ENVIRONMENTAL RESEARCH
CLIMATE

PURPOSE-LED PUBLISHING™

Part of the ENVIRONMENTAL RESEARCH SERIES

LETTER • OPEN ACCESS

Hosting downscaled decision-relevant community data products in ESGF2-US

To cite this article: Elias C Massoud *et al* 2026 *Environ. Res.: Climate* **5** 011002

View the article online for updates and enhancements.

You may also like

- Climate-driven emergence of fire danger in South America under global warming
Fernando P Forgioni, Miguel A Lovino, Gabriela V Müller et al.

- An observation-based causal framework linking Barents–Kara sea ice to European heatwaves
P Vaideanu, D R Nichita, L Boboc et al.

- Compound climate risks to European ports
A Fernandez-Perez, J Verschuur, J L Lara et al.

This content was downloaded from IP address 172.58.150.209 on 10/01/2026 at 03:10

IOP Publishing Environ. Res.: Climate 5 (2026) 011002 https://doi.org/10.1088/2752-5295/ae27ea

ENVIRONMENTAL RESEARCH
CLIMATE

Check for updates

OPEN ACCESS

RECEIVED
4 September 2025

REVISED
27 October 2025

ACCEPTED FOR PUBLICATION
4 December 2025

PUBLISHED
29 December 2025

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

LETTER

Hosting downscaled decision-relevant community data products in ESGF2-US

Elias C Massoud[1,*], Sasha K Ames[2], Nathan Collier[1], Paul J Durack[2], Daniel Feldman[3], Peter J Gleckler[2], Forrest M Hoffman[1], Jitendra Kumar[1], Fred Lipschultz[4], Karl E Taylor[2] and Paul Ullrich[2]

[1] Oak Ridge National Laboratory, Oak Ridge, TN, United States of America
[2] Lawrence Livermore National Laboratory, Livermore, CA, United States of America
[3] Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America
[4] Independent Researcher, Berkeley, CA, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: massoudec@ornl.gov

Keywords: downscaled data, ESGF, metadata standards

Abstract

As regionally-relevant high-resolution Earth system data is increasingly relied upon across scientific, policy, and practitioner communities, there is an urgent need for coordinated and federated infrastructure to store, manage, standardize, and distribute decision-relevant community data products. Substantial effort is required to ensure that these products, which are often critical for regional impact assessments and decision-making, are findable, accessible, interoperable, and reusable. The Earth System Grid Federation US project (ESGF2-US) is addressing this challenge by expanding its open-source, distributed platform to support the hosting and dissemination of downscaled Earth system datasets. This expansion includes aligning new downscaled datasets with developing community standards for metadata and file structure, consistent with existing ESGF archives. This includes ensuring CF-compliance, applying CMORization where appropriate, and developing tools to streamline user access. In this paper, we highlight the technical and coordination work required to bring downscaled data into ESGF2-US and aim to inform the broader Earth system data user community about the growing availability and utility of these curated resources.

1. Background

As demand for high-resolution Earth system information continues to grow, a wide array of downscaled data products, recently coined decision-relevant community data products (DRCDPs), have emerged to meet the needs of scientists, planners, and policy-makers (Wootten *et al* 2020, 2023, U.S. Department of Energy 2024, Ullrich *et al* 2025). These products translate coarse-resolution global model or reanalysis output into finer spatial scales and tailored formats more suitable for local or regional applications. DRCDPs typically include both statistical downscaling (Gutmann *et al* 2022, Thrasher *et al* 2024), which applies empirical or machine-learning techniques to adjust historical and projected data based on observed relationships, and dynamical downscaling (Rahimi *et al* 2024, Spero *et al* 2025), which nests regional models within global simulations to explicitly resolve smaller-scale processes. These downscaled products are critical because they provide high-resolution data that captures the full range of uncertainty in climate projections, thereby enabling more accurate localized impact assessments, guiding resilient infrastructure planning, and supporting informed decision-making across sectors sensitive to environmental change (Wootten *et al* 2025). However, despite their growing importance, DRCDPs are often delivered in inconsistent formats, scattered across disparate servers, and burdened by incomplete or non-standard metadata. This fragmentation presents a serious barrier to reproducibility, synthesis, and integration with broader Earth system data workflows.

