

Summary of Data Management Principles LZ Experiment

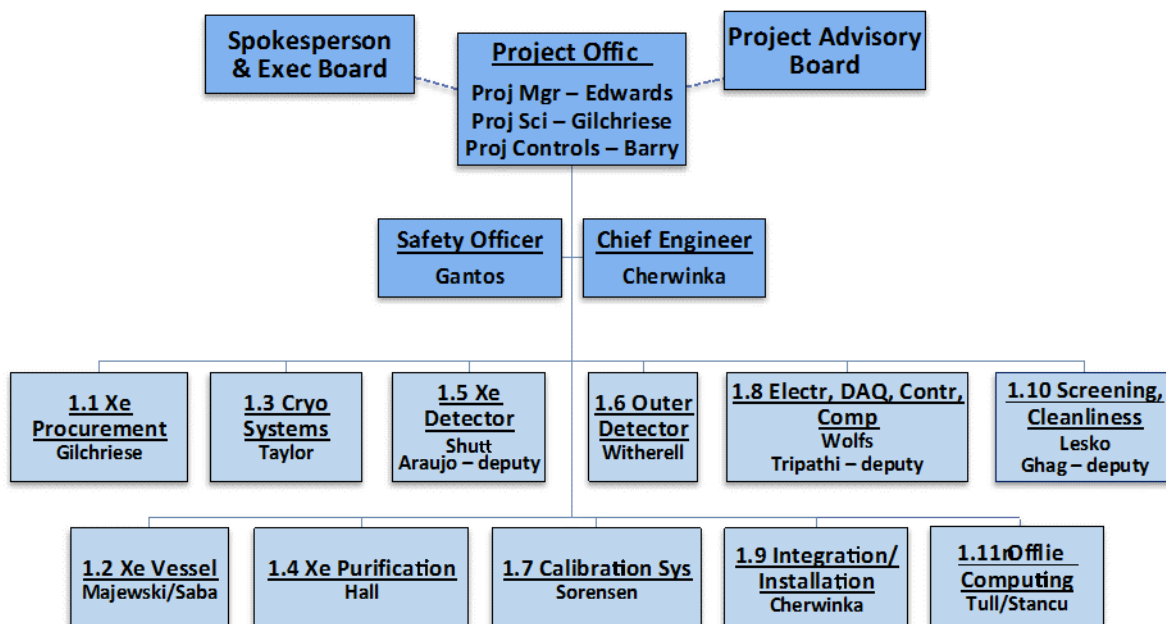
Experiment description: The LUX-ZEPLIN (LZ) experiment searches for direct dark matter signals using a liquid Xenon target and dual-phase time projection chamber. LZ is located at the Sanford Underground Research Facility (SURF) in Lead, SD, USA. The LZ DOE construction project is expected to complete in 2019 and begin collecting research quality data in 2019 after a ~6 month commissioning phase.

DOE’s roles in the experiment: The DOE is the lead agency for LZ for design, fabrication, installation, commissioning and operations.

Partnerships: The United Kingdom is a major partner in the LZ experiment. The South Dakota Science and Technology Authority (SDSTA) is also a major partner. Smaller contributions are expected from Portugal and Russia.

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

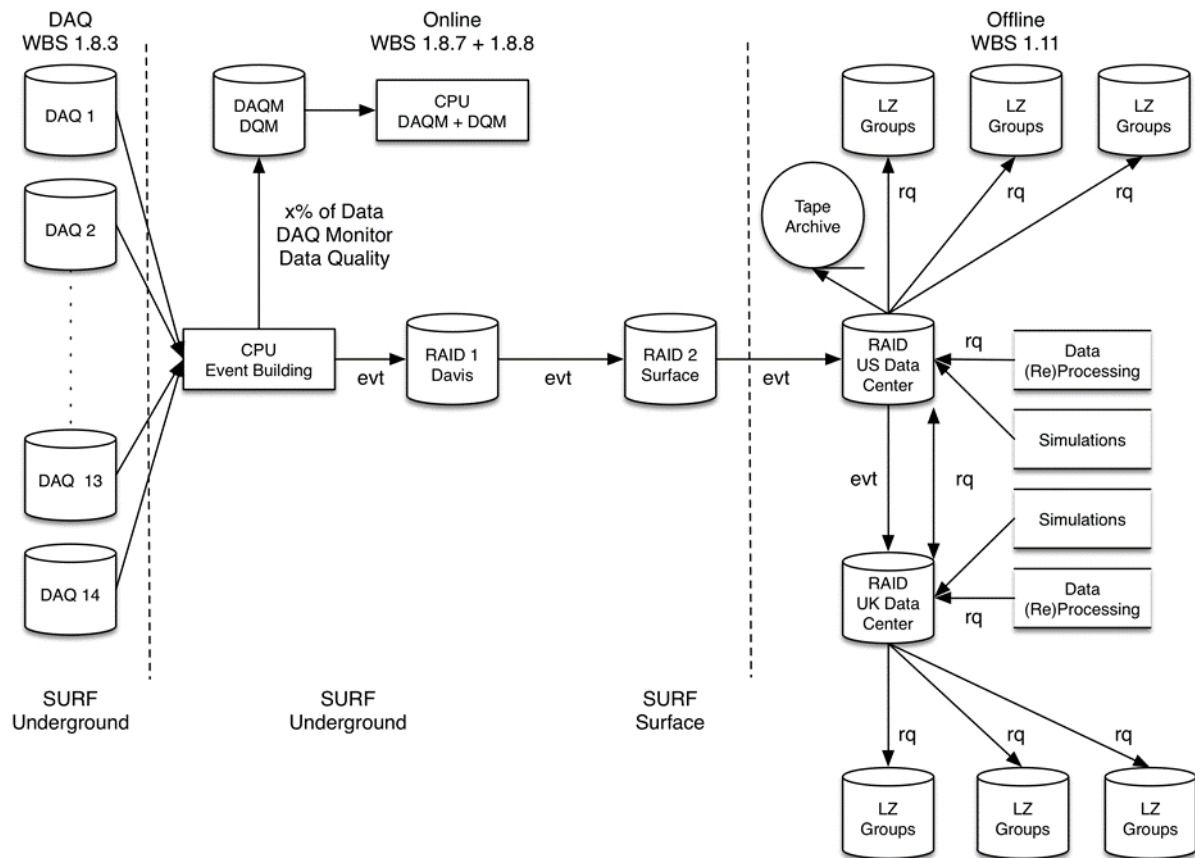
Organization – Experiment level: See the LZ Project organization chart below. The LZ collaboration is led by an elected spokesperson who is advised by Executive and Institutional Boards.



