be less complex than universities.

On the technical side, there are many options to start with.

- Online (commercial) backup providers. Easy in a way as in theory only requirements need to be defined. However, this solution does not put UNINETT on the evolutionary path towards a national cloud, and commercial providers have their own particular disadvantages to deal with;
- CDP, which works, but hands-on experience is lacking, the solution might be complex and not in the least it will be pricey;
- Deduplication promises benefits: the amount of storage space at UNINETT can then increase with less then the amount of space at the organisations being backed up. Deduplication performs especially well with VMware installations, office documents and MS Exchange environments;
- Traditional backup software can be used where RTO matters less, but has the disadvantage of usually being rather complex.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d2-1-backup.pdf>>

Jan asked about the similar activities among the participants. HEAnet mentioned that they have a backup service only for web contents. Rutherford Lab. is also offering similar services.

TDC used to do that but quit. The general feeling is that the NRENs are worried about the reliability that has to be provided for such services.

- **Poste Restante service, Jan Meijer (UNINETT)**

Jan (UNINETT) presented the current status of the Poste Restante service development activity. During the Riga meeting all the existing implementations (CSC, HEAnet, UNINETT, SWITCH/Univ.Basel) have been demonstrated, after which the group of interested NRENs decided to go for the best developed solution, the DocExchange implemented by the University of Basel. SWITCH was asked to investigate the possibilities of Basel University opening the Java-based source code, with SWITCH acting as a godfather for such an open source project. The process has seen little progress and is now awaiting a decision by Basel University which at its earliest will be taken in April, but likely later.

Jan introduced an alternative option he has been working on with HEAnet and AARnet: develop open source flash based Poste Restante software. Through AARnet a competent Flash developer has been found who will be engaged for the project. The estimated cost is about 60K Australian dollars which gives all desired features (including multi-language support).

Jan asked who'd be interested in joining this project from the start. CESNET and Belnet indicated their interest. It was mentioned that a small and simple agreement is enough to set up the project.

***Action Tsec(09)007-1:** on Jan (UNINETT) to discuss with CESNET, Belnet and any others that are interested, about them joining the Flash based, open source, Poste Restante software.*

- **Federating SSH access, Cándido Rodríguez Montes (RedIRIS)**

Cándido (RedIRIS) presented an architecture for federated SSH access proposed by RedIRIS. Based on the architecture the public keys are stored on LDAP server. The question is then how to get the keys through LDAP. There are two different approaches:

- a patched OpenSSH which is able to get keys through LDAP
- develop a software which somehow connects OpenSSH to LDAP

The existing OpenSSH patches are quite big ones. RedIRIS has developed a light weight patch (only 10KB) for SSHD able to get keys through LDAP in real time.

The second approach is more promising without modifying OpenSSH. FedSSH is a web based application (using PHP and JQuery) to deal with the federation developed by RedIRIS. It allows users to upload their keys into LDAP. It supports groups of nodes. The features of the FedSSH application were demonstrated during the meeting. The main result is that the keys and entitlements can be sent to the SSH application in a quick and efficient way via web services.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d2-3-FedSSH.pdf>>

- **Shibbolized iRODS, David Corney (Rutherford Appleton Lab.)**

David (Rutherford A. Lab.) gave a talk about their Shibbolized iRODS solution. First, the structure of the research organization and the underlying data infrastructure has been introduced. The architecture of the BBSRC archiving system and the offered archive/management services, policies have then been detailed.

David presented the ASPiS project. The project uses iRODS as data storage and provides a

Single Sign On login via Shibboleth. SSO allows using single password managed by home institution. The home institution can provide user attributes and ASPIs can use those for access control and for provenance.

The proposed solution does not change iRODS itself; the system is Shibbolized outside the iRODS code. You can check the slides for the detailed architecture.

<Slides: <http://www.terena.org/activities/tf-storage/ws5/slides/d2-4-IRODS.pdf>>

- **Overview on TF-Storage Work Items**

The last agenda item was a brief overview on the actual status of TF-Storage Work Items (proposed in the Terms of Reference document).

- A) *Knowledge dissemination*: Interesting talks have been organized. We have had a significant growth in terms of participants' number since the start of the task force.
- B) *Overview of (national) activities and deployments*: On-going activity.

**Action Tsec(09)007-2**: on Jan (UNINETT) to update the Storage Wiki with the latest information.

- C) *Poste Restante service*: It is progressing. (See the notes above.)
- D) *Collection of best practices and service requirements*: PSNC is working on a cookbook for benchmarking storage performance. UNINETT and Rutherford A. Lab. are also interested in that work. The basic strategy is to define some simple use cases (e.g., LHC, Media storage, TV archive, etc.) and try to answer the main questions using those experiences. It could be useful to identify organizations (outside of TF-Storage) and invite them to talk about their benchmarking approaches. With regards to legal best practices the suggestion was voiced to contact organisations that have worked on this for many years, such as CODATA of ICSU.

**Action Tsec(09)007-3**: on David Corney (STFC) to identify organisations and people who worry about the (legal) aspects of cross-domain data sharing.

- E) *Taxonomy of storage (virtualization) middleware*: HEAnet has already started to work on. The basic idea is to provide information about dCache, iRODS, etc. to NREN community. Make a list on "who does what and who uses it" in this area and identify the common features of systems, tools, solutions, etc.
- F) *Taxonomy of storage technologies*: The aim is to exchange experiences on evaluation of storage technologies. It was noted the storage technology is rapidly evolving. Such storage taxonomy is only useful if it is up to date. We need a plan for that.

