

Summary of Data Management Principles

Daya Bay Experiment

Experiment description: The Daya Bay Reactor Neutrino Experiment, a multinational collaboration that operates in the south of China, announced on March 8, 2012, its measurement of the disappearance of electron antineutrinos emitted by nuclear reactors operated by China Guangdong Nuclear Power Group. U.S. and Chinese physicists built six underground detectors buried in the mountains near the nuclear reactors. This was the first time that the disappearance of reactor neutrinos has been measured. The experiment provided the first measurement and the best measurement of the neutrino oscillation mixing angle θ_{13} . The experiment started data taking in December 2011, and expects to continue data taking through September 2017.

DOE's roles in the experiment: The DOE is the lead US agency for Daya Bay for design, fabrication, installation, commissioning and operations.

Partnerships: The multinational collaboration includes researchers from China, Chile, the United States, Taiwan, Russia, and the Czech Republic. The US side of the project is funded by the US Department of Energy's Office of High Energy Physics.

Organization – Agency/Lab level: Lawrence Berkeley National Laboratory (LBNL) is the lead laboratory for Daya Bay.

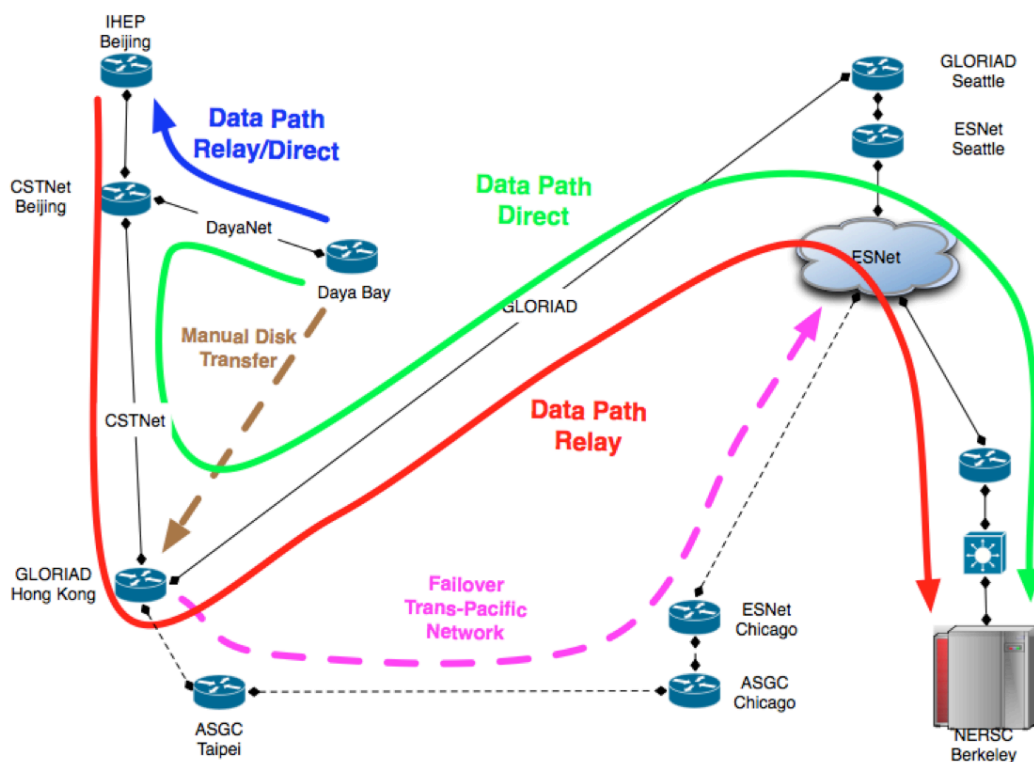
Organization – Experiment level: See the Daya Bay Project organization chart below. The Daya Bay collaboration is led by elected spokespersons from the USA and China who are advised by Executive and Institutional Boards. Scientific leaders on the Analysis Coordination Committee (ACC) and technical leaders from the project Technical Advisory Board provide guidance and leadership to specific subsystems including: Antineutrino Detectors; Muon Veto (Water Pool), Online Systems, and Offline Computing & Software.

