

An Urgently Needed Repository for Planetary Atmospheric Model Output

Submitted to the Decadal Survey on Planetary Science and Astrobiology
Co-Chairs Robin Canup and Phil Christensen

Authored by:

Claire E. Newman (<i>Aeolis Research</i>)	Vladimir Airapetian (<i>NASA-GSFC/American Univ.</i>)
J. Michael Battalio (<i>Yale University</i>)	Stephen Bougher (<i>University of Michigan</i>)
Adrian Brown (<i>Plancius Research</i>)	Shawn D. Domagal-Goldman (<i>NASA-GSFC</i>)
Siteng Fan (<i>Caltech</i>)	Scott D. Guzewich (<i>NASA-GSFC</i>)
Nicholas G. Heavens (<i>Space Science Institute</i>)	Derek Jackson (<i>Ulster University</i>)
Melinda Kahre (<i>NASA-ARC</i>)	Michael A. Mischna (<i>JPL</i>)
Tim McConnochie (<i>SSI/Univ. of Maryland</i>)	Lynn Neakrase (<i>New Mexico State University</i>)
Alexey Pankine (<i>Space Science Institute</i>)	Jorge Pla-García (<i>CAB, CSIC-INTA</i>)
Mark Richardson (<i>Aeolis Research</i>)	Isaac Smith (<i>York Univ./Planetary Science Inst.</i>)
Anezina Solomonidou (<i>ESA-ESAC</i>)	Alejandro Soto (<i>Southwest Research Institute</i>)
Anthony Toigo (<i>Johns Hopkins Univ./APL</i>)	Daniel Viúdez-Moreiras (<i>CAB, CSIC-INTA</i>)

Endorsed by:

Shiblee Barua (<i>NASA-GSFC/USRA</i>)	Meiring Beyers (<i>Klimaat Consulting Inc.</i>)
Bruce Bills (<i>JPL</i>)	Peter B. Buhler (<i>Planetary Science Institute</i>)
William C. Danchi (<i>NASA-GSFC</i>)	Claus Gebhardt (<i>UAE University</i>)
Andrew Gunn (<i>University of Pennsylvania</i>)	Sarah Hörst (<i>Johns Hopkins University</i>)
Mathieu Lapotre (<i>Stanford University</i>)	Christopher Lee (<i>Aeolis Research</i>)
Mark Lemmon (<i>Space Science Institute</i>)	Timothy A. Livengood (<i>Univ. Maryland / GSFC</i>)
Rosaly M.C. Lopes (<i>JPL/Caltech</i>)	Juan Lora (<i>Yale University</i>)
John E. Moores (<i>York University</i>)	Marc Neveu (<i>University of Maryland / GSFC</i>)
Mark Panning (<i>JPL</i>)	Gioia Rau (<i>NASA-GSFC</i>)
Kevin Roback (<i>Caltech</i>)	Deanne Rogers (<i>Stony Brook University</i>)
Suzanne Smrekar (<i>JPL</i>)	Nick Teanby (<i>University of Bristol</i>)
Melissa G. Trainer (<i>NASA-GSFC</i>)	Regner Trampedach (<i>Space Science Institute</i>)
Roland Young (<i>NSSTC/UAE University</i>)	

Contact: Claire E. Newman (*Aeolis Research*), claire@aeolisresearch.com, (626) 344-0593

Summary:

Output from planetary atmospheric models is crucial for a diverse range of NASA-supported activities: from assessing risks for spacecraft and human exploration, to studying the geology of ancient Mars or magnetospheres of giant planets, to engaging the public. It is right for NASA to fund work that includes the development and use of planetary atmospheric models. However, many NASA-supported activities could be enabled by use of existing model output, produced by previously-funded work, avoiding duplication of effort and saving NASA time and money.

Understanding this, NASA increasingly encourages researchers to make model output publicly available. The problem is in facilitating awareness and access to this valuable resource.

We recommend NASA invest in a community-recognized, discipline-specific repository for planetary atmospheric model output. This will increase the efficiency with which engineers, scientists, and EPO professionals utilize such output, and reduce costs to NASA.

