WGISS-54 ESA EO Archive Technology and Future Trends

Daniele Iozzino, ESA

Agenda ID: 2022.10.04_15.15

WGISS-54

Tokyo, Japan (JAXA)

3-7 October 2022

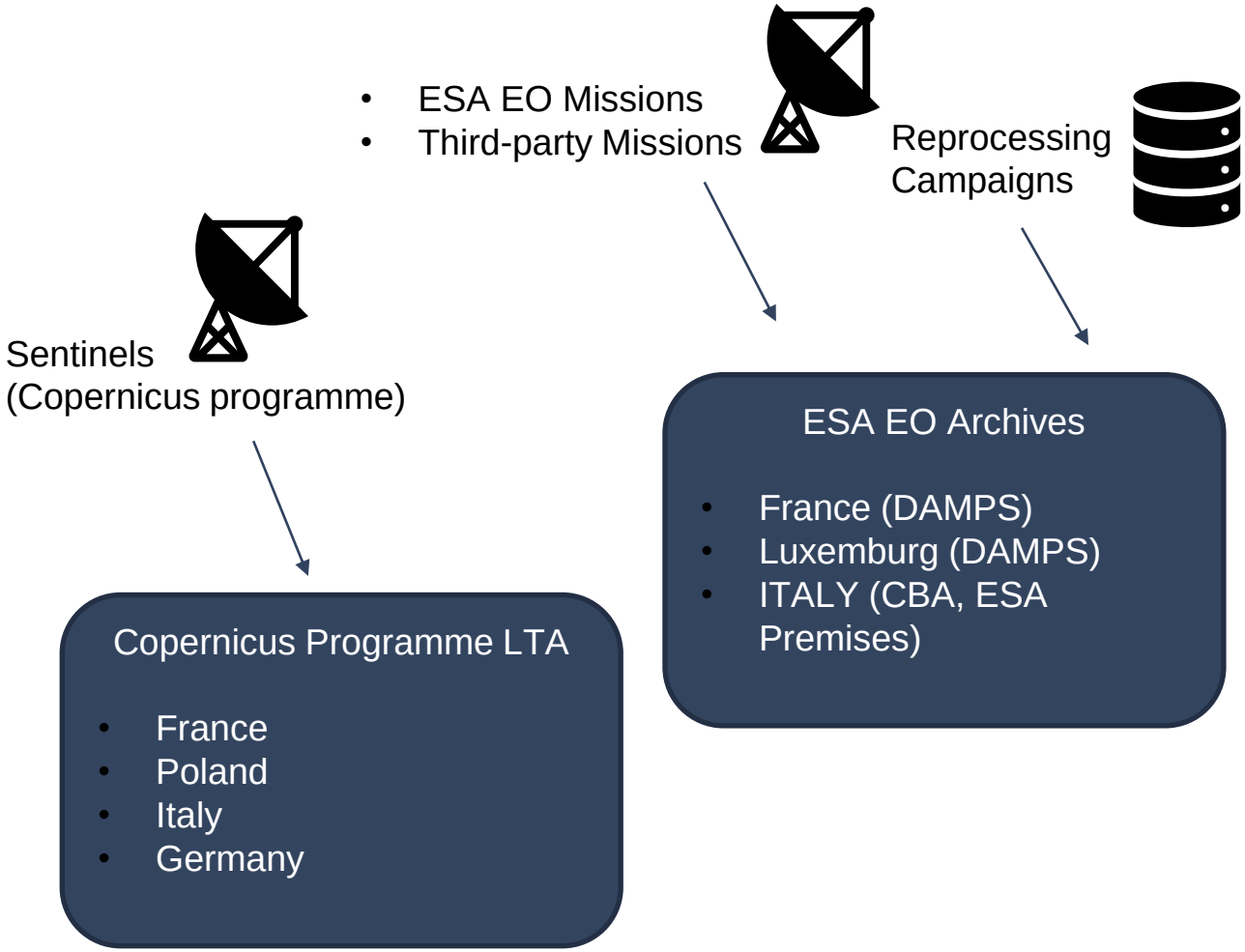
CEOS
Committee on
Earth Observation Satellites

