Integrated Data Archiving Strategy

IBM Spectrum Archive and ELETTRA RESTful Middleware

Marco De Simone

INAF USC-C General Assembly, Trieste 11.03.2026

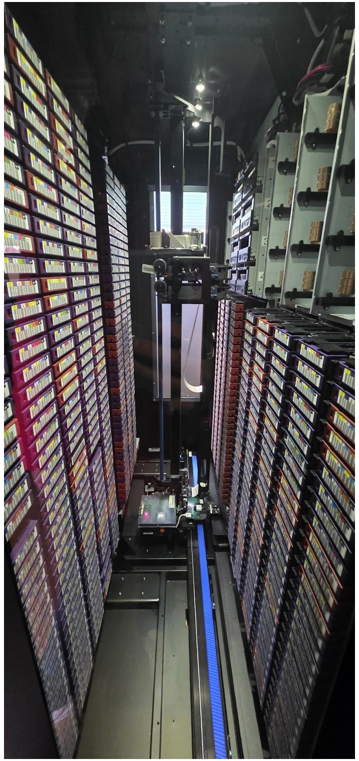
Elettra Sincrotrone Trieste

The Challenge & Agile Evolution

- **Data Pressure:** Exponential growth of experimental datasets vs. finite primary storage.
- **Agile Development:** System evolved through iterative phases to meet operational urgencies:
  - **Phase 1, (2021 Emergency Backup):** Initial focus on securing 200k+ raw datasets.
  - **Phase 2, (2023 Archival & Lifecycle):** Offloading investigations to free up high-performance storage.
- **Elettra Data Policy:** Full compliance with FAIR principles (Findable, Accessible, Interoperable, Reusable) and internal long-term preservation requirements (10y).

Hardware & Connectivity

- **Central Orchestrator:** The EDDIE server (Linux Centos 7) manages the entire archiving stack.
- **Tape Library:** IBM Spectrum Archive T4500.
- **Connectivity:** 6x LTO drives (4x LTO-7M, 2x LTO-8) connected via high-speed Fiber Channel (FC), 10 GB ethernet to NFS4 servers.
- **Inventory:** 810 total tapes (510 LTO-7M, 300 LTO-8), providing a multi-petabyte cold-storage tier



Scaling with LTFS-LE

- **LTFS Library Edition:** acts as the interface between the IBM T4500 hardware and the operating system
- **Logical Mount Points:** * It maps the 810 tapes to a centralized directory structure `/offline/TAPESERIALNUMBER`
- **Smart Media Management**: Automated Mounting into one of the 6 FC drives.
- **Parallel I/O**: Manages the distribution of data streams across the 6 physical drives, allowing for concurrent write/read operations.
- **Operational Robustness**: Manages the lifecycle of tape cartridges (Mount/Unmount/Inventory) without manual operator intervention.
- **POSIX like interface** over a sequential access media type

Data Access & Security Architecture

- **Network Integration:** Mounting "`online4`" experimental nodes via `NFSv4` and `autofs`.
- **Bypassing `root_squash`:**
  - The application runs with a dedicated service user.
  - Advanced ACLs and specific UIDs allow secure access to user data across all beamlines without compromising security.

Tape Storage Allocation & 2x2 Mirroring

- **Dual-Copy Policy:** Every dataset or investigation is simultaneously written to two separate physical tapes.
- **1:1 Redundancy:** Tapes are logically and physically paired (2x2); each pair contains identical data to ensure 100% recoverability in case of media failure.
- **Capacity Breakdown (Effective Mirrored Capacity):**
  - **Raw Data Pool:** 2.6 PB (Managed via 560 LTO-7M & LTO-8 tapes).
  - **Investigation Pool:** 0.95 PB (Managed via 250 LTO-7M tapes).
- **Capacity Normalization:** A 200MB safety buffer is subtracted from the nominal capacity of every tape to handle "tiny" LTFS formatting variances (MB range) and ensure write stability.
- **Reliability:** This "Twin-Tape" strategy protects against physical media degradation or drive-related write errors.