1. Introduction and motivation.

NASA currently funds a wide range of planetary atmospheric modeling activities. In recent years, these have covered such disparate topics as very-high-resolution simulations of landing sites on Mars, moderate-resolution simulations of lake circulations on Titan, global climate models of Venus, Pluto and exoplanets, and global ionosphere-magnetosphere models of Jupiter. The results of such models are valuable across all stages of mission life cycles: from conception and development, to orbital insertion or landing, to data analysis and hypothesis testing, with the latter sometimes occurring decades after the mission itself has ended. The development of new models, or the significant update of previously developed ones, can lead to a revolution in understanding of significant aspects of a planetary body, often going far beyond its atmosphere.

Typically, planetary models are developed, maintained, and run by small groups of scientists in order to answer specific research and/or engineering questions. While simple 1D or energy-balance models may sometimes be sufficient, most questions require high-fidelity models, which must typically simulate an *entire*, integrated planetary system in order to predict *one aspect* of it. A single group may be funded to address several distinct scientific questions at any given time; yet, even in the case of a group with broad interests and collaborations, it is likely that only a fraction of model output generated is of value to their particular investigation(s). The result is a wealth of information that would certainly be applicable to other investigations, but which is often unseen by other potential users in the wider science community, and hence may go unused.

1.1. Produced for one purpose, useful for ten others.

NASA has recently funded several models to predict the *near-surface* meteorology that will be *observed by Mars 2020 sensors* in Jezero crater *along the rover's traverse*. These predictions will soon be published; yet these models also output *more variables than will be observed*, and do so *up to at least 50 km altitude and over a far wider area*. Thus, if the complete model output generated were made available to the engineering, scientific, and education and public outreach (EPO) communities, it could support a huge number of additional projects. Examples include: risk assessment for the Mars Helicopter on Mars 2020; comparisons with sedimentary rock and aeolian features imaged across the region; detailed studies of planetary waves and thermal tides; predictions of trace gas variations over the mission; and outreach in the lead-up to landing.

As another example, consider the huge interest in sending balloon missions to the denser atmospheres of Venus and Titan. A key need for studies of such missions is to assess whether a sufficient fraction of the primary science destinations could be reached by a balloon that drifts in the winds and, if not, how much powered navigation is required. Predictions of the circulation of Venus and Titan are therefore crucial¹. Yet - despite NASA having funded decades of Titan and

Venus global atmospheric model research and dozens of peer reviewed papers, to explore topics ranging from astrobiology to chemistry to geology to circulation dynamics - it appears that output from only one such Titan model and no Venus models is publicly available worldwide. Note, too, that the development of the Titan and Venus Global Reference Atmospheric Models (GRAMS), which are widely used by many NASA engineers, would be far easier if output from the latest, state-of-the-art models for Titan and Venus were made publicly available.

Finally, while *observations* of planetary atmospheres include data gaps (due to the spacecraft orbit, scheduled downtimes, data glitches, etc.), planetary atmospheric *models* produce seamless datasets. Indeed, models are increasingly utilized to fill in such data gaps (via data assimilation, for example). This makes their output especially well-suited for education and public outreach (EPO) activities that involve visualizations of planetary phenomena or NASA research, such as “Science on a Sphere” <https://sos.noaa.gov/Gallery> or “Hyperwalls” <https://svs.gsfc.nasa.gov/hw>

In summary, terabytes of model output may be generated that are incidental to the funded engineering or science needs for which they were produced, but which can be highly beneficial to *other* engineers, scientists, and EPO professionals who may be actively seeking that exact information. With little effort, one can think of dozens more potential uses for the model outputs described above, ranging from topics in astrobiology to geology to human exploration - and hundreds more if one considers a wider range of models, simulation types, and planetary bodies.