© 2025 The Author(s). Published by IOP Publishing Ltd

Efforts to address these challenges are increasingly grounded in community standards that promote findability, accessibility, interoperability, and reusability (FAIR principles, Wilkinson *et al* 2016). Among these, the climate and forecast (CF) metadata conventions (Eaton *et al* 2006) provide a generalized framework for describing CF model data, defining standard representations for coordinates, time, grids, and physical variables. Building on CF, community-specific conventions, particularly those developed through successive phases of the Coupled Model Intercomparison Project (CMIP), define stricter requirements for variable naming, directory structure, global attributes, and controlled vocabularies (Durack *et al* 2025). The Climate Model Output Rewriter (CMOR) library (Mauzey *et al* 2025) supports this by formatting Earth system data to align with these CMIP conventions, enabling consistency and discoverability within Earth System Grid Federation (ESGF). These layered standards are critical for enabling seamless integration of new datasets into ESGF infrastructure (Massoud *et al* 2024), and recent efforts (e.g. Taylor *et al* 2018, Gleckler *et al* 2024, Nikulin *et al* 2024) have codified these conventions via Zenodo-hosted documentation to support reproducibility and community-wide adoption.

The ESGF (Bernholdt *et al* 2007, Williams *et al* 2009, Kershaw *et al* 2020) has long served as the central infrastructure for distributing Earth system data, supporting international efforts such as CMIP (Durack *et al* 2025) and observational data collections including obs4MIPs (Waliser *et al* 2020) and CREATE-IP (Potter *et al* 2018). Building on this foundation, the ESGF–United States (ESGF2-US) project aims to extend ESGF's capabilities to support DRCDPs with the same level of rigor and accessibility afforded to global Earth system simulations, such as CMIP. This includes ensuring that downscaled datasets conform to CF and are coordinated with CMIP standards, integrating them into ESGF's established CMIP-style directory structure, and exposing them through familiar tools such as the MetaGrid web interface and the intake-esgf Python catalog (discussed below). These coordinated efforts enable a consistent, reliable, and scalable platform for accessing downscaled Earth system data products alongside established global model and observational datasets.

Well-established protocols exist for projects delivering data via ESGF, including CMIP6 (Taylor *et al* 2018), obs4MIPs (Gleckler *et al* 2024), and the Coordinated Regional Climate Downscaling Experiment (CORDEX; Nikulin *et al* 2024). There are clear advantages for striving to keep new projects aligned with existing ESGF-projects as much as possible. The goals of the present effort have many similarities with CORDEX, but include metadata refinements (e.g. version control of downscaling methods) and coordination with the community's preparation for CMIP7. Realizing this vision has required coordinated technical development and community collaboration. ESGF node operators, data curators, software developers, and DRCDP producers have worked together to define new CMOR tables specific to downscaled products, adapt directory hierarchies and metadata conventions, and validate many terabytes of output before publication. These efforts represent a significant investment not just in infrastructure, but in standards adoption, tooling, and quality control.

This paper documents that work. Our aim is to inform a broad swath of the Earth system community, including scientists, data providers, evaluators, and decision-makers, about the new capabilities for hosting and distributing downscaled data products (DRCDPs) within ESGF2-US, the standards and tools that enable their integration, and the scope of effort required to make them findable, usable and interoperable. In doing so, we hope to lower the barrier to future contributions, foster feedback, and expand access to high-quality downscaled Earth system data.

2. File structure and metadata standards

The utility and longevity of Earth system data depend not only on its scientific quality but also on how it is formatted, organized, described and accessible. ESGF2-US builds upon established community practices to ensure that all hosted datasets, particularly downscaled regional products, are consistent, discoverable, and interoperable with a wide array of tools. This section describes the key structural and metadata standards used in ESGF2-US: directory and file naming conventions, compliance with the CF metadata conventions, and formatting using the CMOR tool.

2.1. Directory and file naming conventions

To ensure consistency, traceability, and ease of use, ESGF2-US adopts a standardized directory and file naming system based on community conventions from CMIP6 (Taylor *et al* 2018) with modifications to support the unique requirements of DRCDPs. These conventions facilitate automated data discovery, support reproducibility, and enable scalable workflows across tools and institutions (CESSDA 2021, Harvard Biomedical DMP 2023).