Collaboration: At present, the LZ collaboration is composed of 31 member institutions and includes about 175 scientists, students and engineers.

Data Policy Management: The LZ project management team is responsible for establishing the computing infrastructure and software organization in consultation and cooperation with the LZ collaboration. This includes data storage, processing, archiving and data releases.

Data Description & Processing: The offline computing organization provides the software framework, computing infrastructure, data-management system, and analysis software as well as the hardware and networking required for offline processing and analysis of LZ data. The system will be designed to handle the data flow starting from the raw event data files (the so-called EVT files) on the SURF surface RAID array, all the way through to the data-analysis framework for physics analyses at collaborating institutions, as illustrated below.



The LZ raw data will be stored, processed and distributed using two data centers, one in the United States and one in the UK. Both data centers will be capable of storing, processing, simulating and analyzing the LZ data in near real-time. The SURF surface staging computer ships the raw data files (EVT files) to the U.S. data center, which is expected to have sufficient CPU resources for initial processing. The National Energy Research Scientific Computing (NERSC) center at LBNL will contain the resources to act as the LZ U.S. data center. Current estimates of total data volume from LZ should be considered preliminary, but are of order 1.1-1.2 PB per year starting in FY2019.

Data Products and Releases: The type of data produced in LZ includes raw data from testing, calibration, and measurements at participating institutions and at SURF 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 LZ software (including both online and offline code) will be centrally maintained through a software repository. The first physics integration release is planned for August 2016. This version includes all necessary modules for real-time processing (i.e., hit-finding algorithms, calibration constants modules, S1/S2 identification, event reconstruction), as well as a fully integrated simulations package (i.e., from event generation through photon hits, digitization, trigger, and data-format output). The first mock data challenge (February 2017) will test both the data flow (transfers, processing, distribution, and logging), as well as the full physics analysis functionality of the framework, separately, while the second data challenge (December 2017) will be dedicated to testing the entire data chain. The third data challenge (June 2018) will also test the entire data chain and is expected to validate the readiness of the offline system just before the LZ cool-down phase.

Plan for Serving Data to the Collaboration and Community: To effectively and promptly share our research results, the LZ collaboration will take the following steps. The LZ collaboration is committed to provide access to the LZ scientific results.

- (1) We will use our LZ TWiki website to document all major research activities, project status, computing and storage resources. The information will be backed up daily; the entire website will be synchronized every 30 minutes with mirrors that are installed on the surface and underground at SURF, ensuring continuous access to LZ documentation.
- (2) Document newly developed instrumentation, simulation, and analysis tools (including user instructions) on time so that other researchers may use them.
- (3) Collaboration members, including junior researchers and students, will be given the opportunity to join professional workshops and conferences to present the LZ research work and results.
- (4) All LZ collaborators will have read access to raw and derived data, databases, and software releases necessary to analyze LZ data and reproduce LZ science results.

Funding support does not exist and is not planned to provide software tools for community access to LZ data.

Plan for Archiving Data: Raw data will be archived at NERSC and at the UK data center along with appropriate databases and software releases. NERSC's HPSS (High Performance Storage System) will keep the data available for at least 10 years. A similar time period is expected at the UK 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 LZ 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.