1.2. Facilitating collaborations and discoveries.

In many cases, engineers, other scientists, and EPO professionals form collaborations with modeling groups to gain access to the model output they want. Note that, in some cases, this is absolutely required because the model output sought does not yet exist and must be generated. Even if the model output already exists, collaborations may still be very valuable, especially if the user is not an atmospheric scientist: modelers bring expertise that may aid in interpreting results, and such collaborations often lead to new hypotheses and new modeling projects.

In other words, we are not suggesting that such collaborations should cease. Rather, we argue that - especially given current NASA R&A funding limitations - any means by which *existing* resources can be leveraged by researchers is a cost-effective measure. Particularly if all that stands between a research project or mission and the atmospheric model output required to enable or improve it is knowledge of the existence of - and access to - said output.

We also note that public availability of model output, in a repository that enables easy data access for a range of users, may actually make it *easier* for modeling groups and end-users to collaborate on more projects, as it relieves modelers of the overhead of extracting, distributing, and facilitating usage of the desired output for new users. All of this strongly motivates making more output from NASA-funded planetary atmospheric models publicly available, and in such a way that engineers, other researchers, and EPO professionals can both find *and* easily use it.

2. The problem and proposed solution.

NASA already encourages the sharing of data generated by NASA-funded research in its Announcements of Opportunity (AOs), even *requiring* this in some Data Management Plans.

The central problem is that no community-recognized, discipline-specific repository exists for planetary atmospheric model output.

Note that the Planetary Data System (PDS) *does not* currently accept model output.

2.1. The value of a community-recognized repository.

The lack of a community-recognized repository means that whatever planetary atmospheric model output *is* made publicly available is placed wherever each researcher decides to put it, and in whatever format they choose or are required to use by a journal or funding source. The several

thousand repositories available - from university repositories, to open public repositories (e.g. Figshare), to NASA's High End Computing Data Portal (data.nasa.gov) - make it all-but impossible to locate a desired set of planetary atmospheric model output, *unless* its existence *and* location are already approximately known. This leaves few outside the immediate modeling group even cognizant of the availability and value of a particular set of model output.

NASA could designate a suitable repository in which all planetary atmospheric model output should be placed, then advertise this to modeling groups and potential users via AOs, as the PDS and other data archives are advertised already. One option may appear to be NASA's Open Data Portal (data.nasa.gov), which allows users to search for datasets that are either archived locally *or* in other NASA archives. However, data.nasa.gov has barriers to either hosting model output generated outside NASA centers or linking to output not archived in other NASA repositories.

2.2. The greater value of a discipline-specific, community-recognized repository.

Establishing a discipline-specific repository for planetary atmospheric model output would be far more powerful. It would provide a single entry point that enables users to see the range of outputs available, almost at a glance, and to quickly identify which, if any, meet their needs. Unlike NASA's Open Data Portal, it would be easily hierarchical and organizable by planetary body (Mars, Pluto, etc.) and model type (global, limited-area, large-eddy). It would also have the ability to search by the above or other factors - for example, by modeling group, model name, research area (e.g. paleo Mars), mission tie-in (e.g. MSL), etc. It should be available to modeling groups not based at NASA centers, and could even access model output already held in non-NASA repositories (see end section 3.2), including output generated under non-NASA funding.

Another major advantage of a discipline-specific repository for planetary atmospheric model output is that datasets contained therein will be structurally similar. All atmospheric model outputs have up to three spatial dimensions and (usually) a time dimension. If a basic data format (e.g. netCDF) and metadata are mandated, this allows generalized tools to be developed to access, download, visualize, and manipulate the output, without asking each modeling group to provide such tools *or* expecting users to write their own and risk introducing errors, especially if they are unfamiliar with such datasets (e.g. if they are not atmospheric scientists or modelers).

These generalized tools should include a Graphical User Interface (GUI) that allows novice users - or those with simple requirements - to straightforwardly print, plot, or download only what output they need, as well as Extraction, visualization, and analysis code that can be run by more experienced users who are willing to download the full output dataset. These would increase the accessibility and, thus, usage of model output held in the repository; result in time and cost savings at the user's and/or provider's end; and reduce the risk of introducing errors.

