GFDL Fair Use Policy for Code and Data

V 1.0
Effective Date: 10/9/2020

SUMMARY

The GFDL Fair Use Policy is a set of guidelines for sharing code and data (models and model output) within GFDL, with collaborators, and with the wider research community. GFDL is committed to making its products available to maximize their scientific and societal impact, while also striving to ensure that this is done in a fair manner that recognizes the diverse contributions that make these products possible.

Models include the source code and documentation, input and configuration files, scripts and instructions to enable researchers to independently run the model for their own scientific purposes. *Model output* consists of datasets and documentation to enable researchers to independently analyze the output of GFDL models. Scientists producing original research in the form of model enhancements, experiment design, or analysis methods should retain first access to these innovations. Collaborators should not use such innovations and associated data without the consent of the authors. Collaborators, particularly those within the GFDL community, can request pre-publication access to experimental products, and negotiate directly in matters such as publication precedence (e.g., a deferral agreement), or co-authorship in return for specific contributions. Such collaborations are generally encouraged. Following publication of results from original research in peer-reviewed literature, these products can be shared with the wider community.

It is the general expectation that the products of publicly funded research will be made available to the wider community after quality assurance and peer review, consistent with NOAA and academic journal policies, and taking into consideration partner policies as well. Individual research projects and collaborative projects may follow different use policies, which may be less restrictive than this document (e.g., "open development"), but cannot be more restrictive. Scientists who leave GFDL can retain access to their pre-publication data for a defined period. Publicly shared research products will receive support (e.g., help with model configuration, scientific advice for novel configurations) on a best-effort basis and as resources permit, or through a formal collaboration.

BACKGROUND

GFDL's models, code, and data are valuable resources to our community, as well as to the general public. These resources should be used and shared in a legal, ethical, and responsible manner in order to uphold scientific integrity.

OBJECTIVE

Develop and implement a Fair Use Policy for accessing, using, sharing, and supporting code and data created while employed at, or in collaboration with, GFDL, that facilitates broad dissemination of GFDL products while ensuring such dissemination occurs in a fair manner.

METHODOLOGY

This GFDL Fair Use Policy for Code and Data was created using results from an internal survey in order to drive the process from the bottom-up and provide a measure of transparency.

DEFINITIONS

Data includes (but is not limited to) model input, model output, data-derived products, videos, visualizations, and figures.
Code includes (but is not limited to) model code, analysis scripts, configuration files (e.g., XMLs), parameter settings, and configuration software (e.g., gridspec generation).
Stakeholders for GFDL's code and data include:

- **Community** - includes GFDL Federal Employees, Cooperative Institute researchers actively collaborating with GFDL, UCAR employees, USGS employees and government contractors engaged to work on/with GFDL code and data.
- **Partners** - includes NOAA, other federal agencies (e.g., DOE, U.S. Navy), visiting scientists, visiting faculty, academic staff partners (e.g., NCAR), and summer interns who are actively working on/with GFDL code and data.
- **Academia** - individuals or groups at academic or research institutions, including international institutions, who are not actively collaborating with GFDL or contributing to GFDL efforts.
- **External** - any partnerships with the private sector, non-profit agencies, NGOs, foundations, etc.
- **Public** - general public including users of CMIP repositories, GitHub contributors, journals.
- **Special** - any single instance requests (e.g., British Petroleum oil spill).

FAIR USE POLICY

1) The Fair Use Policy addresses accessing, using, sharing, and supporting both published and developmental code, and associated data.

2) As a start, the document exists to address who has access to what and when.

3) The Fair Use Policy should be viewed as a living document. If updates or changes are warranted, a new version of the Fair Use Policy will be created. Each version of the Fair Use Policy document will be assigned a unique version number and a date.

