



## DUNE Introduction to Data Management Projects

Mike Kirby

DUNE US Computing Consortium Startup Workshop

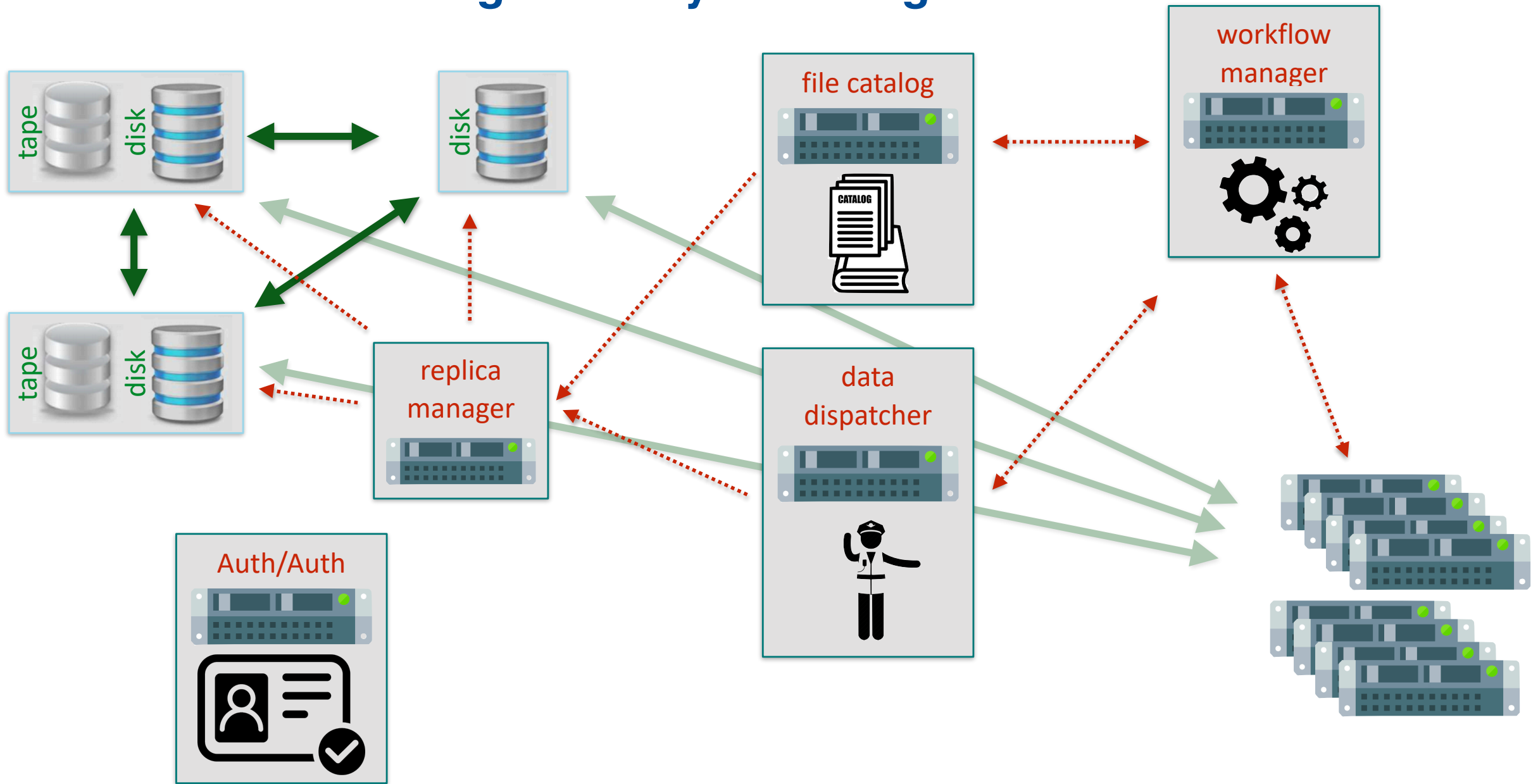
14 Dec 2021

# Introduction to DUNE Data Management

- Overall system diagram
- Data distribution strategy and tactics for both disk and tape
- Projected data volumes over the next decade - ProtoDUNE and DUNE
- Current status of project development
- Fancy monitoring screen shots
- Opportunities for contributions



