DMSP Example

By Crsito Leon, Ph.D. 9/1/2026

General example for a DMSP

1. Title of Data or Research Product Category

Recommended entry:

Computational Models, Simulation Code, and Model-Generated Data

This fits the form's expectation that the title identify a broad class of research output rather than a single dataset.

2. Description of Data or Research Product Category

Recommended text:

This category includes mathematical and computational models developed to characterize the stochastic multiscale dynamics of mucus barriers, associated numerical simulation code, neural surrogate models, model parameters, configuration files, and model-generated datasets. Research products may include stochastic and deterministic model formulations, numerical solvers, parameter-estimation procedures, uncertainty-quantification routines, trained surrogate-model configurations, validation workflows, and scripts used to generate figures and analyses reported in publications. Model-generated data will include simulated trajectories, parameter sweeps, response surfaces, and derived quantitative measures describing mucus viscoelasticity and responses to environmental perturbations.

This is the level of detail the DMPTool requests for a custom category.

3. Access Policies and Limitations

Selection:

Not Applicable

Additional Information:

The mathematical models, source code, model-generated datasets, and supporting documentation associated with published project results are expected to be publicly shared. No restrictions are anticipated for computational products generated solely from mathematical modeling and simulation.

Because this is a **GOALI** project, I would add one internal note for the PI: if an industry collaborator will provide proprietary parameters, unpublished experimental data, confidential

© 2026-2029 CL Digital Mentoring | Creative Commons CC BY-NC-ND | Atribución-NoComercial-SinDerivadas

manufacturing information, or other restricted inputs, then **"Not Applicable" should not be selected for those specific products**. Instead, the applicable legal, confidentiality, or intellectual-property limitation should be identified and justified. The form explicitly allows these exceptions.

4. Data Standards and Metadata

Because the generic form does not provide predefined choices, select:

Add New

Data/Metadata Standard:

Open computational research formats and repository metadata

Additional Information:

Computational research products will use open, nonproprietary formats whenever practicable. Tabular numerical data will generally be shared as CSV or plain-text files; structured parameter and configuration information will use JSON, YAML, or equivalent human- and machine-readable formats; mathematical and numerical code will be maintained in its native source format; and documentation will be provided in Markdown and/or PDF. Each public release will include a README describing the purpose of the model, software dependencies, parameter definitions, units, file organization, input/output relationships, and instructions for reproducing principal analyses. Version information and release history will be maintained through Git.

This matches the form's requirement to describe formats and metadata when a formal disciplinary standard is unavailable.

5. Data or Research Product Provenance

Selection:

New Data Collection

Although "collection" sounds experimental, the form defines this option as data expected to be newly generated by the proposed project.

Additional Information:

The models, software, parameter sets, neural surrogate models, and simulation datasets in this category will be newly generated during the project. Computational outputs will be produced through mathematical modeling, numerical simulation, parameter estimation, uncertainty quantification, and surrogate-model development conducted by the project team. Where external or previously published datasets are used for model calibration or validation, those resources will be identified separately with their original source, persistent identifier, and applicable reuse conditions.

© 2026-2029 CL Digital Mentoring | Creative Commons CC BY-NC-ND | Atribución-NoComercial-SinDerivadas

That final sentence is useful because the mucus project may eventually combine newly generated simulations with existing biological or experimental datasets.

6. Public Archiving

For this field, I would **not use GitHub alone as the preservation mechanism**.

Selection:

Add New

Repository:

GitHub with archival releases deposited in Zenodo

Additional Information:

GitHub will be used for version control, collaborative software development, documentation, issue tracking, and public dissemination of computational research products. Stable versions associated with publications and major project milestones will be released through GitHub and archived in Zenodo or an equivalent persistent research repository. Archival releases will receive a persistent identifier (DOI) and will include the source code, model definitions, configuration files, documentation, and supporting model-generated datasets necessary to reproduce the corresponding published results.

This creates a clean division:

GitHub → active computational research environment
Zenodo → persistent archival record and DOI

The DMPTool asks where the products will be made publicly available and anticipates an online repository.

7. Timeline for Public Accessibility

Selection:

At time of publication

Additional Information:

Research products supporting a peer-reviewed publication will be made publicly accessible no later than the time of publication. Where feasible, software and documentation may be released earlier through the project's GitHub repository. Versioned archival releases corresponding to published results will be deposited in the designated archival repository at the time of publication.

© 2026-2029 CL Digital Mentoring | Creative Commons CC BY-NC-ND | Atribución-NoComercial-SinDerivadas

This follows the guidance you provided, which states that data supporting NSF-funded publications are to be shared at the time of publication unless an exception is justified.

8. Data Availability

I recommend:

Indefinitely

Additional Information:

Publicly archived versions of the computational models, software, documentation, and supporting model-generated datasets will be retained according to the preservation policies of the selected archival repository. The project intends for these research products to remain publicly accessible indefinitely through persistent repository records and associated DOIs.

"Indefinitely" is available as an explicit option in the form, and repository retention policies can govern the actual preservation period.

9. Will you adhere to the standard retention policy of the chosen repository?

Selection:

Yes

No additional explanation should normally be necessary because the form only calls for justification when the answer is "No."

10. Accountability

Selection:

[PI Name]

If the PI has a co-PI specifically responsible for computational infrastructure or data management, that person could instead be selected.

Example supporting language, if a text field is available:

The PI will have primary responsibility for ensuring that computational models, source code, model-generated datasets, metadata, and associated documentation are maintained, versioned, reviewed for release, and shared according to the procedures described in this DMSP. Project personnel generating computational research products will be responsible for maintaining appropriate documentation and transferring final materials to the designated public repositories.

© 2026-2029 CL Digital Mentoring | Creative Commons CC BY-NC-ND | Atribución-NoComercial-SinDerivadas

The DMPTool specifically expects one of the project's Senior Personnel to be identified as responsible for each category.

For this particular grant, I would actually recommend that the PI ultimately have at least three Research Product Categories rather than putting everything under one heading:

1. **Computational Models, Simulation Code, and Model-Generated Data** — the category above.
2. **Neural Surrogate Models and Machine-Learning Research Products** — trained models, architectures, hyperparameters, training/validation scripts, checkpoints as appropriate.
3. **Experimental or Biological Data Used for Model Calibration and Validation** — especially important if the GOALI partner or another collaborator provides empirical mucus measurements.

The third category matters because its access restrictions, provenance, formats, and sharing requirements could be fundamentally different from the mathematical/computational products. Separating them keeps the DMSP much cleaner and prevents an industrial confidentiality restriction from unnecessarily appearing to restrict the project's publicly shareable mathematical models and software.

© 2026-2029 CL Digital Mentoring | Creative Commons CC BY-NC-ND | Atribución-NoComercial-SinDerivadas