Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Archiving Scientific Data Outside of the Traditional HEP Domain, Using the Archive Facilities at Fermilab

M. Diesburg, M. Gheith, R. Illingworth, M. Mengel, A. Norman

Fermilab, Scientific Computing Division
Scientific Data Management

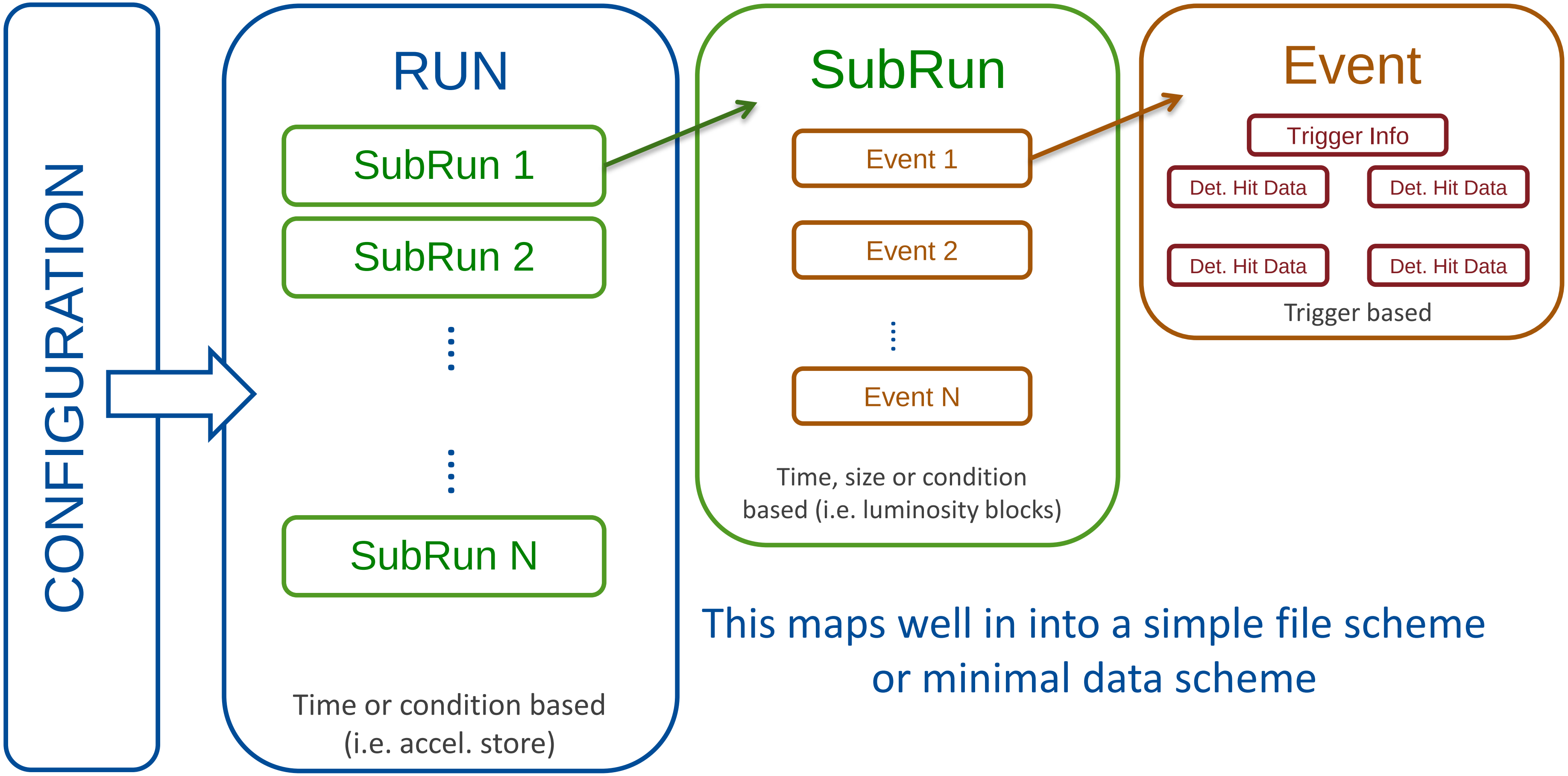
Archiving Problem

- HEP & Non-HEP experiments can produce large O(PB) datasets which need to be archived, managed, retrieved
  - Actual data varies in size, structure, format, complexity, etc…
  - Want to capture all the details and interconnections
  - But need to interact with traditional archival and mass storage system
- Need not only tools but a general strategy for how you perform this mapping into storage

Can this strategy be applied to a wide variety of experiments?