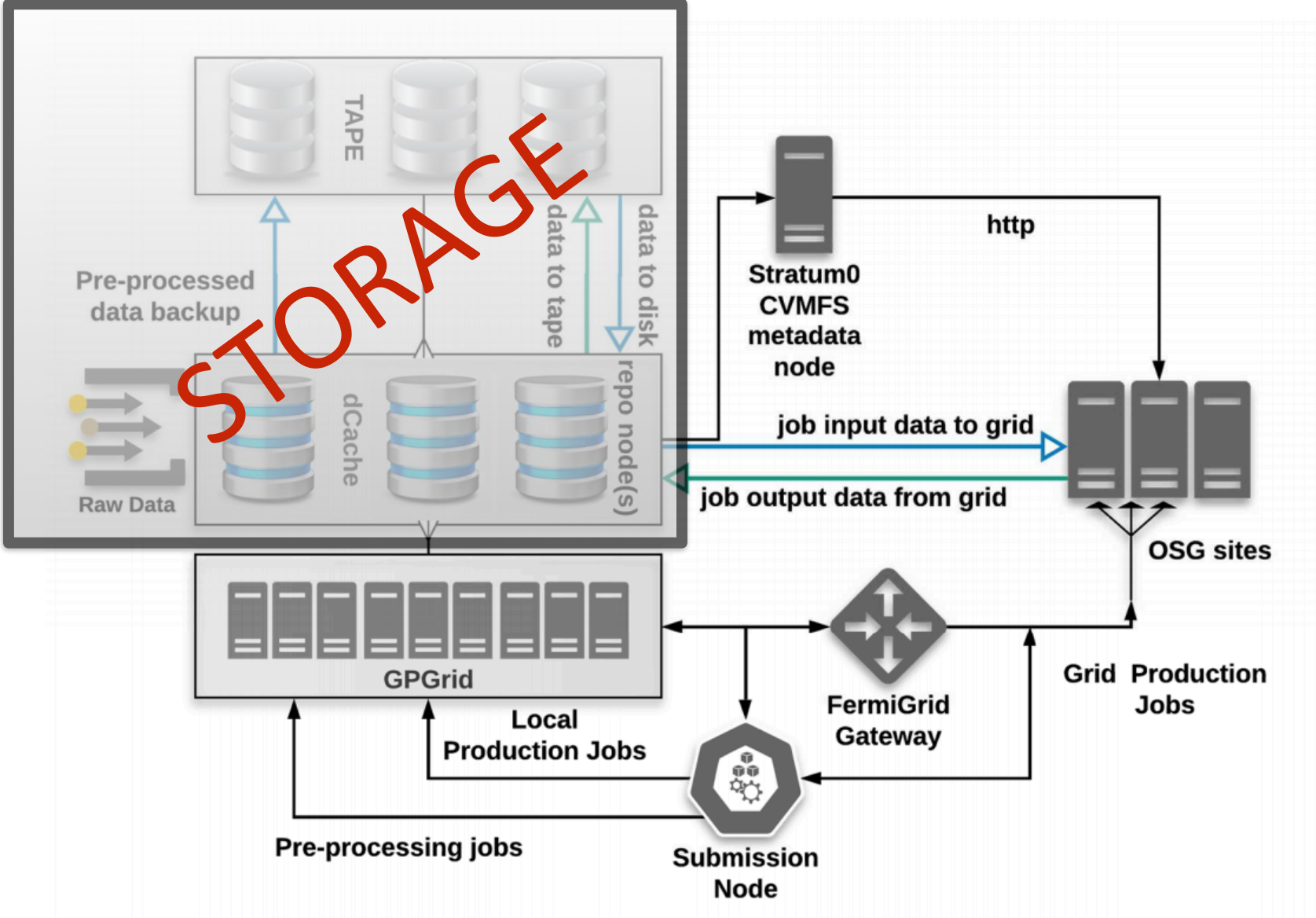
# Overview DUNE Data Management System Diagram





