NATIONAL ACADEMIES Sciences Engineering Medicine

GULF RESEARCH PROGRAM

DATA MANAGEMENT POLICY

VERSION 3.0; EFFECTIVE AS OF OCTOBER 1, 2024

The Gulf Research Program (GRP) of the National Academies of Sciences, Engineering, and Medicine funds research and other activities that generate and disseminate knowledge. Most funded activities produce data. The Federal government defines data in Title 48 of the Code of Federal Regulations (CFR) Section 27.401 as "recorded information, regardless of form or the media on which it may be recorded." The Office of Management and Budget (OMB) further defines data in *Memorandum M-13-13* on Open Data as "structured information" which is to be "contrasted with unstructured information (commonly referred to as "content") such as press releases and fact sheets."[1]

This Data Management Policy is primarily concerned with research data, considered a subset of data defined in Title 2 CFR Section 200.315:

> "Research data means the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: Preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples). Research data also do not include:
>
> (i) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
>
> (ii) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study."

For the purposes of this policy, "data" will refer to "research data" and to other structured information, but not to the items excluded under the definition of research data. "Non-data information products" or "information products" will refer to unstructured content that is created by the funding recipient as part of the funded project.

Examples of data include, but are not limited to:

- Data and data sets
- Databases

[1] Continued explanation: "As described in the *Digital Government Strategy*, content may be converted to a structured format and treated as data. For example, a web-based fact sheet may be broken into the following component data pieces: the title, body text, images, and related links."

- Survey responses
- Images (e.g., maps, data visualizations, photo sets for research purposes)
- Digital models, simulations, and/or software code
- Recorded indigenous knowledge (as defined in 43 CFR 6101.4(h))

Documents such as scholarly publications or peer reviewed journal articles are not considered data, however, the GRP does expect to receive copies or direct links to open-access locations for such documents. Requests for these documents are usually included in project reports over the course of the funding cycle.

Non-data information products are exempt from this policy. Projects that have only non-data information products must indicate this exemption in lieu of a data management. Specific language for such exemptions can be found in *Section 2*. Notwithstanding this exemption, the GRP adheres to Findable, Accessible, Interoperable and Reusable (FAIR) guiding principles[2] for all types of data and information products, and strongly encourages all recipients of GRP funding to submit all information products to a suitable repository, library, or website for long-term availability.

Examples of non-data information products include, but are not limited to:

- Multi-media (e.g., videos, online tutorials, manuals, handbooks)
- Curriculum and lesson plans
- Media and communication products

The GRP supports federal guidance to increase interoperability and openness as described in *Memorandum M-13-13*[3], and supports principles of openness that ensure quality and encourage compatibility as described at resources.data.gov[4] and the Open Knowledge Foundation.[5]

The GRP currently maintains a contract with GRIIDC[6], a Gulf of Mexico science data repository, to help manage and store data and non-data related information products produced by recipients of GRP funding. This contract supports the GRP's goal of making data FAIR, and allows recipients of GRP funding to submit data and information products to GRIIDC at no additional cost. Further information can be found in *Section 2*.

[2] To learn more about FAIR guiding principles refer to: National Academies of Sciences, Engineering, and Medicine. 2018. Open Science by Design: Realizing a Vision for 21st Century Research. Washington, DC: The National Academies Press. https://doi.org/10.17226/25116.

[3] "OMB M-13-13: Open Data Policy – Managing Information as an Asset." May 9, 2013. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf

[4] https://resources.data.gov/categories/data-management-governance/

[5] https://opendefinition.org/od/2.1/en/ - "The Open Definition makes precise the meaning of "open" with respect to knowledge, promoting a robust commons in which anyone may participate, and interoperability is maximized. Summary: KNOWLEDGE IS OPEN IF ANYONE IS FREE TO ACCESS, USE, MODIFY, AND SHARE IT — SUBJECT, AT MOST, TO MEASURES THAT PRESERVE PROVENANCE AND OPENNESS."

