

Feasibility study to verify key technical and operational aspects of a proposed European Object Storage architecture

GN5-2 - GÉANT Project Above-the-Net Services (WP4 T4)

Incubator run by PSNC, CSC, SURF, DeIC and SIKT



GÉANT GN5-2 WP4 T4 Incubator

- SAGE - Sovereign Architecture for Geo-distributed European Object Storage
- Participating institutions:
 - long-running cloud computing & storage providers, NRENs and academic institutions
 - deal with 100PB+ -scale data infrastructure
 - deal with physics (the prose of the data storage)



Why Ceph? (positioning, adoption)

- Ceph has a large community that owns the product - it is a multi vendor project
 - more stable and a safer choice for long term projects than any other alternatives including “open source” ones
- Ceph can serve different purposes including:
 - block storage for cloud: OpenStack, Kubernetes, Proxmox
 - file storage: CephFS + cloud orchestration (Manila)
 - NFS through Ganesha
 - **object storage (with S3 and Swift interface)**
- It is the software defined storage project:
 - can implement aggressive erasure coding (good storage efficiency, while high reliability)
- We all run Ceph (major cloud/HPC sites, NRENs..)
 - incubator participants operate Ceph since 15+ years
 - Ceph might be run at your sites!



Why Ceph S3?

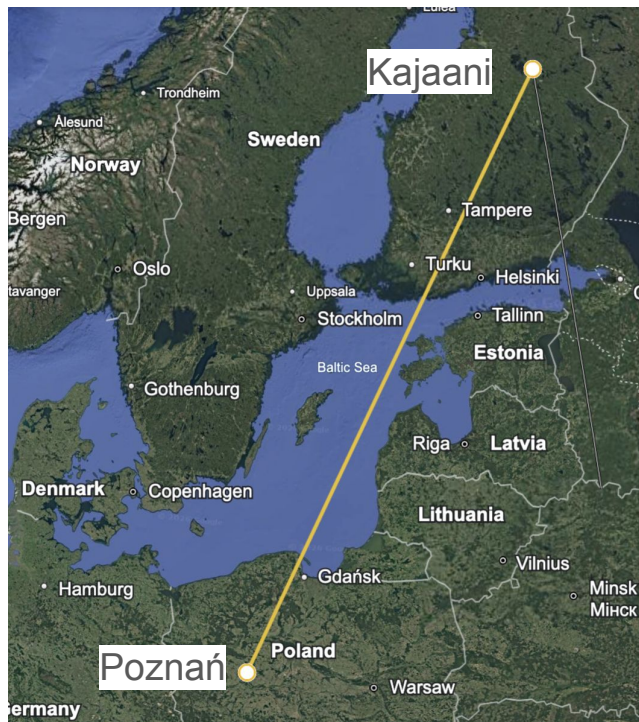
- The most feature-complete open source S3 API implementation
 - Bucket Versioning
 - Object Lock
 - Glacier like cold storage
 - IAM user management
 - **Multisite replication**
- All serious free S3 alternatives are single vendor “open source” projects vulnerable to Minio/ElasticSearch-style “commercialization”