Traditional HEP Data

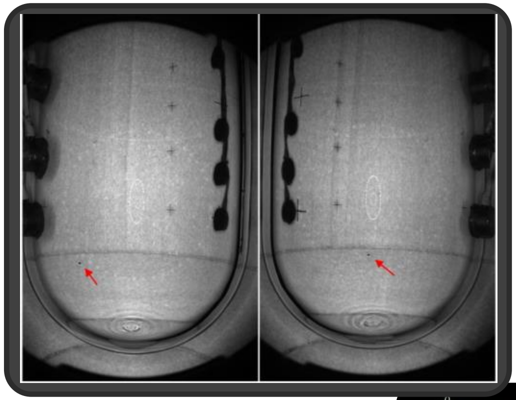
- The traditional HEP experiment use a well defined "Run/Event" model to define and organize their data



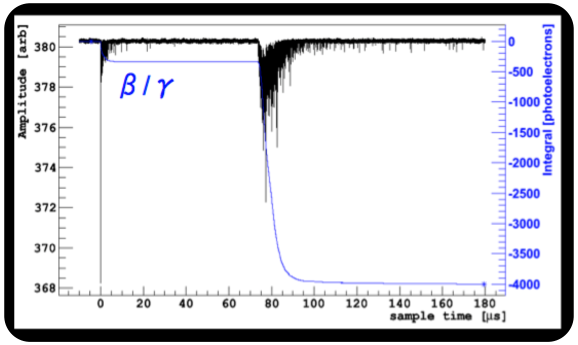
This maps well in into a simple file scheme or minimal data scheme

Non-Traditional HEP Data

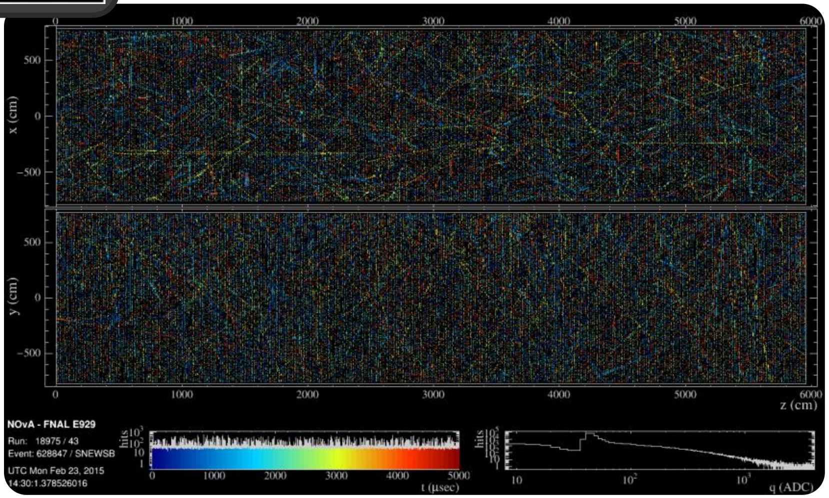
- In contrast newer & non-traditional HEP experiment use all sorts of schemes to record data
  - Image data
    - discrete photos
  - Image series
    - framed time windows
  - Time series data
    - long continuous wave forms



Digital Image Readout (COUPP)



Waveform Readout (Darkside)



Continuous Framed Readout (NOvA)

