Long Term Archives technology trends and evolution

D. Iozzino (Rhea for ESA) PV 2023 May 2–4, 2023 - CERN

Executive summary

Snapshot of surveys on long term archive sessions addressing:

- Archive infrastructure and technology
- Monitoring tools
- Archive volume, missions, coverage
- Archiving flows and processes
- Data format/packaging for long term archive
- Management of the relevant associated information, documentation, software and knowledge
- Trends and future plans, main issues and challenges

Challenges collection The main challenge of the archive is the sheer volume of data, the copious number of files, the diversity of formats and the yearly increase of data produced.

The archive is constantly enhanced to facilitate the archival process by using technological innovation and practices. To save energy, thus making the archive and its hardware environmentally friendly, green computing technologies are favorable.

ESA EO Archives

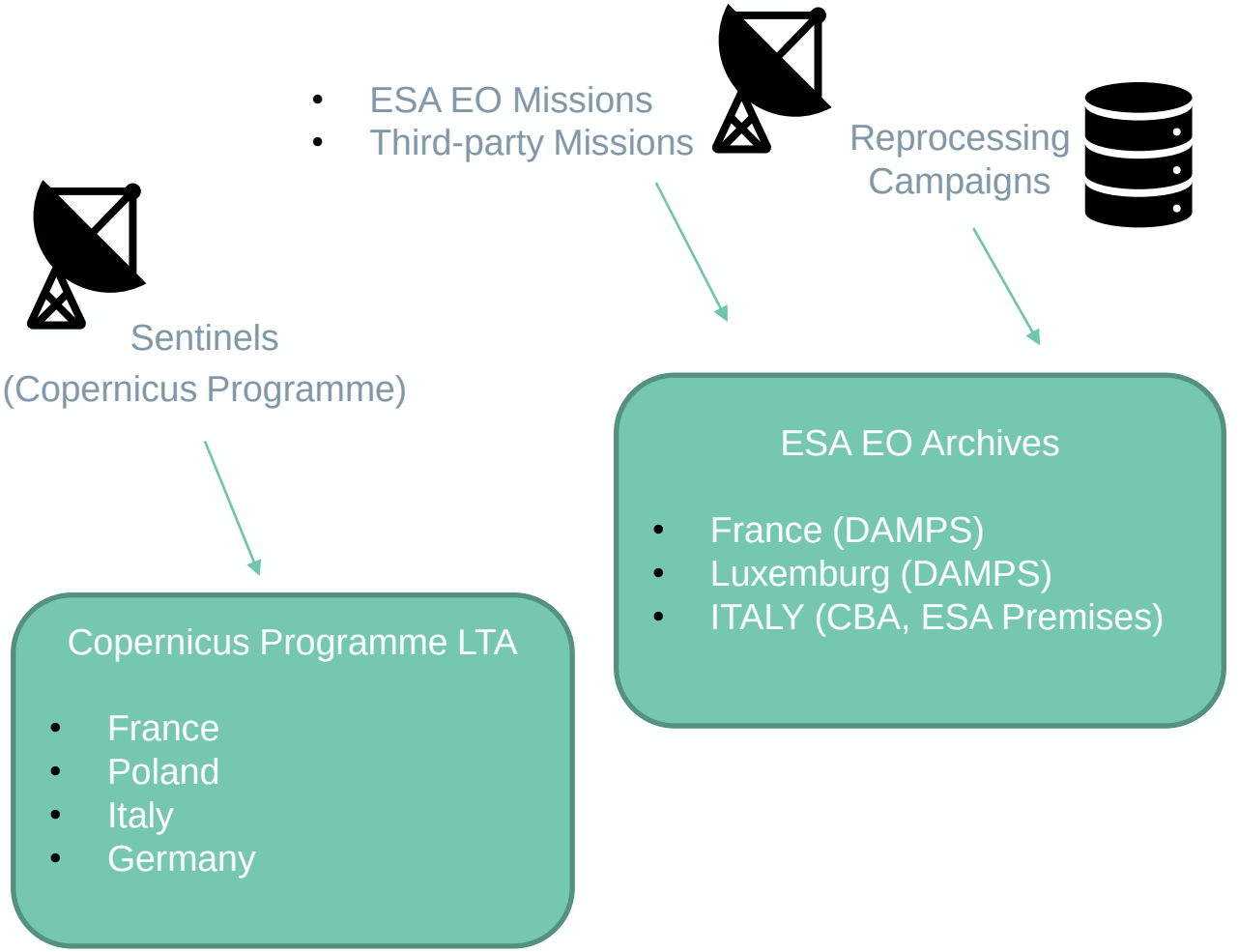
Dedicated Sentinels LTA using different technology:

- Cloud
- Disk
- Tape
- Optical medium

ESA EO Archive, two archives, two copies of the data each:

- Master Archive (DAMPS – External Service)
- Space Data Preservation Archive (ESA/ESRIN, Cold Back-up)

Dissemination to users through Cloud based dissemination services



Space Data Preservation Archive

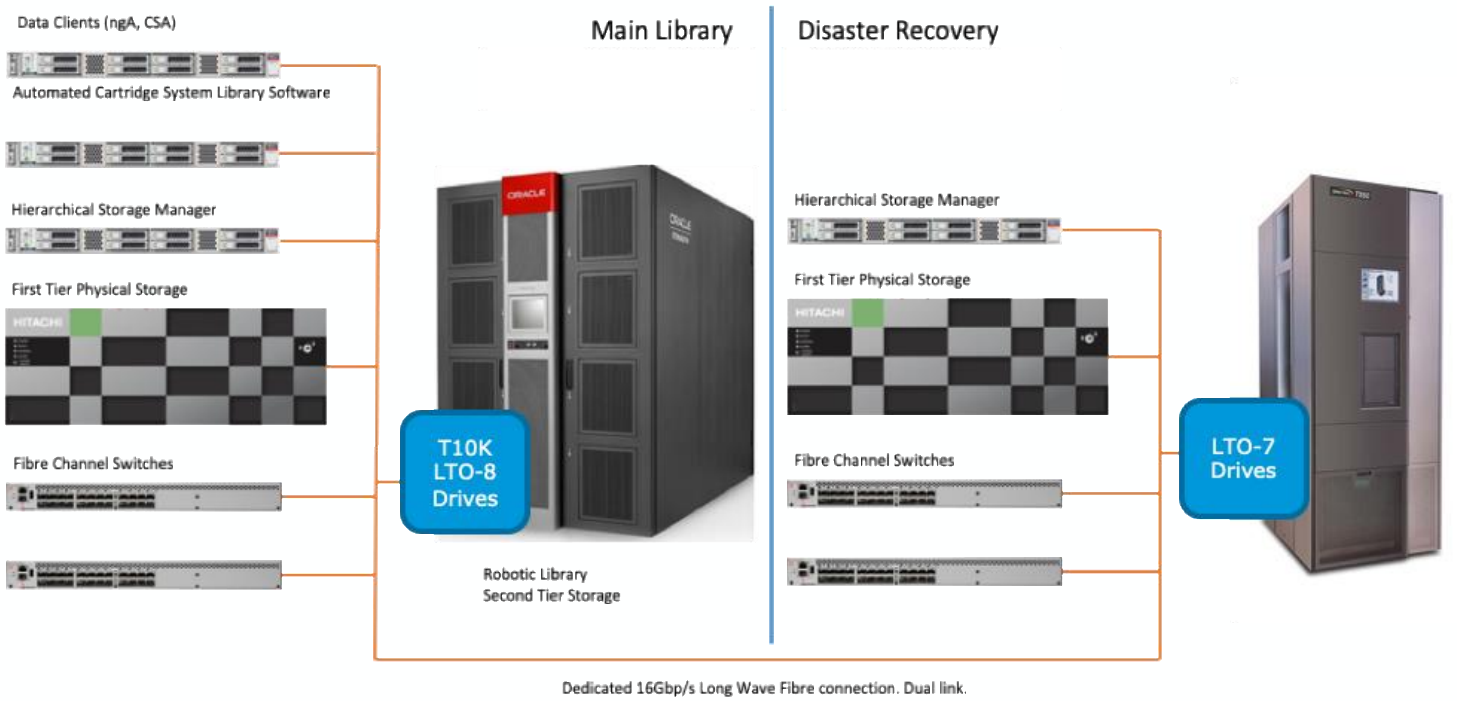
The Space Data Preservation Archive is a data archive deployed at ESA/ESRIN premises to allow seamless archiving and extraction of data. The archive is the front-end of the following long term archiving services:

- Second copy of the master archive
- Long Term Archive of ESA EO datasets not part of the “Master” suite (e.g. unconsolidated data, flagged data)
- Unique RAW data
- Direct connection to missions ground segment
- Direct ingestion of ESA and third-party (TPM) Live Missions
- Disaster Recovery of data from other ESA directorates (ESA-SCI, ESA-HRE, ESA-OPS) using a specific front-end circulator

The main challenge of the archive is the sheer volume of data and the diversity of formats. The archive is operative 24/7 and receives automatically live data acquired by ESA and TPM live EO missions.

Archive Infrastructure

- Main tape library capable of storing 36 PB of uncompressed near-line data
- Incoming data written on fast SSD physical storage
- Data eventually written on magnetic tapes
- Files seen by clients on a virtual file-system
- Data duplicated over a disaster recovery tape library
- Second copy tapes moved to an external site



What do we preserve

A vast variety of data:

- Native format
- Auxiliary data
- Documentation
- Binary files

However, the favorite format is the EO Submission Information Package (EO-SIP), a ZIP archive that includes additional data:

- When received it’s not just copied but metadata is extracted
- Different workflow applies
- Data is ingested automatically
- Half of the archive is EO-SIP, more data will be produced in EO-SIP in the future

We mostly archive data, we rarely extract data, data is written once, read rarely.

Archive requirements

- Seamless archiving of ESA EO live missions (24/7)
- Quick bulk ingestion of reprocessing campaigns
- Data can be near-line or off-line
- Data is only bulk extracted
- No access to users
- Power efficient solution to be favorite
- Data duplicated
- Archive capable of handling increase of data produced (5PB increase estimated in 2024)

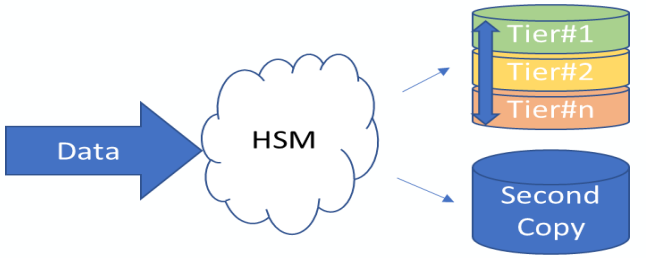
Hierarchical storage management

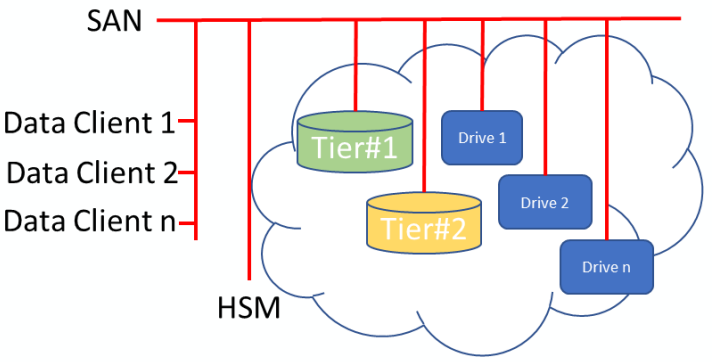
The archive front-end is a hierarchical storage management system with data moved into storage tiers depending on user defined rules. In our case, final tier is always tapes of the two tape libraries. With the data always ending onto tapes, the archive has no actual size limitation.

The tiers are seen by data clients as a single logical archive location and can be accessed using classical network protocols (NFS, Samba) or using a specific client where metadata is mounted over network but any Read and Write operations are routed through 16gbp/s fibre optic.

To do so, the client needs to be in the same storage area network of the HSM and the tape library drives.

Clients therefore can list and read the data no matter what tier it is, in the background, the HSM moves the data across the tiers accordingly.