We therefore recommend that NASA support the development of a new repository for planetary atmospheric model output, to include:

- i. Local hosting and remote serving of planetary atmospheric model output with a mandated data format and metadata guidelines.**
- ii. General tools, including a web interface, for accessing, visualizing, and downloading desired data from any of the model output datasets.**

3. A repository for planetary atmospheric model output.

In this section, we provide a better sense of how a repository for planetary atmospheric model output might be established and structured, and how it might work at a technical level.

3.1. Examples of existing repositories that contain atmospheric model output.

Many discipline-specific repositories already exist to benefit the scientific community. For

Earth climate modeling, large repositories permit users to investigate the predictions of many different weather and climate models. These repositories are typically federally-funded via institutions such as the National Center for Atmospheric Research (NCAR) or the National Oceanographic and Atmospheric Administration (NOAA). Figure 1 shows the landing page of the IRI/LDEO Earth Climate Data Library (<https://iridl.ldeo.columbia.edu>). The benefits of a discipline-specific repository are immediately apparent, with datasets shown as well-organized and searchable, and with many tools available to visualize, download, and manipulate them.

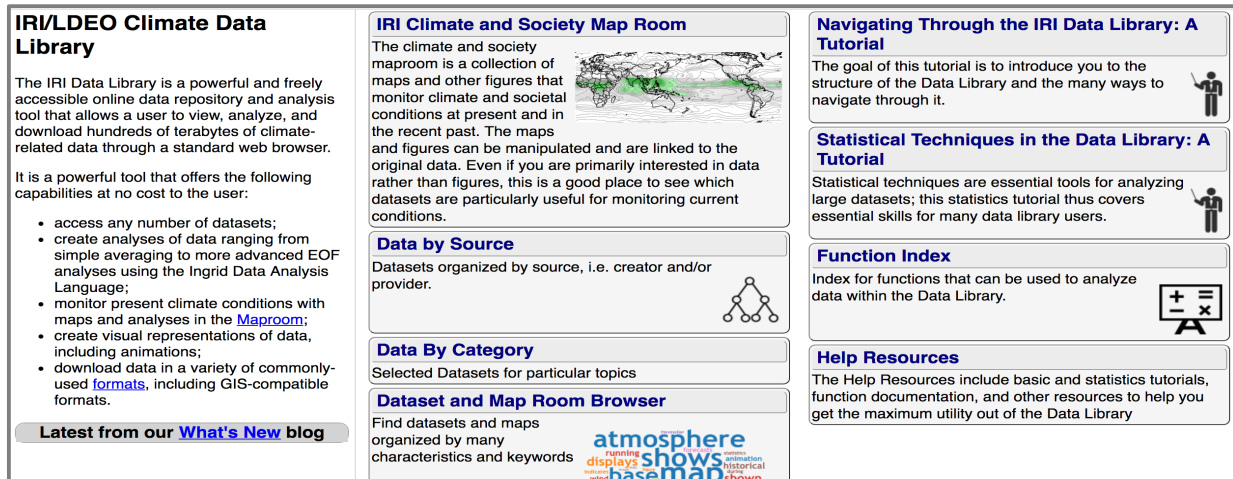


Figure 1: Landing page of the IRI/LDEO, a discipline-specific repository of Earth Climate Data.

Presently, the only repository of *planetary* atmospheric model output that exists in a similar form is the Mars Climate Database (MCD: <http://www-mars.lmd.jussieu.fr>). However, the MCD provides public access to output from *only* the Laboratoire de Météorologie Dynamique (LMD) present-day, global Mars atmospheric model. Figure 2a is a screenshot from the MCD website, showing Mars global model output plotted on demand by the MCD’s web interface. The screen also displays a link to download an ASCII file containing the model output used to make this plot. In addition to the web interface, users may also request the full 4-D dataset, which is then provided along with tools to access, manipulate, and visualize the model output.

