

TERENA Storage Collaboration Services, 2nd Meeting 14 December 2007

Meeting Notes, 14-01-2008, Valentino Cavalli, Jan Meijer

1. Round of introductions

The meeting was convened and chaired by Jan Meijer, from UNINETT. After welcoming the participants, Jan briefly introduced the goal of the meeting and started a roll call to identify the participant's interest.

DFN has just begun with storage services in 2007, starting up a project to establish a backup service. Ulrich Kaehler of DFN remarked that mostly smaller institutes would find this a valuable service; bigger ones would be able and willing to fulfill their own needs. Jan said a similar situation exists in Norway, where the planned backup service is likely to be most interesting for smaller institutions.

In Slovenia, ARNES is planning to offer backup services in the future; these would include disaster recovery. It is not yet clear exactly what, how and when.

SWITCH are interested in watching the developments.

GRNET is providing a storage facility for universities, mainly targeting end users.

In the UK, the Rutherford Appleton Laboratory is managing a storage group, which provides, amongst other things, a tier-1 facility, an SRB service and a digital curation center. RL is actively looking at iRODS (the future SRB replacement). Their interest is in data storage, archiving and brokers.

SURFnet is looking into offering storage services for their own organisation and customers. It is expected that larger universities will offer services themselves, so they will probably look at the smaller customer institutions; the type of services may include backup, data services, and surfgroepen type of services. One could use web service file system for data storage. They want to address the need of reducing power consumption; they are also looking at virtualisation; more services like spam filtering need storage; additionally, SURFNet will support a couple of large projects: CDN and Cinegrid, which need lot of storage and fast access. This is thought to mimic the requirements of the HPC environment.

The University of Porto is interested in disaster recovery and in internal storage solutions to support their own Grid services.

In Iceland they are interested in seeing what's going on in other countries.

PSNC in Poland are working on a national data storage service; user requirements work was finished already. A presentation was given later during the meeting.

UNINETT runs the NorStore project, which will provide (production) large capacity storage services to the Norwegian (natural) sciences as of April 2008. Exactly what kind of storage services will be provided to the rest of the Norwegian higher

education environment (universities and university colleges) is not yet clear but a number of pilot projects (including backup/disaster recovery) will be ran in 2008.

CSC are interested in storage services to support their LHC activities (dCache); they are also interested in supporting a CMS experiment at tier-2 level. CSC/FUNET have been working on storage for a while and are looking forward to the first services to be provided soon. Last year they worked on the ePoste Restante service.

T-Systems are running services for the German Aerospace Centre; they are looking for concepts and ideas on backup and archiving.

RedIRIS do not provide any storage services, but they are interested in collaborating in integrating such services with AAI.

CESNET are running some data services. They are looking at new services, brokers and catalogues.

NIIF/HUNGARNET are running storage services for academy, libraries and videoconferencing archives; they are interested in getting experience with lightweight distributed storage solutions in an EC project.

In Latvia, SIGMANET are interested in storage services to support the Baltic Grid EC project.

The Nordic Data Grid Facility (NDGF) is a tier-1 distributed storage facility; they are looking at how-to documentation and information exchange.

HEANet are interested in storage to support web hosting, videoconferencing and streaming archiving. They just started a tender for additional services in order to test high performance services. UNINETT have just closed a tender for similar services and would find it interesting to share experiences.

Jan summarised at the end of the roll call and pointed to the observed areas of interest:

- Backup & disaster recovery
- Storage for NREN services with a focus on provisioning for video streaming
- Storage from an eScience/Grid point of view (tier-1)
- Exchange of knowledge and ideas

2. Jari Miettinen, CSC/FUNET, presents progress on ePoste Restante and ePoste Payé service

Jari provided a presentation on the ePoste Restante and ePoste Payé service developed by FUNET. This service has been created to enable the easy exchange of large files between customers. He explained what happened in past few months since the previous storage meeting: Jari recapped his presentation from the past meeting and mentioned significant progresses in testing the ePoste Restante service; some special interest group meetings were also organised. There were plans to test

the ePort Payé service early next year and then start the service very soon after the process was finished.