Archive figures

Average EO data ingested/day: 13 TB

Peak EO data read/day: 20 TB

EO data volume (Mar 2023)	EO number of files (Mar 2023)
17,3 Petabyte	220,000,000

Archive recently used as destination of Landsat-8 full reprocessing. Data bulk processed was sent as fast as possible.

17 TB of additional data was received daily for 12 months.

Future ESA EO mission will increase daily data received considerably.

Mission	Uncons.	Cons.	Native	EO-SIP	LIVE
ERS-1/2	X	X	X	X	
Envisat	X	X	X	X	
Cryosat-2	X	x	X	X	X
SMOS	X	x	X		X
SWARM	X		X		X
AEOLUS				X	x
Goce	X	X	X	X	
Oceansat	x		x	X	x
Adeos	X		X		
ALOS	X	X	X	X	X
GOSAT	X		X		X
Ikonos-2	X		X	X	
IRS-P3	X	X	X	X	
JERS-1	X	X	X	X	
Kompsat-1	X		X		
Landsat-1,8	X	X	X	X	X
MOS-1,1b	X	X	X		
Nimbus	X		X		
NOAA-7,19	X	X	X	X	
ODIN	X		X		X
PROBA-1	X		X		X
QuickSCAT	X		X		
Rapideye	X		X		
Scisat	X		X		X
Seasat	X	x	X	X	
SeaStar	X		X		
SPOT1-2	X	X	X	X	
TerraSAR	X		X		
WorldView-1,3	X		X	X	
Windsat	X		X		

Main challenges observed

- Lots of I/O affects performances due to the file-system being constantly updated
- Performances are still scarce for small files even with file-system on separate partition/RAID
- Media failures recovery takes time and uses resources (drives, storage)
- Lots of false media failures alarms (caused by external factors)
- Current archive manager forces us being vendor (and OS) locked
- Support tickets requires lots of data collection even for obvious failures
- Support tickets often result in requests to update firmware
- Firmware updates are complicated time consuming
- Current HSM solution is cheap compared to alternatives

Main challenges observed (2)

- Archive manager migration complicated, time consuming, expensive
- Longevity of technology has side-effects (maintenance costs, obsolescence of HW)
- Tape prices are not predictable. Prices of previous generation tapes used in archive doubled when its vendor declared the end of life of the tape family. LTO-8 tape shortage in mid 2019 due to patent infringement battle meant LTO-8 tapes where unavailable for a long period of time
- With the end of some tape families and HSM archive managers, it is feared that the main vendors are moving away from tape archive solutions
- We are ready for lots of data, less ready for billions of files
- Extracting data from tapes is slow and complicated

Scarce performances for small files

Due to lots of I/O necessary to update the file-system metadata, performances when archiving small files are scarce. The archive has been updated to cope with such limitation and the metadata has got a dedicated partition on two mirrored SSD disks.

Performances are still scarce and were recently challenged by the bulk ingestion of ESAC Scientific Archive. The Archive has ~17 PB of data for a total of 220 Millions of files. Only the first three NAS received contained ~200 TB but over 1 Billion of files.

Consequences of such high number of files have been assessed:

- Degraded file listing/queries
- iNodes (file-system metadata) are reaching 500 GB of size
- Increased time to dump/backup the iNodes

Huge time required to copy the data from NAS into Archive:

- 300,000 files per day per NAS devices
- Maximum of 2 NAS copied in parallel so far
- 4 NAS in parallel may impact archiving operations requiring more than one drive dedicated to the activity
- ~4 years to complete the activity and copy the data into the current archive with 2 NAS in parallel

Media failures

The Main tape library has 8 drives:

- 5x STK T10000D drives
- 3x LTO-8 drives (3 additional drives are being purchased)
- 1 LTO drive is dedicated to live data archival
- 1 LTO drive is dedicated to bulk ingestion of reprocessing campaigns
- 1 LTO drive is used according to extraction and/or additional live or bulk ingestion workload

When a tape is declared faulty, data archived onto the tape needs to be read and copied onto a new tape. This requires time, is not achievable automatically, uses two drives (one reading, one archiving).