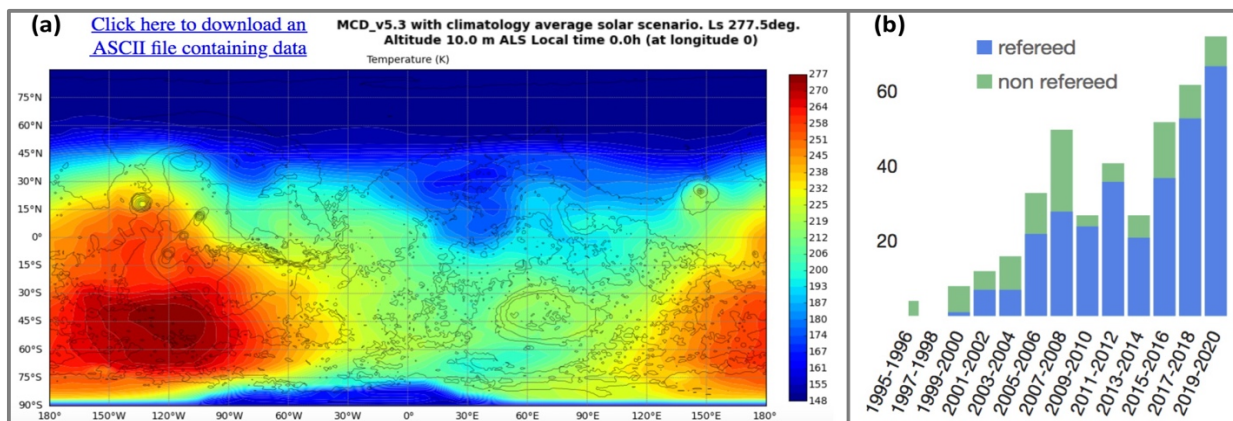


Figure 2: (a) Mars global model output, plotted on demand by the Mars Climate Database’s web interface; (b) number of publications referencing the Mars Climate Database since 1995.

Notably, the MCD has been centrally funded by the European and French Space Agencies (ESA and CNES) since it was created twenty-five years ago with the intended purpose of supporting those agencies’ Mars science and exploration efforts². However, the MCD is now

widely used by Mars scientists across a wide range of disciplines and countries.

According to the Astrophysics Data System (<https://ui.adsabs.harvard.edu>), since 1995 there have been over 300 refereed and 100 non-refereed publications that reference the MCD. These include studies of recurring slope lineae³, spacecraft EDL design⁴, and polar dunes⁵, in addition to the more expected range of direct studies of the atmosphere. As shown in Figure 2b, the number of refereed publications has increased over time, with nearly 70 in the 2019-2020 period alone. Note that this large number of references is for output from only *one global, low-resolution* model of *present-day* Mars, run for a limited set of conditions. Imagine the impact of a repository comparable to the MCD but disseminating output from the *dozens* of simulations produced by NASA-funded planetary atmospheric models, covering *multiple* planetary bodies.

3.2. A possible technical approach for a repository of planetary atmospheric model output.

Feature-rich analysis packages are increasingly available as open source products, especially within the Python, Ruby, and Javascript language and library systems. The goal of any viable output analysis system is, then, to find the lowest possible level-of-effort way of getting data into a pipeline that can use these resources, and of providing the simplest “front-end” to make these resources user-friendly. Here, we suggest how the proposed repository might be set up to do this.

Python is one of the most widely used programming languages, is very well-supported, and is free to all. It also has a very rich library of extension modules that allow it to accomplish a wide range of services, from file handling, to mathematical processing, to data visualization, to web page serving, and it is very easy to drive external libraries from within Python code.

Network Common Data Format (netCDF) was created as a self-defining data format, such that the data contain the metadata necessary to make it useful. For example, a 4D field such as air temperature (3 spatial plus time dimensions) can be labeled in metadata to describe where each data point is, and what the axes mean (e.g., one axis might be latitude in degrees). The file format is self-describing, such that a netCDF library can be called (by a wide range of languages, including Python) to deal with interfacing with the file. NetCDF is now a global standard data format used in the majority of Earth and planetary atmospheric models; for other models, it is straightforward to wrap or convert output into netCDF, with many conversion tools available.

