Summary of Data Management Principles
SuperCDMS Soudan Experiment

Experiment description

The SuperCDMS Soudan experiment used cryogenic germanium crystals to detect and measure the predicted rare scattering of dark matter particles on nuclei. The detectors searched for both ionization and athermal phonon signals from dark matter interactions to achieve a very low recoil energy threshold and exceptional discrimination between a nuclear recoil signal and electron recoil backgrounds.

The SuperCDMS Soudan experimental setup consisted of 9 kg of germanium (15 detectors) installed at the Soudan Underground Laboratory in Minnesota. The detectors were in operation from December 2011 until Fall 2015.

DOE's roles in the experiment

The SuperCDMS collaboration operated the SuperCDMS Soudan experiment. Fermi National Accelerator Laboratory (FNAL) managed the operations and decommissioning of the Soudan experiment, under contract with the DOE. The DOE funding covered shift travel for scientists to work onsite at the Soudan Underground Laboratory. It also covered technical labor and material supply (M&S costs for the operation and decommissioning periods. A small portion of the budget paid for computing and administration of the Soudan dataset, which is hosted at FNAL.

DOE operations funding was provided directly to FNAL, and all M&S purchases were done through the FNAL's purchasing department. DOE-supported travel was handled through the FNAL travel office, even for DOE-supported university groups.

Partnerships

The DOE Office of High Energy Physics (OHEP) worked in close cooperation with the National Science Foundation (NSF) to support the SuperCDMS Soudan experiment. SuperCDMS Soudan operations were managed by the DOE, and jointly funded by the DOE and the NSF. FNAL, managed by the Fermi Research Alliance (FRA), was the host laboratory for SuperCDMS Soudan, and provided management and safety oversight for the Soudan Underground laboratory, which included the SuperCDMS Soudan experiment.

As described above, the DOE funding supported the operation of the experiment through operations management, which includes shift travel for scientists at DOE-funded institutions. For the Soudan experiment, the NSF funding supported several of the universities within the Collaboration and was primarily used for travel for shift work at the Soudan Underground Laboratory. Scientist support is provided through base funding from either the DOE or the NSF at each institution. There are no overarching agreements between the DOE and the NSF regarding data management for SuperCDMS Soudan.

Organization – Agency/Lab level

The SuperCDMS Soudan collaboration is funded through a partnership between the DOE and the NSF. The program managers for the SuperCDMS Soudan experiment were Michael Salamon at the DOE OHEP and James Whitmore at the NSF. The SuperCDMS Collaboration is an international collaboration with members from 23 institutions including 3 national laboratories (FNAL, PNNL, SLAC). Most institutions have base grants funded either by DOE or NSF. FNAL is the host laboratory for SuperCDMS Soudan.

Organization – Experiment level

There are four main systems that provided support for SuperCDMS Soudan (Figure 1). An operations team, chaired by the operations manager, met weekly and oversaw day-to-day operations tasks. Soudan Laboratory staff monitored the cryogenics system, overseen by both the operations manager and a FNAL engineer. FNAL technicians maintained the electronics, data acquisition and computing hardware at Soudan. The Data Acquisition and Data Quality groups included scientists from both lab and university groups, and together they monitored the data being taken.