The directory structure encodes a rich set of metadata attributes about each dataset, organized hierarchically to support systematic cataloging and access (Taylor *et al* 2018). As with other Model

Table 1. Hierarchical structure of ESGF2-US directory paths. Each row lists a metadata field used in dataset organization, along with a representative example and a brief description. These standardized directory components enable systematic cataloging, discovery, and access.

Field	Example	Description
Activity_id	DRCDP	High-level project or activity name (e.g. DRCDP, CMIP6)
Region_id	NAM	Geographic region (e.g. NAM = North America)
Institution_id	UCSD-SIO	Institution producing the downscaled dataset
Source_id	LOCA2-1	Downscaling method or technique, including version identifier
Driving_mip_era	CMIP6	CMIP generation used as the basis for the driving model (e.g. CMIP6, CMIP5)
Driving_activity_id	CMIP	Intercomparison project of the driving model (e.g. CMIP, ScenarioMIP)
Driving_experiment_id	Historical	Driving model experiment (e.g. historical, ssp245)
Driving_source_id	ACCESS-CM2	Name of the global model driving the downscaling
Driving_variant_label	r1i1p1f1	Realization/ensemble member of the driving model
Frequency	Day	Temporal resolution of the output (e.g. day, mon)
Variable_id	Tasmax	Variable short name (e.g. pr, tasmax)
Version	v20250216	Dataset version with release timestamp (YYYYMMDD)

Intercomparison Projects (MIPs), this structure follows a data reference syntax (DRS), where each element of the directory path corresponds to a defined metadata field (Durack *et al* 2025). For DRCDPs, ESGF2-US has adopted a slightly modified path ordering that prioritizes region- and product-specific metadata earlier in the path, reflecting the needs of regional users and facilitating better discovery.

A generalized template and an example directory are shown below (and explained in table 1):

Directory path template:
<activity_id>/<region_id>/<institution_id>/<source_id>/<driving_mip_era>/<driving_activity_id> /<driving_experiment_id>/<driving_source_id>/<driving_variant_label>/<frequency>/<variable_id> /<version>/

Directory path example:
DRCDP/NAM/UCSD-SIO/LOCA2-1/CMIP6/CMIP/historical/ACCESS-CM2/r1i1p1f1/day/tasmax/v20250216/

This revised structure retains compatibility with existing ESGF conventions while improving clarity and grouping for regionally specific downscaled datasets. While the current region_id structure in ESGF2-US uses simplified geographic identifiers (e.g. 'NAM' for North America), we recognize the value of aligning with established regional frameworks such as CORDEX to enhance consistency and interoperability. The ESGF2 team is actively evaluating opportunities to better integrate CORDEX regional definitions in future releases, while maintaining backward compatibility with existing datasets and workflows. The placement of the region_id, institution_id, and source_id earlier in the path ensures better support for region-centric workflows and cataloging needs, particularly for datasets not associated with a single driving model or MIP activity. More detailed definitions of each path element are provided in table 1 and the project's GitHub repository: https://github.com/PCMDI/DRCDP.

The DRCDP file name format mirrors many of the directory path elements, using a standardized, flat string that encodes key metadata attributes. While not hierarchical like the directory structure, the filename remains both human-readable and machine-parsable, enabling consistent validation, provenance tracking, and streamlined reprocessing workflows. This naming format is adapted from CMIP6 conventions (Taylor *et al* 2018) but reordered to better suit regionally downscaled data products. Notably, the filename prioritizes region, institution, and source information earlier in the string, reflecting the same DRCDP-centric logic used in the updated directory structure.

A generalized template and an example file name are shown below (and explained in table 2):

Filename template:
<variable_id>_<region_id>_<institution_id>_<source_id>_<driving_mip_era>_ <driving_experiment_id>_<driving_source_id>_<driving_variant_label>_<frequency>_<StartTime>-<EndTime>.nc

Filename example:

pr_NAM_UCSD-SIO_LOCA2-1_CMIP6_historical_ACCESS-CM2_r1i1p1f1_day_19500102-19500131.nc

More detailed definitions of each filename element are provided in table 2.

Table 2. Structure of ESGF2-US dataset filenames. Each row details a metadata field encoded in the filename, along with an example and description. These standardized filename elements support human readability, provenance tracking, and automated processing.