4) Access and Use
    a) Published code and associated published or unpublished data

● Code and data are open to anyone to access and use once in the public domain.
● Typically, those involved in lab-wide, small group, cross-division, or individual developmental efforts prior to code and data entering the public domain are offered credit through co-authorship on "documentation papers", and data citation author lists.
● Anyone using published code and/or associated published data should reference relevant DOI(s). If using unpublished data, credit/recognition should be offered to contributors when publishing or presenting results from unpublished data.
● Credit/recognition should give due weight to all the individual (scientific, software engineering, technical) contributions involved in producing the published code and/or associated published or unpublished data.
● Credit/recognition should be given to all the individual (scientific, software engineering, technical) contributions involved in producing the developmental code and associated data.
● Access and use of developmental code and data is governed by the Fair Use Policy for experimental models (hereafter FUP/EM, Appendix I).

b) Single-division or cross-division, group or individual projects
● While code is under development for a small group, within one division or cross-division, or individual endeavor, typically only collaborators engaged in the effort should be accessing the code and associated data. An example of this type of project is developing a prototype model.
● Anyone in the Community or any Partner who is not a collaborator or contributor to the effort, and who wishes to access and use developmental code and/or associated data in these cases, should request permission from individuals involved in the effort, *and* consult with the developers and any Public code contributors to offer appropriate credit (e.g., co-authorship) before publishing or presenting results.
● If permission is not granted to Community or Partners to access and use developmental code and/or associated data in these cases, mediation may be requested from the GFDL Research Council. The final authority rests with the GFDL Director.
● Stakeholders from Academia, External, Public (excluding code contributors), or Special who have a specific need for code and/or data from a smaller group, single or cross-division, or individual developmental project should engage in discussions with the individuals involved in the effort. A Collaboration Agreement Form (CAF) should be drawn up covering the scope of work, timeframes, resources, expectations, risks, and outcomes, as well as credit/recognition for personnel (scientific, software engineering, technical) involved.

c) Active/Ongoing Collaborations
● Community and Partners who are transitioning to external positions should have a robust opportunity to complete their work begun at GFDL. It is anticipated that they will continue to have access to code and data from their individual projects for the length of time needed to complete those projects -- typically one to two years, but sometimes longer. For the case of ongoing long-term collaborations this access can be

extended indefinitely. However, the length of this access will depend on the individual circumstances, and may be impacted by issues such as security and conflict-of-interest considerations. It is recommended to consult with the relevant GFDL Division Head well in advance of departure to discuss arrangements. Note that current NOAA policy prohibits access by foreign nationals from outside the US. Additional access and use of code and data, including access to HPC resources beyond the 2-year timeframe, can be negotiated with Division Heads if active collaborations are desired beyond the 2-year timeframe.
● Duration of access and use for any Academia, External, Public (excluding code contributors), or Special stakeholders should be decided by mutual agreement.

d) Non-active collaborations
● Access and use for Community and Partners who are no longer active collaborators should be terminated by Division Heads by requesting deactivation of HPC and other accounts.
● Resignation, removal, separation from the federal government, and termination of federal employment are special cases. Former federal employees are subject to post-government employment restrictions that may limit the type of work they may perform for their new employer for certain periods of time.
● Post-government employment rules fall into five categories: restrictions on contacting the U.S. Government, restrictions on providing advice or other services, restrictions on compensation and employment, restrictions on using non-public information, and disclosure requirements. Two relevant restrictions are noted here:

✓ Two-Year Restriction [18 U.S.C. § 207(a)(2)]: For matters under your official responsibility during your last year of Government service, you are restricted for two years after you leave Government service from representing any non-Federal entity to any Federal department, agency, or court regarding those matters.
✓ Additional laws apply to former senior employees and former very senior level employees. One-Year Restriction on Communication with One's Former Agency - 18 U.S.C. § 207(c): For one year after leaving senior service, no former senior employee may make, with the intent to influence, any communication to or appearance before the Department or agency in which he or she served in the one-year period prior to termination from senior service. Consult your ethics counselor for certain limited exceptions to this prohibition.

