



SIMONS OBSERVATORY

DATA MANAGEMENT AND SHARING PLAN (DMSP)

SIGN-OFF

Function	Name
Author	Simone Aiola (Data Manager)
Reviewer	Suzanne Staggs (Observatory Co-Director)
Reviewer	Mark Devlin (Observatory Co-Director)
Reviewer	Jo Dunkley (Spokesperson)

REVISION HISTORY

Version Letter	Revision Date	Author: Notes
A	2026-08-05	First version released for comments and approval
B	2026-08-26	Version for public release

REFERENCED & APPLICABLE DOCUMENTS

The requirements in the following documents apply, but this document supersedes if there is a conflict.

Reference used within this doc	Title & Description, including SO Document Key if applicable	Notes, relevant part of document
[1]	SO Data Archival Baseline Document	Internal document
[2]	SO Digital Assets and AI Governance Policy	Public document
[3]	SO Developer Guide	Public document
[4]	NSF PAPPG	Public Document
[5]	DOE FY 2026 Continuation of Solicitation for the Office of Science Financial Assistance Program	Public Document

Table of Contents

1. Scope	4
2. Raw Data Generation, Transport, and Storage	4
3. Production of Science-Ready Data Products	4
4. Data Releases and Policies for Reuse	5
4.1 Data Products Release Exemptions (Data Sharing Limitations)	5
5. Software Release and Licensing	5

1. SCOPE

This Data Management and Sharing Plan (DMSP) outlines the high-level data management principles, policies, and processes that govern Simons Observatory (SO) operations. These activities encompass the acquisition of raw data at the observatory site, the generation of science-ready data products, alongside the dissemination of these datasets, associated metadata, and the underlying software to the international scientific community. We do not expect any data used or produced by this project to contain information which is personally identifiable or otherwise sensitive and of a nature which requires special protection. The plan is designed to ensure compliance with the data management and sharing requirements of the Observatory's principal stakeholders, including the Simons Foundation (SF), the National Science Foundation (NSF) [4], and the Science and Technology Facilities Council (STFC). It also satisfies current requirements from the Department of Energy (DOE) [5].

2. RAW DATA GENERATION, TRANSPORT, AND STORAGE

During full operations, the Simons Observatory (SO) will comprise one Large Aperture Telescope (LAT) and six Small Aperture Telescopes (SATs), collectively housing approximately 125,000 detectors that continuously scan the sky and acquire both calibration and science-grade cosmic microwave background (CMB) observations. Raw time-ordered detector data, together with housekeeping and operational data collected across the observatory site (including, but not limited to, telescope pointing, thermometry, and weather), are co-sampled and packaged into data units, referred to as *books*. Time-ordered data are generated at approximately 200 Mbps.

Books are transferred from the observatory site to three geographically distributed computing facilities—the National Energy Research Scientific Computing Center (NERSC) and Berkeley Lab, Princeton Research Computing at Princeton University, and the Universe computing cluster at the University of Manchester—via dedicated fiber connections. Data are typically available at these facilities within 1–24 hours of acquisition, providing both operational redundancy and broad accessibility for processing and analysis activities.

The total archived data volume is projected to reach approximately 2 PB by 2028, with an additional ~1 PB expected each year thereafter. At Berkeley and Princeton University, disk-based (“spinning”) storage will retain the most recent 12–24 months of observations to support routine processing and analysis workflows, while the University of Manchester will retain approximately 24–48 months of data on disk. Older data will be migrated to tape-based archival storage at Princeton University and NERSC. Procedures for large-scale data recovery and full reprocessing campaigns are described in the *SO Data Archival Baseline Document* [1].

3. PRODUCTION OF SCIENCE-READY DATA PRODUCTS

As part of the Data Management work package, raw time-ordered data are processed, calibrated, and conditioned to produce science-ready data products. These products include single-frequency sky maps generated at multiple angular resolutions, depending on the observing instrument (SATs or LAT), as well as the ancillary datasets required for their

interpretation and scientific use. Such ancillary products include, but are not limited to, instrument beams, noise models, and simulated data realizations.

In addition, single-observation maps from the LAT are used to detect transient events and variable sources, which are further characterized by extracting their time-dependent fluxes (light curves).

All data products undergo validation and verification through established science pipelines and quality assurance procedures before being designated as science-ready. In addition to these core data products, higher-level derived products—including gravitational lensing maps, source catalogs, and other science analysis outputs—are generated by the Simons Observatory Analysis Working Groups (AWGs). These products are distributed to the scientific community through the Data Delivery infrastructure developed and maintained by the Data Management team (see Section 4).

4. DATA RELEASES AND POLICIES FOR REUSE

The Simons Observatory is committed to the public release of data products, enabling the broader community to conduct independent research in accordance with internationally recognized FAIR data principles.

The Science-Ready data products described in Section 3 will be released alongside, and documented through, scientific publications, thus allowing for independent verification and replication of results. The release cadence will be determined by project and collaboration needs and tracked by the SO Publication Panel. The Data Management subsystem has developed the capabilities needed to support approximately annual data releases and will demonstrate approximately weekly releases for the transient-data subset. All data products will be released under the CC BY 4.0 license, as described further in the *Digital Assets and AI Governance Policy* [2]. Such products will be made available through the Simons Observatory data release and visualization services (<https://data.simonsobservatory.org/>) and the NASA LAMBDA website.

4.1 DATA PRODUCTS RELEASE EXEMPTIONS (DATA SHARING LIMITATIONS)

The Data Management baseline plan does not include the public release of raw or calibrated timestream data due to storage and user support constraints. Specific requests may be considered on a case-by-case basis, subject to review and approval by the SO Project Office.

5. SOFTWARE RELEASE AND LICENSING

All Data Management software, including the science pipelines used to validate and verify data products, is hosted in the [SO GitHub organization](#) and follows the guidelines outlined in the [SO Developer Guide](#) [3]. Most repositories are publicly accessible from the time of their creation and are licensed for public use in accordance with the Digital Assets and AI Governance Policy [2]. Public repositories currently include software for telescope control, data packaging, data transfer and management, databases, map-making, low-level time-stream conditioning, power-spectrum estimation, cluster finding, and map-level analysis.