[6] Formerly known as the Gulf of Mexico Research Initiative Information and Data Cooperative

SECTION 1. RIGHTS OF ACCESS TO AND DISSEMINATION OF RESEARCH

SECTION 1.1 COPYRIGHTS

The copyrights in all written materials, photographs, drawings, software, and other works subject to copyright protection created or generated under any grant made using the Endowment[7] shall be owned by the recipient of the grant. NAS will encourage the publication and dissemination and other use of these materials. With respect to such copyrighted works, the United States Government and NAS shall have a royalty-free, nonexclusive, and irrevocable license to reproduce, publish, or otherwise use, and to authorize others to use such copyrighted works for Government or NAS purposes. In addition to any other rights it may have, the United States Government shall have the rights provided in paragraph .36(d) of OMB Circular A-110, as it may be revised from time to time, subject to the terms and conditions set forth in that Circular.

SECTION 1.2 RESEARCH DATA

With respect to research data, which shall include the recorded factual material commonly accepted in the scientific community as necessary to validate research findings (but not any preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues), the researcher shall retain all rights in said data but shall provide timely and unrestricted access to the data to NAS and the United States Government. Without limitation of the foregoing, the United States Government and NAS shall have the right to (1) obtain, reproduce, publish, or otherwise use the research data first produced under any grant funded by the Endowment, and (2) authorize others to receive, reproduce, publish, or otherwise use such data for Government or NAS purposes.

SECTION 1.3 PATENTS

The policies on patents outlined in 35 U.S.C. §§ 200-211, in 37 C.F.R. § 401, and in the Presidential Memorandum on Government Patent Policy dated February 18, 1983, will serve as basic guidance on patent rights. Grantees will have the right to elect title to the patent rights in inventions resulting from work under any grant, subject to the United States Government and NAS each acquiring a nonexclusive, nontransferable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States or NAS, but in the case of NAS, solely in connection with the Program, the invention throughout the world in those inventions for which title is elected, and also subject to the "march-in-rights" of the United States Government as set forth in the above-cited statute and regulation. Without limitation of the foregoing, the license provided herein to NAS shall include the right of NAS to sublicense its rights to contractors and grantees that perform studies, projects, or other activities under the Program, except that NAS shall not have the right to commercialize its rights outside the Program.

[7] As defined in the AGREEMENT BETWEEN BP EXPLORATION AND PRODUCTION, INC. AND THE NATIONAL ACADEMY OF SCIENCES, November 2012

SECTION 2. REQUIREMENTS

SECTION 2.1 DATA MANAGEMENT PLANS

Unless otherwise specified, all grant proposals submitted to the GRP must include a data management plan that does not exceed 1,500 words. The data management plan should follow FAIR guiding principles and include the information outlined in *Section 3*. Where proposed projects anticipate no collection, production, or generation of data, the proposal must contain a brief management plan section that states, "This project will not result in the production of the kind of data necessary to validate research findings" with a short description as to why no such data will be generated.

The data management plan should describe how the applicant will manage and disseminate GRP-funded data in sufficient detail to enable evaluation of the plan during the review process.

In proposals involving sub-awards, the lead applicant(s) is responsible for the data management plan for the entire project (i.e., the plan must cover all data that collaborators plan to collect or generate over the course of the project). The lead applicant is also responsible for reporting on the management and the accessibility of all project data and information products in the periodic and final grant reports.

EDUCATION PROJECTS

For funded projects whose main goal is education of students and whose only data come from student projects, and where those products will not be useful outside of an educational context, applicants are not required to complete a comprehensive data management plan or publicly share their data and information products. Applicants who believe their projects fall under this education exemption should include the following statement in a brief data management plan section of their application: "This project is of an educational nature and will not result in the production of the kind of data necessary to validate research findings."