Executive summary

- Current archiving technology and infrastructure
- Archive volume detailed per mission
- Archiving operations concept, flows and processes
- Data format/packaging for long term archive
- Management and archiving of the relevant associated information
- Archive technology and media evolution and future trends
- Challenges and Needs
- AoB

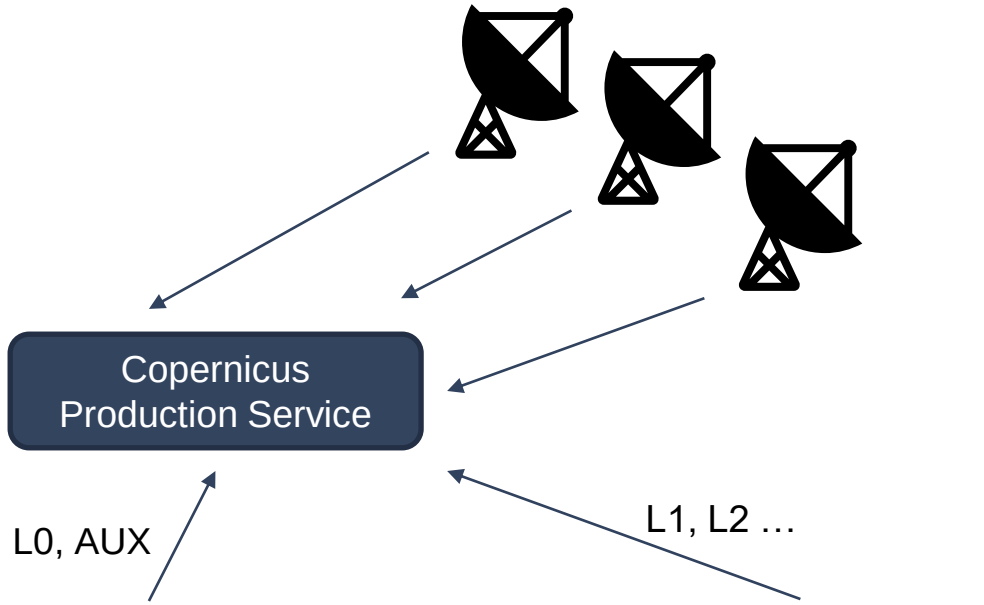
ESA EO Archives