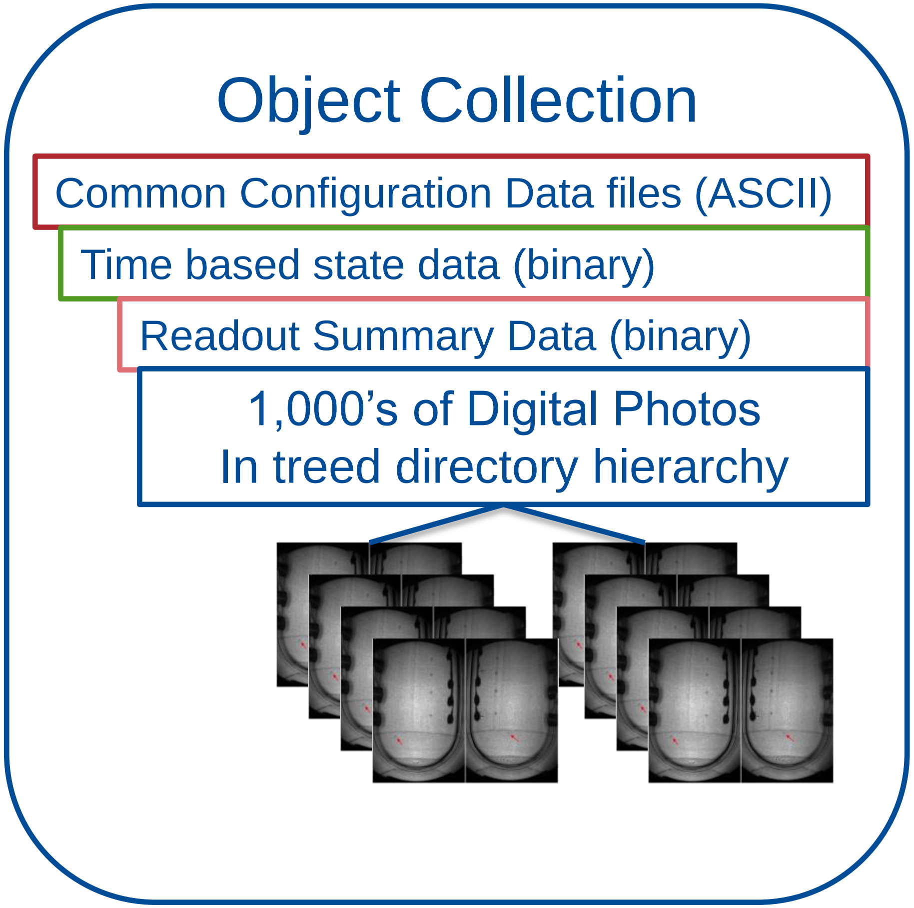
- Organization sometimes fits in a “run” model, but more often is more of a generic “time window”

Non-Traditional Data Relations

- Trying to capture "collections" of related objects

- Need to preserve all of this information and the associations between them

- Need a method that is automatable and can scale
  - e.g. archive 400,000+ collections from a DM experiment



Storage/Catalog Structure

- Storage of these data require
  - Maintain the structure, interconnections & hierarchy of the data
  - Maintain the configuration and ties to the data
  - But want to be able to locate/retrieve/analyze individual data items quickly (i.e. retrieve a single photo)
  - Most experiments encode this via the file system layer

Problematic:

- Requires knowledge of the data layout
- Requires knowledge of the storage system
- Not portable (i.e. replication to second site using different storage types)

Instead want an efficient organization of data based on its characteristics which is agnostic to the actual underlying storage:

Cataloged, Hierarchical, Pseudo-Object Store

Hard to deal with (for physicists) →

Fermilab

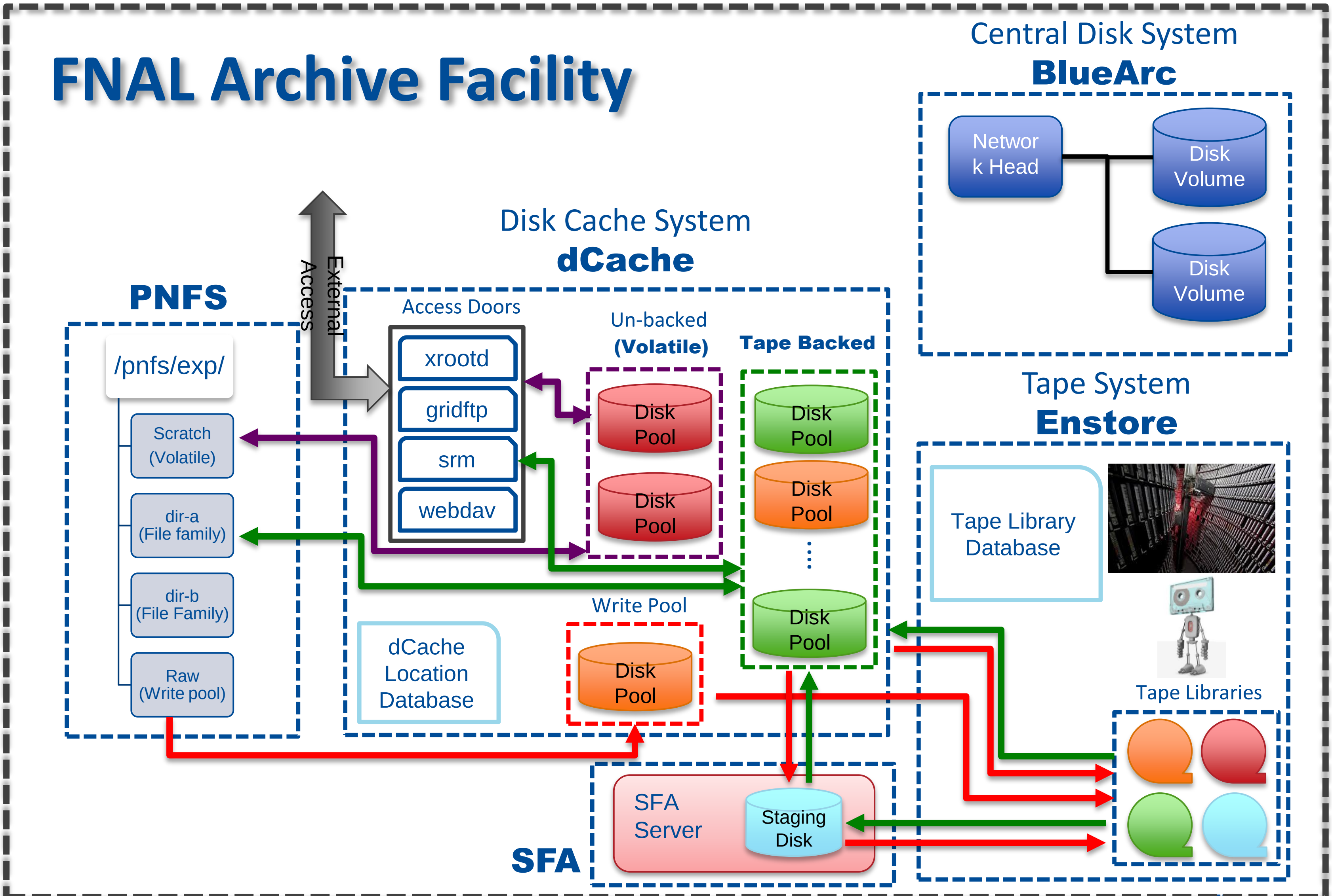
Goal

- Goal is to provide non-traditional experiments with:

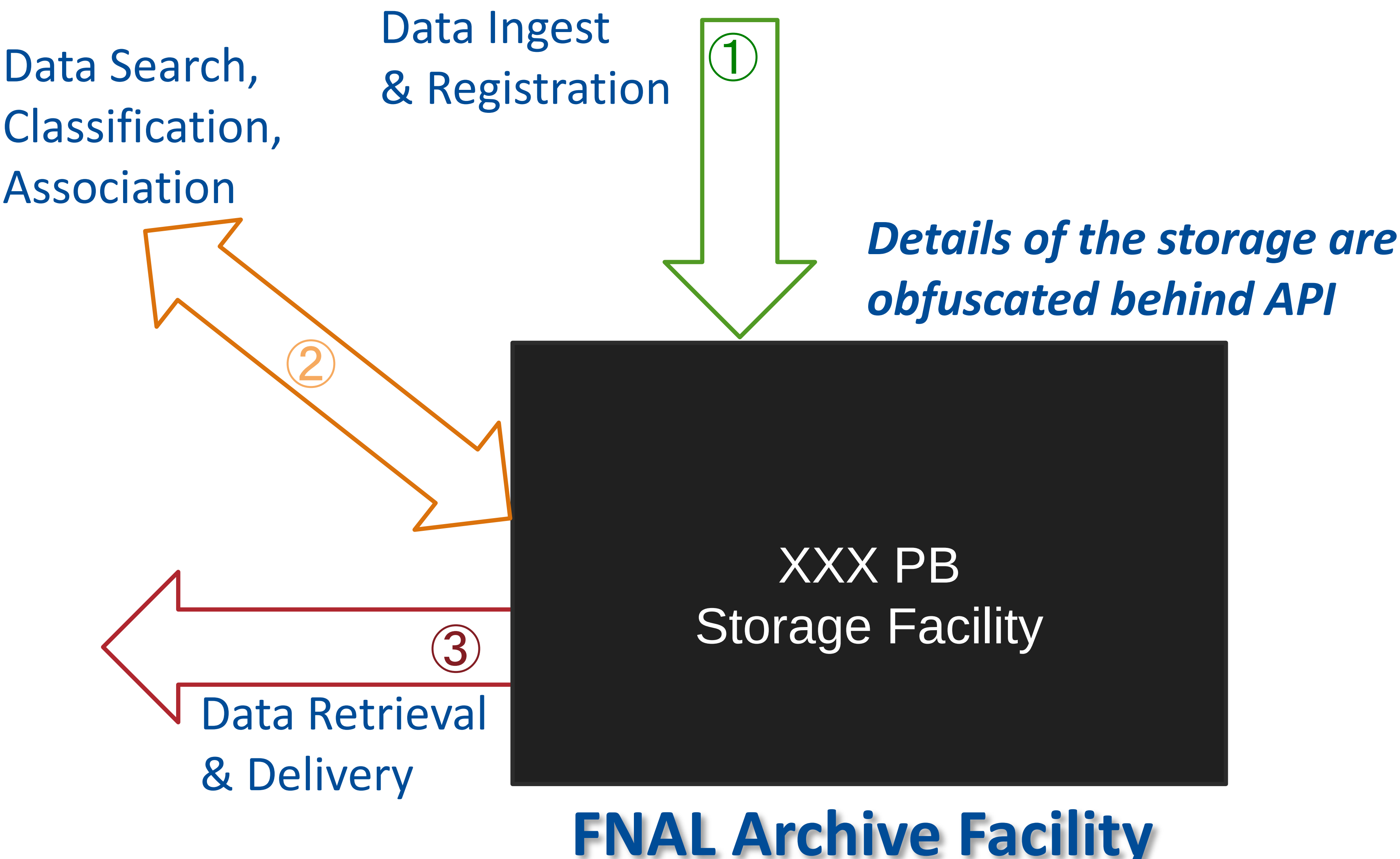
  1. A simple method for data to enter the storage facility (without knowledge of the details of facility operation)
  2. The ability to specify meta information attached to the data (descriptions, associations and hierarchical relationships)
  3. The ability to locate/retrieve the data from storage systems (based on meta data not of knowledge of the storage facility)
  4. A mechanism for delivery of the data to a user specified locations (reconstitute everything for analysis)

Automated, easy to use, at scale etc….

We don't want people to need to understand….



Instead it should be a black box…



Tool Set

- We have built a data handling tool set which consists of three main components:
  - Ingest: Fermi File Transfer Service (F-FTS)
    - Simple interface for transferring, registering & injecting arbitrary data into the storage facility. Supports arbitrary data types and fully customizable meta data. [Asynchronous client side daemon]
  - Catalog & Search: SAM (Sequential Access via Metadata)
    - Integrated metadata and replica catalog with storage facility aware caching and “project bookkeeping” for optimized data delivery.
  - Retrieval/Delivery: IFDH (Intensity Frontier Data Handling)
    - Modular transport protocol abstraction layer with integration into data and replica catalogs. “Fetches data to the user”

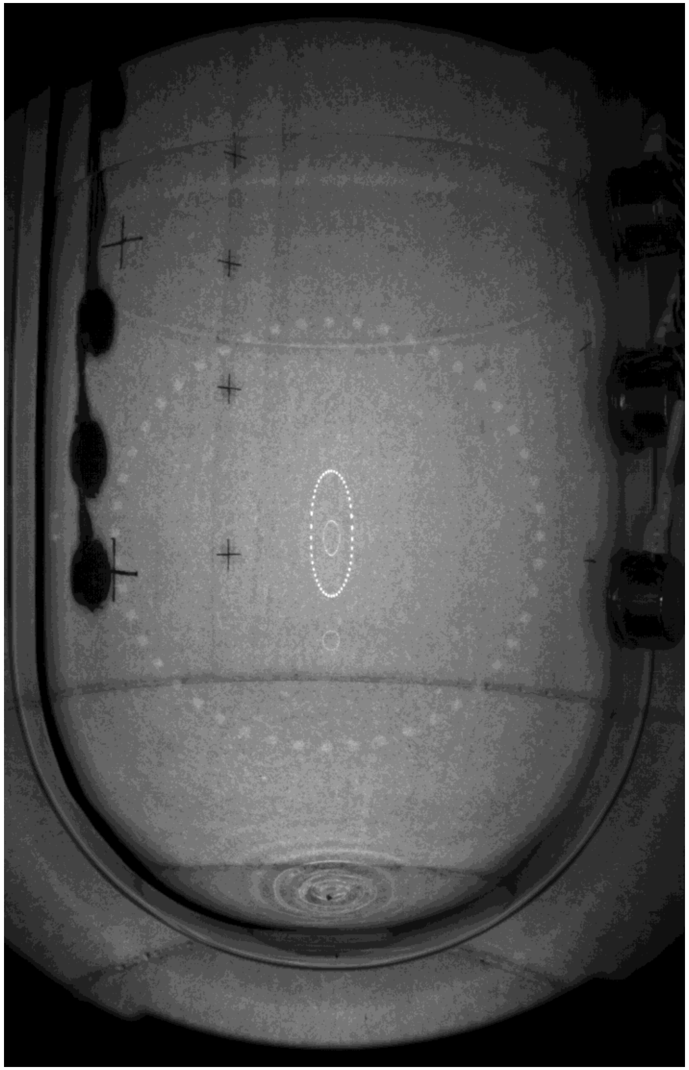
Data Ingest (F-FTS)

