# The ESA Earth Observation Payload Data Long Term Storage Activities

**Gian Maria Pinna[1], Francesco Ferrante[2]**

***[1]ESA-ESRIN***

*Via G. Galilei, CP 64,00044 Frascati, Italy*

*EMail: GianMaria.Pinna@esa.int*

***[2]SERCO Italy***

*Via Sciadonna24/26, 00044, Frascati, Italy*

*EMail: fferrante@serco.it*

**PV-2009**

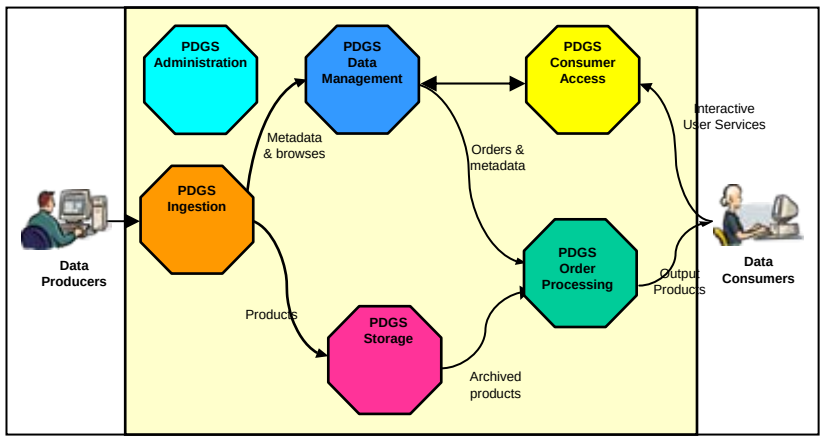
**1-3 December 2009**

# Content

- ESA PDGS and MultiMission Facility Infrastructure
- Storage Technologies
- Hierarchical Storage Management
- Disk-based archives
- Distributed File System
- Data Exchange among archives
- Conclusions

# PDGS Archives Harmonization

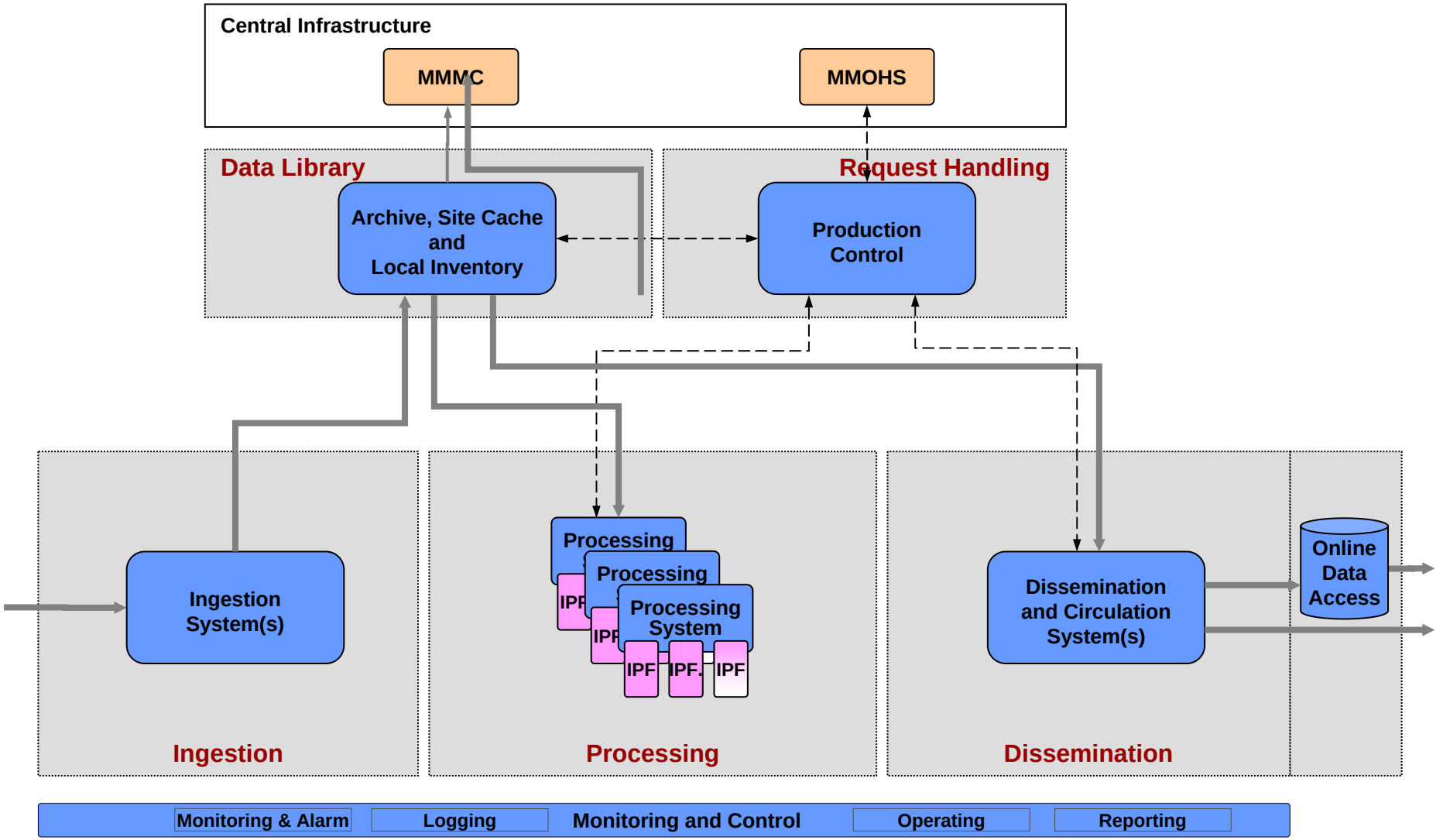
- Harmonization of the design, technologies and operational procedures used in the PDGS
- Requires adaptation for each mission operated
- Implementation by re-use and configuration proved to be very well suited for the harmonization and cost-effectiveness of the long term data preservation
- ESA PDGS approach was to develop a common infrastructure named MMFI (MultiMission Facility Infrastructure)
- The MMFI was developed following the OAIS (Open Archival Information System) Reference Model, a CCSDS/ISO standard



