2DCC-MIP Data Management Policy

Overview

Operational philosophy: The 2DCC-MIP (hereafter referred to as 2DCC) is a national user facility funded by the National Science Foundation which does not charge fees for non-proprietary use by U.S. academic or government researchers. The 2DCC data mission is to create a *community knowledge base* in synthesis of 2D materials, including detailed, comprehensive synthetic protocols, characterization measurements, and simulation results generated through non-proprietary research in the 2DCC. It is an obligation of all non-proprietary users of the 2DCC to reasonably facilitate this process, consistent with shared norms of scientific conduct.

Data covered by this policy includes that generated directly on 2DCC equipment (thin-film and bulk growth, integrated characterization measurements, 2DCC computational resources) and also that generated at support facilities for characterization (Materials Characterization laboratory), fabrication (Nanofabrication Laboratory) and computation (Materials Computation Center), when 2DCC resources have supported use of those facilities.

NSF best practices and resources for data curation in materials research are evolving. Therefore, the Data Management Policy is a living document that is reviewed regularly by the 2DCC executive leadership team and updated as appropriate to serve the broad interests of the scientific community.

Data Types

2DCC user data cover a broad range, including but not limited to: *simulation data* (e.g., databases of material electronic or structural properties, simulations of kinetic parameters, etc.), *experimental data* (e.g., characterization data from STM, TEM, optical measurements, spectroscopy data, structural information, etc.) and *experimental process information* (e.g. detailed synthetic protocols and processing steps).

In the normal course of research, it is common for much data obtained to be non-archival in quality – a sample may be degraded, instrumental settings incorrect, noise levels too high, or simulation settings improper. Following best practices, it is the intention of the 2DCC to save all data generated, since even low-quality data can be crucial to informing later decisions in process optimization or instrument debugging.

Data is classified into two main categories: (1) Work-in-progress Data – data generated at intermediate stages in an investigation, of uncertain quality and with substantially incomplete contextual information and (2) Archival Data – contextualized data worth saving in a community repository for later reference and reuse. The scientific judgment of the investigators involved, guided by 2DCC policy and goals, will inform what data is deemed archival in quality. In cases where consensus cannot be obtained, the ELT will make this determination.

Data and Metadata Standards

Standards for data vary by the method, equipment, and software used for its generation, and evolve over time. This includes raw data in the native format of an instrument, processed data in standard inter-change formats, and metadata necessary to establish meaning and context for the associated measurement or simulation. The issue of data format is complex and evolving –

the philosophy of the 2DCC is to facilitate progress towards data exchange formats and platforms that are accessible, interoperable, and easy to use.

Data Sharing and Access

Data generated at the 2DCC are under the ownership of the user(s). Users, as owners of data, have full control over decisions regarding publication, in accord with shared standards of scientific conduct. Users are responsible for adhering to the policies and procedures of their funding agencies. The user PI grants the 2DCC a non-exclusive license to host non-proprietary data on a community platform as designated by the 2DCC, subject to certain restrictions – designed primarily to protect the users' publication priority – as outlined below.

Non-proprietary data including databases, software and metadata that were produced using NSF funds are expected to be shared publicly after publication. Data that is not indicated by the user PI as intended for publication may be shared publicly by 2DCC no sooner than one year following the end date of the User Project under which it was generated. It is the intention of the 2DCC to provide a central data repository for ensuring community access to non-proprietary data generated under support from 2DCC resources. In cases where publication is intended to take place more than one year after the end date of the project, Users may request a delay of data sharing, subject to approval by the ELT.

Data Re-use or Redistribution

User data will not be shared publicly by the 2DCC prior to publication – or one year following the end date of the associated User Project, as indicated above – without the permission of the user PI. Public sharing includes cases where data reported to NSF becomes part of the public record. Data may be shared internally with NSF program officers as part of NSF oversight, with notice given to the user. Users are free to redistribute their data, following NSF policies, given that **acknowledgement** of the 2DCC in providing resources for generation of those data is made.

Users performing non-proprietary projects at the 2DCC are expected to analyze and submit their results for publication on a reasonable timescale, with proper **acknowledgement** of the facility.

Storage and Archive

The 2DCC will maintain non-archival user data and sample storage for a period of at least five years following the close of the user project. The 2DCC intends to maintain archival user data indefinitely through a central data repository managed by 2DCC or in partnership with external entities that adopt the mission of long-term data curation for the materials research community.

Proprietary or confidential data that is obtained or generated through participation with users will be maintained and secured locally as a copy of that shared with the user and is subject to a non-disclosure/confidentiality (and intellectual property) agreement administered by the Penn State Office of Technology Management between the user institution and Penn State University.