DCB Guidance Related to the Elements of an NIH DMS Plan

For applications with receipt date BEFORE **May 25th, 2026**

Element 1A

Data types: List ***ALL data types*** (not only omics data) proposed in the Research Strategy section of the grant application.

Sample types: List the corresponding sample types (e.g., tumor tissue, cell lines, PDX, organoids, primary cell lines, etc.). Sample types must also include the species from which data are being generated.

Sample number: If any omics data are being generated, providing sample number will help determine if the GDS policy will apply to the proposed research project.

Does the GDS Policy apply to my research?
Determine if your research will generate large-scale genomic data and if the GDS Policy applies.

*If **human genomic data** is being generated and the GDS policy applies:*

- A institutional certification (IC) must be submitted with the application or with Just-in-Time (JIT) documents.
- Human genomic data must be registered in dbGaP AND data must be deposited in an **NIH-supported repository within 9 months** of complete dataset collection and quality control.

*If **non-human genomic data** is being generated and the GDS policy applies:*

- Datasets should be available no later than the date of initial publication or end of the award, whichever comes first.

Sample volume: List the amount/volume of data that will be generated. This will help evaluate the budget required towards the DMS plan.

Additional information related to Sample Volume can be found in the DCB Guidance for Estimating the Volume of a Dataset in NIH DMS Plans.

NIH provides discounts on partner cloud services like Amazon Web Services (AWS), Google Cloud, and Microsoft Azure via the NIH STRIDES initiative for cloud computing, storage, and related services.

Element 1A can be presented in a ***tabular form*** that includes the data types, sample types, sample number, and sample volume of all the data to be generated in the proposed research project. ***An example table can be found in DCB Information Related to Element 1A in NIH DMS Plans.***

Element 1B

List the data types generated in Element 1A that will be **preserved and shared**.

Justification of why certain data or data types will not be preserved and shared must be provided. Ethical, legal, and technical factors should guide the extent of data preserving and sharing.

Principal Investigators may consider whether raw or processed data will be shared and in which commonly accepted/agreed upon data formats the data will be shared.

Element 1C

All data that are preserved and shared should be accompanied by their **metadata** and other associated relevant documentation.

Metadata are data about how a dataset or resource came about and how it is internally structured (e.g. the unit of analysis, collection method, sampling procedure, sample size, categories, variables, etc.).

Metadata have to be gathered by the researchers according to best practices in their research community and should be published together with the data.

If no metadata standards are defined for the data types/research field, provide minimum information that someone would need to know to be able to work with the dataset without any further input from you. It is recommended to think as a consumer of the data, not the producer.

Examples of typical metadata elements
Biological material (e.g., species, genotypes, tissue type, age, health conditions)
Biological context (e.g., specimen growth, entrainment, samples preparation)
Experimental factors and conditions (e.g. drug treatments, stress factors)
Primers, plasmid sequences, cell line information, plasmid construction
Specifics of data acquisition
Specifics of data processing and analysis
Definition of variables
LOT numbers
Accompanying code, software used (version number), parameters applied, statistical tests used, seed for randomization

Element 2

State whether **specialized tools, software, and/or code** are needed to access or manipulate shared scientific data. If so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

The use of open-source code and tools is highly encouraged.

Element 3

State what common **data standards** will be applied to the scientific data and associated metadata.

Data standards are pivotal for enabling interoperability of datasets and resources. A data standard is defined as a type of standard, which is an agreed upon approach to allow for consistent measurement, qualification or exchange of an object, process, or unit of information.

Widely accepted research standards should be used, and it is recommended to use the data standard requirements of established repositories where the data is planned to be submitted.

If no consensus standards exist in the scientific field, this should be indicated.

Examples of some community data standards for various data types:

Data Type	Standards	File Formats
Sequencing (RNA, DNA, & next gen)	MINSEQE	BAM, FASTQ
Microarray	MIAME	
DNA hypersensitivity or methylation assays and immunoprecipitation (IP) of proteins followed by sequencing	ENCODE	
Proteomic datasets	MIAPE	
Flow cytometry	FCS	.fcs
Imaging (Microscopy)	OME	PNG, TIFF
Imaging (Electron Microscopy)	EMPIAR	
Medical Imaging (CT, PET, Ultrasound, MRI)	DICOM	DICOM

Element 4A

Selecting a Data Repository
NIH guidance on how to evaluate and select appropriate data repositories.

List the repository or repositories where scientific data and metadata generated will be archived. It is encouraged to preserve and share data through established repositories.

Here are lists of repositories where scientific data generated from an NIH-funded award can be deposited and archived:

1. NIH Supported Repositories
2. Generalist Repositories

*NIH encourages the use of **domain-specific repositories** where possible; however, such repositories are not available for all datasets. When researchers cannot locate a repository for their discipline or the type of data they generate, a generalist repository (which accepts data regardless of data type, format, content, or disciplinary focus) can be a useful place to share data.*

Generalist Repository Comparison Chart
This chart is the latest version (v.4) of the original comparison chart from the NIH Workshop on the Role of Generalist Repositories to Enhance Data Discoverability and Reuse in 2020.

Desirable attributes of repositories where scientific data generated from an NIH-funded award can be deposited include:

- Unique persistent identifiers
- Long-term sustainability
- Metadata
- Curation and quality assurance
- Free and easy access
- Broad and measured reuse
- Security and integrity
- Confidentiality
- Provenance
- Retention policy

Human omics data *that meet the GDS policy threshold of large-scale data should be archived in an NIH-supported data repository.*

Non-human omics data *that meet the GDS policy threshold of large-scale data can be archived in any established data repository.*

Element 4B

Data archived in repositories must be findable and identifiable. An established repository will assign accession numbers, digital object identifiers (DOI), or unique persistent identifiers to deposited data.

Mention how the data will be findable and identifiable. The recommended format in publications is citation of repositories, trackable IDs, and associated URL locations where applicable.

Element 4C

State when and for how long the data will be shared. It is expected that data will be made available at the time of publication or before the end of the award, whichever comes first.

*However, **human omics data** that meet the GDS policy threshold for large-scale omics data must be registered in dbGaP and deposited in an NIH-supported data repository within 9 months of all data collection and quality control (after an initial round of analysis or computation to clean the data and for quality control).*

Repositories usually set time limits on data availability.

Element 5A

Broad sharing of scientific data is highly encouraged.

List, if any, factors that will affect subsequent access, distribution or reuse of scientific data. Provide justification if broad data sharing is not possible.

Element 5B

State if the shared scientific data will be open access or controlled access.

Data generated from human subjects including from patient derived xenografts, primary tumors, organoids, and primary cell lines are recommended to be deposited with controlled access to protect patient identity, even if samples are de-identified.

Element 5C

If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

It is highly encouraged to obtain informed consent from human subjects that includes explicit allowance of broad research use of biospecimens. Additional information can be found at the Considerations for Obtaining Informed Consent webpage.

Element 6

Describe how compliance with this DMS Plan will be monitored and managed at your institution and by whom (e.g., titles, roles).