# Overview DUNE Data Management System Diagram

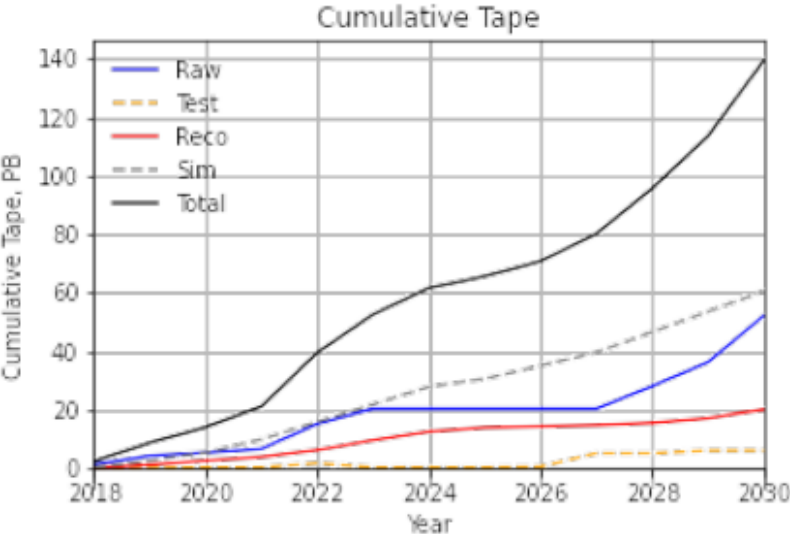
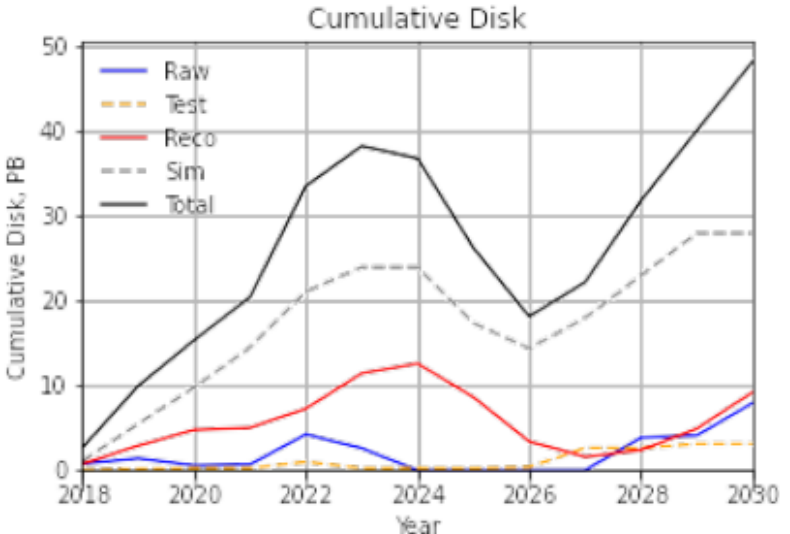
Hopefully make it closer to the situation that the production or analysis users just see “storage”



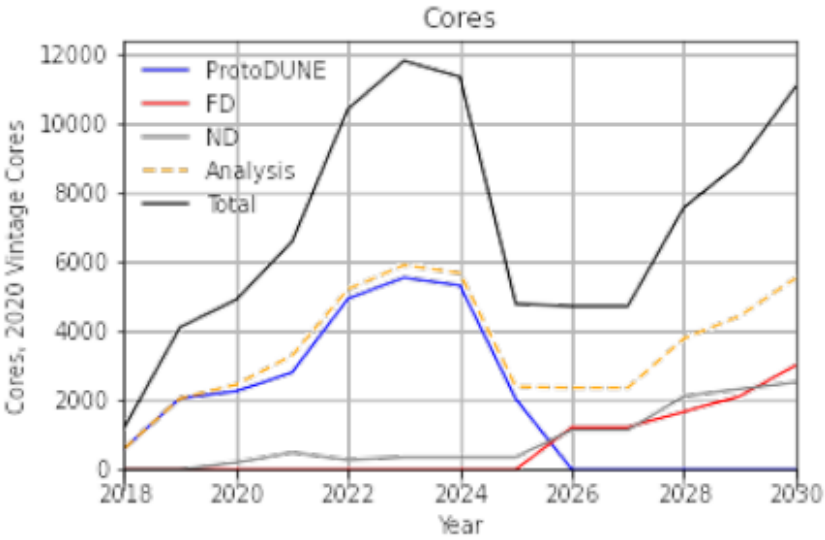
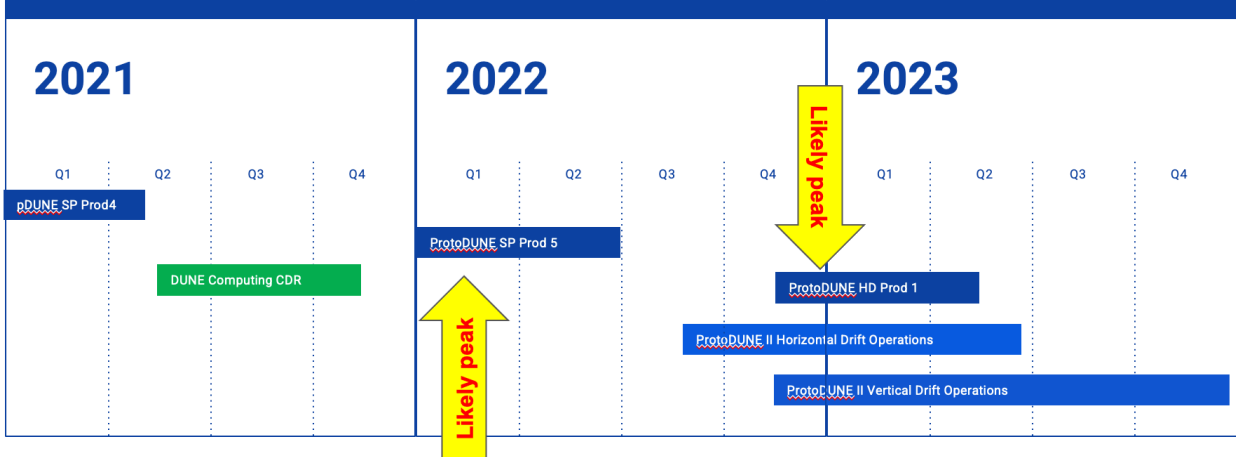
# Strategy and Tactics for Data Distribution