Software Stack: Docker Compose Orchestration

- **Deployment:** Fully containerized via Docker Compose for environment consistency and isolation.
- **Microservices Architecture:**
  - **FastAPI:** High-performance RESTful API.
  - **Celery:** Asynchronous task queue for I/O operations.
  - **Redis:** the tasks broker
  - **Flower:** web gui for monitoring celery jobs across workers
  - **PostgreSQL:** metadata, execution logs, and quota tracking

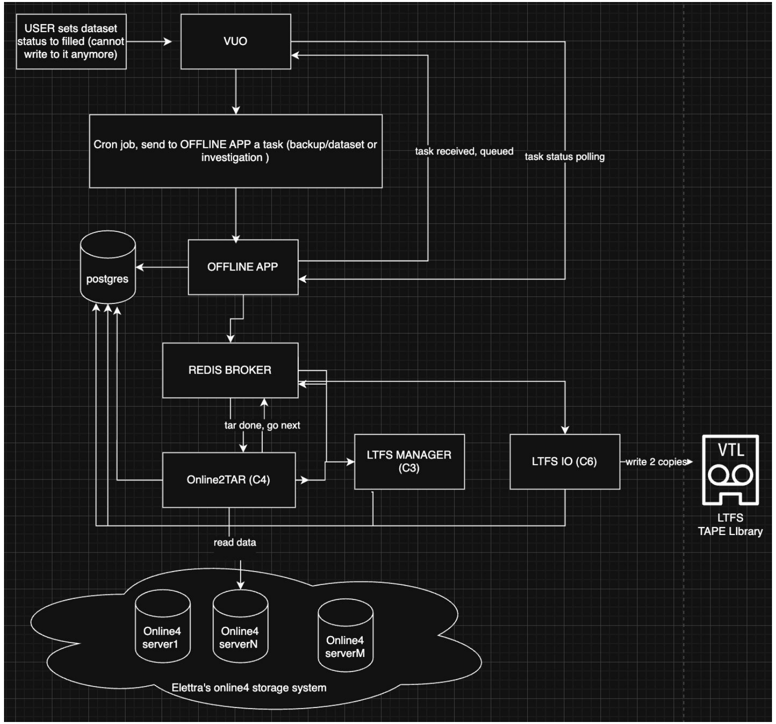
Technical Pipeline: The Celery Workers

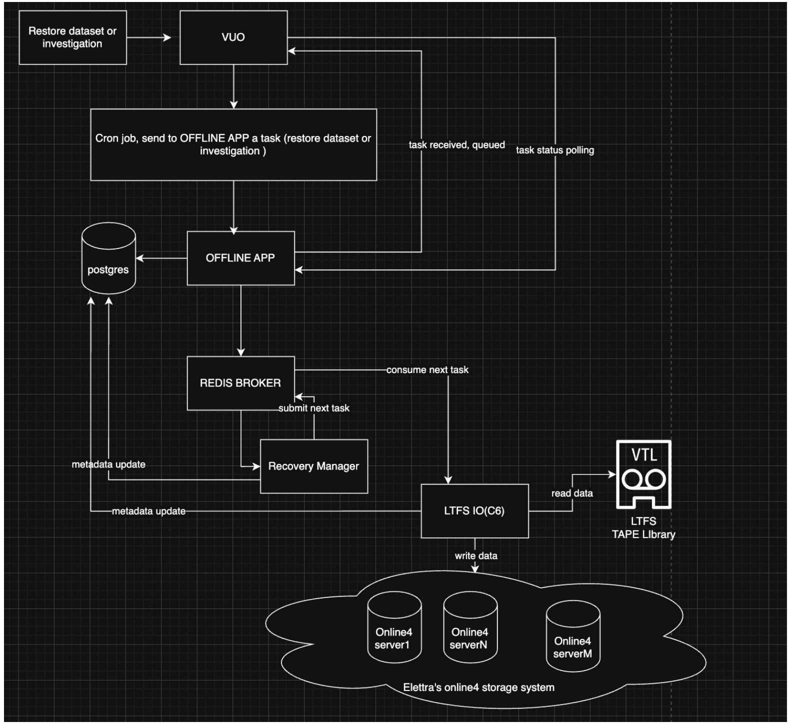
- **Worker 1 (online2tar):** *Concurrency* 4. TAR creation with **parallel compression**, file listing, and **SHA512** generation.
- **Worker 2 (Tape Manager):** *Concurrency: 3*. Orchestrates the logic for **paired tape writing** (2x2 mirroring). It manages the "twins" and ensures both copies are finalized.
- **Worker 3 (Recovery Manager):** *Concurrency 1*. It handles data restoration from offline storage, governed by a **"Pre-flight Capacity Check"** (Restore only starts if destination space is guaranteed).