- 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



Copernicus programme Archives

- LTA takes the Data from Production Service
- Four centers, different technology
- Rest of ESA EO Archives on different programme, detailed in the presentation



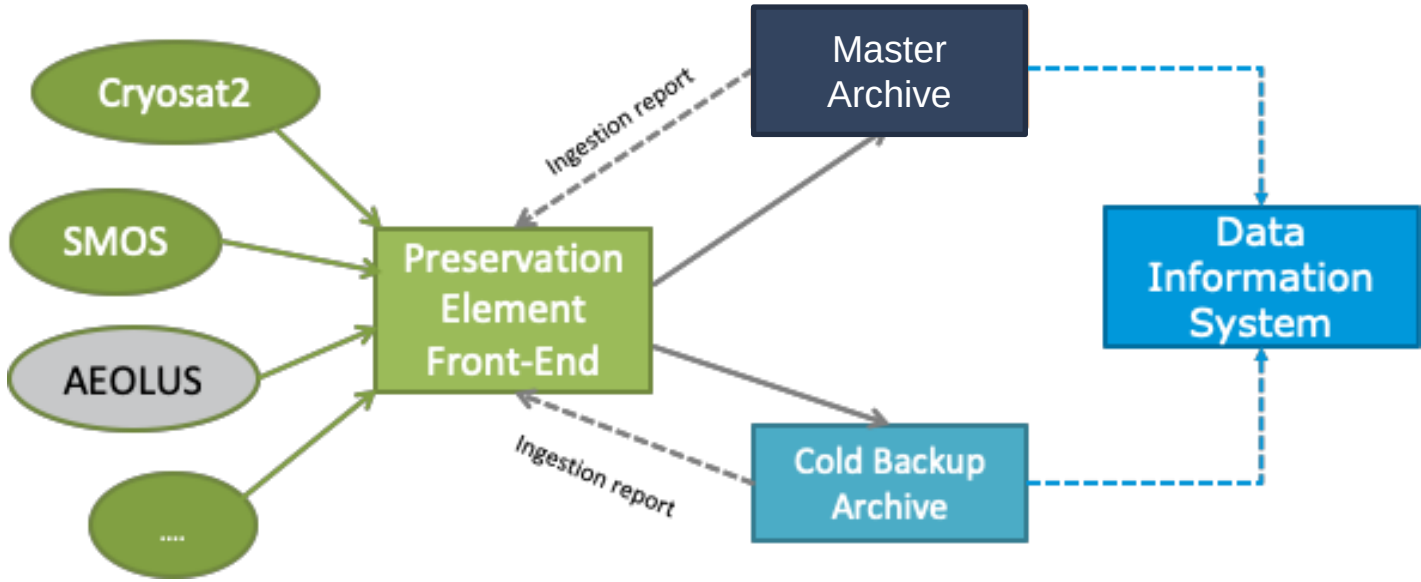
Location of LTA	Technology
Sophia Antipolis, France	Disk/Tape Backup
Warsaw, Poland	Cloud
Palermo, Italy	Optical Medium
Frankfurt, Germany	Cloud/Tape Backup