An open-source library of *netCDF operators (NCO)* is also available that allows files to be searched and arbitrarily subset (e.g., specific time periods or variables to be extracted). This fast, well-supported library means that custom data extraction code is not needed.

Given all this, the repository could operate as follows: Python’s web file handling generates a table-of-contents for all available model output; for a given planetary body, a map could also be generated showing what output (models, resolutions, etc.) is available for any location, similar to the PDS Geosciences Node’s Orbital Data Explorers (<https://ode.rsl.wustl.edu>). When an output dataset is selected by a user, Python’s netCDF library is used to get variable names, dimensions, etc. The user then selects a subset of variables and time/spatial range for the desired output. These requirements are converted into NCO commands which are executed by the Python code to create a user-desired subset. Results are downloaded or visualized in-browser (e.g. Figure 2a).

With this or a similar approach, note that even model output held in another repository could be provided via the proposed repository, *provided* that it is in the required format and held in an OPeNDAP (which stands for “Open-source Project for a Network Data Access Protocol”) server.

3.3. Proposed establishment of an oversight committee.

An oversight committee made up of modelers, output users, developers, and NASA, would be vital to establish the philosophy and technical approach. This may include: whether the output should be held forever (enabling a Digital Object Identifier, DOI, to be assigned, as required by

most journals); what output may be accepted (e.g. from any simulation described in a peer-reviewed publication, or whether other metrics should be required); the metadata and format needed (e.g. netCDF files); and the additional documentation that would be needed (e.g. a README file describing caveats, such as ways in which the output is known to not match reality). The committee might also decide how users should acknowledge their use of output, or whether users should be encouraged to contact modeling groups for project-specific guidance.

3.4. Potential location of a repository of planetary atmospheric model output.

While the PDS does not currently accept planetary atmospheric model output, the PDS Atmospheres Node is aware of the strong need for the repository proposed here and already hosts a partial list of Mars atmospheric model output archived in other repositories: https://pds-atmospheres.nmsu.edu/data_and_services/atmospheres_data/MARS/external_sites.html

For the main PDS Atmospheres Node, no general tools are provided to facilitate access to the datasets. This is in large part due to the disparate nature of *observational* datasets, which range from 1-D entry profiles and met station time-series to cloud images or spectra. Even observations with time and spatial dimensions - such as global ‘mapping’ of air temperature - are not obtained as complete, gridded datasets. By contrast, as noted above, atmospheric *model* output has a rather standard structure, making it easier to provide general tools that can access different datasets.

For this reason, as well as the very large data volumes and potential differences in longevity and documentation requirements, the PDS Atmospheres Node itself is not the most suitable repository for atmospheric model output. **Instead, we suggest that the planned repository - including tools to access the model output - be established as a PDS Atmospheres Annex.**

4. Who benefits from the proposed repository of planetary atmospheric model output?

Below, we describe the benefits of the proposed repository to users, modelers, and NASA.

4.1. Benefits to users of model output.

The primary users of planetary atmospheric model output may be broken down into three groups: engineers, scientists, and EPO professionals. To demonstrate the range of applications possible, we provide here some examples of how these groups frequently utilize such output:

ENGINEERS:

1. *Entry, Descent and Landing (EDL)*: Density, temperature, aerosol, and wind profiles and uncertainties, for times and locations of landing, to design EDL systems and minimize risk.
2. *Surface mission planning*: Long-term surface meteorology to evaluate survivability for nominal mission duration; variability in radiative fluxes for solar-powered missions.
3. *Aerobraking planning and operations*: Upper atmosphere density and temperature profiles and uncertainties for all possible conditions (e.g., solar cycles) at time of aerobraking.
4. *Aerial vehicle planning and operations*: 3D winds to explore reachability of unpowered vehicles (or assess power requirements); 3D winds, density to assess safest times for flights.
5. *Mission operations teams*: Near-surface environmental conditions at landed location (esp. if local data are unavailable) to find best times to transfer samples, search for surface frost, etc.
6. *In Situ Resource Utilization (ISRU)*: Near-surface and subsurface environmental predictions at potential landing sites to assess ISRU required operating ranges and likely performance.
7. *Human mission risk assessment*: Aerosol abundances and properties to assess risks to humans and their equipment; wind to assess potential transport of biological material.