# MMFI Goals

- The MMFI project at ESA aims at:
  - **Evolving** the ESA EO ground segment toward a harmonized multi-mission capable infrastructure
  - Maximizing the **reuse** of existing operational elements in a coherent standard architecture (based on the OAIS-RM)
  - **Standardizing** the MMFI external and internal **interfaces**

# MMFI Abstract Model

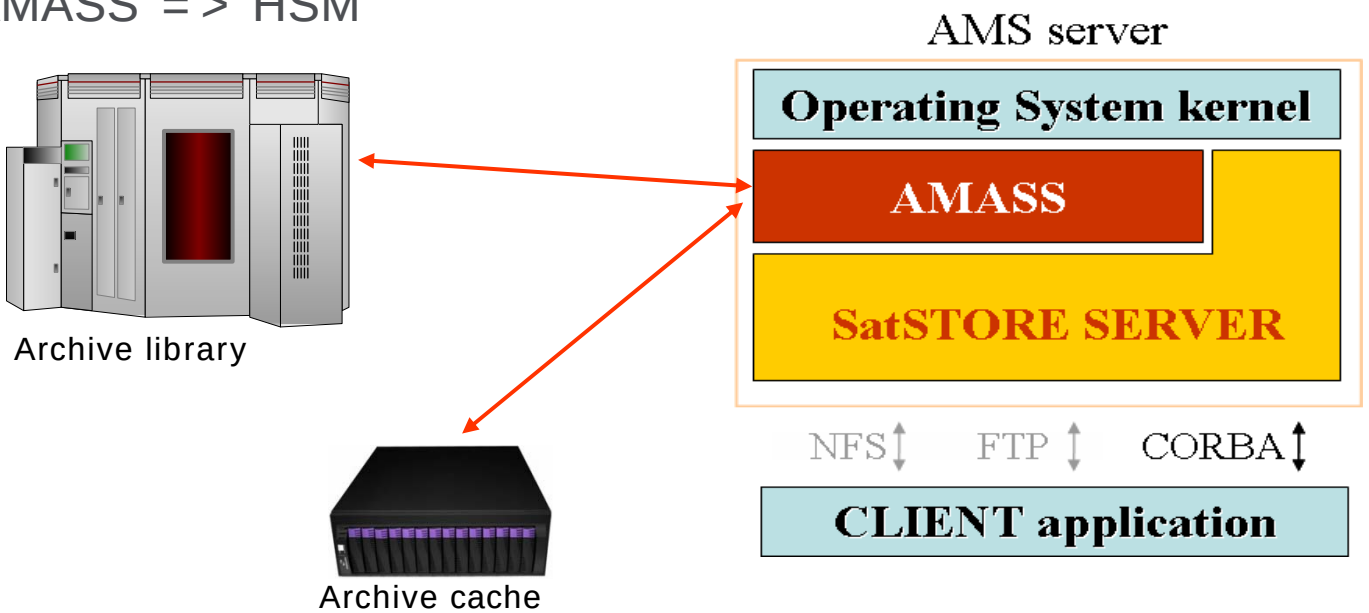


# Hierarchical Storage Management

- Long Term Data Preservation is normally achieved using multiple tiers of storage media:
  - Tape storage is the basic technology utilized to decrease the cost of storage
  - The need for faster access, both in reading and writing, to the data requires a disk-based storage layer to be dimensioned based on the required I/O performance
  - It is also common in certain architectures that other levels of storage, either disk- or tape-based, are utilized to increment the performance and reduce the overall storage cost