Technical Pipeline: The Celery Workers

- **Worker 4 (Notification):** Real-time feedback to the **VUO portal** and diagnostic/failure alerts via email.
- **Worker 5 (LTFS-IO):** *Concurrency: 6*. Mapped 1:1 to the **Fiber Channel drives**. Handles the physical data stream to the 810+ independent LTFS filesystems.
- **Worker 6 (Flower/Monitor):** Dashboard for real-time process monitoring, job tracking, and cluster health.

Technical Pipeline: The schema





Metadata & "Always-On" Visibility

- **Dual-Layer Metadata Tracking:** All archival data, including SHA512 checksums and Tape IDs, is stored in a centralized PostgreSQL database for rapid querying.
- **Self-Describing Media:** A strict naming convention is applied to every TAR file, embedding critical info: Dataset/Investigation ID, online4 Server Name, and the Full Source Path.
- **Disaster Recovery Preparedness:** By keeping filenames under the 256-character limit, the tapes remain human-readable and recoverable even without database access.
- **Online Listing:** File listings remain on the primary storage; users can browse archived investigations without mounting tapes.
- **Integrity Guarantee:** SHA512 hashes are cross-referenced during every restore operation to ensure zero data corruption.

Mastering the LTFS Scaling Challenge

- **Fragmentation:** Managing 810 mount points with variable available space.
- **Automation:** Tape selection, mounting, indexing, and unmounting are fully automated (Zero Human Intervention).
- **Multi-tape Spanning:** Distribution of 2 TB chunks across multiple tapes for investigations or datasets exceeding physical media limits.

Introducing the VUO Portal 1/2

VUO (Elettra Virtual User Office), is the institutional backbone, our source of truth, that manages the entire lifecycle of a scientific experiment for Fermi & Elettra.

Proposal-Driven Archiving:

- Beamtime & Proposals: every archival task is bound to a specific Proposal ID. The data is always linked to a specific research project and its experimental timeframe, data is secured by Unix ACLs.

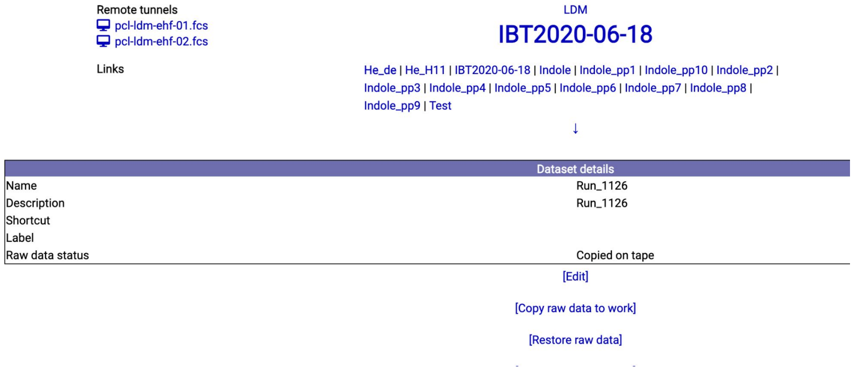
Introducing the VUO Portal 2/2

- Identity & Ownership (The Policy Engine):
    - Principal Investigator (PI) & Collaborator Rights
    - Automated ACL Mapping

- Data Policy Governance:
    - Embargo Management: The VUO tracks the 3 year (+2) embargo period, the data is private to PI & collaborators.
    - FAIR Transition to Open access after the embargo expires.

User Integration (VUO Portal)

- **User Experience:** Scientists trigger archiving directly from the Virtual Unified Office (VUO).
- **Workflow Integration:** Seamless transition with no learning curve for researchers.
- **Transparency:** Real-time job status and data availability are always visible to the user.