However, if student-produced data is likely to be useful outside of an educational context, then applicants are still required to submit a full data management plan and elaborate on how those data will be shared publicly. In such instances, it is not required for funding recipients to share student data through GRIIDC or in another national repository; instead, applicants should anticipate highlighting or sharing the student data in a more localized fashion, such that the data are findable and accessible, but also properly contextualized within the educational nature of the project. For example, student data and/or final papers could be shared through their college or university's library. If applicants believe their projects fall under this category, they should include language in their data management plan that says, "This project is of an educational nature and will be maintaining student data in an appropriate repository which is not GRIIDC or another national repository."

INTENT TO PUBLISH

If applicants are planning to publish the results of their GRP-funded project in a peer-reviewed scientific or technical journal, then they are required to submit a data management plan. Funding recipients whose projects fall under the educational exemption above are subject to this requirement regardless of the

educational nature of the project and should construct the data management plan around their anticipated published research findings.[8]

SECTION 2.2 EVALUATION PLANS

Some applications for funding opportunities may request submission of an evaluation plan and a data management plan. Where both plans are required, applicants do not need to repeat information between the data management plan and the evaluation plan. If a significant amount of information would be repeated, applicants may include all required information in the evaluation plan and under the data management plan simply state, "The relevant information is included in the evaluation plan." Furthermore, evaluation data that will only be used for internal purposes and are not needed to validate research findings, are not required to be shared publicly. If this is the case, the following statement should be included in either the data management or the evaluation plan: "These data and information products will only be used for internal evaluation purposes and are not necessary to validate research findings".

SECTION 2.3 HUMAN AND ANIMAL WELFARE

All applicants are required to abide by human subjects research laws and policy, as applicable for their projects and institutions. If a proposed project involves research[9] on human subjects[10], the applicant shall comply with the Department of Health and Human Services Regulations (45 CFR Part 46) regarding the protection of human research subjects, unless that research is exempt as specified in the regulation. Submission to an Institutional Review Board is indicated during the application stage as a separate item from the data management plan but can be referenced in the data management plan or evaluation plan if appropriate.

All applicants are further required to abide by animal welfare and wildlife and fisheries research laws and policy, as applicable for their projects and institutions. Applicants are required to obtain the correct permits, licenses, or other documents as needed for their proposed activities.

SECTION 2.4 LONG-TERM STORAGE AND ACCESSIBILITY

The GRP requires data management practices that curate data for future use and that make information widely available for public discovery and access. Recipients of GRP funding must deposit data and metadata in a digital repository, data center, and/or other suitable curation facility that facilitates access to these products and that ensures quality assurance quality control (QAQC) and long-term curation. The GRP's grantees, fellows, and other collaborators are expected to make data generated by GRP-funded projects, initiatives, and activities publicly discoverable and widely accessible in a timely manner, subject to any Institutional Review Board (IRB) or legal restrictions. See *Section 2.5* for required timelines.

---

[8] This requirement follows the Federal definition of published research findings and research data, as found in 2 CFR 200.315(e).

[9] "Research means a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge. Activities that meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program that is considered research for other purposes." (45 CFR Part 46)

[10] Human Subject means a living individual about whom an investigator (whether professional or student) conducting research obtains: (1) Data through intervention or interaction with the individual, or (2) Identifiable private information.

Grantees have the option to put all or part of their data and information products in GRIIDC or submit their data to another approved national repository that also adheres to the GRP's requirements. GRIIDC maintains a list of approved national data repositories which can be found under the FAQ section on their website. If the applicants desire to use a repository that is not on the approved list, then it must be reviewed and approved by the GRP prior to use as the project's long term data storage. The request to use a different repository should be submitted after the application stage and the data management plan as submitted in the application should anticipate using GRIIDC or one of the approved repositories.

Regardless of the primary data repository selection(s), the project must also have its own page at GRIIDC on GRP's specific portal (grp.griidc.org). This page will be the GRP's digital "home" for the project's data, and include links to data, information products, and publications as appropriate. Acknowledgement of this requirement should be indicated in the data management plan.