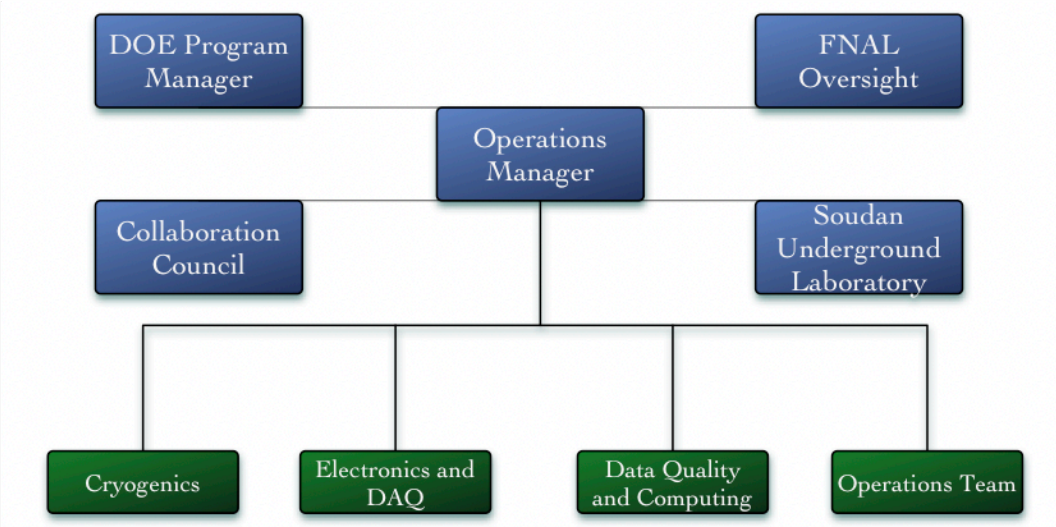

Figure 1: SuperCDMS Soudan experiment organization

Collaboration

The SuperCDMS Collaboration consists of approximately100 scientists from 23 institutions in 3 countries. The Collaboration Spokesperson is elected from the collaboration and may be from either a university or one of the national laboratories. The current spokesperson is Prof. Priscilla Cushman (U Minnesota). Day-to-day affairs of the collaboration are managed by the Executive Committee, which reports to the collaboration Council, which includes the PIs and working group chairs, and which was chaired during the Soudan experiment by Prof. Sunil Golwala (Caltech). Working groups and the experiment operations support organization within the Collaboration are shown in Figure 2.

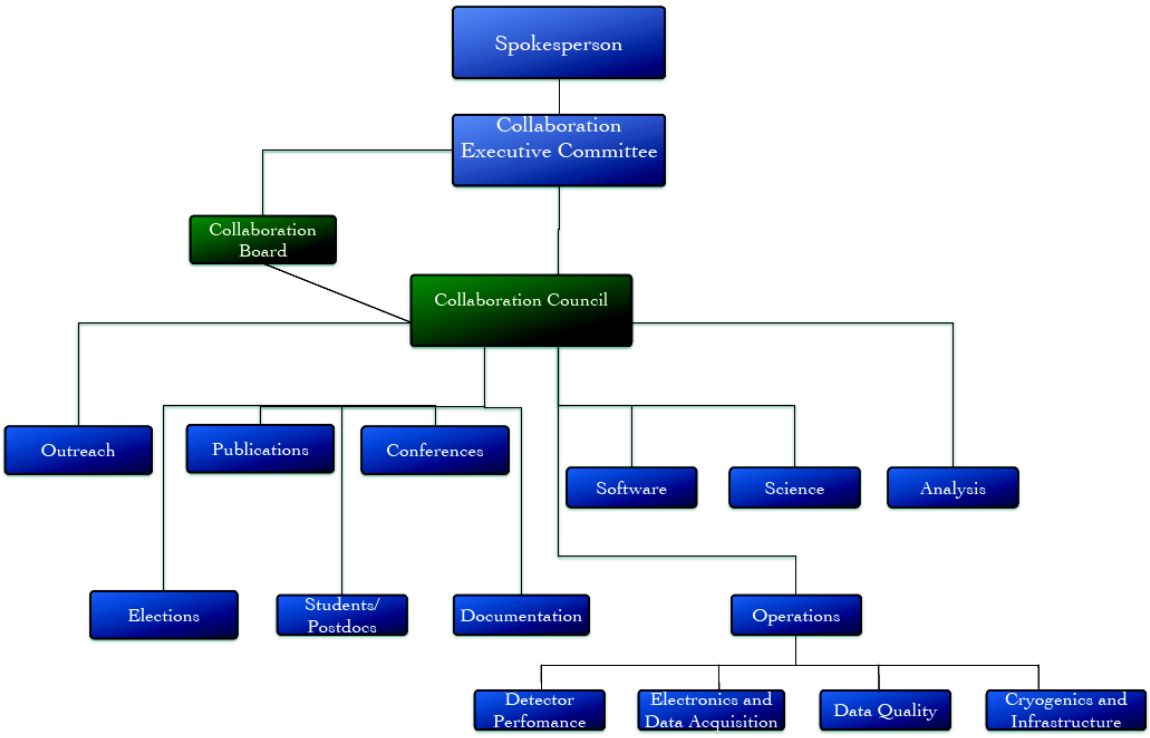

Figure 2: SuperCDMS Collaboration organizational chart

Data policy management

The SuperCDMS Collaboration Council sets the data management policy for the collaboration. Collaboration policy documents are available through https://supercdms.slac.stanford.edu/, the home website of the Collaboration.

