Computed Tomography Dataset Management & Sharing Plan

Introduction

Following the U.S. federal guidelines on data management, this plan delineates protocols for the proper management of computed tomography datasets created throughout the project. This vital component of the proposal is designed to align with the intellectual merits and broader impacts relevant to the scientific community. The Non-Clinical Tomography Users Research Network (NoCTURN) has formulated the subsequent recommendations specifically for CT derivatives unrelated to human health and healthcare data.

Section 1: Data Classification

- **Data Types and Quantity**:
    - Detail the types of data and estimate data sizes (e.g., in storage volume, specimen number, or other informative metric) to be generated, including 2D slices, 3D volume renderings, surface files, etc. Identify which data types will be preserved and shared, along with the reasoning behind these decisions (e.g., facilitating further research, encouraging educational use, museum guidelines).
- **Metadata and Documentation**:
    - Enlist metadata, other relevant data, and accompanying documentation (e.g., scan parameters, sample origin, material properties) to be shared, facilitating the interpretation of the scientific data.

Section 2: Technical Requirements

- Mention any specialized tools, software, or codes required to access or manipulate the shared scientific data (e.g., DICOM viewer software, 3D rendering tools), detailing how they can be accessed (e.g., open-source platforms, institution-provided software).
- For CT-specific vocabulary, see the NoCTURN Terminology Resource.

Section 3: Data Standards

- Visit the FAIR Sharing resource to search for common data and metadata types. Describe the common data standards to be implemented on the scientific data and metadata, facilitating interoperability of datasets and resources (e.g., DICOM and TIFF for image formatting, Unicode-8 plain text files for metadata and "ReadMe" documents).
    - If you are unsure about the common data types you will be using, confer with the submission requirements of your preferred repository (see Section 4).

- Specify how these data standards will be applied to the scientific data generated in this project to ensure quality control and reproducibility

- Indicate if there are no consensus standards available, proposing solutions or remedies for absent or inadequate standards (e.g., developing new protocols, adapting existing standards).

Section 4: Data Accessibility and Preservation

- **Repository Information**:

    - Specify the repository(ies) where scientific data and metadata will be archived (e.g., institutional repositories, public data banks). See NoCTURN recommendations for CT data and metadata repositories, as well as the Registry of Research Data Repositories (re3data) for a curated list of repositories by data type (e.g., for non-CT datasets).

    - Detail mechanisms to ensure data findability and identification (e.g., unique dataset IDs, metadata tagging, persistent identifiers like DOIs, indexing of the repository by search engines, citing the dataset in any associated publications). Generally, the more findable the dataset, the more improved its accessibility.

    - Define the timeframe for data availability to other users to match expectations of your funding source (e.g., immediately upon publication). Also, specify the duration of data preservation, if known (e.g., a minimum of years post-project completion). Refer to the policies of your preferred repository, while factoring in costs (if appropriate).

- **Usage and Distribution Policies**:

    - Detail the various factors that might affect how the data generated during the project can be accessed, distributed, or reused after its initial creation and usage. For example, are the data fully open access, under a Creative Commons license, or subject to restricted access (e.g., confidentiality agreements, intellectual property rights)? If appropriate, describe provisions for reuse, redistribution, and the creation of derivatives (e.g., mesh files, coordinate points, 3D surface models).

Section 5: Compliance and Oversight

- Narrate how compliance with the plan will be monitored and managed, specifying the frequency or time frame of oversight and identifying the responsible personnel at your institution (e.g., data stewards, ethics committees, PIs). Refer to your funding agency for guidance about required compliance details.

Section 6: Additional Considerations

- If your work involves bespoke software, custom R functions, or project-specific Python code, also include details about how those aspects of the project methodology will be findable and reusable.

Appendix of Additional Resources:

Data Management Plan Tools and Templates

- DMPTool: Templates are available here for creating DMSPs that comply with US federal grant application requirements.

Federal Guidance on DMSPs

- NSF DMSP Guidance
- NIH DMSP Guidance

Considerations for Human-related Samples

- IFAA Recommendations for the Ethical Use of Anatomical Images
- Operationalizing the CARE and FAIR Principles for Indigenous Data Futures
- Traditional Knowledge (TK) and Biocultural Labels (BC) offer Indigenous communities a tool to add cultural and historical context and authority to digital heritage content.