- F- FTS presents experiments with a simple “dropbox” interface to storage systems and file catalog.
  - Rules based declaration and transport engine

- F-FTS automates:
  - Recursive scans designated “dropbox” directories for new files
  - Extracts/Generates metadata for each file
    - Rules based on file type
    - Either user supplied json files or plugin scripts.
  - Queues files for transfers (LAN or WAN, multi-hop & chaining)
  - Verifies successful transfer of data to final storage locations.
  - Catalogs files w/ replica information
  - “Cleans up” successfully transferred files
  - Detailed monitoring

Case Examples (COUPP)

- COUPP is a bubble chamber dark matter search experiment running at SNOLAB.
- Data is organized on disk based on:
  - Running configuration
  - Collection date
  - A subdivision within each date based on running period
  - Individual events
- Data consists of multiple formats:
  - Text based data
  - Binary numeric data
  - BItmap image data
- Common config files exist for each configuration and date.

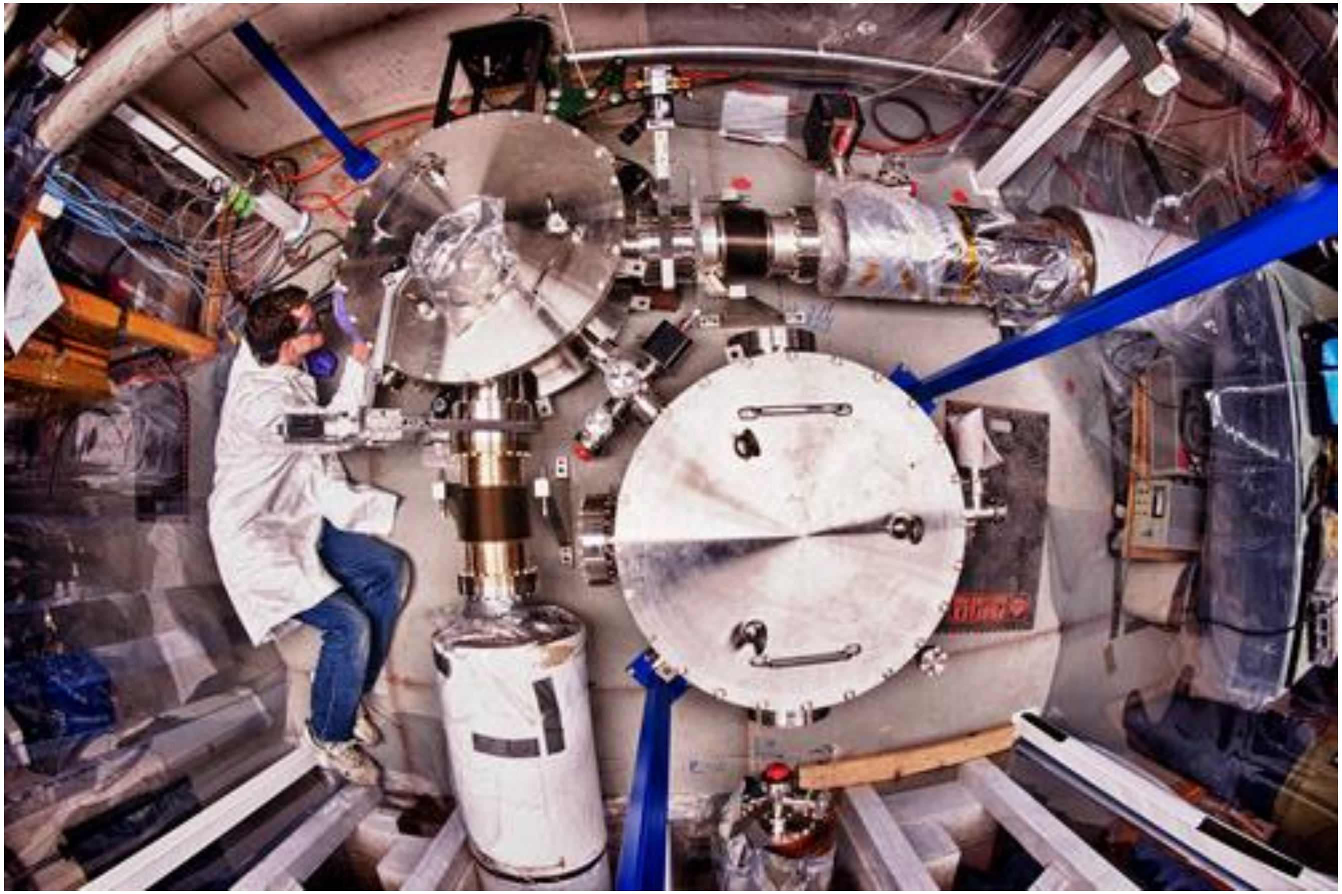


Case Examples (COUPP)

- The following metadata was defined to match the above organization:
  - Data_stream : Identifies global running configuration
  - Data_tier : Idenitfies data type, i.e. configuration data, running condition data, or event data
  - Run_number : Identifies the date when data was collected