Due to external factors (network unavailability, fiber optic unavailability) tapes may be declared faulty even if fully functioning and the root cause for the false alarm is not always evident.

Avoid vendor locked solutions

The current HSM used in the archive, requires some software pre-requisites:

- Archive manager running specific operating system (not open source)
- System library software running on specific operating system (not open source)
- Client module used by data clients only running on old release of Linux dated 2017
- NFS is used as a workaround meaning that the common 10G Network is used instead of the dedicated 16GBp/s fibre optic
- Alternative archive managers are either not compatible with file-system metadata (data needs to be read from tapes) or have got “complicated” pricing schemes
- Is migrating data onto new tapes considered archiving ? How about the initial population of the archive ? This is a key factor when choosing replacement archive manager

Longevity vs obsolescence

The main tape library of the archive is an enterprise-class robotic tape library released in ~2005. To this date the library is still being sold and its end-of-file has not been reached nor declared.

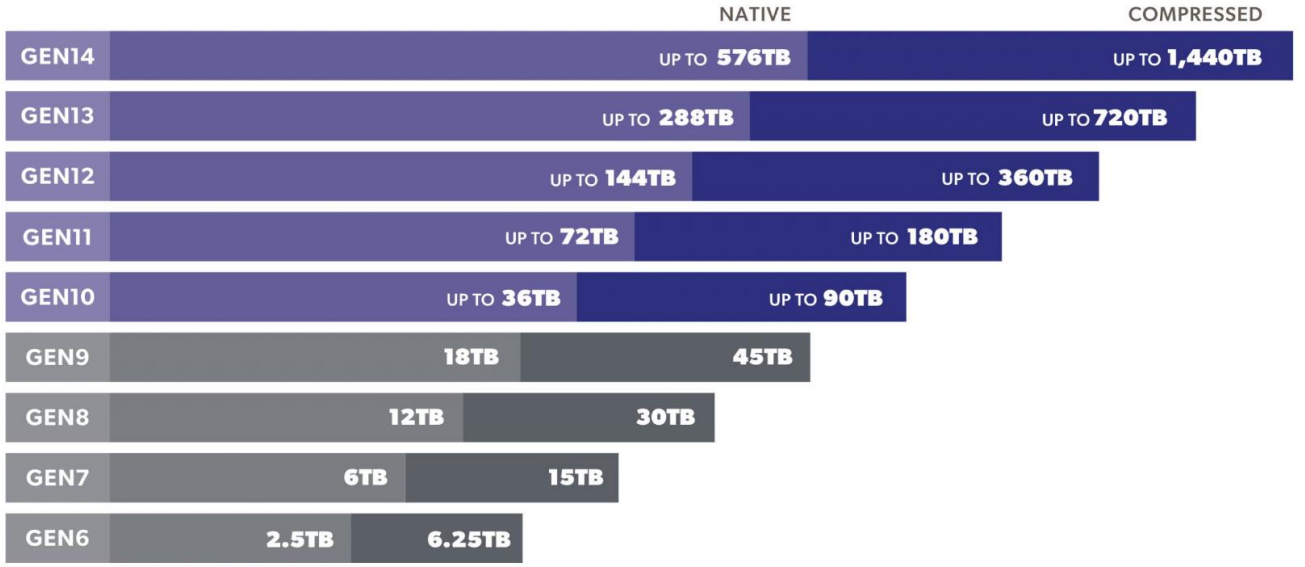
Even if the Library has been updated over the last decade, most of the HW is dated and requires regular support.