Field	Example	Description
Variable_id	Pr	Variable short name (e.g. pr, tasmax)
Region_id	NAM	Geographic region (e.g. NAM = North America)
Institution_id	UCSD-SIO	Institution producing the downscaled dataset
Source_id	LOCA2-1	Downscaling method or technique, including version identifier
Driving_mip_era	CMIP6	CMIP generation used as the basis for the driving model (e.g. CMIP6, CMIP5)
Driving_experiment_id	Historical	Driving model experiment (e.g. historical, ssp245)
Driving_source_id	ACCESS-CM2	Name of the global model driving the downscaling
Driving_variant_label	r1i1p1f1	Realization/ensemble member of the driving model
Frequency	Day	Temporal resolution of the output (e.g. day, mon)

2.2. CF compliance

The CF metadata conventions (Eaton *et al* 2006) are a community-developed standard for describing Earth system datasets in NetCDF format. CF-compliant files include rich metadata that describes each variable's meaning, units, spatial and temporal coordinates, and encoding, all in a machine-readable, self-describing, commonly-used format. By adhering to the CF conventions, ESGF2-US ensures that its datasets are compatible with popular analysis tools such as xarray or the netCDF Operators. This enhances usability across diverse scientific workflows and minimizes the need for manual interpretation or preprocessing. Tools such as the cf-checker, xclim, and xarray's built-in CF conventions support validation and interpretation of CF metadata, enabling both data providers and users to maintain data integrity.

While the CF standards are commonly used to provide the foundation for data description, it is important to note that how these conventions are applied can be application dependent. For instance, the directory and filenames templates defined for the DRCDP project are not identical to those used in CMIP6 and earlier versions of CMIP. One reason for this is that there is not always a one-to-one mapping of the data that needs to be carefully described across projects. However, despite application dependent data description, efforts to coordinate metadata across projects facilitates their overall coordination. CF compliance and project specific application of the conventions improves searchability within ESGF's internationally federated system, as metadata fields are exposed to indexing services and APIs. In this way, the CF conventions serve as a backbone for data interoperability and transparency across Earth system model and observational datasets.

2.3. CMORization

To assist in the creation of standardized, ESGF-ready NetCDF files, ESGF2-US leverages the CMOR tool (Mauzey *et al* 2025). CMOR is a C-based software library that automates the transformation of raw model or downscaled output into a format that conforms with CF and CMIP or other project conventions. CMOR enforces the use of controlled vocabularies and CMOR tables for variable names, units, axis definitions, and metadata fields. It also ensures that datasets pass strict validation checks before publication to the ESGF system. This process, commonly referred to as 'CMORization', is critical to maintaining consistency across a federated, multi-model archive. Recognizing the unique requirements of downscaled datasets, this effort has extended CMOR to include new tables tailored to DRCDPs (c.f. https://github.com/PCMDI/DRCDP). These tables incorporate metadata fields specific to statistical and dynamical downscaling methods, regional grids, and high-resolution formats.

CMORization not only improves dataset integrity and usability but also enables automated ingestion into ESGF search and access services. Figure 1 illustrates the CMORization workflow demonstrated in a Jupyter notebook, highlighting the key steps from loading configuration and input data to executing the CMORization script and confirming the standardized output. The resulting CMORized NetCDF file is fully CF-compliant and adheres to the DRCDP-specific directory and filename conventions described above, ensuring it is ready for publication within the ESGF federation. Several demos, including this one, are available on the DRCDP GitHub repository to assist users in preparing their own DRCDP-compliant datasets.

In addition to producing CMORized NetCDF files, ESGF2-US is actively exploring approaches to improve accessibility and performance for large-scale and cloud-based workflows. This includes evaluating chunking strategies during NetCDF generation and investigating the use of Kerchunk to provide Zarr-like access to existing NetCDF files without data duplication. For ESGF, which hosts petabytes of existing NetCDF data, Kerchunk provides a way to offer Zarr-like cloud-optimized access without