  - Subrun_number : Identifies subdivision within a date
  - Event_number : Identifies each event within a subdivision
- The above allows the actual directory structure of the original organization to be reconstructed.
- This allow selections of data subsets such as all data in a configuration, all data within a range of days, specific sets of individual events.

Holometer

- Laser Interferometer
- "Data" is a series of interference patterns



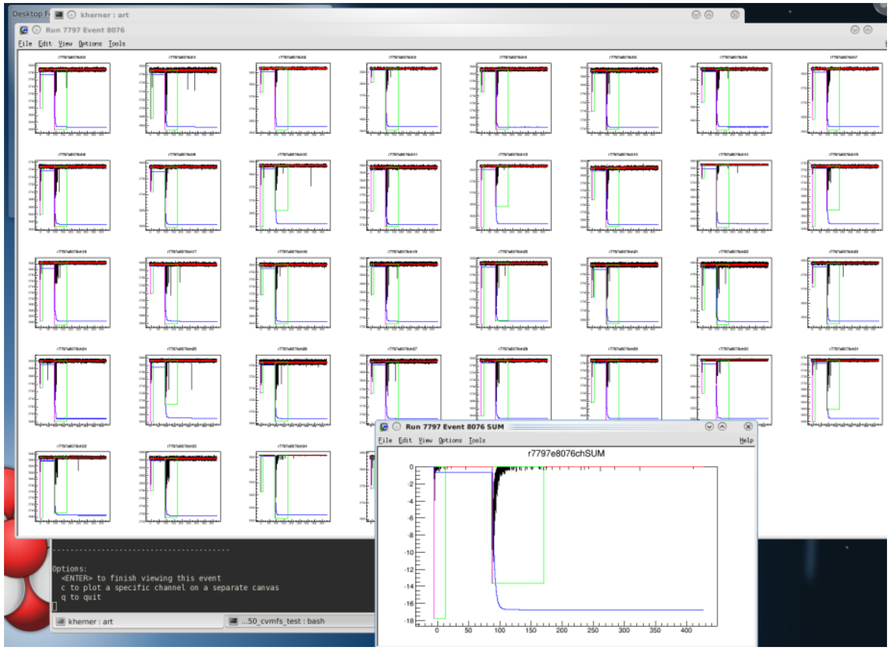
Fermilab

Case Examples (Holometer)

- Data collected on combination of machines running either embedded Linux or MS Windows.
- Data exported via NFS and CIFS shares to offline machine where F-FTS runs
  - Allows F-FTS to provide DAQ data storage independent of OS compatibility issues.
  - F-FTS maintenance and operation is independent of DAQ operations.
- Data is stored in multiple formats including
  - .gwf files (Gravitational Wave Frame)
  - .h5 files (Hierarchical Data Format v5)

Case Examples (DarkSide)

- DarkSide liquid argon TPC dark matter search is located at Laboratori Nazionali del Gran Sasso.
- F-FTS is used to transfer neutron veto data from Italy to Fermilab.
- F-FTS runs on DAQ machines and initiates gridftp transfers over wide area network into the disk cache from end of the Fermilab tape storage facility.
  - ~500 TB of data in ~100K files have been transferred over the wide area network.



