Appendix E: Dissemination and Sharing of IceCube Research Results and Data

This Appendix defines the IceCube Neutrino Observatory (ICNO) strategy for providing access to research results and data. NSF policies and guidance promote efforts by grantees to produce timely publication of results and to make data and software available to other researchers. In addition, the Parties to the Antarctic Treaty agree that, to the greatest extent feasible and practicable, scientific observations and results from Antarctica shall be exchanged and made freely available.

The NSF supports a wide range of approaches to the release of science data from large facilities. For example, within the astrophysical community there are two different models: the particle physics model where data are exclusively available to members of the collaboration and the astronomy model where data are readily available to other scientists and public. The Large Hadron Collider experiments follow the particle physics model and the Atacama Large Millimeter/submillimeter Array (ALMA) follows the astronomy model. The Pierre Auger Cosmic Ray Observatory and NASA's Wilkinson Microwave Anisotropy Probe (WMAP) follow an intermediate model where data releases are periodic, often tied to scientific publications, and represent different degrees of data processing.

Currently, the IceCube Collaboration follows the intermediate data release model, but will, together with the IceCube Science Advisory Committee, periodically review the efficacy of the IceCube Data Sharing Policy and may initiate changes in the policy (with appropriate consultation with NSF and foreign funding agencies) that might become needed.

There are three pathways to access the IceCube data:

1) IceCube Collaboration Membership;
2) Associate Membership; and,
3) Direct Access to IceCube Public Data Pages.

The IceCube Collaboration consists of scientists at the respective Constituent Institutions operating under the Governance Document; the condition for "Collaboration Membership" and respective institutional recognition is that the particular group makes significant contributions to IceCube maintenance and operation. These groups have full and immediate access to all ICNO data – raw, processed, and value added. Scientists outside the IceCube Collaboration who have a concept for a particular analysis can apply for Associate Membership for the purposes of performing a particular analysis or class of

analyses within the Collaboration. These members have limited access to ICNO data – mostly to the partially processed data for a defined period.

Direct Access to the IceCube Neutrino Observatory Data

The ICNO raw data are securely stored and backed up, consistent with the overall NSF Data Policy requirements. Extracting science from these data requires the use of elaborate hardware and software tools developed by the Collaboration members. Like any other particle physics detector, data directly relevant to a scientific issue are obtained after a series of calibrations, data quality tests and analyses that typically require coordinated efforts of several members of the Collaboration.

In order to be responsive to both the scientific community's need for usable scientific data and to the NSF requirement for public access to data, IceCube plans to release data in two ways described below. The first dataset is meant to comply with the NSF requirement for public access to data. The second dataset is a value-added dataset meant to be first used by the IceCube scientific community. It provides in-depth information that is indented to lead to scientific publications by the Collaboration.

1. **Release of IceCube primary event data transferred to the University of Wisconsin from the South Pole over communication satellites.**

The IceCube Public Data Web site http://icecube.umd.edu/PublicData/I3OpenDataFormat.html provides access to the IceCube Observatory primary data for the general scientific community and the public without requiring any internal or proprietary software. This Web site provides information on how to access all primary IceCube data sent over the satellite (both "neutrino events" and "non-neutrino events"). There are basically two types of data sent north from South Pole that represent the entire dataset for all IceCube analysis. The type of data for each is described here:

Data Summary

This dataset is comprised of a brief summary of every event recorded by the IceCube detector; it stores a summary record of the UTC time, the number of Digital Optical Modules (DOMs) hit, and reconstructed muon direction. This dataset is used, for example, in cosmic-ray anisotropy and Moon shadow studies. The dataset size is approximately 2TB (terabytes) per year.

Event Data

This dataset is comprised of full information (see below) for events that are identified as possible "neutrino events" based on software triggers. For each of these events, a record is sent north without event reconstruction, consisting of the run number, trigger information (i.e., time and type), and the array of the sensors with photoelectron signals. The muon-generated Cherenkov emission photons strike the surface of photomultipliers in the DOM creating photoelectron signals. The signal's amplitude ("charge" q), position of the DOM that was struck (in the IceCube x,y,z coordinate reference frame), and the

photon's arrival time (t, in the IceCube time reference frame) are recorded in the DOM's data output as the following quantities:

Run/Event header with trigger information, event date and time (as defined by the IceCube data format)
Array of DOM signals with calibrated position, time, and charge (x,y,z,t,q)

This dataset size will be anywhere between 10 and 40 TB per year.

Starting in 2011, the ICNO data described above and recorded by the entire 86-string detector will be grouped together by year (usually beginning in April as austral summer ends) and made available through the Public Data Web site annually following the data quality assessment. It will typically take six to nine months to fully verify data quality following the end of a year-long data collection run. Instructions on how to obtain these datasets are provided through the Public Data Web site. Along with the full datasets, some selected neutrino event samples and test datasets could also be provided.

