Data Management for Phase III of the Project 8 Experiment

A. Genawi, P. Slocum

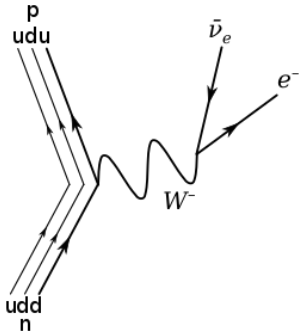
Wright Laboratory

PROJECT 8

Project 8 Experimental Overview: *Objective*

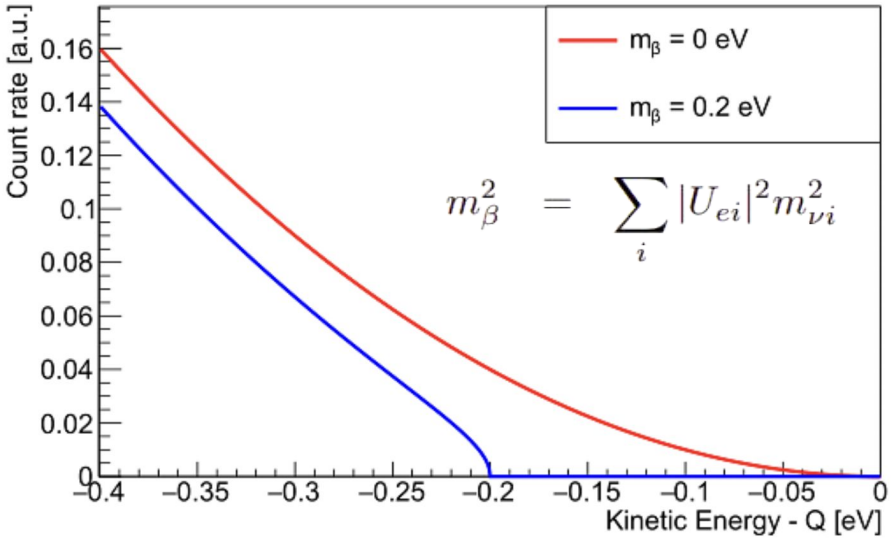
Objective: Constraining m_β through precise spectroscopy of electrons emitted from tritium beta decay.

$$^3H \rightarrow {}^3He + \beta^- + \bar{\nu}_e$$
$$(1/2^+) \rightarrow (1/2^+)$$

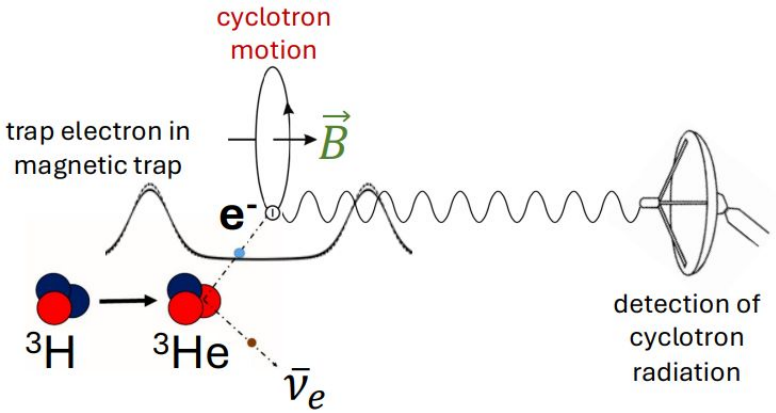


- **Allowed Fermi decay, Q** = 18.6 keV
- ***β*- spectrum, E_{max}** = 18.6 keV
- Present state of the art in direct measurement is $m_\beta \leq 0.45$ eV/c2 by KATRIN, Science 388:6743, 180-185.

- Distortion near E_{max} is affected by m_β, which reduces the maximum energy available to the electron and causes a characteristic distortion near the endpoint.



Project 8 Experimental Overview:
Cyclotron Radiation Emission Spectroscopy (CRES)



Hannah Binney

To measure

$$f_c = \frac{1}{2\pi} \frac{e\langle B \rangle}{m_e + E_{kin}/c^2}$$

Average magnetic field experienced by electron

Cyclotron frequency

Kinetic energy

- A novel approach invented by Monreal and Formaggio (PRD 80 (2009) 051301) and pioneered by the Project 8 Collaboration
- Detects the faint radio-frequency radiation emitted by the tritium electrons as they spiral in a magnetic field
- Electron energy is then determined directly and non-invasively via frequency measurement
- Mapping the electron energy spectrum near the endpoint reveals information about the neutrino mass.