- **primary task - maintain access to all raw data and be certain that no data is lost**
  - utilize tape and disk storage to both store data and provide access to files
  - keep multiple copies of files when appropriate
  - make sure there are no “dark” files/data - things that we don’t have cataloged
- **secondary task - provide tools to locate and efficiently deliver files for processing and analysis**
  - catalog all files and metadata for both discovery and delivery of files
  - keep copies of relevant files on disk for efficient access
- 2 copies of raw data on tape at separate locations
- 1 copy of “test” data stored for 6 months
- 1+ copy of reco/sim on tape
  - currently assume 1 reco and 1 sim pass/year but reco passes go over previous years for each data set
- 2 disk copies of current reco and sim
  - assume reco/sim resident on disk 2 years
  - impose shorter lifetimes on tests and intermediate sim steps
- every data file put onto storage is registers with Rucio (minimum) and Metacat

# Projected Data volumes going forward



3/25/21 Computing

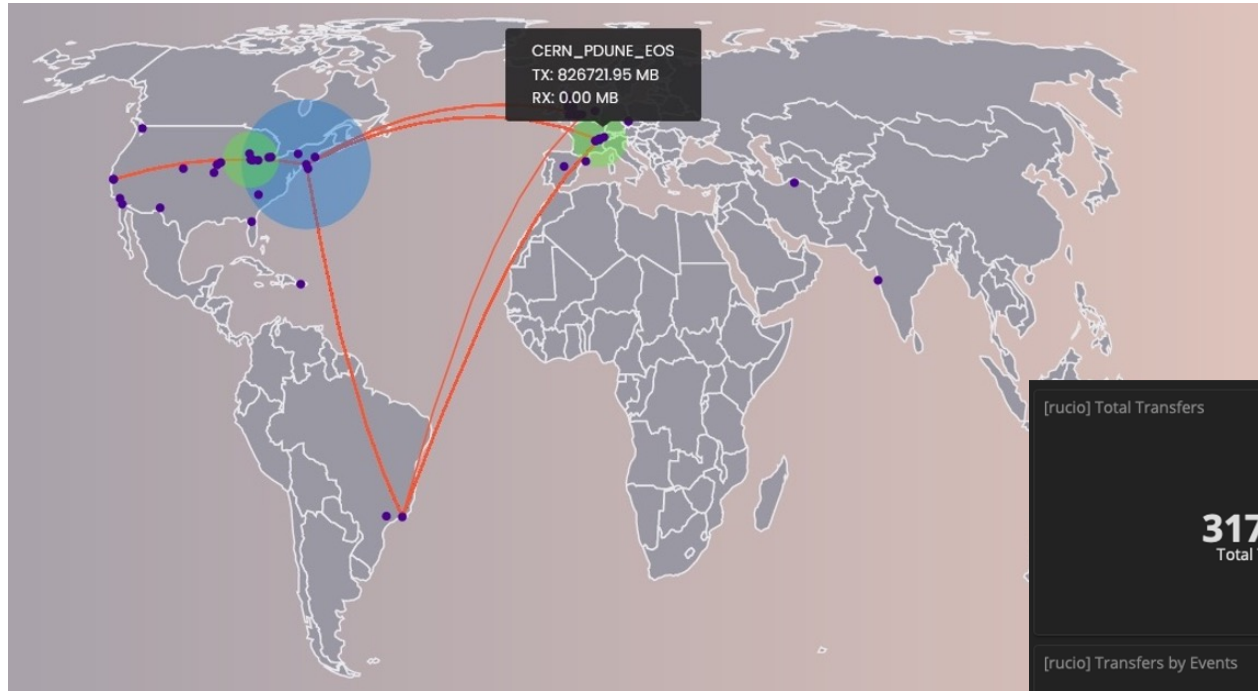


