DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on [sharing.nih.gov](https://sharing.nih.gov/). The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no “form page” for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Public reporting burden for this collection of information is estimated to average 2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering, and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0001 and 0925-0002). Do not return the completed form to this address.

**Element 1: Data Type**

1. **Types and amount of scientific data expected to be generated in the project:**

*Summarize the types and estimated amount of scientific data expected to be generated in the project.*

1. **Scientific data that will be preserved and shared, and the rationale for doing so:**

*Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.*

1. **Metadata, other relevant data, and associated documentation:**

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

**Element 2: Related Tools, Software and/or Code:  
  
Analyses will be conducted in MATLAB and Python. Model Pages (“the website”) will be built on HTML, CSS, Python, and Javascript. Each of these tools is available directly from their respective company. In addition, interactive versions of all models will be hosted on a permanent sub-domain of the lab’s website [https://ambersmithlab.com/], outlining the implementation while providing a demonstration with adjustable parameters (Figure 1).**

A screenshot of a computer

Description automatically generated

**Figure 1.** PyScript website showing an example Model Page. Model from 10.3389/fmicb.2018.01554.

**Element 3: Standards:**

The models will be constructed in MATLAB and/or Python as appropriate and will follow standards of the respective languages.

**Element 4: Data Preservation, Access, and Associated Timelines**

1. **Repository where scientific data and metadata will be archived:**

All code will be provided via GitHub, along with dependent datasets under 2 GB. For larger standalone datasets (2 GB – 2 TB), storage and DOI identifiers will be handled via ImmPort.

1. **How scientific data will be findable and identifiable:**   
     
   Each Publication will have a persistent and unique DOI link, with usage tracked via views, downloads, and citations, as provided by the publisher.

GitHub will provide a persistent location and identifier (repository name) for models and secondary analysis code. These repositories are publicly searchable. Use of the code will be tracked via GitHub’s usage statistics (Pulse), including Forks (changes to the original project), Clones (download of repositories for any further use), Visitors, and Stars (community attention metric).

Model Pages (“the website”) will be indexed by search engines like Google and will include internal navigation and search functionality. The website will track visitors and traffic sources via Google Analytics. Each of these three components (Publication, Repository, and Model Page) will link to one another.

1. **When and how long the scientific data will be made available:**

The repository will be made available at the end of the project period or the time of the publications as appropriate and will remain available on GitHub indefinitely.

**Element 5: Access, Distribution, or Reuse Considerations**

1. **Factors affecting subsequent access, distribution, or reuse of scientific data:***NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See* [*Frequently Asked Questions*](https://sharing.nih.gov/faqs#/data-management-and-sharing-policy.htm) *for examples of justifiable reasons for limiting sharing of data.*
2. **Whether access to scientific data will be controlled:**

**The repositories will be publicly available, applying to both code and data.**

1. **Protections for privacy, rights, and confidentiality of human research participants:**

*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).*

The publicly available data used in this secondary analysis is fully deidentified such that no connection to the original participants is possible.

**Element 6: Oversight of Data Management and Sharing:**

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