Project 8 Experimental Overview: *Phased Approach*

Phase I *(Completed)*

- Proved CRES can detect GHz radiation from single electrons (using Krypton-83m instead of Tritium).

Phase II *(Completed)*

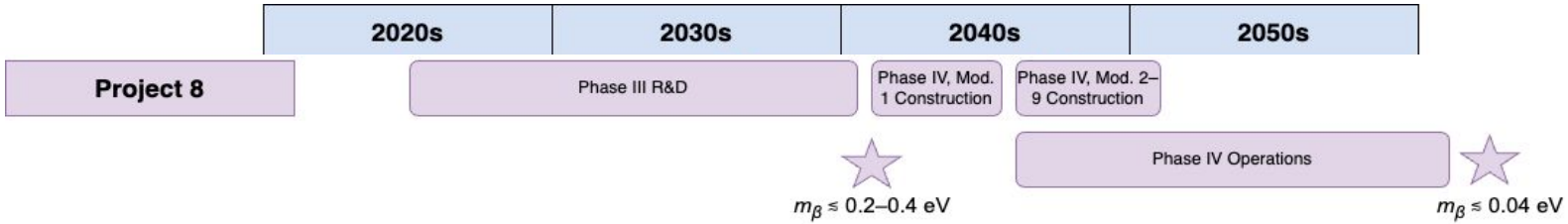
- Proved that CRES can work with the continuous energy spectrum of beta decay electrons from molecular tritium

Phase III *(In Progress)*

- **Objective:** Scale up from a small trap to a large-volume detection region capable of collecting enough events for meaningful neutrino mass limits.

Phase IV *(Future)*

- **Objective:** Final precision measurement; reach sensitivity to the neutrino mass at the level of ~40 meV or better



Data requirements

FAIR data principles: Findability, Accessibility, Interoperability, Reproducibility.

Experiment:

- Data rates of 10 MB/s.
- Data volume up to several PB.
- Heterogeneous data from multiple test stands, demonstrators, and simulation.
- Dynamic requirements of exploratory R&D, as well as automated stable production running.

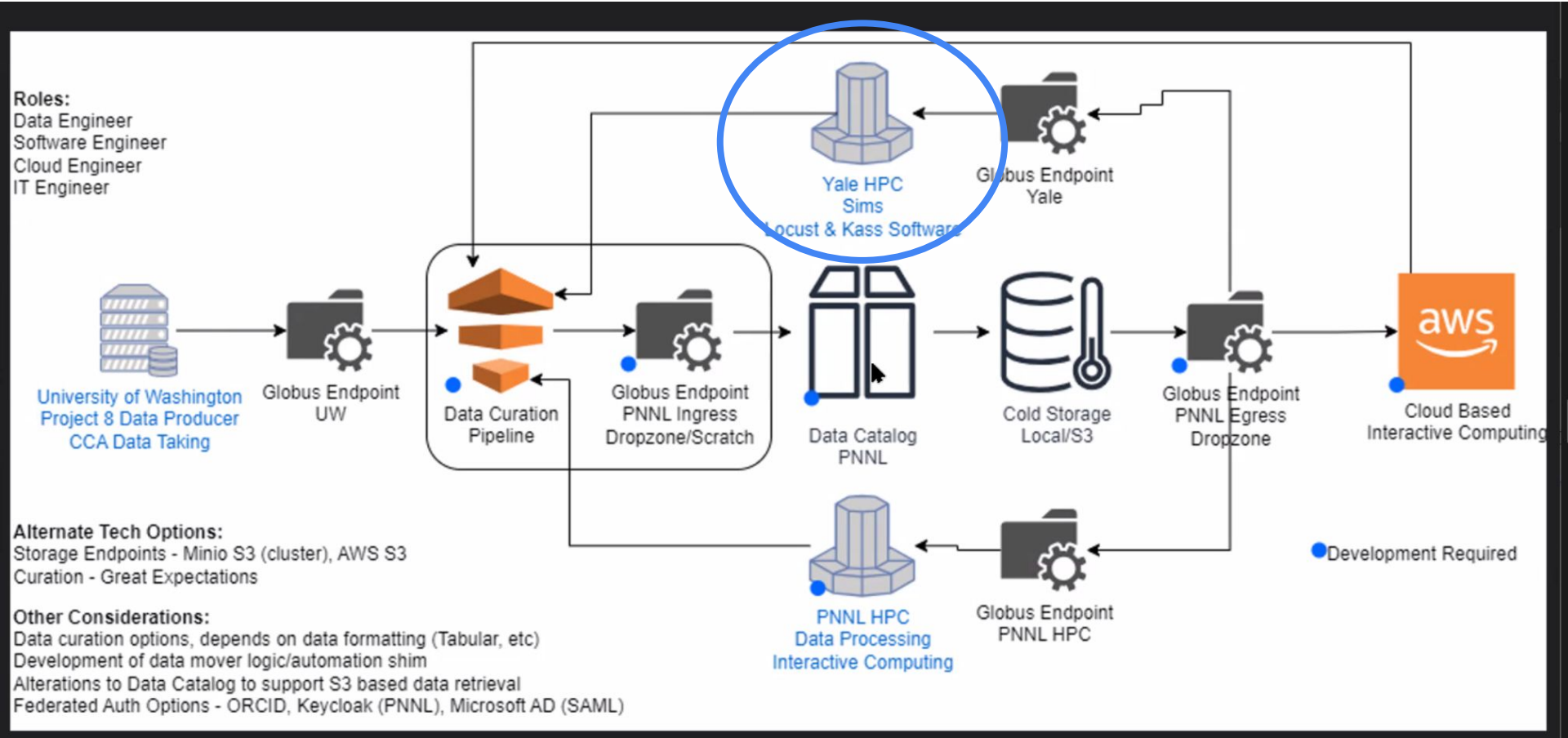
Catalog and compute:

- Need for an accessible catalog for viewing/querying the above data to understand what is available, what analyses have been done or could be done, etc.
- Need access to compute resources to analyze the above data, producing derived data products, both in coordination with arrival of new data, and with already existing data.

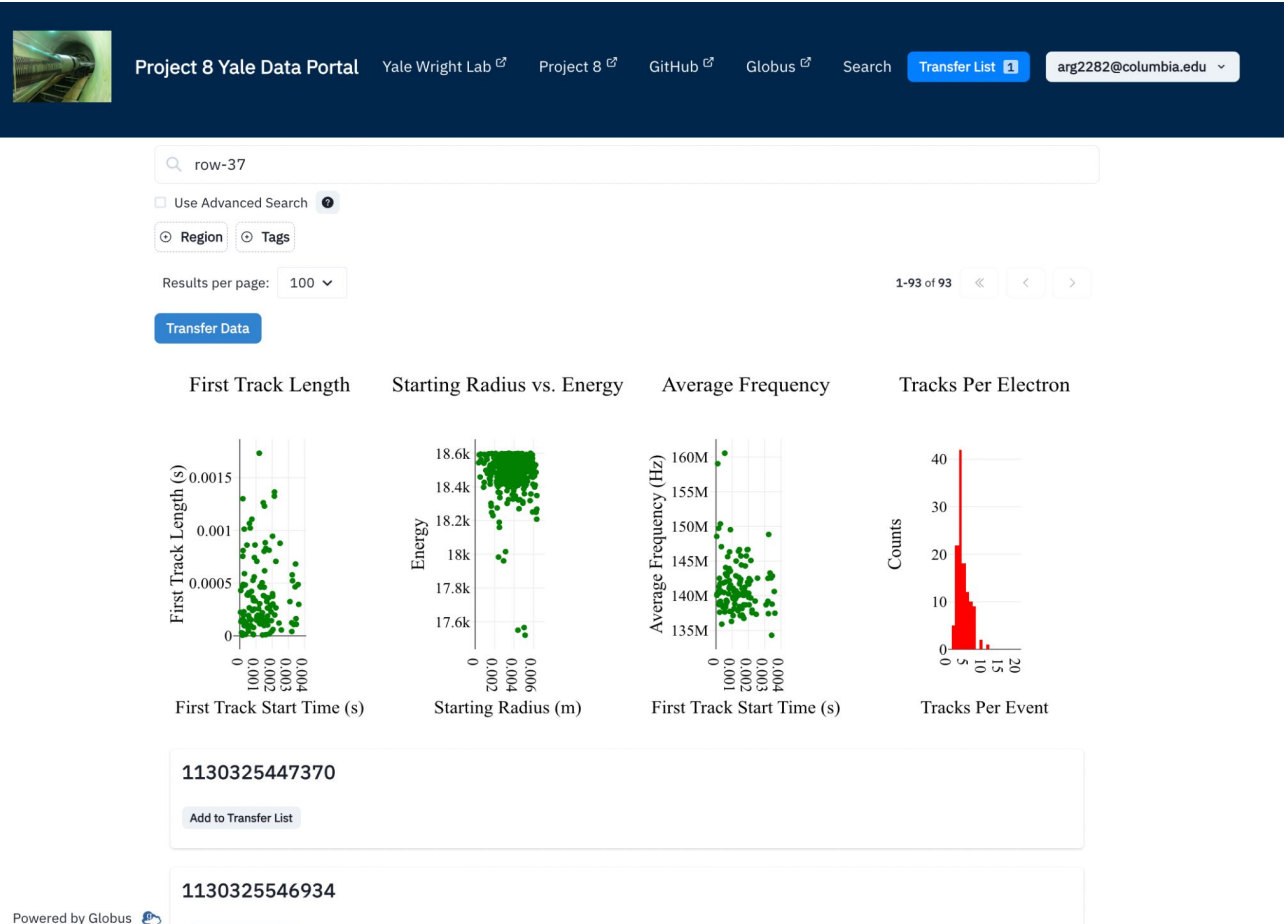
Globus as a Candidate Solution

- **Globus is a cloud-based service for secure, high-performance data transfer and management.**
- **Why Globus for the Project 8 Collaboration?**
  - Solves the problem of efficiently moving specific datasets between remote HPC storage and local compute resources.
  - Integrates easily with existing Linux-based systems through the Globus Command Line Interface, providing a foundation for building automated, reproducible workflows.