Requesting support became somehow complicated over the years, sometime time consuming even for apparent obvious failures:

- Drive firmware updated no longer under support responsibility
- Specific vendor monitoring tools are often required to gather logs
- “Did you turn it off and on again” approach with firmware upgrade required before even assessing the issue
- Logs/dumps are often complicated to gather
- It often takes more than 4 days to have a field engineer dispatched for faulty part replacement
- Firmware updates and replacement drives with recent firmware means that all other drive needs to be updated to keep drives coherent

Advantages of Tape Libraries

- Easy to expand
- Power efficient
- Non enterprise tapes are cheap
- Tapes evolution are on par with disk evolution
- Malicious encryption of data is just not possible
- Malicious deletion of data is not achievable



Archive Upgrades

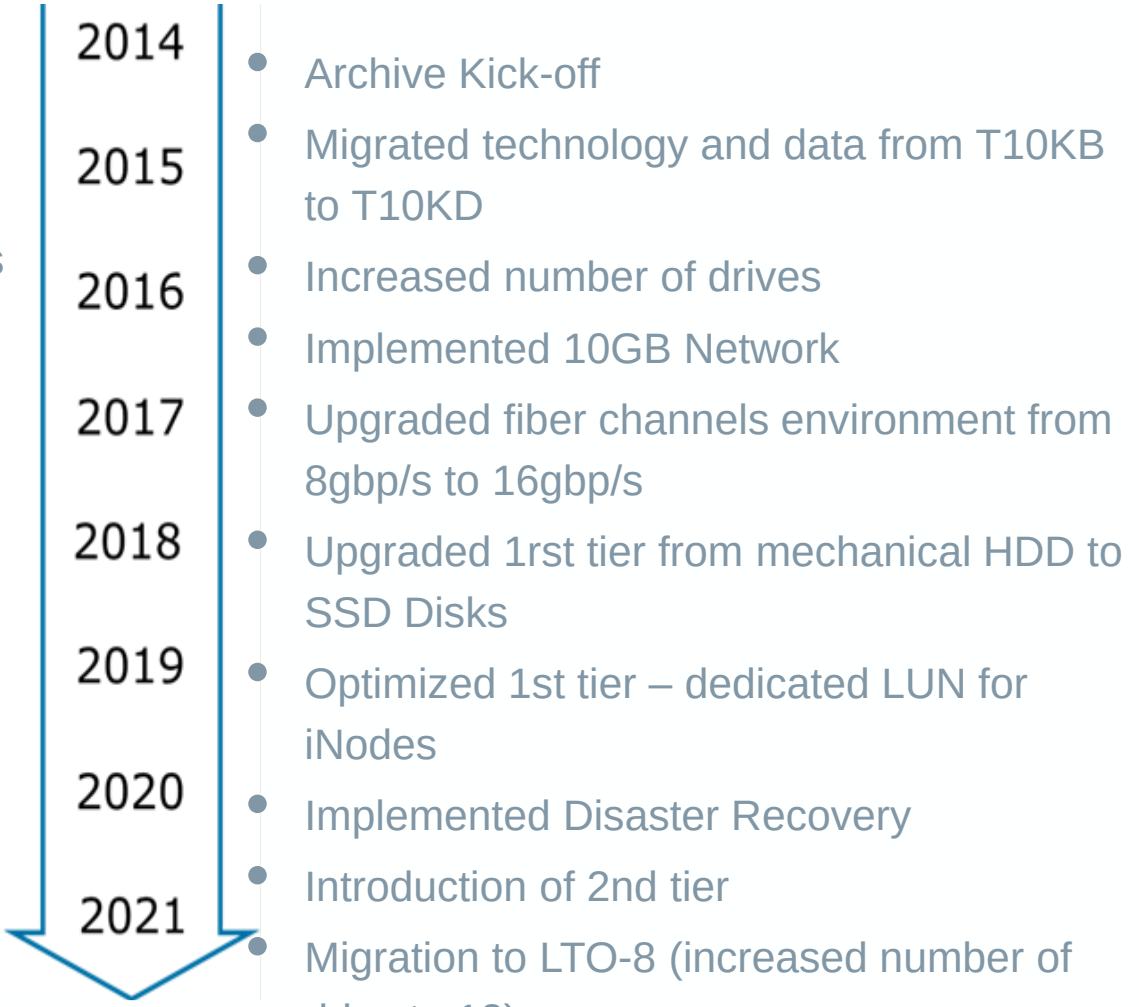
To keep up with the always growing amount and variety of data produced, the archive is constantly enhanced to facilitate the archival process.

Over the course of the last 10 years, several HW upgrades have been performed to the core of the Archive, the tape library.

