



DIII-D National Fusion Facility Guidance Document	Title: Data Management Plan	
	Date: January 30, 2024	Number: D3D-G08
	Process Owner: DIII-D CSS Coordinator	
	Distribution: Public	

REVISION HISTORY

Rev.	Date	Description
A	January 30, 2024	Initial release as DIII-D Guidance Document. Previously released publicly through DIII-D program websites.

Table of Contents

ACRONYMS..... **2**

GLOSSARY..... **2**

 1. *PURPOSE AND SCOPE*..... 3

 2. *REFERENCE DOCUMENTS*..... 3

 3. *ROLES AND RESPONSIBILITIES*..... 3

 3.1. *Computer Working Group (CWG)*..... 3

 3.2. *Technical Management*..... 3

 3.3. *DIII-D CSS Coordinator*..... 4

 4. *DATA MANAGEMENT POLICIES*..... 4

 4.1. *Types of Data*..... 4

 4.2. *Storage of Data*..... 4

 4.3. *Access to Data*..... 5

 4.4. *Tools and Resources*..... 6

ACRONYMS

Acronym	Description
CWG	Computer Working Group
DMP	Data Management Plan
DOE	U.S. Department of Energy
GA	General Atomics
SC	Office of Science
TB	Terabyte

GLOSSARY

Term	Description
Team Member	Any person who contributes to any aspects of the DIII-D program.
Users	Personnel who directly engage in the Experimental Science program through participation in recognized projects (i.e., projects that are submitted for the annual DOE SC User Report). All users are also team members.

1. PURPOSE AND SCOPE

The DIII-D National Fusion Facility is a large international program, with more than 100 participating institutions and a research team of approximately 800. GA operates DIII-D for the Department of Energy, Office of Science, through the Fusion Energy Sciences program. DIII-D is an Office of Science User Facility. As part of DIII-D's mission, approximately 500 TB of research data has been acquired and/or generated since its inception. Management and access to these data are described below.

The DIII-D Data Management Plan (DMP) is a critical component of the program's fundamental research data pipeline, ensuring data standards, validation, security, and integrity. The DMP describes the types of data that are measured or produced through analysis, and it also describes the resources available for the data management and preservation during the course of research operations.

2. REFERENCE DOCUMENTS

- Data Usage and Publication Policy Agreement (DUPPA)

This agreement describes the DIII-D program policies related to the usage and sharing of research data, including the review process for materials intended for public dissemination (all journal publications are considered public dissemination). Users must sign this form to acknowledge awareness and acceptance of these terms. Access to unpublished DIII-D research data is only provided after this form is signed by the User.

This agreement is available as a downloadable PDF at the following publicly available link, https://d3dfusion.org/wp-content/uploads/diii-d_data_usage_policy_ca.pdf.

- DIII-D Guidance Document 04 (D3D-G04): Review Process for Research Products

This document describes program policies related to the internal review process for research products. These research products frequently provide the first public dissemination of DIII-D research data.

3. ROLES AND RESPONSIBILITIES

3.1. Computer Working Group (CWG)

Established by Director memorandum (March 10, 2021), this is an internal DIII-D group with leadership selected through an Expression of Interest process and participation open to all DIII-D users. The goal of the CWG is to improve efficiency within the DIII-D team by carrying out activities that will help obtain and maintain computer resources and information technology needed to carry out the DIII-D research plan in a secure manner.

The CWG is responsible for collecting feedback on relevant issues from across the DIII-D team, overseeing the analysis codebase of the program, providing/improving documentation, setting priorities for computing needs, reviewing proposals for database studies, tracking digital asset responsible officers, and minimizing disruptions to workflows during upgrades and implementations.

3.2. Technical Management

For the DIII-D National Fusion Facility, the Computer Systems and Science Group (CSS) is responsible for 1) the data acquisition, instrumentation, and controls systems unique to the facility's operation, 2) the underlying computer infrastructure encompassing user-support services, computing, networking, and data storage, 3) support for effective and efficient data analysis, and 4) cybersecurity. As such, technical management of DIII-D's data management plan is under the auspices of the CSS Group.

3.3. DIII-D CSS Coordinator

Assigned by the CSS group to maintain communication with the DIII-D team and CWG. Participates in DIII-D program planning and resource management. Responsible for coordinating activities undertaken by the CSS team as part of their technical management described in Section 3.2 above.

4. DATA MANAGEMENT POLICIES

4.1. Types of Data

Typically, research data is obtained during plasma operations of the tokamak. These data originate from a suite of diagnostics (scientific instruments) and analysis codes. Four main types of data are stored: raw, calibrated, derived, and metadata.