- The migration among the different levels of storage, from the faster (and more expensive) to the slower (cheaper) storage levels, is normally managed by a hierarchical storage management system

# EO Archive Management System

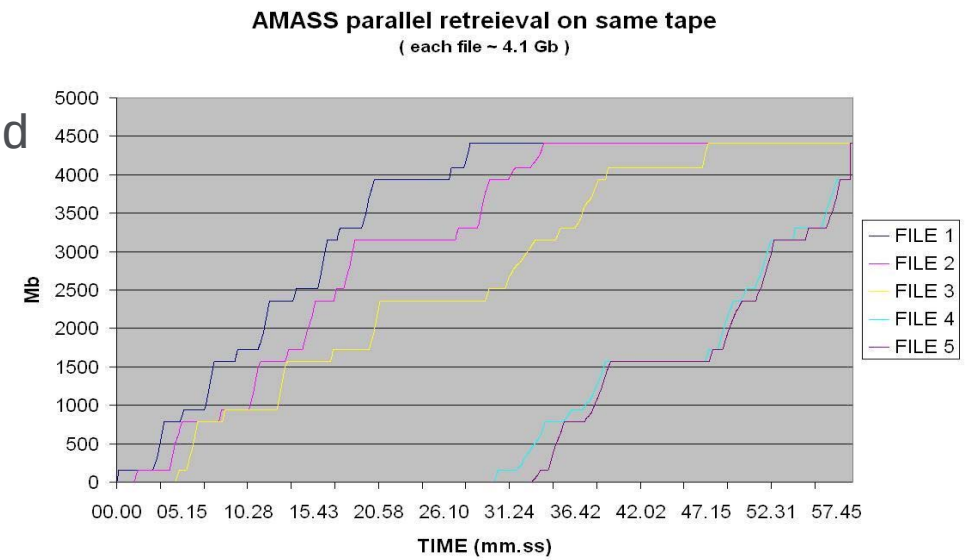
- 1996, start of the study of an ESA EO Archive Management System
- EADS SatSTORE + ADIC AMASS identified to manage the ESA EO AMS:
  - SatSTORE => data management and clients acl
  - AMASS => HSM



- Advantages for slow tape drive (respect to the current available technologies)
- Advantages for concurrent reading operations from different tapes (the concurrent reading form the same tape was managed by SatSTORE)
- Read/write performed by :

  4 x blocks = 4 x (<file size>/<configured block size>)

- Tape fine-tune to specify the tape block size (256 Kb) to achieves good performances



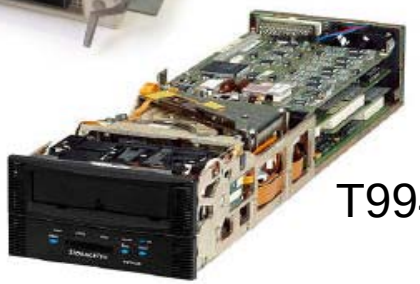
# AMASS
(vs Tape technologies)

- DLT4000/7000
  - average speed of 1.5/4 Mb/s
  - average access time 65 s
  - 20/40 Gb size
- T9940B
  - average speed of 29 Mb/s
  - Average access time 50 s
  - 200 Gb size

