

Breakout Session 7:

Medical Image Processing and Structured Storage

Dr. Govind Bhagavatheeshwaran
Staff Scientist, NINDS/NIH

Medical Image Processing and Structured Storage

PI: Daniel S. Reich, NINDS, NIH

Software Engineer: John Ostuni

Presented by: Govind Bhagavatheeshwaran

AWS support: Tom Fonseca et. al.

Funding

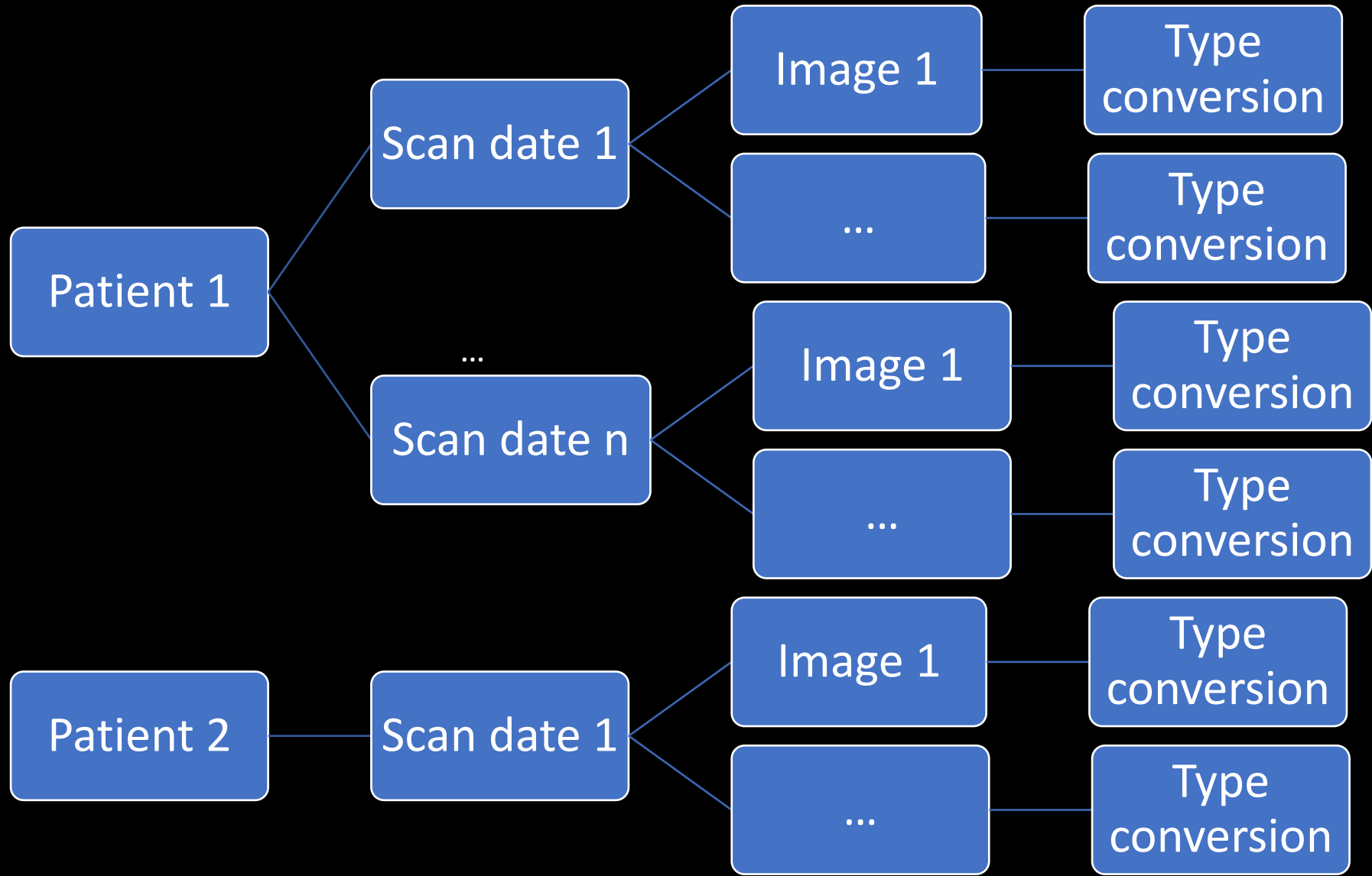
FY21 ODSS High-Value Datasets award – till 9/2023