Future upgrades shall consider technological innovation and practices and, to save energy thus making the archive and its hardware environmentally friendly, green computing technologies shall be favorite.

Evolution roadmap has been drawn to address current and future challenges and reach several goals:

- Better ingestion and extraction performances
- Wider the audience using the archive
- Decrease the power consumption of the whole environment
- Use of innovative technology and archiving trends



Evolution requirements

- Study made to help on a possible evolution of ESA ESRIN Archives
- Archive with current or future technologies
- Analysis on hardware, software and cloud services
- Review of other ESA and external agencies tools
- Environment-friendly approach
- Must guarantee the correct preservation of the data regardless of the medium used

Approach

Market survey of:

- Hardware (Tapes, Folio, Glass Drives, Disk Storage, OS Orchestrators, etc.)
- Software (Alluxio, Atempo, Versity, MinIO, LABDRIVE, nageruHive, etc.)
- Cloud Providers survey (on prem, public offers, etc.)
- Standards
- Existing ESA Tools (Copernicus LTA, Planetary Science Archive, etc.)
- Other Organization (NASA, CERN, NOAA – NCEI, Eumetsat, etc.)
- Sustainability (Energy consumption, Cooling Requirements, etc.)
- Costs (HW and SW)

Initial Key findings

- There are solutions to keep using current ESA/ESRIN storage technologies. You don't need to change the entire infrastructure because of HSM end of life
- Hardware providers are developing hot, warm and cold storage solutions that can complement the current infrastructure and, defining a roadmap, could substitute it in the long-term
- Some WORM storage options are already available or being developed that are based on nonmagnetic media, like optical disks or crystal based. There are some that could help ESA on the long-term perspective without needing to migrate as the media lasts for a longer time
- More and more, organizations are moving to cloud-based services because of their flexibility and cost effectiveness in the short-term. However, the step to cloud needs to be planned because long-term costs and vendor locking could be an issue in the future

Initial Key findings (2)

Complementary to pure cloud solutions, the hybrid-cloud approach suggestable. A mix between public access on the cloud plus cold on-premise storage would benefit in the long-term to address purecloud flaws

The preservation software market is very active and growing. In addition to the classic providers in this area, some new players are showing interesting solutions also with a very open-minded approach to avoid extra costs or cloud locking. SaaS is a trend on this market but private cloud solutions are also available.

Security and sustainability are also worries in this market. Many providers are showing some environmentally friendly messages but it is not clear the extent of such statements.

OAIS is the standard for preservation and needs to be followed. In addition, other standards could improve ESA/ESRIN archive operations in different directions.

Big cloud providers are charging for requests or operations which makes it almost impossible to make a budget in advance. Some cloud providers are not charging for the recovery or the model is more manageable.

Upgrade Roadmap

- Review the tape library status
- Replacement needed ?
- If kept, new management infrastructure should be acquired to replace current HSM. There are several options to do this
- Decide which cloud services are required (if any)
- Check if some hybrid-cloud approach could be feasible
- Define which standards need to be followed
- Prepare requirements for the systems to follow and checklists for providers analysis
- Evaluate archiving software alternatives between those existing on the market or build a proprietary solution
- Calculate accurately the current energy consumption and heating dissipation needs for the current solution
- Open RFIs with selected providers to get a deeper technical information (and pricing)

Conclusions

After analyzing the market, the current situation of the archive and the reference guidelines the best solution for updating the archive is a hybrid cloud model, which integrates the current tape storage infrastructure with a flexible model that facilitates access and improves its online capabilities.

The current tape library work-life has been considered sufficient to cover the needs of the coming years and therefore, the recommendation is to continue using it for some years now. The archive manager shall be replaced by any of the identified alternatives or by an integration into cloud services.

One understated advantage of continuing to use tape-based archives is the always increasing security concerns. Malicious encryption or deletion of the data is not possible or highly complicated. Moreover, to reduce energy consumption and carbon emissions using tape storage instead of hard disk is an effective solution.

ESA is defining a roadmap for the evolution of its ESRIN archive. The roadmap is the first step in the direction of assuring seamless archival, extraction, valorization, and preservation of the always increasing ESA and Third-Party Missions EO data records and associated information.