Recipients of GRP funds who are conducting projects that utilize publicly available data should be aware that significant changes or alterations to the publicly available data will be considered as production or generation of new data and subject to all listed data management and storage requirements. Significant changes or alterations of data should be determined by the replicability of the project with or without the altered data; if the experiment or research would not be replicable without the altered data then that change is considered significant. Additionally, data that are collected, produced, or generated at intermediary stages of a project and then significantly altered at later stages of a project should be treated as separate sets of data. For models with an extensive number of input and output datasets, please consider what is necessary for verification and replicability of the results.

SECTION 2.5 REQUIRED TIMELINES

GRIIDC AND DATASET INFORMATION FORMS

The GRP requires that GRIIDC be contacted within six months of project start date to initiate the creation of the project's homepage on the GRIIDC website. Within those six months, at least one Data Information Form (DIF) should also be submitted to GRIIDC. The DIF is a planning document that is created and submitted on the GRIIDC website and identifies a single dataset for the project; it is likely that more than one DIF will be submitted for each project. This requirement applies whether the data itself will eventually be stored in GRIIDC or in another approved repository.

DIFs submitted to GRIIDC should be updated throughout the length of the project, at least every 6 months or more frequently where appropriate, to reflect any changes to the data that will result from the project. The DIF must be updated and accurate by the time data is submitted.

RELEASE AND SHARING OF DATA

The GRP requires the timely release and sharing of data to be as soon as is practical for each dataset, and no later than the project end date.

If funding recipients so choose, project data may be embargoed during the project timeline such that it is submitted to an appropriate repository or curation facility and is available to the GRP and project collaborators

as soon as the data is complete but is not available to the general public. If submitting to GRIIDC, funding recipients can place download restrictions on any dataset such that it can be reviewed by the repository but will not be publicly available until the restrictions are lifted, which will automatically occur at the end of the project unless other arrangements are specified and approved by the GRP; other repositories likely have similar options.

Data must be accessible to the public with an issued DOI by the project end date. If data are stored at an approved repository other than GRIIDC, the link to the data and issued DOI must also be available in GRIIDC by the project end date.

Requests for extensions for release and sharing of data may be submitted to the GRP, must include a detailed rational, and need to be formally approved in writing. If an extension is granted, it will not impact the project end date and data compliance requirements detailed elsewhere in this policy and any costs associated with said extension will not be covered by the project's contracted budget. No-cost extensions to the project's period of performance are governed by the funding contract and automatically extend the data submission deadline to the new project end date.

SECTION 3. DATA MANAGEMENT PLAN FORMAT

Data management is a vital part of the entire research project process, from project design to completion and dissemination. Applicants and project personnel should manage project information and data through all data life-cycle stages: planning, collecting and assuring (quality and security), describing (i.e., creating metadata and documentation throughout all life-cycle stages), processing and analyzing, preserving, and publishing, curating, and sharing data.

The GRP recognizes that different disciplines may have their own best practices and standards for management of information products, particularly data. Moreover, the GRP understands that accepted norms will likely change as science and other information practices continue to evolve and collaborations across disciplines increase. Therefore, each data management plan should be appropriate for the data the project expects to generate. The data management plan should adhere to (1) data management best practices that are widely adopted across many disciplines, or (2) those data management standards and best practices that are specific to the area(s) of research or activities proposed, as long as neither are in conflict with the data compliance requirements specified in this policy.

Data management plans should be no longer than 1,500 words and contain all of the following six sections and subsections:

1. Data: Planning
    1.1. Describe the kinds of data the proposed project will create or capture (e.g., environmental or ecological data, oceanographic data, climatic data, health or social well-being data, oil and gas safety management, risk assessment data, quantitative modeling data, etc.).
    1.2. Identify the different formats of the proposed data capture, creation, and storage (e.g., tabular data sets, relational databases, geospatial or visual media [i.e., maps], simulations and related data, software code, file formats, etc.).