Circulation of Data

ESA and Third-Party active mission data are circulated by the Preservation Element Front-End (PE- FE).

Once the data has been ingested and validated, confirmation reports are sent to the Mission Payload Data Ground Segment via standard network transfer protocol.

Bulk dissemination of data coming from reprocessing campaigns is circulated on storage devices.

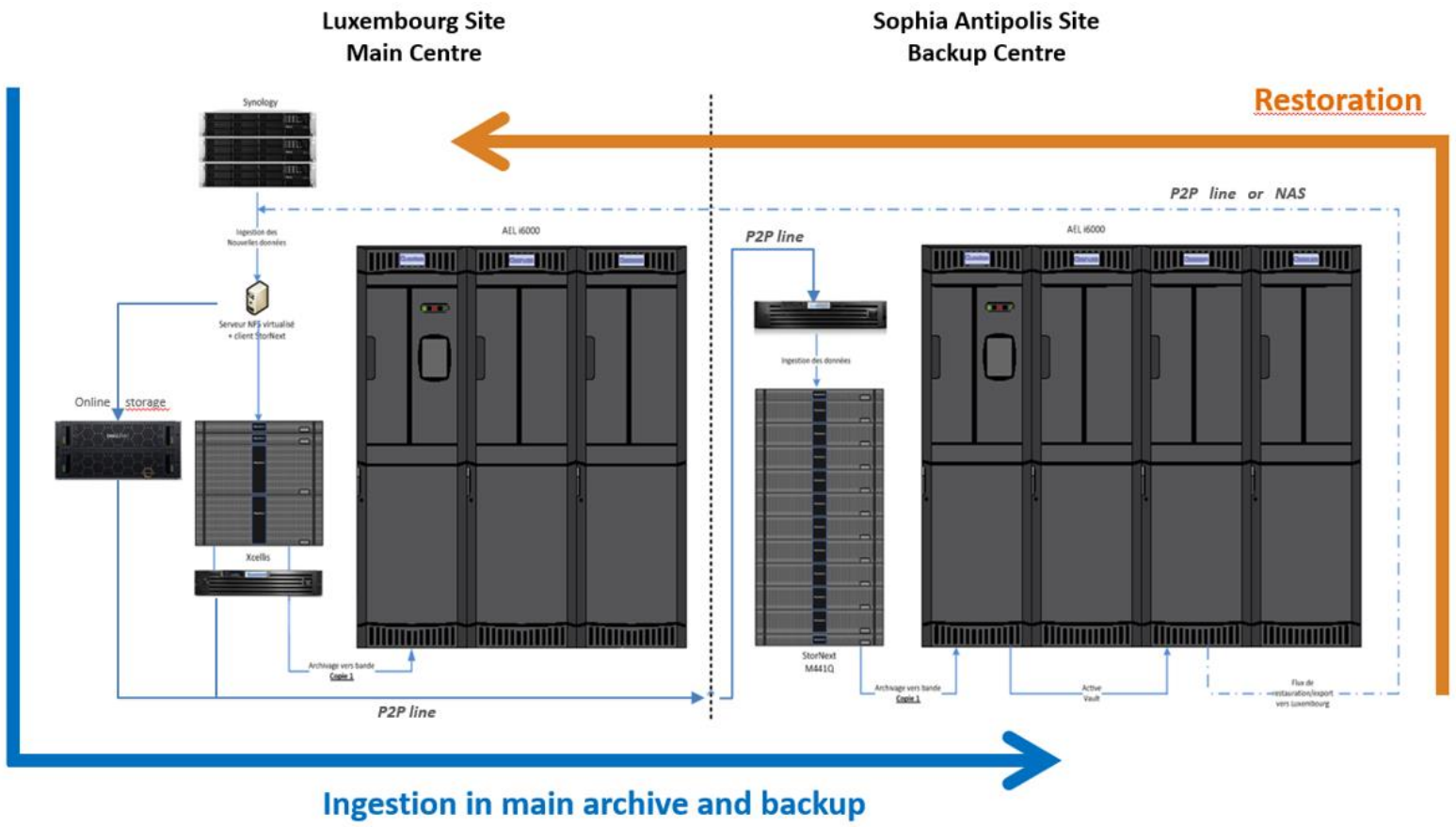


ESA EO Master Archive

- ESA EO Master Archive - PDGS Data Archival, Management & Processing Services (DAMPS)
  - Implements the ESA Master Archive through a dedicated service awarded to industry through an open ITT, contract Kicked-off in 2016
  - Two copies of the data held > 200 Km apart (France, Luxemburg)
  - Connection to the ESA EO WAN
  - Checks of the data at all stages from ingestion to delivery
  - Scalability required for growth of contents
  - Management of the data information (MetaData) in fully redundant Databases