(next generation)

- T10KB
  - average speed of 120 Mb/s
  - Average access time 49 s
  - 1 Tb size

DLT700



T9940B

T10KB

- End Of Life (EOL) for:
  - AMASS
  - Sun Storagetek ‘PwoderHorn’ libraries – Q4 2010
  - Sun Storagetek T9940B
- Block-read is not efficient with high I/O speed tape drives.
- Block-read is not efficient with high speed repositioning and load tape drives.
- T10KB not supported by AMASS (neither by PowderHorn libraries)

# SAMFS
(next MMFI HSM)

- Support storage tiering
- Support multiple copies of the same file
- Support to new library/tape drive technologies
- File based and block based reads (run time configuration per file/directory).
- Data handling based on regular expression on path and files
- Tape can be read without SAMFS.
- Enhanced Web-based management portal
- Open source code

- Agreement with Sun of a wide/global 40 PB license for 10 ESA MMFI sites

# SAMFS
(vs tape drive & Libraries I)

- The new HSM lead the possibility to renew the HW technologies:
  - Sun Storagetek SL8500, located in ESA ESRIN (Frascati, Italy).
  - Sun Storagetek SL3000, located in Matera (I), Maspalomas (S), Kiruna (SE), Farnborough (UK), Oberfaffenhofen (D).
  - Sun Storagetek SL1400, located in Tromsoe (N) and Sodankyla (Fi).
  - 50 T10KB drives spread ESA centres.

# SAMFS
(vs tape drive & Libraries II)

SL8500

- Up to 7 libraries inter-connected
- from 1448 to 10000 slots
- 64 tape drives
- Drive support (any combination): T10000B, T10000A, T9840D, T9840C, LTO 4,LTO 3
- No downtime for library management/maintenance

SL3000

- from 200 to >5000 slots
- 56 tape drives
- Drive support : T10000B, T10000A, T9840D, T9840C, LTO 4,LTO 3
- No downtime for library main tasks management / maintenance

# Disk-based Archives

**The long data preservation together with a modern HSM can support the whole operational infrastructure to speed up operations**

T10KB drive speed ~120 Mb/s, with 6 drives storage needs to serve ~720 Mb/s

An Mid/Enterprise storage is needed to give to the archive such I/O:

- Disk-Array architecture is important
- Disk types (FC, SATA, SAS) is important
- Number of disks is critical

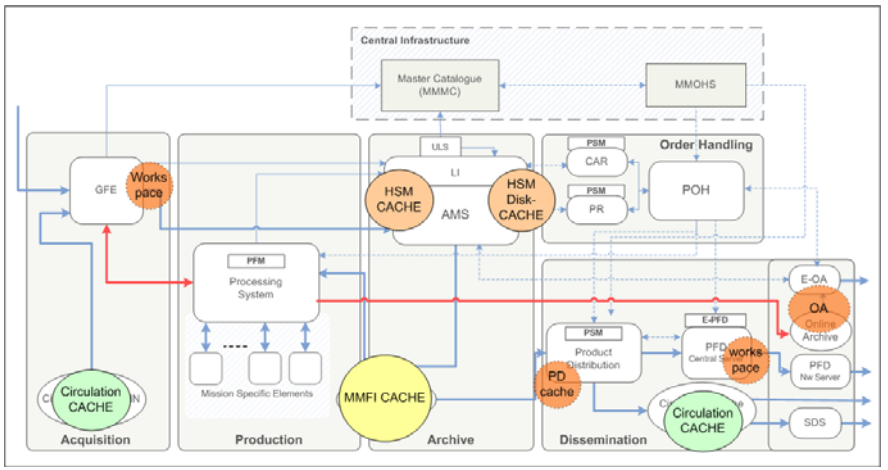
# Disk-based Archives
( synergy in MMFI I )

MMFI shared caches:

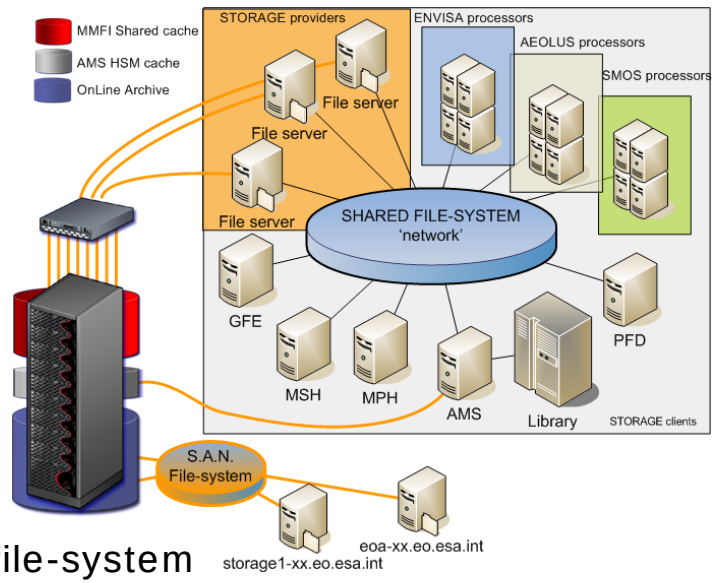
- Central Cache
- AMS disk archive
- HSM cache
- MMFI On-Line Archive

MMFE caches:

- GFE Workspace
- PD cache
- PFD Workspace




NAS/NFS

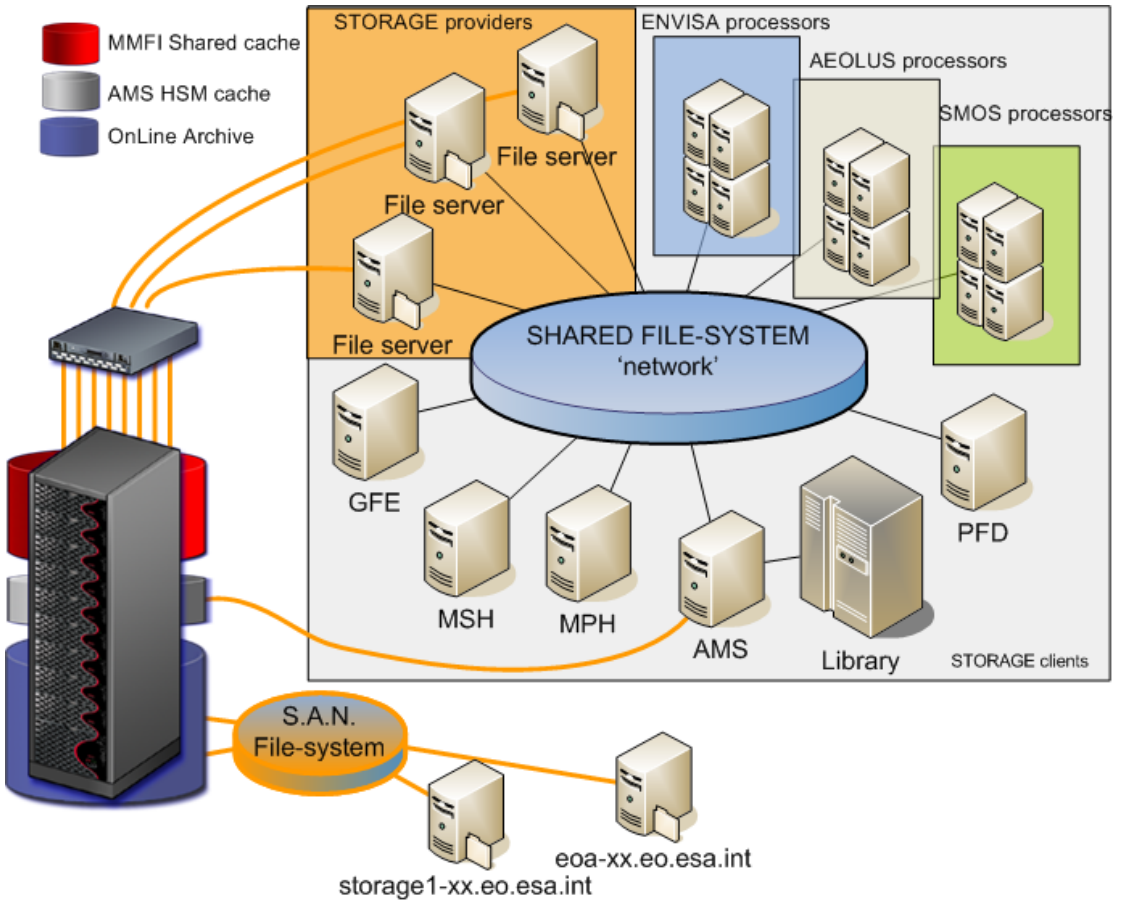



SAN/Distributed file-system

# Disk-based Archives
( synergy in MMFI II )

MMFI caches:

- HSM cache
- MMFI On-Line Archive
- MMFI cache (Central Cache, AMS disk archive, MMFE caches)



# Disk-based Archives
## ( lustre )


- Scalability Individual nodes: cluster size and disk storage are all scalable ( > 100.000 clients, #PB of shared storage.
- Performance: Lustre delivers over 240 GB/sec aggregate in production deployments. On single non-aggregate transfers, Lustre offers current single node performance of 2 GB/s client throughout (max) and 2.5 GB/s OSS throughput (max).
- POSIX compliance
- High-availability: shared storage partitions for OSS targets (OSTs) and a shared storage partition for MDS target (MDT)
- Security:TCP connections only from privileged ports are optional. Group membership handling is server-based. POSIX ACLs are supported
- Open source.

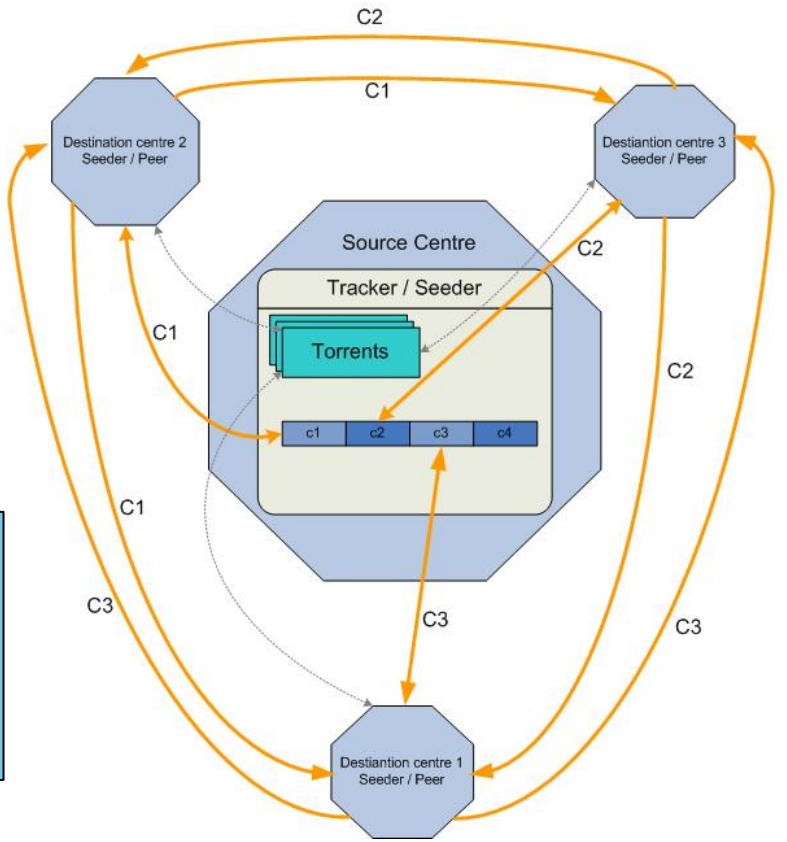
# Exchange data between archives

- ESA VPN between EO archives
- Full duplex links
- Same data archived in more that one centre
- Current infrastructure have centers that upload (Stations) and centre that download (PACs) same products

**BitTorrent-protocol**

- Geant topology does not allow high speed traffic with standard BitTorrent clients
- Using a modified client (<1 Mb of 'network block size') ~ ftp nominal throughput

**BitTorrent-modified-protocol**



# Conclusions

- The MMFI is in operations at ESA since several years
- Proven to be very effective on one side in improving the performance of the PDGS of all missions, and on the other to decrease the overall operational costs for the ESA EO payload data exploitation
- Constant evolution to:
  - implement new missions requirements
  - improve the overall performance
  - decrease the operational costs to a more affordable level
  - This is normally done by upgrading specific MMFI components or further standardizing and harmonizing the infrastructure.
- The archive function is one of the most important for the MMFI and will require continuous monitor and evolution toward better and cheaper architectures and systems