

## **Summary of Data Management Principles For the Nearby Supernova Factory Experiment**

**Experiment description:** The goal of the Nearby Supernova Factory (SNfactory) is to develop a large sample of well calibrated nearby supernova for cosmology. Two unique aspects of the SNfactory are its focus on “nearby Hubble flow” supernovae and an observational technique that obtains calibrated spectra for supernovae as they brighten and fade. During the period 2005-2008 the SNfactory conducted a vertically-integrated supernova search and follow-up program. During this phase, 600 supernovae were found and spectroscopically confirmed, making it at that time the most successful nearby supernova search and follow-up program in history. During the period 2010-2013 a second phase of the project was begun, following supernovae from external sources. The SNfactory built and operates the SuperNova Integral Field Spectrograph (SNIFS) on the University of Hawaii 2.2-m telescope on Mauna Kea, Hawaii. The resulting SNfactory dataset is by far the largest collection of spectrophotometric supernova time series in the world.

**DOE’s roles in the experiment:** DOE supported the development of the three CCD cameras used as the detectors for the SNIFS, used to obtain the primary SNfactory dataset. During the SNfactory-I phase, DOE supported the supernova search operations at LBNL. Currently DOE supports some of the data analysis efforts at LBNL.

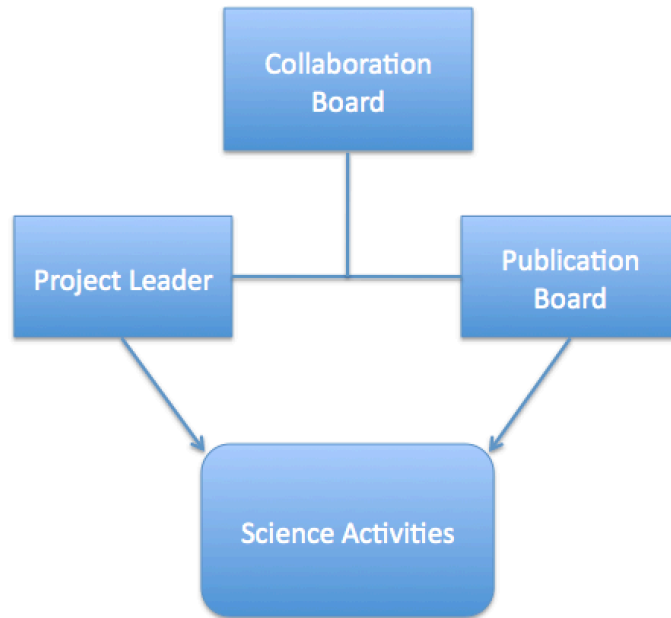
**Partnerships:** There is an MOU between the member institutions, but not at the agency level.

**Organization – Agency/Lab level:** The science collaboration is independent of DOE and LBNL. LBNL represents DOE’s interest in the form of the Project Leader, membership on the Collaboration Board, and the chairmanship of the Publications Committee.

**Organization – Experiment level:** The SNfactory has completed its operations phase, so its organizational structure, given below, is quite simple. At the top level there is a Collaboration Board (CB), with membership from each institution. The CB sets the overall policy. There is a rotating chair, who convenes meetings of the CB, maintains the list of collaboration participants, and reports on CB activities to the wider collaboration. The CB so far has succeeded in governing by consensus. The Project Leader oversees the operations and scientific direction of the collaboration, subject to the policy prescriptions adopted by the CB. The Publications Committee consists of four members, including a chair, who implement the details of the publication policy provided by the CB to scientific analyses as they work their way toward becoming scientific papers.

**Collaboration:** The SNfactory collaboration includes LBNL, Yale University, LPNHE, IPNL (Lyon), CRAL (Lyon), CPPM (Marseille), Humbolt University, Max Planck Institute for Astrophysics, and Tsinghua University. A total of 22 scientists are involved. Since the operations phase is concluded, Institutions pursue scientific investigations with the data, including reprocessing of the data, as dictated by their coordinated interests. Institutions chose their

representatives to the Collaboration Board. The Collaboration Board appoints a new chair every few years. The Project Leader position does not rotate. The composition and chairpersonship of the Publication Board is changed as needed.



**Figure 1. SNfactory Organizational Chart - Post Operations**

**Data policy management:** The Nearby Supernova Factory CB sets the data management policy. The CB has representatives from all the institutions involved in the SNfactory. Policies can be modified by unanimous consent of the CB. The collaboration is small, and dominated by institutions outside the United States, therefore, no separate data management board has been established.

**Data Description & Processing:** The raw and processed data include science images and the associated databases for the SNfactory supernova search. For the initial supernova search there is a series of SQL interfaces meant for machine interrogation only. There is also a SQL relational database and GUI that has been used to monitor supernovae from discovery through screening, follow-up, and completion. Likewise, there is a separate collection of raw and processed images and spectra from the SuperNova Integral Field Spectrograph (SNIFS). The SNIFS dataset includes engineering and calibration data in addition to the science data. There is a Twiki with documentation on the operations of SNIFS, and meta information concerning operations and analysis. There is also a CVS repository where software is archived, and a Mantis web interface for reporting and tracking bugs. Search data were processed at NERSC and at LBNL. All SNIFS data are processed initially on-site, and slightly later at LBNL, to monitor data quality and follow-up progress. Final processing, including saving the intermediate data products, is done at CCIN2P3. For these processed spectra there are webpages that show the

data, provide summary statistics, and consolidate metadata such as weather conditions. Data are reprocessed once or twice a year as algorithms and calibration are improved.

**Data Products and Releases:** The SNfactory collaboration releases data in coordination with its scientific publications. These are included in the publications and/or provided at <http://snfactory.lbl.gov/snf/data/index.html>

**Plan for Serving Data to the Collaboration and Community:** The SNfactory received funding from the Gordon and Betty Moore Foundation to develop the “Legacy Supernova Catalog” for the scientific community, based on SNfactory data. We have developed the “SNfactory Zoo” interface for interactive access to the data. This interface currently is used internally, but will become available to the outside world as larger fractions of the data are published. Meanwhile, we offer our data published so far at <http://snfactory.lbl.gov/snf/data/index.html>

**Plan for Archiving Data:** Data are archived on spinning disk at the IN2P3 Computing Center (CCIN2P3) in France and at the NERSC High Performance Storage System facility in the United States. We expect to maintain the storage on tape at NERSC, subject to funding.

**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. Publications will be posted on a free website (e.g., arXive/astro-ph).

**Responsiveness to SC Statement on Digital Data Management:** It is the intent of the SNfactory 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.