Goals

1. Deep-storage of all primary MRI data acquired in the Reich lab (historically of trial participants seen through the Neuroimmunology Branch)
2. Catalog the data, associate with meta-data
3. Run common preprocessing steps to make data user ready

NINDS – Neuroimmunology Branch/Clinic

- Runs clinical trials for 3 labs (branch established in 1975)
 - Multiple sclerosis (MS)
 - HTLV-1 Associated Myelopathy (HAM/TSP)
 - Progressive Multifocal Leukoencephalopathy (PML)
- MRI Data
 - Participants of clinical trials (from the early 1990s – 40+ years of data)
 - Animal models
 - Postmortem imaging data
- Recent addition of small nuclear RNA data (5 years)



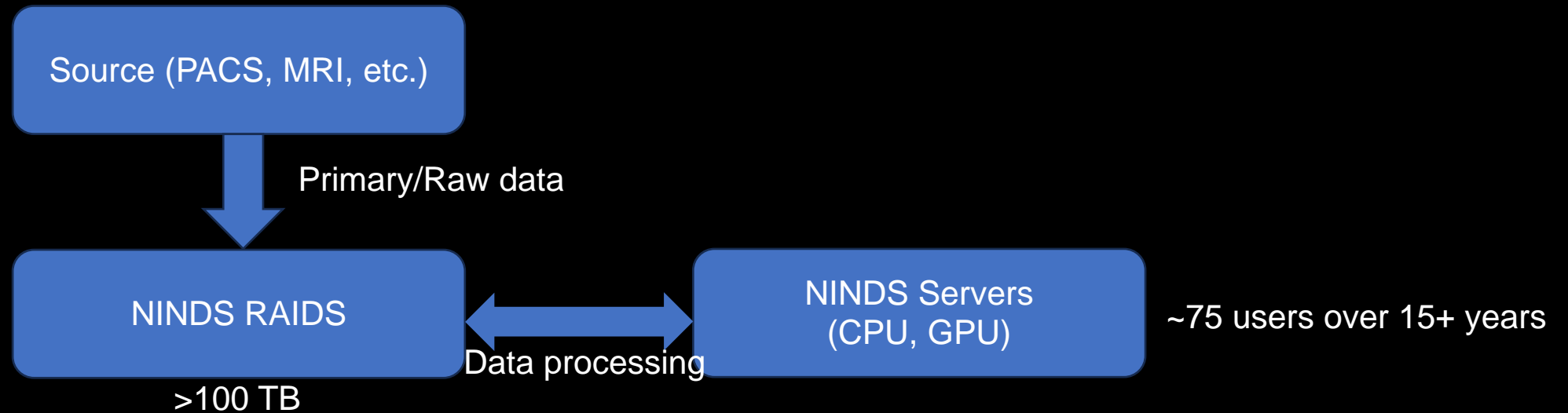
Meta data – say, demographics

clinical status

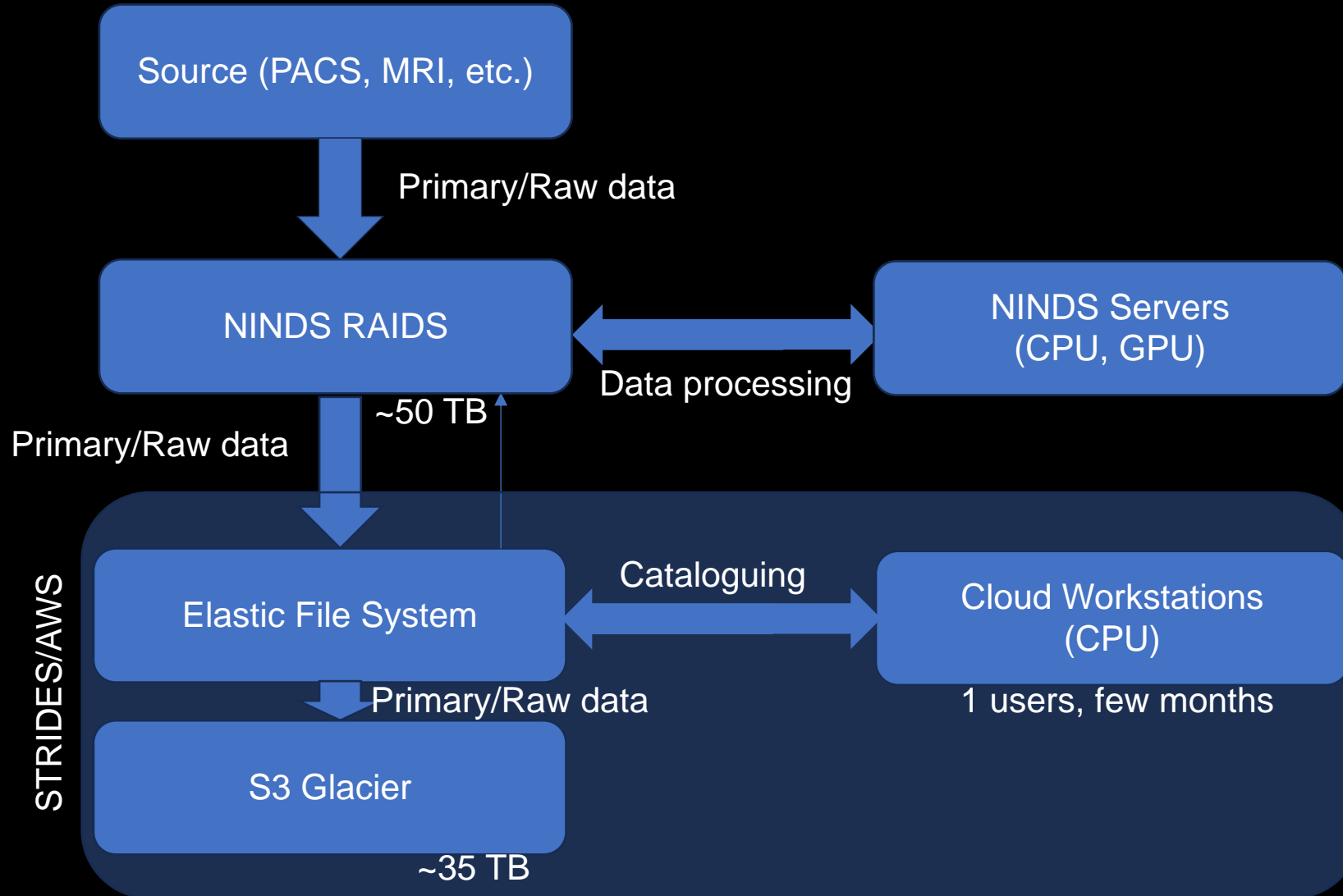
Scanner version,
software version,
sequence version