**Action Tsec(09)007-4**: on all (TF-Storage) to think about how to establish a maintainable storage taxonomy document on the Storage Wiki.

- G) *Measuring storage performance*: After the great work on the storage performance measurement cookbook by Stijn Eeckhaut not that much work was done. During the

iPODS presentation there was a call to supply the iPODS researchers with usage patterns (i.e., I/O traces), needed for accurate storage modelling. It was also noted that storage performance measurement know-how would be useful for defining acceptance tests with tender evaluations.

- H) *Backup and/or disaster recovery services*: It is progressing. (See the notes above.)
- I) *Storage and AAI*: Clear scenario is needed to solve the non-Wed AAI issues. There will be a BoF organized during the TERENA Networking Conference (8-11 June, 2009 in Malaga, Spain) about non-Web based AAI solutions. It would be good if the TF-Storage makes its desires for federated authentication known at that meeting. Call for participation!
- J) *Energy consumption considerations*: At a previous meeting it was decided to drop this item.
- K) *Liaise with other communities*: On-going activity.

## 6. Date of next meeting, aob and close

The next TF-Storage meeting will be co-located with the NORDUnet Conference ([http://www.nordu.net/conference/ndn2009web/home\\_.html](http://www.nordu.net/conference/ndn2009web/home_.html)) held on 16-18 September, 2009, in Copenhagen, Denmark. The exact date of the TF-Storage meeting will be decided later. The most likely option is the day before the NORDUnet Conference (15 September, 2009). A social dinner invitation was offered by Attoware.

Jan Meijer (UNINETT) thanked HEAnet for hosting the meeting and all the attendees for the active participation, then closed the meeting with a call for a free discussion about the future directions.

During the AOB a lively discussion took place around the concepts of cloud storage and cloud computing and what to do with it. Especially those participants responsible for very large (multi petabyte) installations or those planning large datacentre consolidations were interested: they keep adding more storage and compute resources to their installations and are in need of better tooling to manage this process. It was also remarked that cloud computing and cloud storage is very likely to seriously impact the way we do things, so it would be the right time to gather some experience with the more practical aspects of this slightly foggy concept.

Pol MacAonghusa's presentation showed promising directions. At the end of the discussion a decision was made to see if a meeting could be organised with the IBM Cloud Research Centre and a number of TF-Storage participants to discuss a possible collaboration project. Interested participants: Luke Drury, David Corney, Lars Fischer, Jan Meijer, Brian Boyle, Sajid Qureshi. David takes it upon him to organise the meeting.

**Action Tsec(O9)007-5:** on David Corney (STFC) to organise a meeting with IBM Cloud Research Centre and interested participants to discuss collaboration possibilities.

### Action list

Reference	Who	Action	Deadline
Action Tsec(09)007-1	Jan Meijer (UNINETT)	Discuss with CESNET, Belnet and any others that are interested, about them joining the Flash based, open source, Poste Restante software.	Next meeting
Action Tsec(09)007-2	Jan Meijer (UNINETT)	Update the Storage Wiki with the latest information.	Next meeting
Action Tsec(09)007-3	David Corney (STFC)	Identify organisations and people who worry about the (legal) aspects of cross-domain data sharing.	Next meeting
Action Tsec(09)007-4	all	Think about how to establish a maintainable storage taxonomy document on the Storage Wiki.	Next meeting
Action Tsec(09)007-5	David Corney (STFC)	Organise a meeting with IBM Cloud Research Centre and interested participants to discuss collaboration possibilities.	ASAP

### List of participants

First name	Last name	Affiliation
David	Antos	CESNET
Chris	Ariyo	CSC - IT Centre for Science
Kurt	Bauer	ACOnet
Brian	Boyle	HEAnet
Maciej	Brzezniak	PSNC
Shahid	Butt	Attoware
Brian	Coghlan	Trinity College Dublin
David	Corney	Rutherford Appleton Lab
Paul	Dekkers	SURFnet
Nick	Dingle	Imperial College London
Dobrisa	Dobrenic	University Computing Centre (SRCE)
Luke	Drury	Dublin Institute for Advanced Studies
Stephane	Dudzinski	Dublin Institute for Advanced Studies
Lars	Fischer	NORDUnet
Lukas	Hejtmanek	CESNET
Justin	Hourigan	HEAnet
Kashif	Iqbal	Irish Centre for High End Computing (ICHEC)
Axel Ramón	Klint	Attoware
William	Knottenbelt	Imperial College London
Kostas	Koumantaros	GRNET

Andreas	Landhäußer	T-Systems Solutions for Research GmbH
Rossend	Llurba	NCF
Pol Mac	Aonghusa	IBM
Jan	Meijer	UNINETT
Paul	Millar	DESY
Paul	Mullen	HEAnet
Gabriele	Pierantoni	Trinity College Dublin
Geoff	Quigley	Trinity College Dublin
Sajid	Qureshi	Attoware
Jean-Christophe	Real	BELNET
Keith	Rochford	Dublin Institute for Advanced Studies
Cándido	Rodríguez Montes	RedIRIS
Jeroen	Roodhart	University of Amsterdam, Informatiseringscentrum
Neil	Simon	Trinity College Dublin
Peter	Stefan	NIIF/HUNGARNET
Peter	Szegedi	TERENA
John	Walsh	Trinity College Dublin/Grid-Ireland/e-INIS
Soraya	Zertal	PRISM lab. University of Versailles