Files are currently stored for up to four days and the suggested limit on the file size is 1 DVD disk. The service uses shibboleth for authentication, should it be provided as a pan-European service this needs to be looked at. EduGain is reportedly not yet ready to run production to provide this.

ARNES, GRNET, SWITCH, HeaNet, SURFnet and UNINETT indicated interest in having the service available for their constituency. After some discussion it was concluded that for the time being most participants consider it most practical to start with running the software themselves, linking it to the local AAI, and get an initial service going. Extra functionality and linkage to EduGain-like functionality would be nice to work on after that first step. CSC was open to bilateral AAI linking agreements.

Jari suggested sending an invitation via the mailing list for people interested in joining the tests. The code needs a bit of cleaning before it is shared with other people. He would try to speed up things and convince the developers it could be released before being perfect.

3. PSNC NDS Project

Maciej Brzezniak from PSNC presented a 2-year project for a national Data Storage service, which started in September 2007. The project is meant to provide a secondary data storage system. Two types of services were envisaged:

- Remote logical file system – accessible through data access applications, such as SCP, FTP etc.
- Backup archive service on commercial servers (Tivoli storage manager). PSNC will develop an own client API.

Target users are: health care, government, educational users, supercomputing and telecommunications. The project should also deal with privacy law issues and other regulations.

Initial requirements are ready. PSNC are now working on an analysis of regulatory issues, data protection laws etc. It was remarked that end users very often do not know about regulations themselves. There are also open issues related to metadata management. The project has secured 2 years of funding.

PSNC are currently working on a formal report for the Polish government. Later on an English version would be published for GridCore (Q1 2008).

Using the Polish activity of legal analysis possible collaboration on legal issues was discussed. What are the legal issues surrounding distributed storage, are there any European laws that apply? Does anyone have any plans to investigate legal issues? Some plans exist, mostly in the area of privacy and data retention laws. It was concluded that a common activity on legal issues surrounding storage activities did not look very viable but it would be quite interesting to have a legal expert speak on the topic at a next meeting. It was remarked that there are many different national regulations. Possibly some institute might be interested in studying these, along the lines of what Bert-Jaap Koops did with the crypto law survey at the University of Tilburg.

4. Discuss collaboration

Jan pointed out that there are basically two ways of continuing collaboration on data storage within the TERENA framework: either in a loose set of workshops or in a TERENA task force. He listed the possible areas for collaboration gathered in preparing this meeting and aimed at identifying the common denominators. The basic idea was to find items to work on together, but continuation could also be limited to exchange of ideas.

Lars Fisher was in favour of at least establishing a forum for information exchange, but said it would be nice to have a couple of pragmatic tasks to jointly work or focus on.

4.1 Ranking list on collaboration areas

a) File sharing

There was wide consensus on working on this area (see point 2 above).

b) AAI issues

There was significant interest in this in Norway and the Netherlands. AAI federations are mostly web based so when you want to use them to authenticate national storage services or Grid services work needs to be done to enable non-browser based applications to use these existing AAI federations.

Christoph Witzig expressed a warning not to try and hack something together for a short-term solution. For shibboleth they are looking into the problem but Jan was not convinced the Shib people could deliver on the promise within a reasonable term of say 2 years. Christoph remarked that any thing you do should fit in the Shibboleth way. The TERENA task force TF-EMC2 was potentially interested in exploring this, but, it was observed, the task force is not really looking beyond web-based applications. This is mostly because there was no push so far to go beyond that: the storage group could be such a push.

The following participants were interested in working in this area:

- Actively contributing:
 - Spain,
 - The Netherlands,
 - Greece,
 - Norway;
- Observers:
 - Ireland was interested but not able to contribute.

People agreed on the need to identify technologies and see how they can be federated. Cándido Rodríguez volunteered to gather information on AAI and storage and present the results at the next meeting.

c) Disaster recovery

It was remarked that this service seems to make a lot of sense because you get economy of scale and also business continuity. In many countries there are commercial companies offering backup/disaster recovery services.