5) Code and Data Sharing

Code should contain a license that covers sharing and redistribution of the code. Data should have an associated Digital Object Identifier (DOI) and should also contain a license that covers sharing and redistribution of the data.

a) Public domain code and data
- If code and data are intended for the public domain, then Community and Partners involved in the project are responsible for working together to serve code and associated data publicly.
- Sharing and redistribution of code in the public domain is unrestricted (or governed by the code license if available).
- Sharing and redistribution of data in the public domain is unrestricted (or governed by the data license if available).
- DOIs must be created for data entering the public domain and should be used to cite data.
- Data quality should be assured before entering the public domain.
- Data and code should not enter the public domain unless there is an associated or forthcoming peer-reviewed manuscript or technical document.

b) Developmental code and associated data
- Sharing and redistribution of developmental code and associated data are inherently covered by the current Fair Use Policy for experimental models (FUP/EM) as developmental activities and projects are collaborative in nature.
- When in doubt, it is advisable to check with respective individuals or groups before sharing or redistributing developmental code or data.
- Developmental code and data should be traceable and research should be reproducible.

c) Manuscripts
- In accordance with the NOAA Public Access to Research Results (PARR) policy, which affects anyone receiving government funds, all peer-reviewed published papers will become public after one year.
- Code and associated data (if not already provided to journals as part of their requirements) should be in the public domain within one year of publication of a journal article, "documentation paper" or other relevant manuscripts. The timeframe is in accordance with the NOAA PARR policy.

6) Code and Data Support
Code and data support (including training) is time-consuming and requires resources. This will be done on a best-effort basis. Code and data support and training should be recognized and valued.

a) Public domain code and associated data
- Community, Partners and Public provide code and data support and training as deemed reasonable by individuals involved, to all stakeholders accessing and using code and data in the public domain.

b) New and ongoing collaborations
- Code and data support is provided when circumstances warrant (e.g., Special), or when such support has been explicitly requested for new and ongoing collaborations with Partners, Academia who transition to Partners, and External stakeholders, through a proposal or Collaboration Agreement Form (CAF). The CAF should cover the scope of work, timeframes, resources, expectations, risks, and outcomes as well as

credit/recognition for all personnel involved in providing support and training.
c) Porting code and data
● Support for porting code and/or transferring data to new compute systems is not typically provided to Community and Partners who are transitioning to new external positions, or are no longer affiliated or collaborating with GFDL.
● If such support is desired from Modeling Systems Division, or any other individual or divisions, a CAF should be developed so that the scope of work, time frames, resources needed, expectations, risks, and outcomes are clearly defined, and credit or recognition is offered to all personnel involved in providing support and training.

REFERENCES
● NOAA Administrative Order 202-735D: Scientific Integrity
● White House Office of Science and Technology Policy Memorandum (issues 2/22/2013), "Increasing Access to the Results of Federally Funded Scientific Research"
● NOAA Public Access to Research Results (PARR)
● NOAA Institutional Repository
    -- https://repository.library.noaa.gov/
    -- https://repository.library.noaa.gov/faqs
● GFDL Co-Authorship Guidance (August 25, 2016)
● CRediT: https://www.casrai.org/credit.html
● Restrictions on Post-Government Employment:
https://www.doi.gov/ethics/restrictions-on-post-government-employment
● Chapter 31 - Separations by Other than Retirement – OPM:
https://www.opm.gov/feddata/gppa/gppa31.pdf
● Balaji, V., Taylor, K. E., Juckes, M., Lawrence, B. N., Durack, P. J., Lautenschlager, M., Blanton, C., Cinquini, L., Denvil, S., Elkington, M., Guglielmo, F., Guilyardi, E., Hassell, D., Kharin, S., Kindermann, S., Nikonov, S., Radhakrishnan, A., Stockhause, M., Weigel, T., and Williams, D.: Requirements for a global data infrastructure in support of CMIP6, Geosci. Model Dev., 11, 3659-3680, https://doi.org/10.5194/gmd-11-3659-2018, 2018.

APPENDIX I

GFDL Fair Use Policy for Experimental GFDL models

Experimental GFDL models refers to models (e.g., a coupled climate model), model components (e.g., parameterization schemes), and model configurations (e.g., the specific arrangement and parameter settings of model components) arising from model development efforts at GFDL, but whose formulation and configuration have not yet been documented in the peer-reviewed literature, as well as outputs from simulations with these models. Those wishing to make use of such data should contact the model developers to discuss potential model use, and request approval from the developers for planned use of the data. It is strongly desired that such use would be in the form of a collaboration with the developers. Any products derived from the model use (papers, presentations, etc.) should give appropriate credit to both the model developers and the

collaborators.