Data Description & Processing

The raw data produced by the SuperCDMS Soudan DAQ system consist of ionization and phonon waveform traces, muon veto data, as well as environmental data such as cryogenic system data and run-time data (e.g. trigger rates and detector state). The raw data are stored in binary files and SQL databases. Two main types of data were taken:

- Calibration data using radioactive sources
- Low background data for dark matter search

Most of the data taking was in the low background mode. A limited amount of calibration data were taken per week. The data volume was dominated by the calibration data.

Real-time processing with a limited number of algorithms and preliminary calibrations is done first at Soudan. These data are used to monitor the stability of the experiment and check the quality of the incoming data in real time. The raw data are then transferred to FNAL for longer-term storage. All the raw data are then re-processed twice a year on the FNAL grid using the full reconstruction package, including blinding of the dark matter (WIMP) search data.

The first level of processing produces ROOT files containing variables calculated from pulse reconstruction analyses, such as optimal filter, and environmental data analyses. A second level of processing produces the calibrated quantities.

Monte Carlo simulation computing is performed mainly at SLAC and Texas A&M.

After analysis of the WIMP search data and publication of results, 3 types of data are produced:

- Candidate data: Information (e.g. charge and phonon energy) about the events passing all the selection criteria.
- Exposures and efficiencies: Final WIMP efficiency for each detector as a function of total phonon energy, after applying all selection criteria, and exposures for each detector.
- Nuclear recoil energy scale: parameters used to calculate the recoil energy of the events.

Data Products and Releases

The data used in a given publication are made publicly available at the time of publication or shortly thereafter. All public data releases are available through the collaboration website: https://supercdms.slac.stanford.edu/public-data. A document with instructions and detailed descriptions of the data release (including any quality cuts applied, efficiencies, exposures and nuclear energy scale) is provided for each data release. An email address at which the collaboration can be contacted regarding any questions about the release also is provided in the documentation.

Plan for Serving Data to the Collaboration and Community

The collaboration is committed to making all experimental data available to collaboration members as quickly as possible. Raw data are available to view immediately after the data have been collected. Processed data take longer to prepare, but have been made available to the Collaboration in a timely manner. SuperCDMS Soudan data are available at data centers located at FNAL, Stanford University and Texas A&M. Collaboration members with proper login credentials can obtain/view/analyze data from any of those locations.

We are not planning to release raw or processed data to the community. Nevertheless, we will provide data from all finished analyses alongside the specific publication. The final datasets used in publications are typically much smaller in size and will not require special software tools to analyze. The decision to not provide all data to the community was made due to cost benefit considerations. We don't have the resources within our collaboration to provide an easy to use dataset, along with analysis tools. Additional resources would need to be invested to accomplish this task, that we feel are better spent elsewhere.

Plan for Archiving Data

Data collected onsite at Soudan Underground Laboratory are copied to a RAID disk array on the surface and then copied to FNAL over a wide-area network. At FNAL, the raw datasets are redundantly archived on a series of RAID disk servers and spooled to tape in a central storage

facility at FNAL (Enstore). Tapes in Enstore have been an in-kind contribution to SuperCDMS, as well as power and space charges for the SuperCDMS servers. At FNAL, the data are readily accessible for further processing on the general purpose FermiGrid and then distributed to the other institutions for analysis. Processed data are also stored on the RAID servers at FNAL. A second copy of the processed dataset is stored at Stanford University on a RAID disk array. Portions of the processed dataset are also stored at other collaborating institutions such as SMU and Texas A&M for science analyses.

Plan for Making Data Used in Publications Available

The collaboration is committed to provide data from all publications to the wider community. We strive to make data relevant to a given publication available at the same time as the publication, but may not achieve that goal in all cases. If we are not able to provide the data from a given publication at the same time as the publication becomes public we will append the publication with the relevant data as soon as possible. Along with the data, we are committed to providing scripts that show how the data can be used and visualized. Data will be provided in a standard format (e.g. text and/or ROOT files) and any scripts provided will be written in a widely used programming language (e.g. Python). The exact data format and script language are left to the analysis leader's discretion.

Responsiveness to SC Statement on Digital Data Management

This data management plan follows SC Statement on Digital Data Management with the exception that the entire data set is not made public. The plan describes our justification for this exception.