The IceCube Public Data Web site requires from the users only some basic knowledge of the programming and visualization language Python (a free, widely available programming and visualization language with open source license; see http://www.python.org). Therefore, the required effort by potential users to understand and visualize the ICNO data is minimal because the public Web-based documents and tutorials provide enough information about the IceCube primary data structure. This means that the ICNO calibrated data stream can be decoded in any language (e.g., C/C++, FORTRAN, IDL, etc.) using any available computer architecture (e.g., 32-bit or 64-bit computers). The Web site provides short scripts written in Python that allow the users to easily understand the primary data format for reading events with properly recorded timing and location and visualize these events as needed. The ICNO Public Data Web site will be undergoing improvements as needed or as required by feedback from this community of users.

This dataset represents a release of all primary ICNO data transmitted north for analysis and should satisfy the NSF requirement for public release of data (for both “neutrino events” and “non-neutrino events”) at a level far exceeding the particle physics or most astronomy models.

2. **Release of value-added IceCube data consisting primarily of Reconstructed Events.**
   An example would be the "neutrino event" data used for a publication on searches for a point source, but may consist of other supporting information for published results as a service to the general scientific community.

The IceCube Collaboration will be processing data for scientific analysis on an annual basis following conclusion of each year-long data run. This processing will result in a dataset that represents reconstructed events identified as "neutrino events". This data set will generally consist of the following quantities:

Event Time (Modified Julian Day)
Direction (Right Ascension, Declination)
Directional Error
Degrees of Freedom in Fit
Energy Estimator
Flags Indicating Event Types (e.g., track like, cascade like, etc.)

Beginning two years after the first full data run is complete (to allow time for IceCube analyses as outlined below), these data will be released annually as an event catalog for the preceding year. The catalog of an earlier year might be revised in later years to update information and include better reconstruction algorithms and filtering processes. In this case, an updated set of data will be offered to the scientific community as the new version of a particular catalog. Based on feedback from the community, more information could be added into later releases to accommodate all types of community requests.

Note that the events not selected as “neutrino events” (for example, "cosmic ray muon background events") are also valid data that could be used for other analyses, such as studying the symmetry or asymmetry of the cosmic ray distribution between the Northern and Southern hemispheres, etc. These events are represented by the above quantities, except the "Flags Indicating Event Types" are set for "cosmic ray muon" or some other appropriate event type. This "non-neutrino events" dataset is subject to a number of specific scientific analyses that are additional to the primary IceCube goal to detect and analyze cosmogenic neutrinos coming from far and energetic sources.

The IceCube scientific working groups will analyze year-long datasets for various physics parameters and conduct specific data analyses. It is expected that ten to fifteen scientific publications will be prepared using the annual data set and completed/published on a time scale of about two years (see below for additional information on the timing of data and publication releases). The intention of the IceCube collaboration is to make the ICNO data utilized for science analyses available upon publication of results. For example, when the initial searches for point sources, neutrinos from transient sources, and diffuse astrophysical neutrinos are published, the relevant event information associated with those analyses will be made available in an easy to read format. Partial event reconstruction information may be made available earlier than the intended publication.

During the operations phase of IceCube it is anticipated that IceCube neutrino data (Category 2 above) will be released within two years (or up to three years with NSF approval of extenuating circumstances) after the completed run in which the data are acquired.

The IceCube Collaboration created a data release webpage that serves as the entry point for future data releases to the scientific community, http://www.icecube.wisc.edu/science/data. Initially, this Web page contains release of the 2000-2006 AMANDA data. The

URL to IceCube data release Web page is an explicit reference in the corresponding journal publication and will remain the same during the IceCube operations [Abbasi et al. (IceCube Collaboration) Phys. Rev., D79:062001, 2009. e-Print: arXiv: 0809.1646].

Important requirements for the data release in Category 2 are:

(1) The IceCube Collaboration's analyses are completed in accordance with the Collaboration's internal approval processes, which include adhering to the principles of blind analyses where practical;

(2) The calibrations and reconstructed event information are of high quality and it is unlikely this information will need to be changed or corrected.

Long-Term Data Management Plans

Once IceCube is in steady state operation, the ICNO will continue releasing the data on annual cycles from the data runs beginning in April.

The sequence from data taking to publication can be summarized as follows:

- Data Taking Run (~12 months)
- Annual compilation of data (starting point of the activities below)
- Release of primary event data (Category 1) without rejection of cosmic ray muons via the Public Data Web site (typically 6-9 months after completion of the year-long data taking run)
- Data Processing for the IceCube Event Reconstructions:
  - "neutrino events"
  - other "non-neutrino" events, including cosmic ray muon types
- Data Analyses for Specific Science Goals
- Preparation of the Final Data Sets
- Perform Final Physics Analyses and Un-Blind Results
- Publish Results with the Data used for specific publications
- Release of value-added IceCube data consisting primarily of Reconstructed Events (Category 2) for both the "neutrino events" and other types (within 24 months, or up to 36 months for extenuated circumstances)