Persistent Identifier Strategies

- Developing a US National Persistent Identifier (PID) Strategy (e.g., ORCID, DOI, etc.)

Licensing and Intellectual Property

- NSF Licensing Policies
- Intellectual Property Guidelines

NSF PAPPG and Public Access

- NSF Proposal and Award Policies and Procedures
- NSF Public Access FAQ

Data Structuring Protocol

- CURATED protocol for data structuring

Acknowledgements: This document was developed by the Non-Clinical Tomography Users Research Network (NoCTURN) Open Science Committee with peer-review feedback from Fernando Rios and Jim Martin.

Example DMSP for a CT-focused Proposal in Evolutionary Biology for adaptation to your project
we recommend limiting your proposal to one page, if possible

Data Classification: The proposed research will generate several types of primary data, including 2D CT slices, 3D volume renderings, surface files, specimen photographs, and analysis files. The estimated data size is approximately 7 TB, with 6 TB allocated for raw data and 1 TB for processed data. These data types will be preserved and shared to facilitate further research, encourage educational use, and adhere to museum guidelines.

Metadata and Documentation: Metadata will include scan parameters, sample origin, and curation information, ensuring comprehensive documentation for data interpretation. See example table of generated data:

Ref.	Data*	Number	Scale	File type	Preservation	Repository
A)	2D CT Slices	3000/scan	3TB	DICOM, TIFF	10 years	Morphosource
B)	3D Volume renderings	75	2TB	JSON	10 years	Morphosource
C)	Surface files	75	1TB	OBJ, PLY, STL, u3d	10 years	Morphosource
D)	Specimen Photographs	1500	1TB	JPEG	10 years	FigShare
E)	Analysis files	50	200 MB	R scripts, TNT, Executables	10 years	TreeBase, Dryad, MorphoBank
F)	Metadata	up to 75	50 KB	unicode-8 plain text, .pca, .xtekct, .csv	10 years	Morphsource, Institutional Storage

Technical Requirements: Accessing and manipulating the shared scientific data will require specialized tools such as DICOM and TIFF (A) viewer software and 3D rendering (B) tools, which are available through open-source platforms and institutional software provisions. For CT-specific vocabulary, the NoCTURN Terminology Resource will be utilized.

Data Standards: To ensure interoperability, common data standards will be implemented, including DICOM and TIFF (A) for images and Unicode-8 plain text files for metadata (F). Scan files including surface files, 2D slices, and metadata (C, A, F) will be archived in several formats (.stl, .u3d, .tif, .dicom, .pca, .xtekct) relevant to publication and sharing. In the absence of consensus standards, new protocols will be developed or existing ones adapted through consultation with NoCTURN.

Data Accessibility and Preservation: 2D CT data slices, 3D volumes, surface files, and photographs (A, B, C, D) will be managed by the relevant PI and redundantly archived locally on a Drobo 5C RAID array with 14 TB of storage. Data sharing will comply with institutional (e.g., museum) restrictions, and will be coordinated during the project via a shared Dropbox account. CT image stacks, metadata, and

derivatives (A–F) will be made available through public repositories like Dryad, Zenodo, and MorphoBank. Unique dataset IDs, metadata tagging, and DOIs will enhance data findability. All data will be available upon publication and preserved for a minimum of ten years, licensed under CC BY 4.0 for open access, reuse, and redistribution.

Compliance and Oversight: Compliance with this plan will be monitored by the lead PI to ensure adherence to all protocols and address any issues that arise. Annual and end-of-project compliance reports will be included in project summaries to the funding agency.

Publications & Additional Considerations: Results will be published in peer-reviewed, scientific journals, prioritizing open-access options, and will include specimen data, visualizations, evolutionary modeling results, and phylogenetic character lists. Phylogenetic and evolutionary modeling data (E), including R scripts, TNT files, and executable files, will be shared via publications and online platforms like TreeBase and Dryad. Phenotypic datasets and character sets (F) will be made available through MorphoBank in .csv format, coinciding with publication.