Niels noted that in the Netherlands there are at least two commercial service providers offering disk space combined with VMware hosting to provide business continuity services. What the cost for this is is not yet known. In thinking this option through SURFnet might consider offering connectivity to the site of a selected provider.

It was remarked that backup services and disaster recovery services are different things. After discussing for quite a bit it was concluded that although interesting disaster recovery is not difficult in itself: solutions are available, it can be done and it scales very well with the amount of money fed to it. Exchanging knowledge and experiences on the topic would be interesting but no other collaborative work items could as yet be identified.

d) Distributed storage infrastructure middleware

Middleware is one of several possible solutions in managing geographically distributed storage.

Andreas Landhäußer said it was interesting for T-Systems to look into metadata: data stored does not have description so it becomes quickly unusable: after a while you can find the data but you do not know anymore what it is or is used for, so one needs to enrich original data with metadata. It was remarked that SARA in Amsterdam should have some projects dealing with enriched video data. Rutherford labs and UNINETT were also very interested in this.

Options for common work in this area:

- In Norway next year will work on evaluating storage middlewares.
- Rutherford Lab: would be useful to have an understanding of what single pieces of middleware can do.
- Lars thinks there is a case for a short-time working group, to catalogue experiences and create a taxonomy of the different middlewares; the NDGF did not volunteer to do this but volunteered to describe what they are doing; it was also suggested to list the projects and people who are working in this.

Michael Gronager (NDGF) and David Corney (Rutherford Labs) agreed to steer work in this area. David could definitely speak about that, possibly also via video link with other world regions. TERENA will make a wiki space available for the work.

e) Best practices, technology watch, product evaluations.

This was found a very interesting area from many participants; information should be on the wiki, but people should be active in providing the source material. David was interested in evaluating various materials: he mentioned a new tape writer from SUN, various pieces of software etc.

The technology watch potential was found very useful for the Norwegian Higher Education community.

David agreed to take responsibility for this area.

f) **Power consumption.**

SURFnet and UNINETT were interested in listing infrastructure stuff and monitoring power consumption. This was also one of the main motivations in looking at virtualisation.

5. Next Meeting

It was agreed the next meeting would be co-located with the NORDUNet conference in Helsinki on Tuesday 8 April 2008.

It was suggested to invite people to prepare presentations. David said it would be interesting to present the range of services they provide in the UK; The University of Ghent volunteered to present their work related to protocols in media environment, including long-distance performance measurement.

6. Actions

20071214-01 TERENA make wiki space available for the storage group.

20071214-02 Valentino Cavalli to subscribe all meeting attendees to storage email list DONE.

7. List of Registered Participants

Henny Bekker	SURFnet
Brian Boyle	HEAnet
Maciej Brzezniak	PSNC
Jim Buddin	TERENA
Paulo Carvalho	Universidade do Porto
Valentino Cavalli	TERENA
David Corney	STFC - Rutherford Laboratory
Paul Dekkers	SURFnet
Stijn Eeckhaut	IBBT - Ghent University
Michael Gronager	NDGF and NORDUnet A/S
Avgust Jauk	ARNES
Saethor L. Jonsson	RHnet - University of Iceland
Ulrich Kaehler	DFN
Kostas Koumantaros	GRNET
Kaspars Krampis	IMCS UL
Andreas Landhäuser	T-Systems Solutions for Research GmbH
Dana Ludviga	IMCS UL
Ari Lukkarinen	CSC - Scientific computing Ltd
Jan Meijer	UNINETT
Jari Miettinen	CSC/Funet
Zsombor Nagy	NIIF (HUNGARNET)
Cándido Rodríguez	RedIRIS

Miroslav Ruda	CESNET
Magnus Strømdal	UNINETT
Niels van Dijk	SURFnet
Brecht Vermeulen	IBBT - Ghent University
Christoph Witzig	SWITCH