1.3. If the proposed project involves the use of existing data, identify the original source of the data. Describe when and how the existing data were collected. Describe any relevant sharing arrangements and whether you intend to significantly alter said existing data. (See *Section 2.4* for clarification on significant alterations of data.)

1.4. If the proposed project involves the use of proprietary data or information products, describe the data and information products and the permissions obtained to use them.

1.5. If the proposed project involves the creation or use of confidential or other sensitive information, describe said information, reasons for needing protection, and methods for protection. When working with sensitive information, the applicant must adhere to federal IRB policies, institutional guidelines, and other professional policies and best practices as applicable. For projects involving human subjects, the applicant is responsible for providing evidence of IRB approval or justification for exemption from all applicable regulation (see *Section 2.3* for further information).

1.6. If the proposed project involves animal, wildlife, or fisheries research, identify the process and timeline by which the applicant anticipates receiving the appropriate permits, licenses, or other documentation that is necessary for the continuation of the research. The applicant is responsible for adhering to all relevant federal policies, institutional guidelines, and other professional policies and best practices as applicable.

1.7. Describe the project's quality assurance and quality control procedures.

2. Short-term Management: Collection and Processing

2.1. Identify the hardware, physical facilities, and cyberinfrastructure that project personnel will use to capture and store data during collection and processing.

2.2. Identify software that project personnel will use to capture, store, and process data.

2.3. Describe procedures and protocols for managing and storing data during the project (e.g., organizing and aggregating data, adding newly collected data, altering and correcting data, version control, daily back-ups on and off site).

2.4. If the proposed project involves confidential or other sensitive information, describe security procedures and protocols. Describe provisions for protection of privacy, intellectual property rights, other rights, and other security issues as appropriate.

2.5. Identify the persons responsible for maintaining and managing data during collection and processing. If the proposed project involves confidential or other sensitive information, identify persons who will have access to these data and persons responsible for maintaining security.

3. Metadata: Describe the Data, Data Collection, and Data Processing

3.1. Identify the metadata standards that will be used and describe the metadata that will be created to document the project, the project's data, and relevant information products.

3.2. Identify any other documentation (e.g., guides and manuals) that project personnel will create to describe data collection methods and data processing/analysis.

3.3. Describe the management and storage of metadata.

4. Data Sharing

4.1. Identify policies and procedures that will govern the timely release of and access to project data and related metadata in compliance with *Section 2.5*.

4.2. Describe plans for making project data and associated metadata discoverable and available to others (e.g., other researchers, decision makers, and the interested public).

4.3. Describe the format for citing data products in publications (e.g. [data originator last name], [data originator first name], [data co-originators]. [title of the dataset] [year dataset registered/published]. Distributed by: [data repository name]. Digital Object Identifier (DOI).) so they can be more easily discovered by others. Researchers are encouraged to cite data in the references section of their publication.

4.4. If the project involves confidential or other sensitive information, refer to human subjects research and/or other professional policies and guidance as applicable. Describe the short-term and long-term management of restricted access to sensitive data and other information products.

5. Long-term Management: Curation and Accessibility
   5.1. Outline a schedule in compliance with *Section 2.5* by which your project page at GRIIDC will be created and updated.
   5.2. Identify a curation facility or facilities (e.g., a national digital repository and/or data center) in which project data and metadata will be deposited.
   5.3. Describe the timeline and process for transitioning from short-term management to long-term management of data.
   5.4. Describe procedures that will govern the continued, long-term management of data.
   5.5. Describe policies and procedures that will govern access to project data and other appropriate information products.
6. Data Management Budget
   6.1. Include an allocation in your budget for data management activities throughout the project life-cycle.

POLICY AUTHORIZATION

The GRP will periodically review this policy and update it as necessary, with any changes communicated to their funding recipients. Funding recipients should ensure they are adhering to the most recent version of this policy.

October 1, 2024

_____

Date

Melissa Forbes, GRP Deputy Executive Director