Darkside 50 Event Data consisting of series of time series

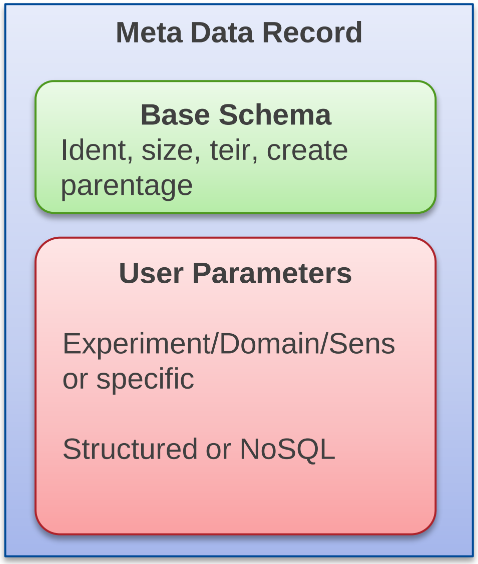
Case Examples (Nova)

- Nova makes extensive use of FTS in all aspects of operation
  - Transferred > 1.6 PB of data and over 12M files over via FTS
  - Raw data, calibration data and logs are transferred from the far detector at Ash River Minnesota to Fermilab.
    - Transfer done with multi-stage FTS transfer
    - One instance transfers via gridftp from Ash River to FNAL disk
    - Second instance transfers form disk into tape storage
    - Second stage also replicates data for immediate use at FNAL via use of multiple transfer destinations
    - Final status of storage to tape is transmitted back to Ash River
  - Production reconstruction and Monte Carlo generation done on Grid resources also use FTS to store output to the Fermilab tape facility.

Data Registration & Catalog

- "Object based" data, replica and project catalog

- Each data object is registered in the catalog along with metadata describing it.
  - Two components to the metadata
    - Base schema – General Object Information
      - identifier, size, data tier, begin/end times, parentage/provenance
    - User parameters – Data content specific fields
      - Detector type, location, trigger stream, etc…
  - Only base schema is required
    - Simplifies registration of foreign/legacy data with catalog systems



- "Datasets" are then defined via queries against the meta data.
  - Evaluate to the set of objects to retrieve/analyze

Data Search Classification/Association (SAM)

- Leverage SAM data handling service
  - Provides full metadata based service for data handling.
- Facilities for:
  - Defining arbitrary string-value pairs which can be associated with each file in the system
  - Storing location information for each file.
  - Searching the database for files which match logical constraints on the metadata.
  - Storing the results of such searches as dataset definitions.
  - Recording the processing history of files accessed via the stored dataset definitions.

All the associations and structure between objects can be captured and searched w/ SAM catalog

Data Retrieval

- Retrieval is through the *IFDH* tools set.
  - Interacts with data and replica catalog to find data elements
  - Handles "last mile" of data movement between storage facilities → "local storage"
    - Can move data between arbitrary elements, e.g. local disks, disk caches such as dCache, or tape libraries.
    - Acts as protocol abstraction layer
      - Will select a transport protocol suitable for the storage elements.
      - Modular support for protocols: gridftp, srm, dccp, aws S3, cp, dd etc.…
    - Can instigate transfers as local copies, copies to or from remote nodes, or as third party transfers between remote nodes.
    - Understands data storage locations as provided by queries to the SAM metadata system.
    - Incorporates load leveling mechanisms in multi-file transfers to prevent overloading of storage resources.

Summary

- We have created a set of tools which are able to map almost arbitrary data into structures that can be:
  - Stored
  - Organized
  - Queried
  - Retrived

- The tools set has been successfully used to perform large scale archiving of data from dark matter searches, neutrino oscillation experiment, astro physics data.

- Opened up use of the Fermilab Mass Storage and archive systems to more experiments