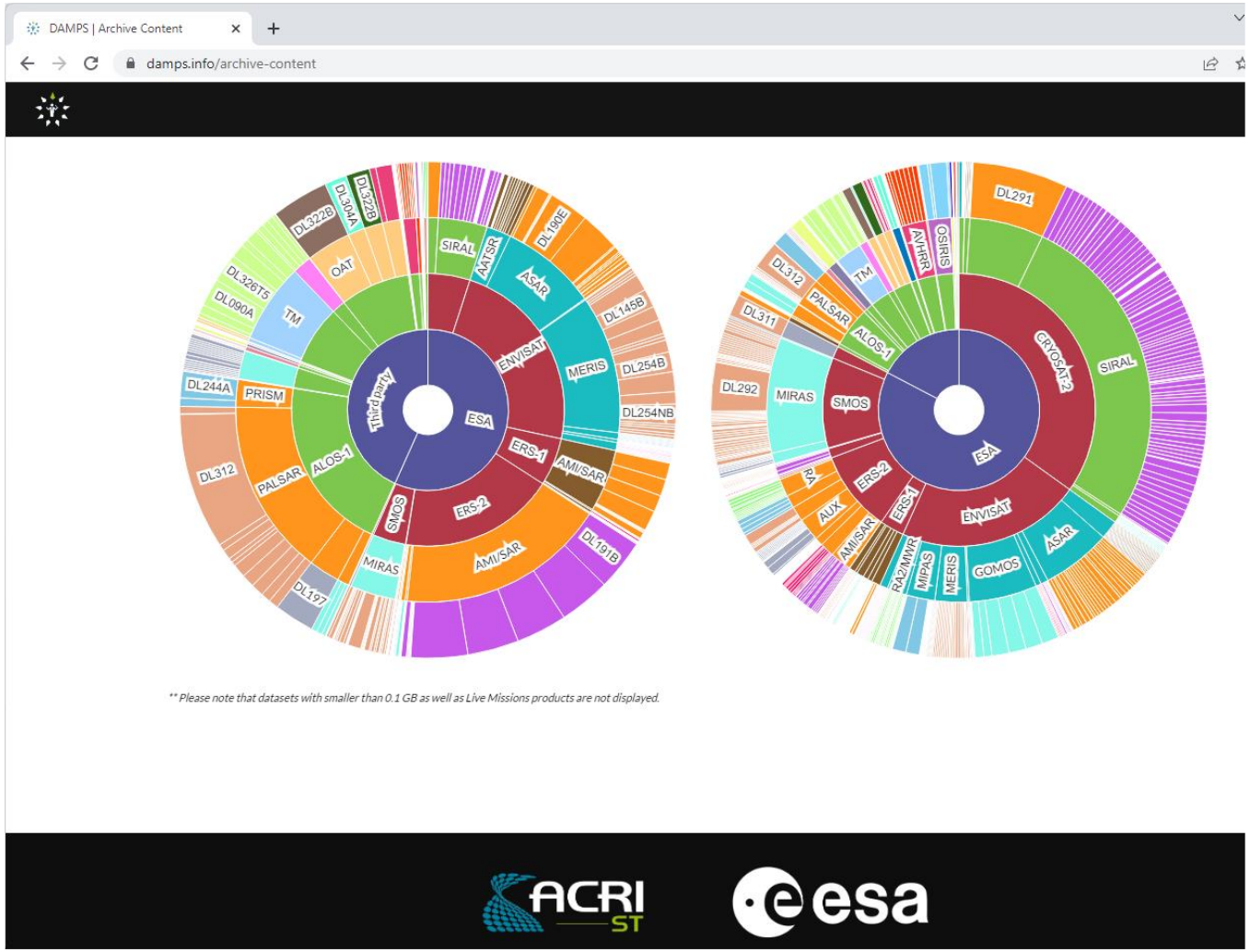
Duplication of data

- Incoming data is only deleted from inbox basket once the Backup Copy is verified
- Transfer of data between sites over P2P (Internet)
- Possibility to process the data on the archive



Archive Figures

- 9.8 PB of EO consolidated Data
- 63.78 Millions of Products
- 61 Missions
- 1314 Datasets
- Archive content browsable at damps.info



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 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 Third Party Live EO Missions