  - Eliminates the need for institution-specific credentials; access is managed through your existing Globus account and federated identity (e.g., university login, ORCID).

Project 8 Data Flow



Globus Data Portal



Advantages:

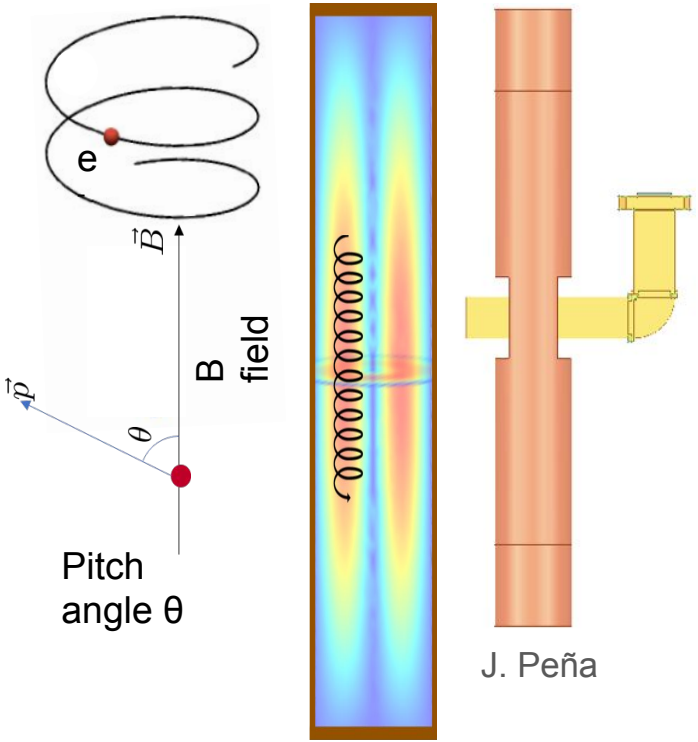
- Visualization of data available in search results.
- Access to raw data for transfers.

Present limitations:

- Requires manual clicking and web navigation; not scalable for repeated tasks.
- Limited batch handling; maximum of 100 files returned per query

Computing and Simulation

- Data sets are being generated at Yale on Grace HPC.
- Simulation output consists of raw data files with the same format as the experiment data files.
- Post-processing on the raw simulation output is required, just as it is in the lab experiment data.
- Post-processing workflows should be portable and accessible to collaborators.
- Present simulated data volume: Several TB.