<https://docs.dunescience.org/cgi-bin/private/ShowDocument?docid=20515>

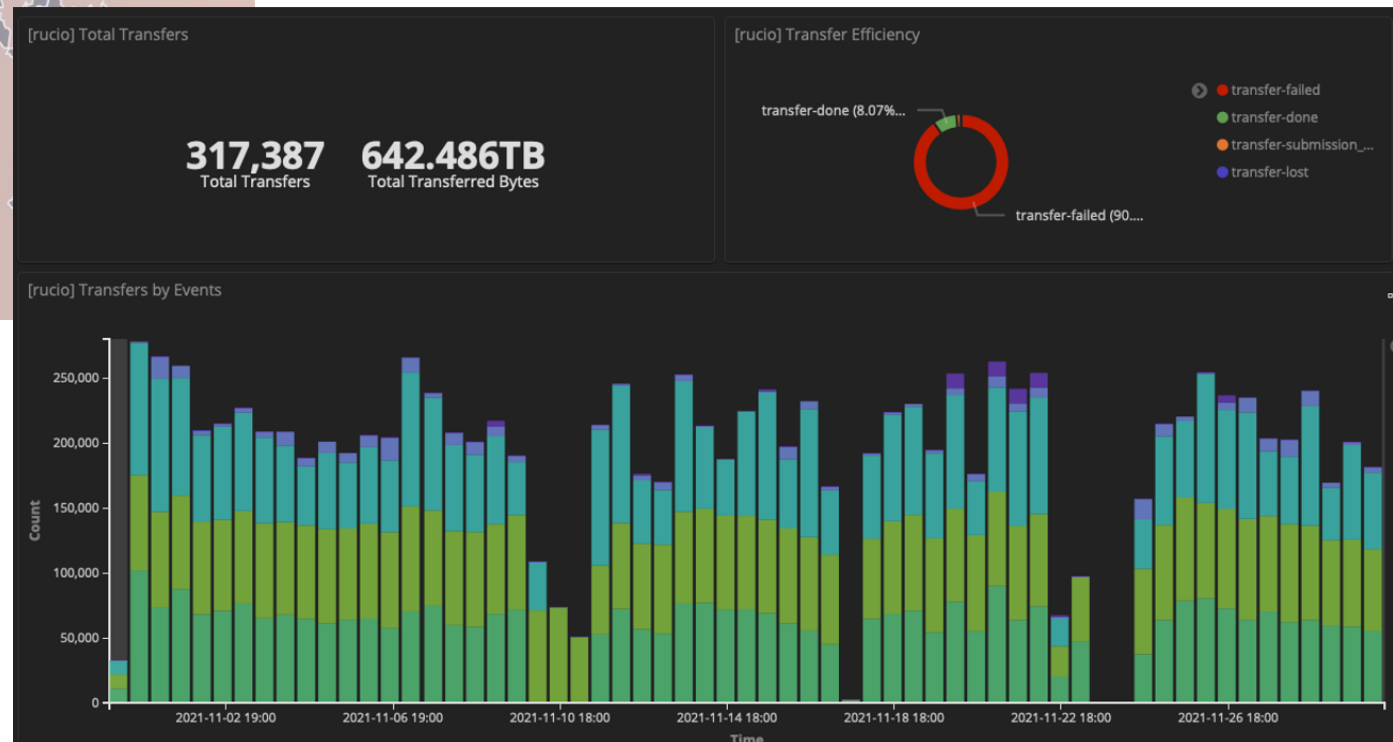
# Current status of Data Management R&D

- Rucio is in production but still in need of some integration
  - https 3rd party transfers
  - token based authentication/authorization
  - user client and interface (???)
  - localization for data delivery
- Data Dispatcher
  - requirements document is in development
  - initial design needs to match those requirements once finalized
  - integrate with workflow and production operations
- New Data Streams from ProtoDUNE II
  - VD (NP02) Coldbox top electronics
  - VD (NP02) Coldbox bottom electronics
  - NP02 cryostat legacy photon detectors (HV tests)
  - NP02 cryostat Arapucas + SiPM photon detectors
  - HD coldbox (APA testing)
  - NP04/NP02 electronics test stand
- Monitoring of data locations and transfers
- Data challenges upcoming for new RSEs
- Effort could be absorbed by all of these efforts

# Monitoring



Averaged 10,000 transfers and 20 TB/day  
~ 1.5 PB total transfer to 5 UK disk SE with Rucio  
(RAL, Manchester, Lancaster, QMUL)



DUNE Transfers FNAL -> UK in Nov 2021