Remote tunnels

pcl-ldm-ehf-01.fcs

pcl-ldm-ehf-02.fcs

Links

LDM

IBT2020-06-18

He_de | He_H11 | IBT2020-06-18 | Indole | Indole_pp1 | Indole_pp10 | Indole_pp2 | Indole_pp3 | Indole_pp4 | Indole_pp5 | Indole_pp6 | Indole_pp7 | Indole_pp8 | Indole_pp9 | Test

↓

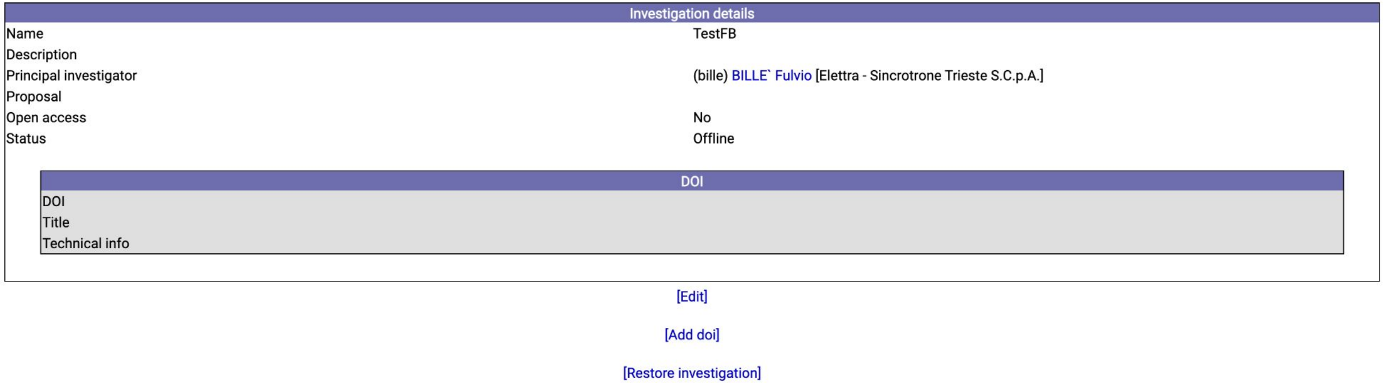
Dataset details	
Name	Run_1126
Description	Run_1126
Shortcut	
Label	
Raw data status	Copied on tape

[Edit]

[Copy raw data to work]

[Restore raw data]

User Integration (VUO Portal)



Investigation details

Name	TestFB
Description	
Principal investigator	(bille) BILLE` Fulvio [Elettra - Sincrotrone Trieste S.C.p.A.]
Proposal	
Open access	No
Status	Offline

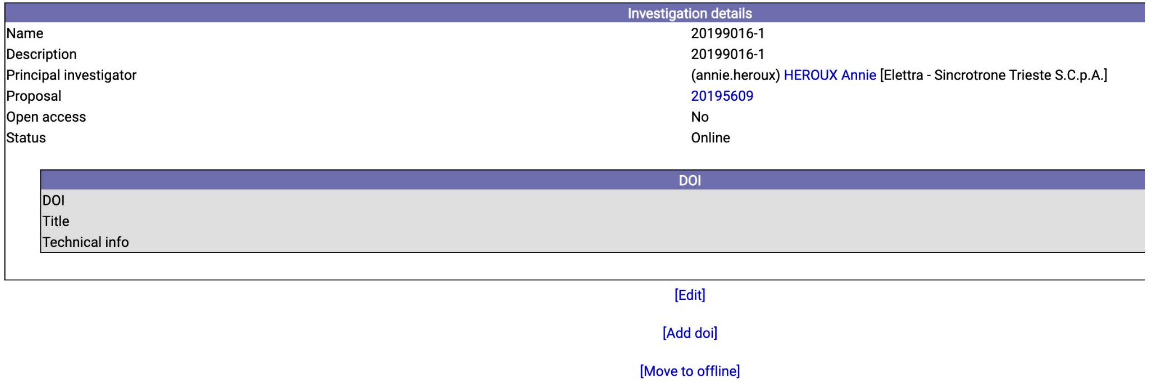
DOI

DOI
Title
Technical info

[Edit]

[Add doi]

[Restore investigation]