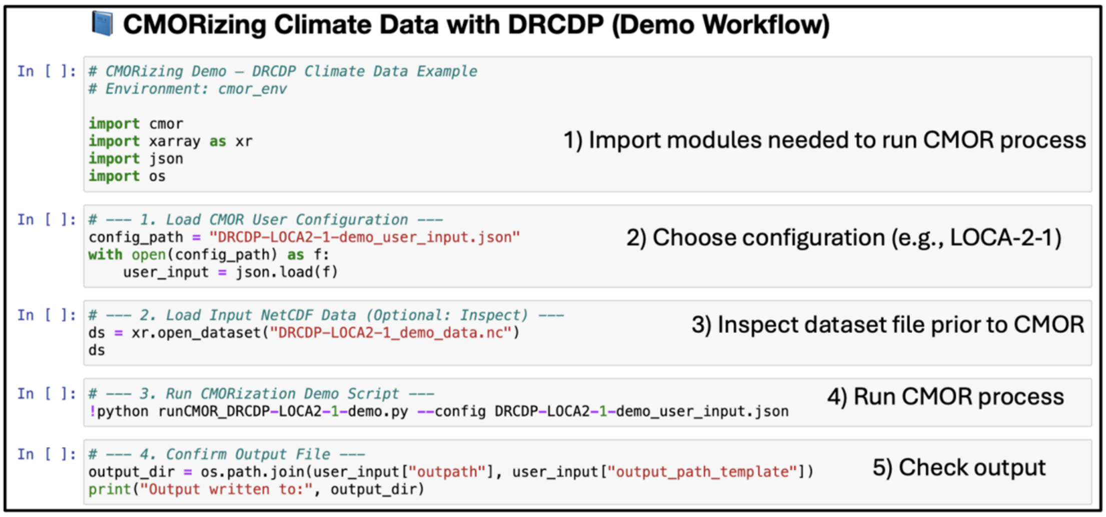


Figure 1. Workflow for CMORizing DRCDP data files in a Jupyter notebook. The process begins by loading the user configuration JSON (in this case the LOCA2-1 configuration), followed by opening the input NetCDF dataset using xarray. Next, the CMORization script is executed to standardize the data to DRCDP/ESGF specifications. Finally, the output directory is confirmed. Other examples like this that can be used for processing alternative data products can be found at (https://github.com/PCMDI/DRCDP/tree/main/DataPreparationExamples/DEMO).

re-encoding or storing multiple copies of the data, making it attractive for scalable analysis while preserving existing file archives. Kerchunk generates lightweight reference descriptions that map the internal structure of NetCDF or HDF5 files to logical Zarr-compatible datasets, enabling efficient, parallel, and cloud-native access through tools such as xarray and Dask. Similar approaches have been adopted within the CMIP6 and ESGF ecosystem, through initiatives such as the ESGF Virtual Aggregation Service (Cimadevilla *et al* 2025), which use Kerchunk-style reference files to expose existing CMIP archives as cloud-optimized, Zarr-accessible datasets. By aligning with these emerging community standards, ESGF2-US ensures interoperability with CMIP workflows while supporting scalable, high-performance analytics and machine-learning applications.

3. Data hosting on ESGF2-US

The ESGF2-US project has published the first DRCDPs within the ESGF infrastructure, a key milestone in making downscaled, decision-relevant data broadly accessible alongside global model and observational datasets. Two flagship DRCDPs, LOCA2-1 as well as LOCA2-0 (Localized Constructed Analogs, version 2) (Pierce *et al* 2023) and STAR-ESDM-v1 (Seasonal Trends and Analysis of Residuals empirical statistical downscaling model) (Hayhoe *et al* 2024), are now fully hosted as CF-compliant NetCDF files and integrated into ESGF's federated catalog. These products conform to ESGF conventions for metadata, directory structure, and file naming, with only minor customization of the DRS during ingestion. This achievement demonstrates the feasibility of incorporating DRCDPs into ESGF at scale and paves the way for additional downscaled products in the near future. For these initial datasets, the ESGF publisher software (esg-publisher) was extended with a user-defined project configuration to encode the DRCDP-specific DRS, without requiring any custom code.