Collaboration: At present, the Daya Bay collaboration is composed of 42 member institutions and includes about 254 scientists, students and engineers.

Data Policy Management: The Daya Bay leads for Offline Computing & Software (Craig E. Tull - LBNL, Weidong Li – IHEP) are responsible for maintaining the computing infrastructure and software organization in consultation and cooperation with the Daya Bay collaboration. This includes data storage, processing, archiving and data releases.

Data Description & Processing: Daya Bay has been in continual data taking and Keep Up Processing (KUP) mode since data taking started in 2011, with interruptions only for the installation of the last 2 ADs in 2012. Data are taken at Daya Bay (north-east of Hong Kong) and shipped automatically across Chinese and US national networks, and trans-Pacific, international networks to 2 data processing facilities. (See Figure below) One facility is in Beijing (IHEP) and

another in Berkeley, CA, USA (NERSC/PDSF). Raw data are stored on both spinning hard drives and on tape at both facilities.



All science critical and irreproducible data produced during the course of the Daya Bay experiment are archived on NERSC's HPSS tape system and will be available for a minimum of 3 years after the end of data taking. Science critical and irreproducible data include raw data (~150 TB/year) recorded by the Data Acquisition system (DAQ), sensor data recorded by the Detector Control System (DCS), derived calibration parameters and reactor information stored in the Offline control database, and any production reconstructed data used for scientific publications and/or public announcements of scientific results.

In the last half of calendar year 2015, data transfer (Spade) and Keep Up Production (KUP) transferred and processed 70,958 DAQ files (58.9 TB) bringing the total DAQ data since 24 December 2011 to 527,379 files (446 TB). NERSC disk space deployed was 1177 TB, of which 1133 TB was in use for data. Daya Bay will continue to generate ~150 TB of raw data per year through the end of data taking. Data stored on disk at NERSC are available for all collaborators. Data Stored on HPSS at NERSC will be available for 3 years beyond data taking ends. At that time, a renegotiation with NERSC storage and management may extend that period.

Data Products and Releases: The type of data produced in Daya Bay includes raw data from testing, calibration, and measurements at participating institutions and at Daya Bay and digital and graphical results from analysis of raw data and Monte Carlo simulations. The

corresponding metadata includes design documentations, detector operating parameters, calibration data, analysis tools, student theses, and materials for education outreach activities. All Daya Bay software (including both online and offline code) is centrally maintained through an SVN software repository at IHEP. Regular (~yearly) production runs of all data taken are preceded by a software release with all newest algorithms and calibrations included.

Plan for Serving Data to the Collaboration and Community: Data on the NERSC and IHEP compute facilities are available for all collaborators in the experiment, through one of 3 methods:

1. Direct access from batch and interactive jobs on the facility compute clusters
2. Direct download from the Offline Data Monitor (ODM) at:
<https://portal-auth.nersc.gov/dayabay/odm/>
3. Spade-negotiated data transfers as arranged by collaborating institutions

Funding support does not exist and is not planned to provide software tools for community access to Daya Bay data. However, if such funds were available, we would plan to stand up a web service which would provide the raw numbers used in any public science results (such as the values of points, fits, and curves displayed in graphs, plots, or tables), and the provenance information necessary to completely identify and reproduce the software release, rollback date of the Offline DB, dataset definition, and production configuration used to derive those scientific results.

Plan for Archiving Data: Raw data are archived at NERSC and at the Chinese data centers along with appropriate databases and software releases. NERSC's HPSS (High Performance Storage System) will keep the data available for at least 3 years beyond data taking. A similar time period is expected at the Chinese data center.

Plan for Making Data Used in Publications Available: After the publication of significant science and engineering results, these results will be made available to the scientific community and public at no cost via a web service maintained by the collaboration. Publications will be posted on a publicly accessible website.

Responsiveness to SC Statement on Digital Data Management: It is the intent of the Daya Bay collaboration to comply with the Statement on Digital Data Management (<http://science.energy.gov/funding-opportunities/digital-data-management/>) within the constraints of limited funding and resources.