Investigation details

Name	20199016-1
Description	20199016-1
Principal investigator	(annie.heroux) HEROUX Annie [Elettra - Sincrotrone Trieste S.C.p.A.]
Proposal	20195609
Open access	No
Status	Online

DOI

DOI
Title
Technical info

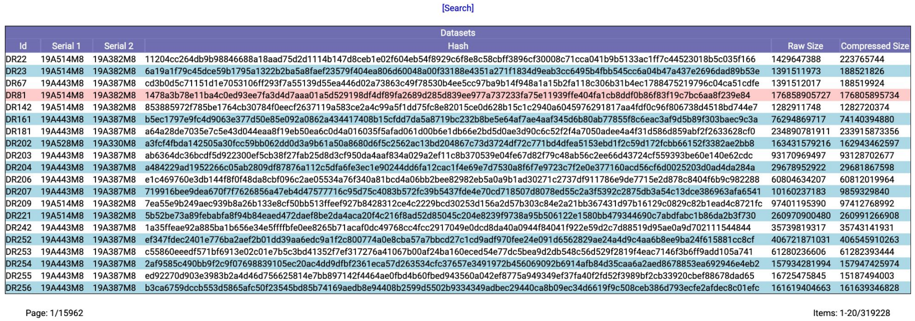
[Edit]

[Add doi]

[Move to offline]

Statistics & Performance (2022–Present)

- **Backup:** 318,942 raw datasets secured, ~700 TB on disk, ~570TB on tapes
- **Archiving:** 401 full investigations offloaded from primary disks.
- **Efficiency:** 150 TB and growing of primary storage recovered and returned to production.
- **Tape occupancy**: 1.4 PB
- **Success Rate:** Zero data loss loss thanks to the "Twin-Tape" redundancy.
- **Faulty tapes**: 4 LTO-7M



[Search]

