Supplemental DRAFT Guidance: Elements of a NIH Data Management and Sharing Plan (Plan)

To assist those who may be subject to a future NIH Policy for Data Management and Sharing, NIH is proposing supplemental DRAFT guidance regarding elements of a Data Management and Sharing Plan (Plan) for public comment. A Plan should describe in two pages or less the proposed approach to data management and sharing that the specific research will employ. If certain elements of a Plan have not been determined at the time of submission, an entry of "to be determined" may be acceptable if a justification is provided along with a timeline or appropriate milestone at which a determination will be made. Note, NIH does not expect researchers to share all scientific data generated in a study. Elements of a Plan should consider:

1. **Data Type:** A description of the types and estimated amount of scientific data that will result from NIH-funded or conducted research, which scientific data will be preserved and shared, and the rationale for these decisions. Descriptions may include any additional metadata, information, or documentation about the scientific data that will be made publicly available (e.g., study protocols, data collection instruments). In describing the data types to be managed, preserved, and shared, consider:

    - Describing data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., exome sequences of20 to 30 gene variants from an estimated 800 cases and tMRI data from ~100 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be).

    - Providing a rationale for decisions about which scientific data are to be preserved and made available for sharing, taking into consideration scientific utility, validation of results, availability of suitable data repositories, privacy and confidentiality, cost, consistency with community practices, and data security.

    - Identifying metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) which will be made accessible to facilitate interpretation of the scientific data.

    - For scientific data derived from human participants or specimens, outlining plans for providing appropriate protections of privacy and confidentiality (i.e., through de-identification or other protective measures) that are consistent with applicable federal, tribal, state, and local laws, regulations, statues, guidance, and institutional policies.

2. **Related Tools, Software and/or Code:** An indication of whether specialized tools are needed to access or manipulate shared data to support replication or reuse, and name(s) of the needed tool(s) and software. Consider specifying how needed tools can be accessed, (i.e., open source and freely available, generally available for a fee in the marketplace, or available only from the research team or some other source).

3. **Standards:** An indication of what standards, if any, will be applied to the scientific data and associated metadata to be collected, including data formats, data identifiers, definitions, unique identifiers, and other data documentation. While many scientific fields have developed and adopted common data standards, others have not. In such cases, the Plan may indicate that no appropriate data standards exist for the data to be collected, preserved, and shared. Provide the name of any data standards or metadata standards proposed for use, considering:

   - Use of existing, widely adopted standards for scientific data and associated metadata. Some examples include: Clinical Data Interchange Standards Consortium, Minimum Information About a Microarray Experiment, Minimum Information about a high-throughput SEQuencing Experiment, and the Office of the National Coordinator for Health Information Technology Interoperability Standards Advisory.

   - Use of common data elements (CDEs) to facilitate broader and more effective use of scientific data and to advance research across studies. For assistance in identifying NIH-supported CDEs, the NIH has established a Common Data Element (CDE) Resource Portal.

4. **Data Preservation, Access, and Associated Timelines:** An indication of the timelines for data preservation and access, considering:

   - Where scientific data will be archived to ensure long-term preservation (i.e., which repository(ies)). If scientific data will be archived in an existing data repository(ies), consider providing the name and URL web address of the repository(ies). If an existing data repository(ies) will not be used, consider indicating why not and how scientific data will be preserved and shared.

   - How the scientific data will be findable and whether a persistent unique identifier or other standard indexing tools will be used, and any provisions for maintaining the security and integrity of the scientific data (e.g., encryption and backups).

- Whether additional considerations are needed to implement the Plan, (e.g., whether permission needs to be sought to use a specific data repository, and from whom).

- Whether scientific data generated from humans or human biospecimens will be available through unrestricted (made publicly available to anyone) or restricted access (made available only after the requestor has received approval to use the requested scientific data). If the scientific data will be shared through a restricted access mechanism, consider describing the general terms of access for the data.

- Anticipated timeframes for preserving scientific data, describing if different timelines will apply to different subsets of scientific data, and when the scientific data will be submitted to specified data repositories.

- When the scientific data will be made available to other users (e.g., researchers and the broader public). In general, scientific data should be made available as soon as practicable, independent of award period and publication schedule. If applicable, consider indicating when scientific data will no longer be available to other users.

5. **Data Sharing Agreements, Licenses, and Other Use Limitations:** NIH encourages the broadest use of scientific data resulting from NIH-funded or conducted research, consistent with privacy, security, informed consent, and proprietary issues. In describing proposed plans for managing data sharing agreements and other types of arrangements, consider indicating:

   - A description of any restrictions imposed by existing agreements that would limit the ability to broadly share scientific data, as well as a summarizing what those limitations on sharing or reuse are.

   - Whether the applicant anticipates entering into any agreements that could limit the ability to broadly share scientific data and describe those agreements.

   - Any other considerations that may result in limitations on the ability to broadly share scientific data.

   - How relevant limitations to sharing are consistent with community expectations, and how scientific data will be shared to the maximum extent possible while honoring these limitations.

6. **Oversight of Data Management:** An indication of the individual(s) who will be responsible for executing various components (e.g., data collection, data analysis, data submission) of the Plan over the course of the research project and the roles of the individual(s) in data management, and a description of the appropriate expertise for oversight.