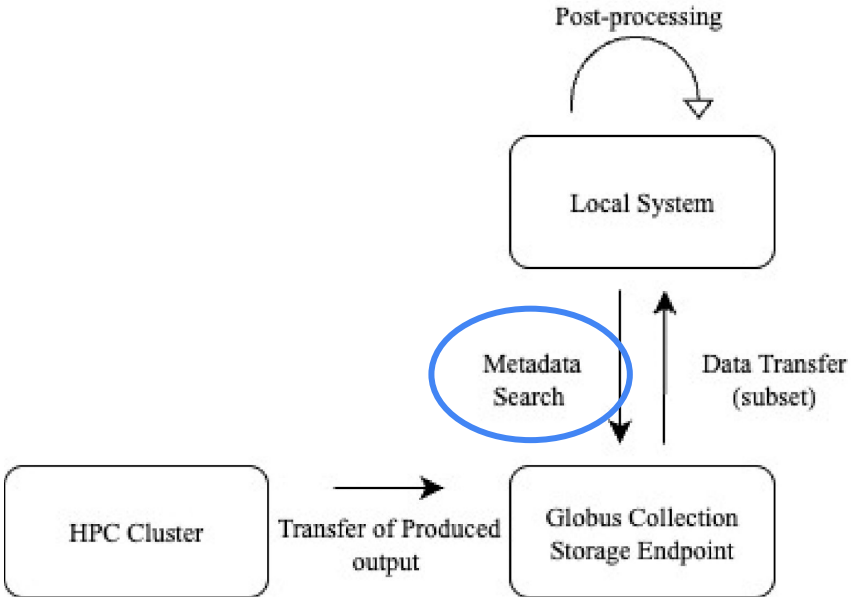


Metadata Search

- **Purpose:** Identifying only the relevant subset of files needed for analysis

- **Metadata Search Query (Globus CLI):**
  globus search query -q "<SEARCH_QUERY>" <SEARCH_UUID> --format JSON
  - **Where:**
    - **<SEARCH_QUERY> =**
    - **<SEARCH_UUID>** = Globus Search Collection UUID

- **Search Output:** A structured JSON file of specific file paths for targeted transfer

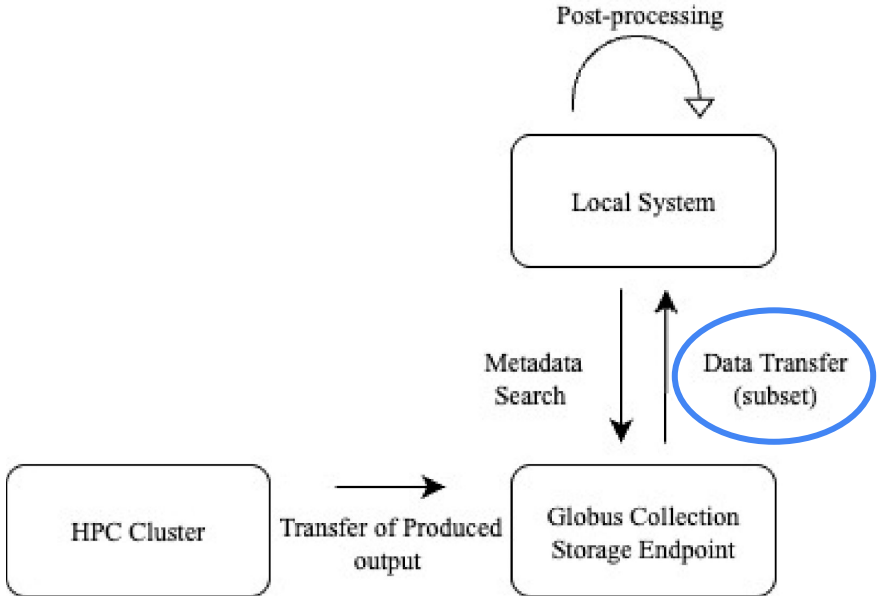


Data Transfer

- **Purpose:** Transferring only the targeted subset of data identified from the metadata search, avoiding transfer of the entire dataset (~500 TB).