SCIENTISTS:

1. *Aeolian [and sedimentary] geologists*: Present-day [past climate] winds, wind stress, and boundary layer structure, to interpret characteristics, orientations, and migration directions of active aeolian features [inactive aeolian features or evidence found in the rock record].

2. *Geologists and hydrologists*: Climatological and hydrological cycles for different epochs, to help understand and interpret evidence of recent or ancient hydrological processes (from evidence of past liquid water on Mars to current methane clouds and rainfall on Titan).
3. *Scientists exploring planetary escape rates and past climates*: Upper atmosphere temperature and trace gas abundance for range of solar cycles, epochs, and events (e.g. Mars dust storms).
4. *Astrobiologists interested in habitability and extreme conditions*: Temperature ranges and probability of liquid brines in the surface/sub-surface (Mars), and atmospheric temperature range (e.g., Titan), to assess where life may exist / have existed on other planetary bodies.
5. *Chemists*: Winds to drive transport of gases in coupled chemistry-dynamical models.
6. *Seismologists/geophysicists*: Surface pressure and wind variation to assess seismic noise; surface temperature and gas/momentum exchange as upper boundary condition for interiors.
7. *Atmospheric scientists*: Theoretical or observational studies require model predictions for data retrieval schemes or to test hypotheses / interpret data. In addition, even modelers may require output from *other* models (e.g. global output to drive a limited-area model; output from a lower atmosphere model to drive the lower boundary of an upper atmosphere model).

EDUCATION AND PUBLIC OUTREACH:

Planetary atmospheric model-related EPO activities are currently very limited. By making such output easily available *and* accessible in the manner described in section 3, the proposed repository would open up a world of possibilities for NASA and others to develop their own tools to display model output. For example, Science on a Sphere displays in museums, etc. could show animations of Mars dust storms or convective events on giant planets: <https://sos.noaa.gov>

4.2. Benefits to modeling groups.

Planetary modelers would also benefit by having a way to better disseminate the fruits of their labor and also satisfy the requirements of NASA Data Management Plans and journal Data Policies, which increasingly demand that model output be made public too. It would remove the strain of facilitating data requests from those wishing to use model output. It would also remove the need to provide code and guidance on accessing the output once it has been provided, thus would be valuable even to modelers who have *already* made output publicly available, provided that the output dataset is in netCDF format and can be served remotely (see end of section 3.2).

4.3. Benefits to NASA.

Planetary atmospheric model output is the product of substantial model development and research, often with corresponding investment by NASA, but may rarely be used fully or after the motivating research is published. In publications, often a tiny subset of a full simulation is made publicly available (that necessary to reproduce the published results), meaning that the remaining output is not widely available except by special request. And what published datasets exist are typically hard to find and access. **Funding the development of a discipline-specific, community-recognized repository for this output will make the results of NASA-funded planetary atmospheric modeling studies more widely available and accessible, increase the efficiency with which such studies can be leveraged by other scientists, engineers, and EPO professionals, and reduce overall costs to NASA R&A programs.** Such a repository could, in addition, provide a template for future repositories of model output in a range of planetary fields, including planetary interiors, plasmas, and ocean worlds, which would be equally beneficial.

References: 1) Fathpour et al., *IEEE Sys. J.*, 8, 1112-1125, 2014. 2) Millour et al., *EPSC Abs.*, 10, EPSC2015-438, 2015. 3) Schmidt et al., *Nat. Geosci.*, 10, 270-273, 2017. 4) Liang & Mease, *J. Guid., Cont., Dyn.*, 42 (7), 1571-1578, 2019. 5) Hansen et al., *Icarus*, 251, 264-274, 2015.