Testing scope: our replication scenarios



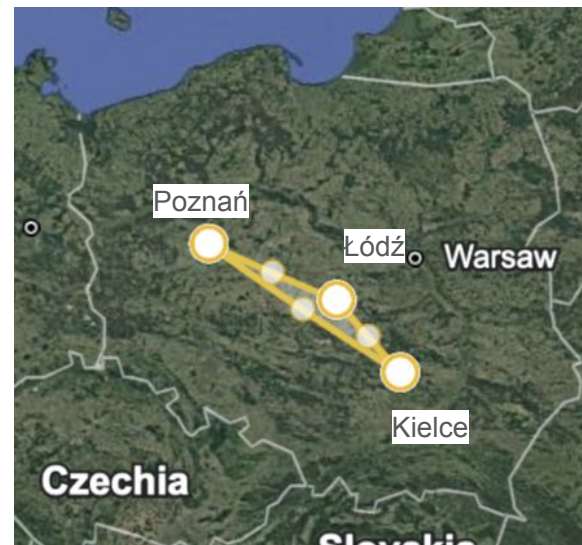
GN5-2 WP4 T4
SAGE incubator



S3 replication:
PSNC<->CSC
Poland<->Finland
distance ~1500 km
Latency ~32 ms



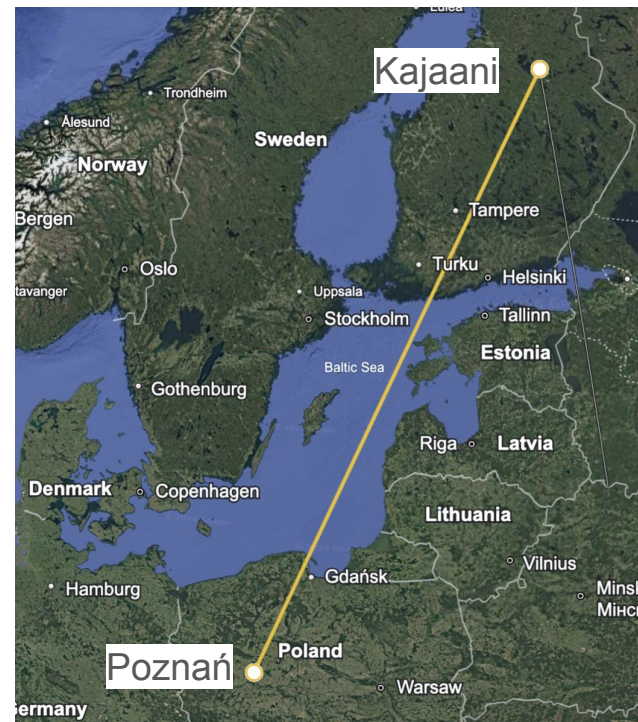
National Data Storage
project technology tests



S3 replication: 3 sites in PL
Distance: 180-300 km
Latency: ~9 ms

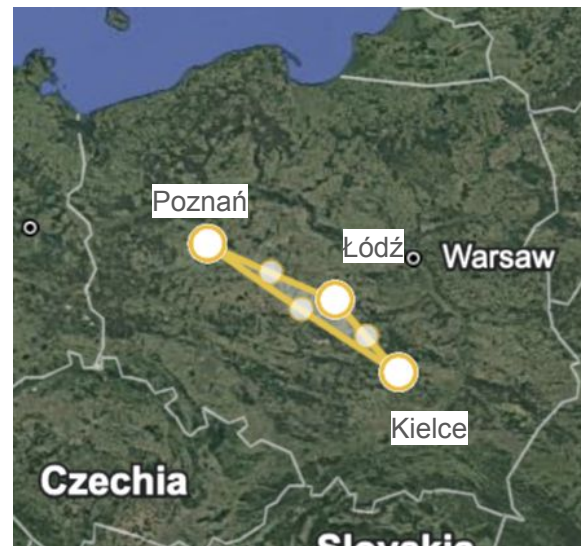
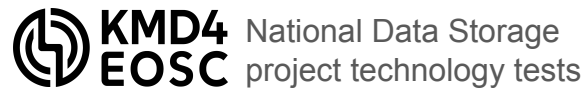
GN5-2 incubator case: cross-border replication)

- Test environment: GN5-2 (WP4 T4) sites:
 - now PSNC and CSC,
 - SURF and DeiC in future
- Current tests: active-active replication among PSNC and CSC
- Long-distance links
- Replication over public internet with TLS encryption
 - Worst case scenario with latency and encryption overhead
 - However, good asynchronous replication should handle it



KMD4*) case: in-country replication, multiple islands

- Multiple “replication islands”
 - 2-site replication islands
 - 3-site replication islands
 - we are considering several variants for production implementation
 - possibly having multiple “islands” in the country
 - such approach could be elevated to the European scale
- Again: replication over public internet with TLS
 - could use dedicated channels in future - it could complicate the setup but provide more performance and reliability



* KMD4 = KMD4EOS - National Data Storage project in Poland (4th edition)

What did we learn - test results (technical)

- Replication within Poland:
 - Works OK in normal “production-like” use cases
 - Replication fails in simple but demanding “synthetic” tests
- Cross-border replication:
 - Almost the same results as within Poland
 - Works OK in normal “production like” use cases
 - Replication fails in simple but demanding synthetic tests
- Immediate use-cases:
 - Proxmox cluster backup system using Restic
 - OpenStack cluster backup
 - Rclone S3 to S3 synchronisation

What did we learn - Ceph S3 (still technical)

- Complicated and difficult to debug
- Documentation is lacking
- Not designed for cross organisational deployments
- Replication is scalable but doesn't guarantee consistency (yet)
- Detection and “fixing” mechanics needed
- Validating data synchronization may be difficult for active buckets

What did we learn? - organisation-level

- RGW is not designed for cross organisational deployments
 - There is some level of trust assumed among sites
 - If you control site A you can probably affect site B
 - Effective RGW debugging requires direct SSH access to other site's ceph cluster - debugging over chat isn't efficient
- Agreeing on common architecture for different institution/teams can be challenging
- => extra security measures needed
- => relevant operational coordination needed

Alternatives?

- Looking into generic S3 replications mechanisms
 - a layer between client and S3 server
 - glueing together different S3 implementations into single replication island

Summary

- Mid-term report for the incubator
 - prepared, final report in 3Q2026
- We might share the final report along with the technical details



GN5-2 - GÉANT Project Above-the-Net Services Incubator
(WP4 T4)

Feasibility study to verify key technical and operational aspects of a proposed European Object Storage architecture

Validating key technical aspects of the proposed European Object Storage architecture

April 2026

PSNC, CSC, DeIC, SURF

Credits

Adam Prycki - PSNC
Eugeniusz Pokora - PSNC
Maciej Brzeźniak - PSNC
Kalle Happonen - CSC
Joni Junni - CSC
Ron Trompert - SURF
Jean-Marie de Boer - SURF
Morten Kjeldgaard - DeIC
Bo Bai - DeIC
Jan Meijer - SIKT