			Datasets		
Id	Serial 1	Serial 2	Hash	Raw Size	Compressed Size
DR22	19A514M8	19A382M8	11204cc264db9b98846688a18aad75d2d1114b147d8ceb1e02f604eb54f8929c6f8e8c58cbff3896cf30008c71cca041b9b5133ac1ff7c44523018b5c035f166	1429647388	223765744
DR23	19A514M8	19A382M8	6a19a1f79c45dce59b1795a1322b2ba5a8faef23579f404ea806d60048a00f33188e4351a271f1834d9eab3cc6495b4fbb545cc6a04b47a437e2696dad89b53e	1391511973	188521826
DR67	19A443M8	19A387M8	cd3b0d5c71151d1e7053106ff293f7a55139d55ea446d02a73863c49f78530b4ee5cc97ba9b14f948a1a15b2fa118c306b31b4ec1788475219796c04ca51cdfe	1391512017	188519924
DR81	19A514M8	19A382M8	1478a3b78e11ba4c0ed93ee7fa3d4d7aaa01a5d529198df4df89fa2689d285d839ee977a737233fa75e11939ffe404fa1cb8ddf0b86f83f19c7bc6aa8f239e84	176858905727	176805895734
DR142	19A514M8	19A382M8	853885972f785be1764cb30784f0eecf2637119a583ce2a4c99a5f1dd75fc8e82015ce0d628b15c1c2940a6045976291817aa4fdf0c96f806738d4518bd744e7	1282911748	1282720374
DR161	19A443M8	19A387M8	b5ec1797e9fc4d9063e377d50e85e092a0862a434417408b15cfdd7da5a8719bc232b8be5e64af7ae4aaf345d6b80ab77855f8c6eac3af9d5b89f303baec9c3a	76294869717	74140394880
DR181	19A443M8	19A387M8	a64a28de7035e7c5e43d044eaa8f19eb50ea6c0d4a016035f5afad061d00b6e1db66e2bd5d0ae3d90c6c52f2f4a7050adee4a4f31d586d859abf2f2633628cf0	234890781911	233915873356
DR202	19A528M8	19A330M8	a3fcf4fbda142505a30fcc59bb062dd0d3a9b61a50a8680d6f5c2562ac13bd204867c73d3724df72c771bd4dfea5153ebd1f2c59d172fcbb66152f3382ae2bb8	163431579216	162943462597
DR203	19A443M8	19A387M8	ab6364dc36bcdf5d922300ef5cb38f27fab25d8d3cf950da4aaf834a029a2ef11c8b370539e04fe67d82f79c48ab56c2ee66d43724cf559393be60e140e62cdc	93170969497	93128702677
DR204	19A443M8	19A387M8	a484229ad1952266c05ab2809df87876a112c5dfa6fe3ec1e90244dd6fa12cac1f4e69e7d7530a8f6f7e9723c7f2e0e377160acd56cf6d0025203d0ad4da284a	29678952922	29681867598
DR206	19A443M8	19A387M8	e1c469760e3db144f8f0f48da8cbf096c2ae05534a76f340a81bcd4a06bb2bee82982eb5a0a9b1ad30271c2737df911786e9de7715e2d878c8404f6b9c982288	60804634207	60812019964
DR207	19A443M8	19A387M8	719916bee9dea670f7f7626856a47eb4d47577716c95d75c4083b572fc39b5437fde4e70cd718507d8078ed55c2a3f5392c2875db3a54c13dce386963afa6541	10160237183	9859329840
DR209	19A514M8	19A382M8	7ea55e9b249aec939b8a26b133e8cf50bb513ffeef927b8428312ce4c2229bcd30253d156a2d57b303c84e2a21bb367431d97b16129c0829c82b1ead4c8721fc	97401195390	97412768992
DR221	19A514M8	19A382M8	5b52be73a89febabfa8f94b84eaed472daef8be2da4aca20f4c216f8ad52d85045c204e8239f9738a95b506122e1580bb479344690c7abdfabc1b86da2b3f730	260970900480	260991266908
DR242	19A443M8	19A387M8	1a35ffeae92a885ba1b656e34e5ffffbfe0ee8265b71acaf0dc49768cc4fcc2917049e0dcd8da40a0944f84041f922e59d2c7d88519d95ae0a9d702111544844	35739819317	35743141931
DR252	19A443M8	19A387M8	ef347fdec2401e776ba2aef2b01dd39aa6edc9a1f2c800774a0e8cba57a7bbcd27c1cd9adf970fee24e091d6562829ae24a4d9c4aa6b8ee9ba24f615881cc8cf	406721871031	406545910263
DR253	19A443M8	19A387M8	c55860eeedf571bf6913e02c01e7b5c3bd41352f7ef317276a41067b00af24ba160eced54e77dc5bea9d2db548c56d529f2819f4eac7146f3b6ff9add105a741	61280236606	61282393444
DR254	19A443M8	19A387M8	2af9585c490bb9f2c9f07698839105ec20ac4dd9dfbf2361eca57d263534cfc37657e3491972b456069092b6914afb84d35caa6a2aed8678853ea692946e4eb2	157934281994	157947425974
DR255	19A443M8	19A387M8	ed92270d903e3983b2a4d46d756625814e7bb897142f4464ae0fbd4b60fbed943560a042ef8775a949349ef37fa40f2fd52f3989bf2cb33920cbef88678dad65	16725475845	15187494003
DR256	19A443M8	19A387M8	b3ca6759dccb553d5865afc50f23545bd85b74169aedb8e94408b2599d5502b9334349adbec29440ca8b09ec34d6619f9c508ceb386d793ecfe2afdec8c01efc	161619404663	161639346828

Page: 1/15962

Items: 1-20/319228

Conclusions & Strategic Value

- **In-house Innovation:** A bespoke solution that eliminates enterprise licensing costs (estimated savings >€50k/y).
- **Future Vision:** Transition towards Logical Chunking to enable partial data recovery and increase system resilience.
- **Takeaway:** A proprietary Elettra digital asset, scalable and ready for the future challenges of next-generation light sources.
- **Looking for collaborations**: Secure data from big

Call for Partners & Collaboration (PRIN -EU)

- We are seeking Partners for Proposal submission (PRIN and EU)
  - **Distributed Pipelines**: Advanced metadata management and modular architectures for massive volumes of scientific and sensitive datasets.
  - **Efficient Data Ingestion**: Integration of Compressive Sensing for efficient acquisition, ensuring scalability, security, and full traceability of datasets.
  - **Data Sovereignty**: High-performance security and access frameworks for long-term preservation and protection of mission-critical data.