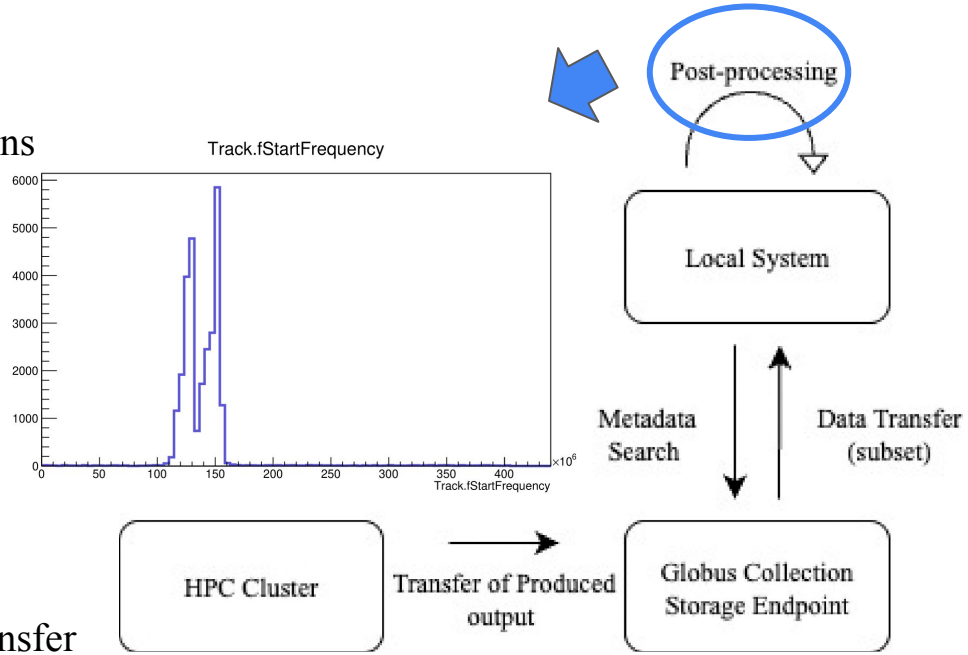
- **Source Endpoint:** Remote HPC storage collection
- **Destination Endpoint:** Local Machine or HPC cluster

- Uses file paths returned from the metadata search to initiate a Globus transfer task between endpoints via the Globus CLI



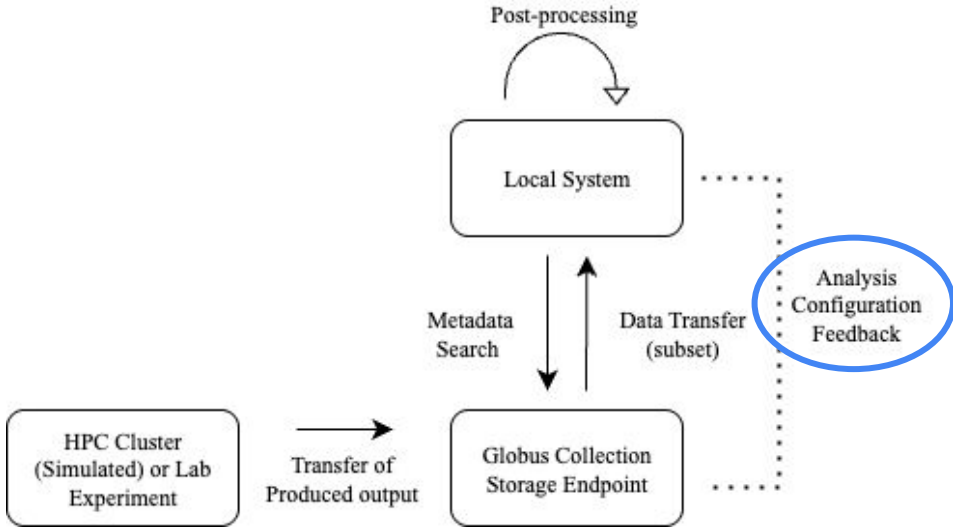
Post-processing

- Utilizes custom Project 8 software packaged inside Singularity containers
- Singularity containers ensure the software runs consistently across different Linux clusters, enabling portable and reproducible analysis without manual software installation
- These containers are integrated into a post-processing script using the Globus Command Line Interface
- The post-processing script runs after data transfer to process the selected datasets.



Next Steps

- Refine post-processing syntax based on goals of Project 8 working groups.
- Integrate new post-processing methods currently being developed within the Project 8 Collaboration, such as:
  - Machine learning models for enhanced analysis
  - Improved event detection techniques
  - Refinements for better energy resolution
- Integrate the post-processing outputs with the data framework, with the goal of building a feedback loop for optimized analysis



Wright Laboratory

Summary

- Initial tests of the Globus data management framework for Project 8 have been completed successfully.
- A modular and portable workflow is now in place.
- The next steps will be to develop and integrate post-processing techniques into the framework.