Acquisition protocol

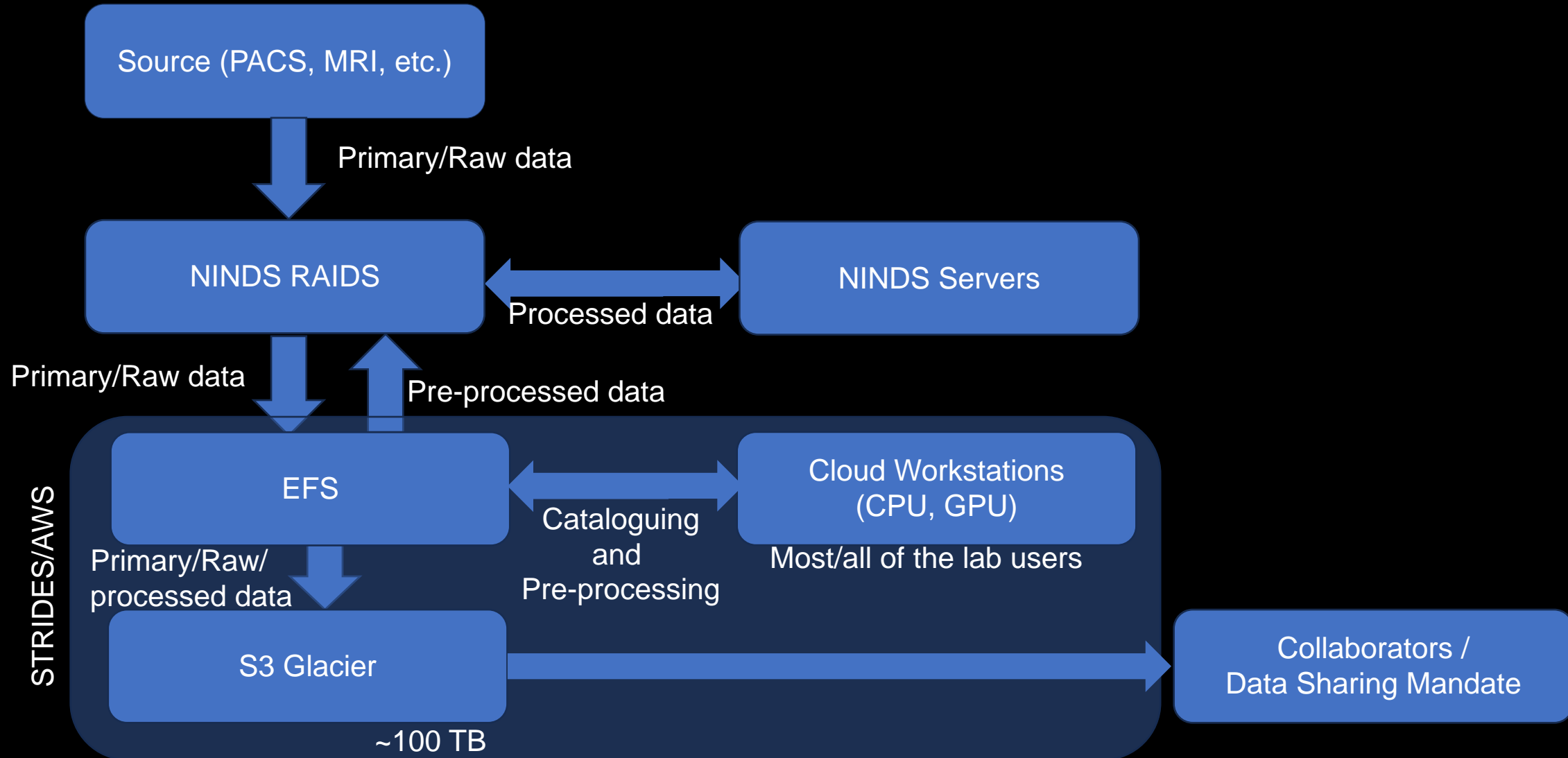
Data Flow - Traditional



Cloud Implementation



Vision for Data Flow Setup (Ideal)



Scientific Computing Processes

- Common pre-processes
 - Cataloguing
 - Skull stripping
 - Brain and spinal cord segmentation
 - Longitudinal subtraction
 - De-identification (for data sharing mandate)
- Implemented using containers
- Will increase NINDS-RAIDS usage, but reduce NINDS compute resource usage in the short term
- Cost is a major concern for full implementation.

Progress

- ~35 TB of primary data transferred to S3 Glacier Deep Archive
- Cataloged, some meta data added.
 - 10 Linux workstations for 2 months + EFS
 - Few months to complete
- Other preprocessing work is pending
- Feasibility of opening up selected data to outside users

Budget

- Estimate
 - \$1,000 to catalogue and deep storage
 - \$15,000 per year for basic preprocessing
- Actual cost
 - < \$100 per month for storage
 - > \$2,500+ per month for processing
 - Not accounting for personnel cost

Lessons Learned

- Shovel ready project, stable implementation is essential
 - The award was short term, required extension
 - Dedicated or partial effort of *software engineer (cloud certification)*
- Moving to the cloud is more expensive in the short term
 - >15 users at any given time
 - Currently, hardware exists in the lab to support use

Lessons Learned

- Availability of latest hardware/software on demand on the cloud is a major advantage
 - Biggest advantage is flexibility in machine specs and storage needs (implemented in minutes)
 - Phased-transition into cloud computing as existing hardware gets outdated (needs further exploration, costing)
- Looking forward to
 - Cloud based collaborations
 - Implementation of GPU based computing on cloud

Thank You!