PROSPECT Data Management Plan

The PI, co-PIs, and senior investigators associated with the PROSPECT experiment are strongly committed to compliance with DOE policies on the preservation, dissemination, and sharing of research data. This data management plan addresses the specific requirements of the PROSPECT project. In addressing the associated challenges we will make use of the extensive knowledge base and support for data management at the collaborating institutions of the PROSPECT experiment.

The PROSPECT Data Management Plan establishes the following goals to respond to the DOE Statement on Digital Data Management [1].

1. Research data should be accessible to the public for all published results.
2. The provenance of all published results should be recorded so that results can be reproduced.
3. The raw, simulated and processed data, code repository, database and associated documentation should be managed, maintained and archived for a reasonable period of time.
4. Long-term preservation of some form of the data is desirable for a reasonable period of time.

1 General Data Management Policies

The principal and senior investigators will take several steps to ensure compliance with the data management policies of the DOE, the university, and their individual home institutions. All collaborators will be informed that the work supported by the DOE must follow the guidelines for data protection and storage, abide by DOE policies and other laws regulating intellectual property, and acquire and maintain licenses for all software and other data used as part of the their activities. We will also inform PROSPECT collaborators that results from the experiment are intended for publication in the open literature.

2 PROSPECT Specific Data Management Policies and Activities

The research work of the collaboration will generate a variety of data in different formats. The collaboration management will provide resources to assist researchers in good management practices of research data and will issue guidelines for the storage, preservation, and dissemination of data. This includes standards on appropriate file formats and the creation of metadata for data preservation, sharing, and archiving.

2.1 Published Results and Supporting Data

Results from the experiment will be published in peer reviewed journals and made available through the arXiv [2] and on PROSPECT's website [3]. Data for publication figures and tables will be supplied as ROOT files and/or flat text files as appropriate. We will use services provided by INSPIRE-HEP [4] and publishers such as Physical Review that support ancillary data preservation.

PROSPECT Collaboration

Achieving the goal of providing reproducible results requires versioning of the simulated and processed data, software and database as well as establishing the relationship between the raw and processed data. It also requires documentation. PROSPECT has established a software repository in GitHub that supports versioning. Currently PROSPECT software uses external packages ROOT, GEANT4, xercesc, python and cmake that are widely used and supported in physics (ROOT, GEANT4) and software communities.

PROSPECT will use a centralized database that enables replication, roll back and access control for slow control, calibration and other data. MySQL is an open source and widely used in general and within the physics community, in particular.

2.2 Documentation

Documentation of the PROSPECT project is maintained in DocDB (technical notes, presentations, etc.), ELOG (electronic logbook), a wiki, and email lists. The provenance of all processed files (original file, software version, database version, etc.) will be recorded in the each file as well as in the database as is done in the Daya Bay experiment, for example. DocDB, elog, the wiki and email lists are maintained and backed up by Yale Wright Laboratory.

2.3 Experimental Data

Data from the detector will be handled and achieved by use of an existing, widely adopted and well-supported library, such as HDF5 or ROOT for processed and summary data files. The long-term storage of data is discussed below.

3 Access, Sharing, and Archiving

Research data and results produced as part of the collaboration's activities will follow the general open access model while protecting the intellectual property of the researchers involved. Patents and inventions follow the rules and regulations of collaborating institutions and federally funded research.

Data will be stored and hosted on collaborating institutions' High Performance Storage Systems (HPSS). Collaborating institutions will have open access to all PROSPECT data, limited only by their ability to accept it.

A system to efficiently transfer select data samples will be employed. At the end of each run the DAQ will notify an offline data movement system, such as SPADE, that will transfer the available files to collaborating institutions where they will be both archived on the HPSS tape storage system and made available on disk. At their request, collaborators will be given direct access to those computing resources. Requests for raw data from the scientific community and the public will be discussed and managed on a case-by-case basis by the collaboration management. Supporting data for published results will be made available along with the publication as described above.

The PROSPECT collaboration will closely work with computing managers at collaborating institutions to set up systems in agreement with University and DOE policies. Data will be archived for 3 years following the conclusion of the research activity and publication of results.

Revised June 22, 2016

PROSPECT Collaboration

PROSPECT data will consist of zero-suppressed PMT waveforms that will be processed to summarize the essential features of the waveforms. Raw, simulated and processed data will be archived at collaborating institutions. Given the expected total volume, only a single copy of the raw data will be archived. Processed summary files of raw and simulated data will be available to all collaborators. An automated system to receive, validate and process each raw data file upon receipt at collaborating institutions will be developed.

PROSPECT will adopt a tiered approach to long-term data storage depending on available resources at collaborating institutions. Archiving data beyond the planned 3-year storage period will depend on available resources. Preservation of published research data has been described above.

References

1. Statement on Digital Data Management, Office Of Science, U.S. Department of Energy; http://science.energy.gov/funding-opportunities/digital-data-management/
2. http://arxiv.org
3. PROSPECT website, http://prospect.yale.edu
4. INSPIRE-HEP, http://inspirehep.net/