4.1.1. Raw Data

These data represent the first acquired elements during plasma discharges. Typically, they consist of measured voltages, current, amplitudes, and images acquired by digitizers, cameras, and similar data acquisition units.

4.1.2. Calibrated Data

Raw data is normally analyzed through stand-alone diagnostic-specific software to yield physical characteristics or parameters of the plasma or associated system, usually referred to as calibrated data. Examples of such data would be magnetic geometry (shape and strength), current profile, density and temperature profiles, power injected, etc.). These calibrated data are validated by individual Responsible Officers (ROs) for each system.

4.1.3. Derived Data

Many of the characteristics/parameters of the plasma are not directly measured and therefore are derived from state-of-the-art codes. Calibrated data are used as inputs to these codes. Derived data, along with calibrated data are used in research, shown in scientific conferences/presentations and published in journals.

4.1.4. Metadata

Metadata refers to the information that helps to explain the raw, calibrated, or derived data. This metadata is sourced from all DIII-D team members through their interaction with the facility.

4.2. Storage of Data

All of DIII-D data is centrally archived at the facility under three main platforms. They are PTDATA (raw and metadata), MDSplus (calibrated, derived, and metadata) and S3 Object Store (camera data).

- PTDATA: Data acquisition and storage system used at DIII-D, mostly for raw data. It controls the shot cycle and initiates the analysis cycle (e.g., MDSplus).
- MDSplus: Set of software tools for data acquisition and storage and a methodology for management of complex scientific data.
- S3 Object Store: Multiple diagnostics are based on video imaging and data is stored in standard video formats available via a Simple Storage Service (S3) interface.

4.2.1. Data Contributed to International Collections

When approved as a DIII-D program project, unpublished research data will be submitted for inclusion in various international projects. The storage and maintenance of this data is the responsibility of the particular project lead, with the technical assistance of DIII-D personnel provided only as agreed through the project definition. Transfer of the relevant data will be determined by agreement of both parties. The DIII-D program does not provide access to this data once it becomes maintained by the international project lead.

4.2.2. Backups and Archiving

Long-term preservation of DIII-D data is accomplished by two redundant copies retained locally and a third copy of the data stored off-site at a location geographically distant from DIII-D.

4.3. Access to Data

Research data is produced throughout the DIII-D program by the combined effort of the entire user base. The subject matter expertise associated with measurements, modeling, and background context is therefore spread widely across the set of individuals and organizations that participate in program activities. Access policies described below are designed to ensure accurate usage of research data, along with appropriate acknowledgement of contributors and the associated sources of funding for any specific data origin. All of this activity is important to complete prior to submitting research results for publication.

4.3.1. Data Sharing for Non-proprietary Projects

It is a policy that all calibrated and derived research data generated through non-proprietary projects at DIII-D is made available to all users. Use of program research data is subject to the terms of the DUPPA.

4.3.2. Data Sharing for Proprietary Projects

The ability to identify research data as proprietary is provided to users. Such projects are individually reviewed by the DOE Contracting Officer for DIII-D. Within that review, specific classifications of which resulting data will be treated as proprietary are required. Users engaging in proprietary provide cost recovery for all facility resources that are utilized in the performance of the project.

4.3.3. Methods of Access

Data are generally accessible from a variety of computer platforms.

4.3.4. Network Connection

Data transfer over the Internet is accomplished through ESnet which services more than 40 DOE research sites and also connects to over 140 research and commercial networks yielding worldwide connectivity.

4.3.5. Data Within Publications

All DIII-D program journal publications are reviewed internally according to the policies of DIII-D Guidance Document 04 (D3D-G04) and submitted to OSTI following publication. Access to data that is published through these journal products determined through engagement between the primary author of the product and the requestor. Access to the publications (in addition to direct

publisher access) is provided through the catalog maintained by OSTI and available at this location, <https://www.osti.gov/>

DIII-D program publications appear in this catalog in three ways, subject to modification at any time by OSTI, which exists outside of DIII-D:

- OSTI provides free download of the Accepted Manuscript, <https://www.osti.gov/biblio/1923376>
- OSTI sends reader straight to free version when publisher provides directly, <https://www.osti.gov/biblio/1897285>
- OSTI indicates when free version will become available, <https://www.osti.gov/biblio/1997400>

4.4. Tools and Resources

A variety of basic data extraction software tools are available to Users and are written in common programming languages (FORTRAN, IDL, Python and others). Visualization tools are also available for more complex data sets. High-level analysis software tools are generally the responsibility of the Users, although general support (licenses, compilers, libraries, job submission tools) is available. A set of computational nodes is available for most common analysis jobs. Users who plan to use significant computational resources (storage, access, computational power, etc.) must contact DIII-D management prior to submitting proposals to any funding agency.