

# Summary of Data Management Principles

## LZ Experiment

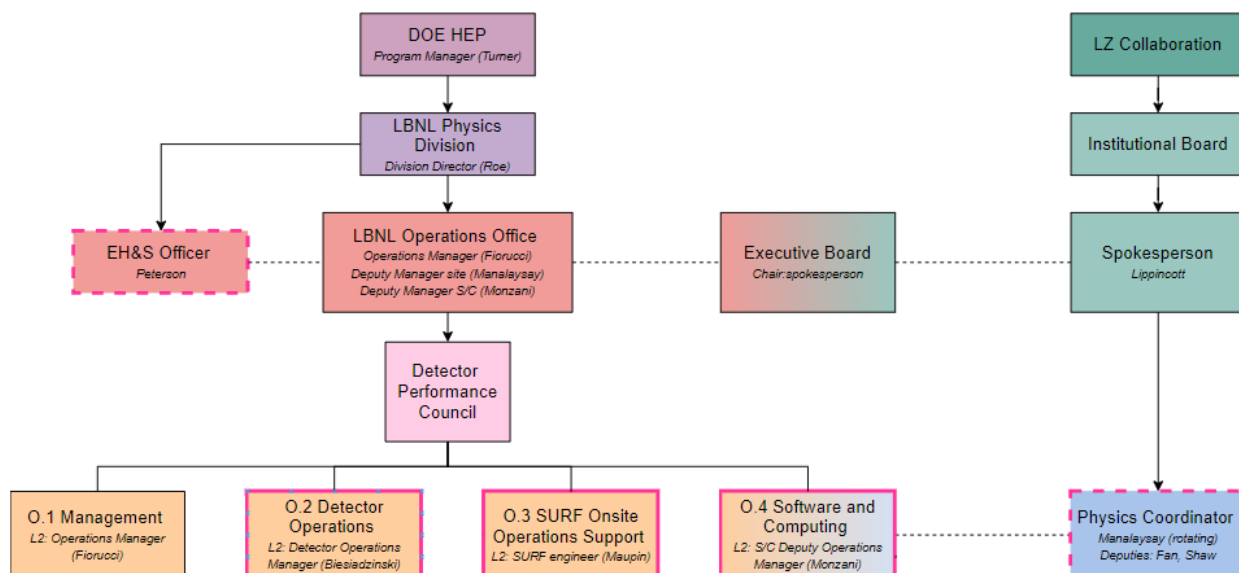
**Experiment description:** The LUX-ZEPLIN (LZ) experiment searches for direct dark matter signals using a liquid Xenon target and dual-phase time projection chamber. LZ is located at the Sanford Underground Research Facility (SURF) in Lead, SD, USA. The LZ DOE construction project was completed in 2020 and the experiment expects to operate through 2025.

**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 provided by Portugal and South Korea.

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

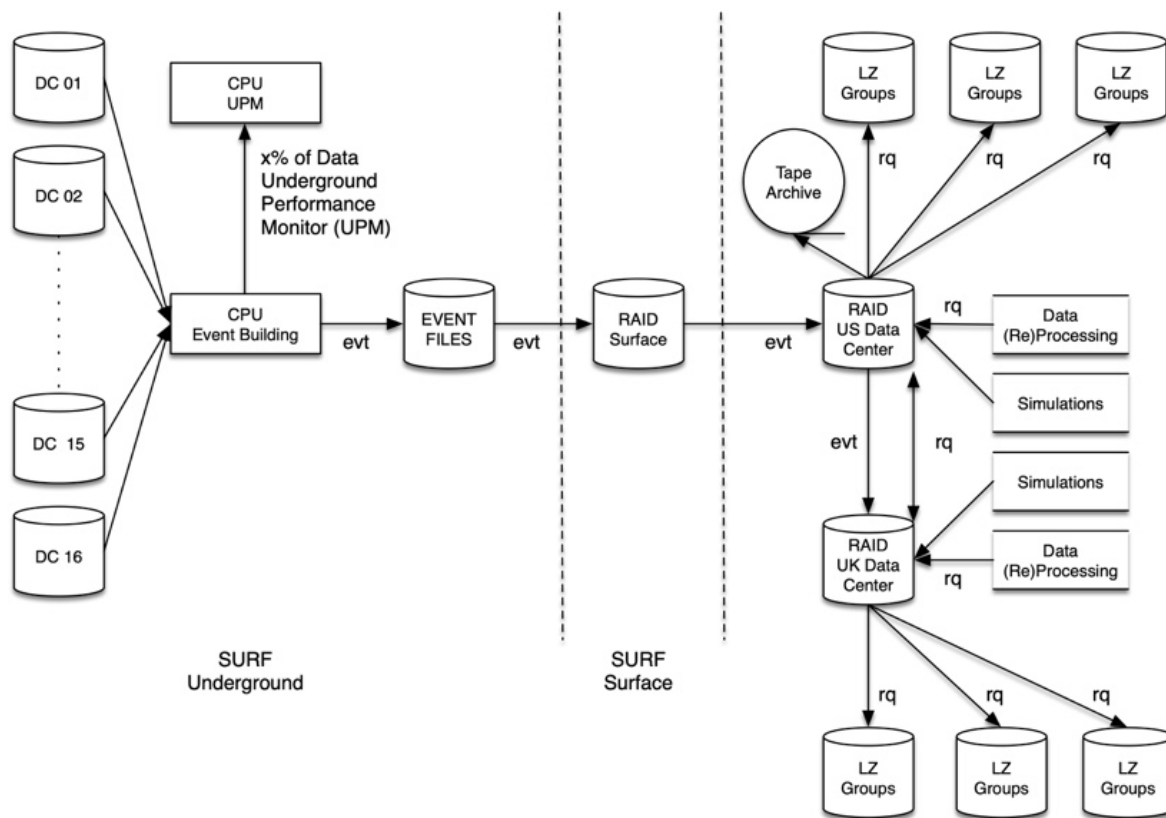
**Organization – Experiment level:** The LZ Operations organization chart is displayed below. The LZ collaboration is led by an elected spokesperson who is advised by Executive and Institutional Boards.



**Collaboration:** The LZ collaboration is composed of 38 member institutions and includes about 250 scientists, students and engineers.

**Data Policy Management:** The LZ operations 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 is 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 is stored, processed and distributed using two data centers, one in the United States and one in the UK. Both data centers are 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 US data center, where initial processing is performed. The National Energy Research Scientific Computing (NERSC) center at LBNL contains the resources to act as LZ’s US data center. Estimates of total data volume from LZ are of order 1.1-1.2 PB per year during science data taking.

**Data Products and Software Management:** The 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) is centrally maintained through a software repository.

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

- (1) Use LZ TWiki website to document all major research activities, project status, computing and storage resources. The information is backed up daily; the entire website is synchronized every 30 minutes with mirrors 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, have the opportunity to join professional workshops and conferences to present the LZ research work and results.
- (4) All LZ collaborators 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 to provide software tools for community access to LZ data.

**Plan for Archiving Data:** Raw data is 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 archival data available for at least 10 years.

**Plan for Making Data Used in Publications Available:** Concurrent with the publication of significant science and engineering results, all the research data displayed in the publication will be made available to the scientific community and to the public. The research data will be provided either in the body of the publication or as supplementary material, according to journal policy. All LZ 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 Office of Science Statement on Digital Data Management (<https://science.osti.gov/Funding-Opportunities/Digital-Data-Management>), within the constraints of limited funding and resources.