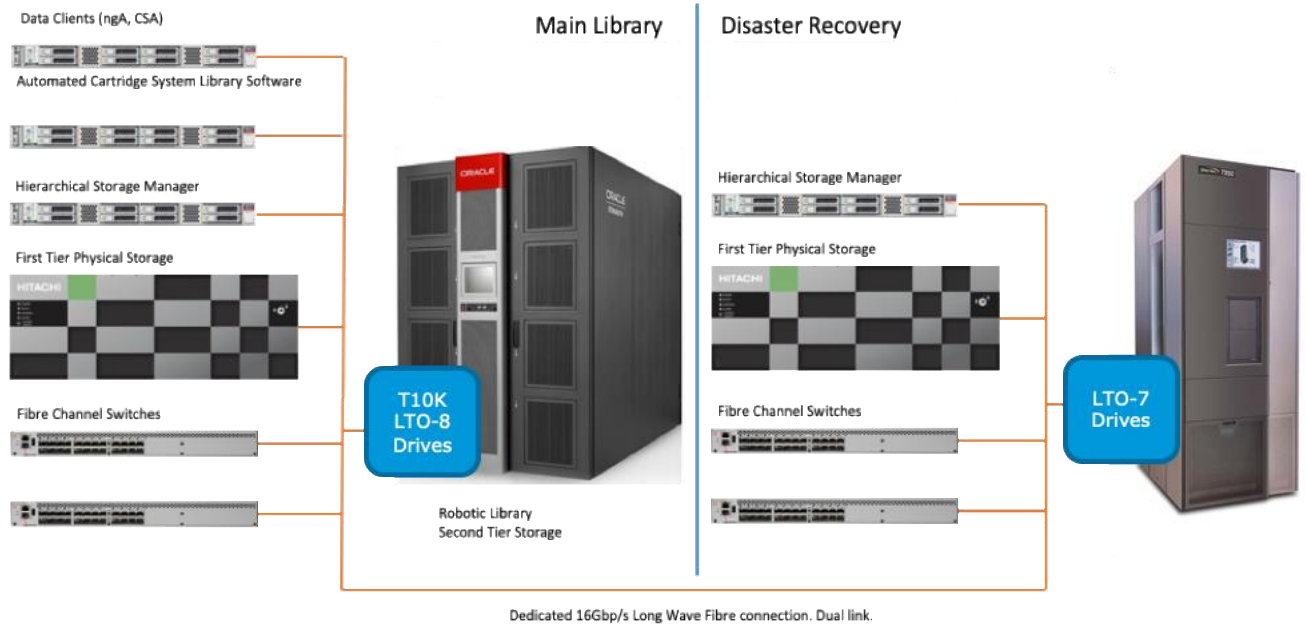
Archive Holdings and Figures

- 15.6 PB of EO Data
- Over 200 Millions of EO files
- Over 1 Billion of ESA/SCI files
- 13 TB of data received a day
- 20 TB of peak data read in a day
- 5 PB of yearly increase
- Data is contained in a Long-Term Preservation Format (EO-SIP)

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-1,2	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		

Archive Infrastructure

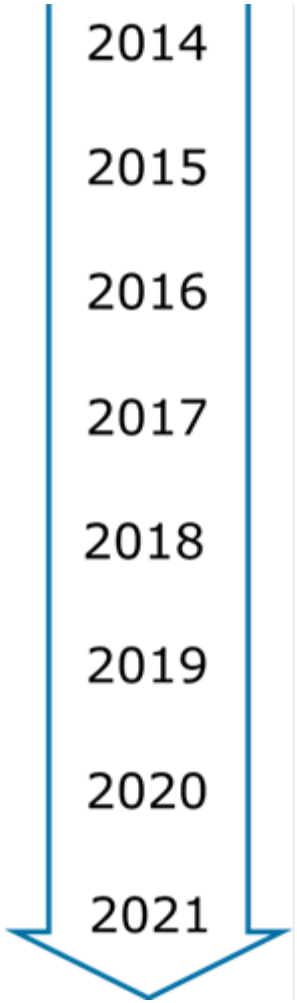
- Main Library capable of storing 36 PB of uncompressed near-line data
- Incoming data written on fast SSD physical storage
- Data eventually written on Magnetic Tapes
- Files seen by clients on a virtual file-system
- Data duplicated over a disaster recovery Robotic Library



Archive Challenges

- Recovery of faulty tapes takes time. When a tape has I/O error, it has to be read sequentially to re-archive all data contained
- Tape prices are not predictable. T10K tapes prices doubled when ORACLE 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 life of Storagetek T10K tape family and ORACLE HSM archive manager, it is feared that the main vendors are moving away from tape archive solutions. At the moment the only Enterprise solution is limited to IBM

