RESTORE Act Centers of Excellence
Data Accessibility and Management Best Practices

Introduction:

Treasury's Office of Gulf Coast Restoration (OGCR) is providing this best practices document to guide the management of data[1] resulting from science funded by Centers of Excellence (COE) Research Grants under the RESTORE Act.

The goals for scientific data management and stewardship are described in the FAIR Guiding Principles, initially developed in 2014 at the Jointly Designing a Data Fairport workshop in Leiden, Netherlands (1). According to these principles, data should be **findable**, **accessible**, **interoperable**, and **re-usable (FAIR)** to facilitate knowledge discovery and innovation.

- **Findable** (or discoverable) means that data are assigned a unique and persistent identifier (such as a Digital Object Identifier (DOI)) and described with rich metadata that are registered or indexed in a searchable resource.
- **Accessible** means that data and metadata are retrievable by their identifier using a standardized communications protocol that is open, free, and universally implementable.
- **Interoperable** means that data and metadata use community agreed formats, language, and vocabularies and metadata is machine-readable.
- **Re-usable** means that data and metadata maintain initial richness, are clearly associated with their provenance, include a clear data usage license, and meet domain-relevant community standards.

In addition, a White House Office of Science and Technology Policy (OSTP) Memorandum "Increasing Access to the Results of Federally Funded Scientific Research" (22 February 2013) states that the results of federally funded scientific research should be made publicly available in a digital format in order to stimulate innovation and accelerate scientific breakthroughs (2). The OSTP Memorandum further instructs federal agencies to encourage the deposit of data in publicly accessible databases and the development of data management plans by extramural researchers receiving federal grants. An Office of Management and Budget (OMB) Memorandum "Open Data Policy – Managing Information as an Asset" (9 May 2013) also directs federal agencies to "manage information as an asset throughout its lifecycle to promote openness and interoperability" (3). The memorandum emphasizes the importance of using machine-readable and open formats, data standards, and common core and extensible metadata for newly generated data in order to facilitate public access and use.

[1] As defined in OMB circular A-110, data are the factual material commonly accepted in the scientific community as necessary to validate research findings. Environmental observations, data collected in a laboratory, and digital audio or video recordings of environmental phenomena are included in this definition. Data may include numerical model outputs, particularly if they are used to support the conclusion of a peer-reviewed publication.

Data Accessibility:

Data resulting from COE awards should be made publicly available in an **open-standard, machine-readable** format and include **appropriate metadata** documentation. Ideally, data and metadata will follow international and US standards, such as International Organization for Standards (ISO) or Federal Geographic Data Committee (FGDC) standards, as applicable. For further guidance and best practices on data standards and metadata documentation, please refer to "Project Open Data" (4) and the National Oceanic and Atmospheric Administration Data Documentation Directive (5).

- **Open-standard** means that proprietary software is not required for the data to be read.
- **Machine-readable** means that the data and metadata are stored electronically in a digital format whose structure is well-described and the data can be read without the aid of a human.
- **Appropriate metadata** means that users can search for, access, read, understand, and use the data.

To meet these criteria for data accessibility, OGCR recommends that data resulting from COE awards be submitted to a data facility that preserves data, properly manages archived data, mints DOIs, and makes archived data available to users in a timely fashion. To facilitate ease of data discovery, state entities and COE are encouraged to select data facilities used by other COEs, such as the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) (6) or the National Centers for Environmental Information (NCEI) (7).

Data are generally expected to be made publicly available at the time of publication of a peer-reviewed article relying on the data or two years after the data are collected. When data resulting from a COE award are made accessible, a brief description of the data and the location (repository, DOI) of the data must be included in the next Performance Progress Report narrative submitted to Treasury per the RESTORE Act Standard Terms and Conditions and Program-Specific Terms and Conditions.[2]

Data Management Plans:

OGCR recommends that state entities require their COE to develop and follow written data management plans in order to consider the management of data and information resources well before the collection or creation of data and to comply with data management best management practices. Data management plans describe the process through which data will move from field

[2] RESTORE Act Standard Terms and Conditions and Program-Specific Terms and Conditions, August 2017, Section D Recipient Reporting and Audit Requirements, 2 Performance Reports, b, i, c-d:

> Section B-3: Summarize any significant findings or events, including any data compiled, collected, or created, if applicable. Section B-4: Describe any activities to disseminate or publicize results of the activity, project, or program, including data and its repository and citations for publications resulting from this Award.

observations to the data user, including the acquisition, quality control, metadata cataloging, storage, and archiving of data. COEs are expected to make data publicly available. Further, state entities are strongly encouraged to have COE consider and prioritize end user needs when designing their data management frameworks. In addition, COEs should engage subrecipients in data management efforts to ensure data are appropriately managed throughout the information lifecycle.

COEs are encouraged to state in RFPs that data management plans are required and to review data management plans in the evaluation of proposals, as applicable. Although OGCR recommends the development of data management plans, submitting them to OGCR is not required.

References:

1) Wilkinson, M.D. et al. 2016. The FAIR Guiding Principles for scientific data management stewardship. Nature 3:1476-4687. https://www.nature.com/articles/sdata201618

2) White House Office of Science and Technology Policy Memorandum "Increasing Access to the Results of Federally Funded Scientific Research" (22 February 2013): https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

3) Office of Management and Budget Memorandum "Open Data Policy – Managing Information as an Asset" (9 May 2013): https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

4) Project Open Data: https://project-open-data.cio.gov/

5) National Oceanic and Atmospheric Administration Data Documentation Directive: https://nosc.noaa.gov/EDMC/documents/DataDocumentationPD-v2.0.0.signed.pdf

6) Gulf of Mexico Research Initiative Information and Data Cooperative: https://data.gulfresearchinitiative.org/

7) National Centers for Environmental Information: https://www.ncei.noaa.gov/