

February 11, 2026

Fermi Data Platform

Scarlet Norberg on Behalf of the FDP team



U.S. DEPARTMENT
of **ENERGY**

Fermi National Accelerator Laboratory is managed by
FermiForward for the U.S. Department of Energy Office of Science

Fermi Data Platform

- Core mission is to host and serve curated data sets for the Genesis mission
- Initially we will host HEP (Treasure), accelerator and nuclear physics data
- We are striving to move towards supporting a broader constituency within the DOE complex
- FDP provides:
 - Storage resources
 - WAN accessible
 - In-kind contributed disk storage
 - Accessible through Fermilab's Multi-Terabit ESnet connection
 - Currently deployed transfer and streaming protocols
 - WebDav
 - XRootD
 - Globus
- We can also help with publishing your data giving it a DOI number

Fermi Data Platform Access

- First get a Fermilab account
- Right now, you need to be related to the overall Genesis mission
- Request to Robert Harris: rharris@fnal.gov
- Current running of code some has happened on NERSC
- Can stream FDP data on to NERSC
- Working on getting Tokens for data access to be able to place data and get access to FDP for current experiments on FDP (eg: Treasure)

- To access the data, the user needs to be mapped to an account on the server. This is currently a manual process done on a case-by-case basis.
 - From globus online:
 - <https://www.globus.org/>
 - Log in with Fermilab credentials
 - Search for the “Fermi Data Platform” collection
 - The data is in the /amsc directory
 - From the command line:
 - `globus transfer [source-endpoint-ID]:[/path/to/source] [destination-ID]:[/path/to/destination]`
 - The Fermi Data Platform endpoint ID is b35955d3-14d1-4aab-a1c9-189989f7d8d0 and the path is /amsc

Webdav

- From browser:
 - <https://amsc.fnal.gov:2880/amsc>
- From command line:
 - `curl -L https://amsc.fnal.gov:2880/\[source_path\] -o [destination_path]`
 - The `source_path` is the path to the data on amsc
 - Using `gfal`, which requires the `gfal2-*` packages:
 - `gfal-ls https://amsc.fnal.gov:2880/amsc`
 - `gfal-copy https://amsc.fnal.gov:2880/amsc/\[source_path\] [destination_path]`

- Requires the *xrootd-client* package
- From the command line:
 - Copy a file:
 - `xrdcp root://amsc.fnal.gov/[source_path] [destination_path]`
 - List a directory:
 - `xrdfs root://amsc.fnal.gov ls -l /amsc`
 - Using *gfal*, which requires the *gfal2-** packages:
 - `gfal-ls root://amsc.fnal.gov/amsc`
 - `gfal-copy root://amsc.fnal.gov/amsc/[source_path] [destination_path]`
- `xrdcp -f root://amsc.fnal.gov/amsc/cms/aoj/data/RunH_batch38.h5 /dev/null`
 - `[2.515GB/2.515GB][100%][=====][107.3MB/s]`



Fermi Data Platform Areas Currently Working On

- Working on a meta data catalog
 - [Open Metadata](#)
 - [MetaCat Documentation link](#)
 - [invenioRDM](#)
- Local authentication and authorization framework
 - Currently only hosting public datasets -> authentication not necessary to access/transfer data
- Websites for information
- Streaming has been done successful with parquet and awkward array code
- Currently we require a Fermilab account to access the FDP, but when the Genesis mission/American Science Cloud include a token, we will work to accept that token to access data



Streaming

- Streaming data is the **continuous flow of real-time data from various sources**
- **We at FDP have worked to provide some code to stream data from FDP to NERSC**
- **Now provide code that using parquet files as input processes with awkward array from FDP to NERSC**
- **Also, now provided code for h5 files using h4py code to stream but has not been tested by treasure yet**

Meta Data Catalogs

- Trying Open MetaData, MetaCat and Zenodo light
- Seems like open metadata or MetaCat may suit our needs better but all are being tested

	Open Metadata	MetaCat	Zenodo
Good searching	✓	⚠	✓
Download stats	⚠	✗	✓
User-defined tags	✓	⚠	⚠ (keywords)
Dataset collections	✓	✓	⚠ (communities)
Data provenance	✓	✓	✗
Generate DOI	✗	✗	⚠
Access control	✓	⚠	✓

- Customizable metadata fields ✓
- Open source ✓
- Ability to connect with central AmSC catalog ✓
- Link to paper/license info ✓



Conclusion

- We are learning and developing the FDP
- We have a lot of experience and knowledge to help and work with teams
- We are here to help you run your workflows in the easiest way possible



Fermilab

Fermi *FORWARD*



U.S. DEPARTMENT
of ENERGY



Streaming

H5py

```
>>> import h5py
>>> import fsspec
>>> with fsspec.open('root://amsc.fnal.gov/amsc/cms/aoj/data/RunG_batch39.h5') as f:
...     with h5py.File(f) as h:
...         for name in h:
...             print(name)
```



Streaming

Awkward array with Parquet

```
import awkward
import os, sys

def iter_files(filename, columns=None):
    metadata = awkward.metadata_from_parquet(filename)
    row_groups = metadata['num_row_groups']
    for rg in range(row_groups):
        data = awkward.from_parquet(filename, columns=columns, row_groups=set((rg,)))
        for d in data:
            yield d

def test(fname):
    print("\n Attempting to read FPD file %s" % fname)
    data = iter_files(fname)
    print("\n First element: ")
    print(next(data))
    print("\n Second element: ")
    print(next(data))
    print("\n Success!")

    return

if __name__ == "__main__":

    test(sys.argv[1])
```



01

Section title