Evolution of the Archive

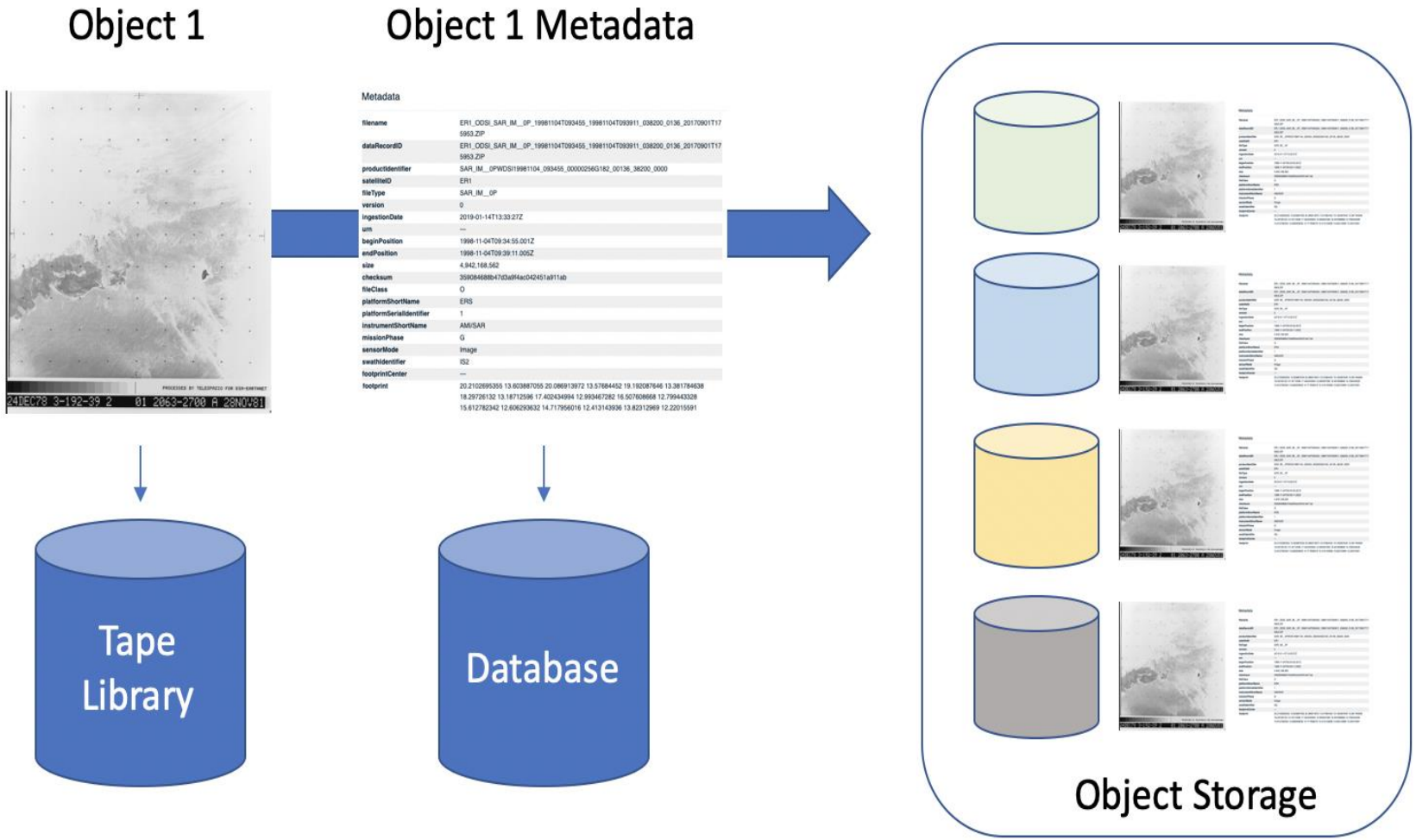


- Archive Kick-off
- Migrated technology and data from T10KB to T10KD
- Increased number of drives to 8
- Implemented 10GB Network
- Upgraded fiber channels environment from 8gbp/s to 16gbp/s
- Upgraded 1rst tier from mechanical HDD to SSD Disks
- Optimized 1st tier – dedicated LUN for iNodes
- Implemented Disaster Recovery
- Introduction of 2nd tier
- Migration to LTO-8 (increased number of drive to 12)

Possible future use cases

Object Storage is being studied as a possible replacement for the disaster recovery Library

- Data is already unstructured and metadata driven
- Limitations of OS are similar to those of the Robotic Libraries



Future of the Archive

- Upcoming Earth Explorers missions Biomass, EarthCARE and Flex are being assessed and consequently, the archive is being benchmarked and fine-tuned to assure seamless integration with the increased amount of data produced
- Expected increase of data produced in 2023 of 5 PB, 10 PB in 2024
- The industry has been engaged to suggest a roadmap to keep the archive on par with technology evolution and allow seamless archiving
- Possible solutions are already being assessed