3.1. Data access workflow: MetaGrid

A central advantage of hosting DRCDPs on ESGF2-US is that they are immediately discoverable and accessible through established ESGF tools, including the user-friendly MetaGrid web portal (https://esgf-node.ornl.gov/search/DRCDP). MetaGrid provides a graphical interface for searching, filtering, and downloading datasets hosted across the ESGF federation (figure 2). Users can initiate a search by selecting the DRCDP project from the 'Project' filter and refine their queries using metadata fields such as realization (e.g. r1i1p1f1), driving source ID (e.g. ACCESS-CM2), variable ID (e.g. pr), and experiment ID (e.g. historical), among others. These search options correspond directly to the directory and file naming metadata fields described in tables 1 and 2. This interface lowers the barrier to entry for decision-makers and scientists who may lack command-line expertise, while still supporting advanced capabilities like REST API queries and direct wget script generation. As an example, figure 2 shows the

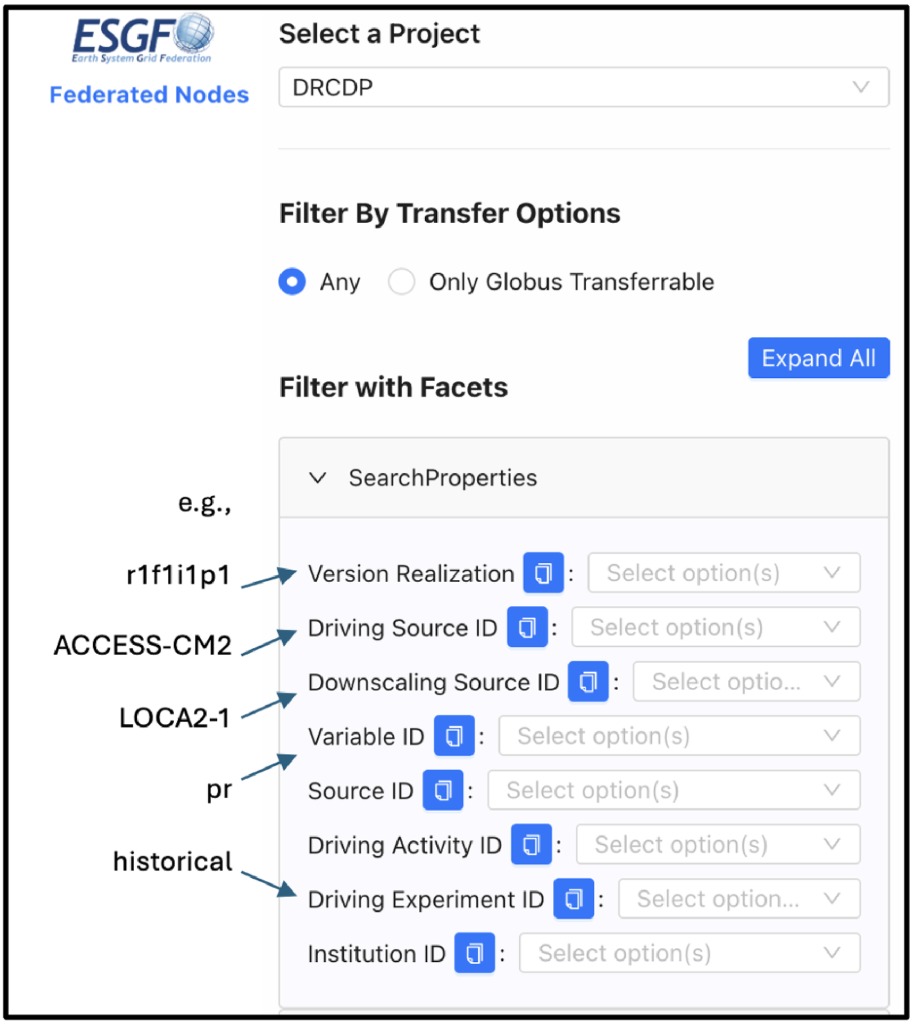


Figure 2. Left panel of the ESGF MetaGrid web interface, showing filtering options with the DRCDP project selected. Users can refine their search using standardized metadata fields such as realization (e.g. r1i1p1f1), driving source ID (e.g. ACCESS-CM2), variable ID (e.g. pr), and experiment ID (e.g. historical). These metadata elements align with the directory and filename structures defined in tables 1 and 2, enabling precise discovery and selection of DRCDP datasets within the ESGF federation. Once the search is narrowed, a list of available data files is presented, allowing users to view key metadata and select specific files for download based on their needs. DRCDPs can be accessed via MetaGrid at: https://esgf-node.ornl.gov/search/DRCDP.

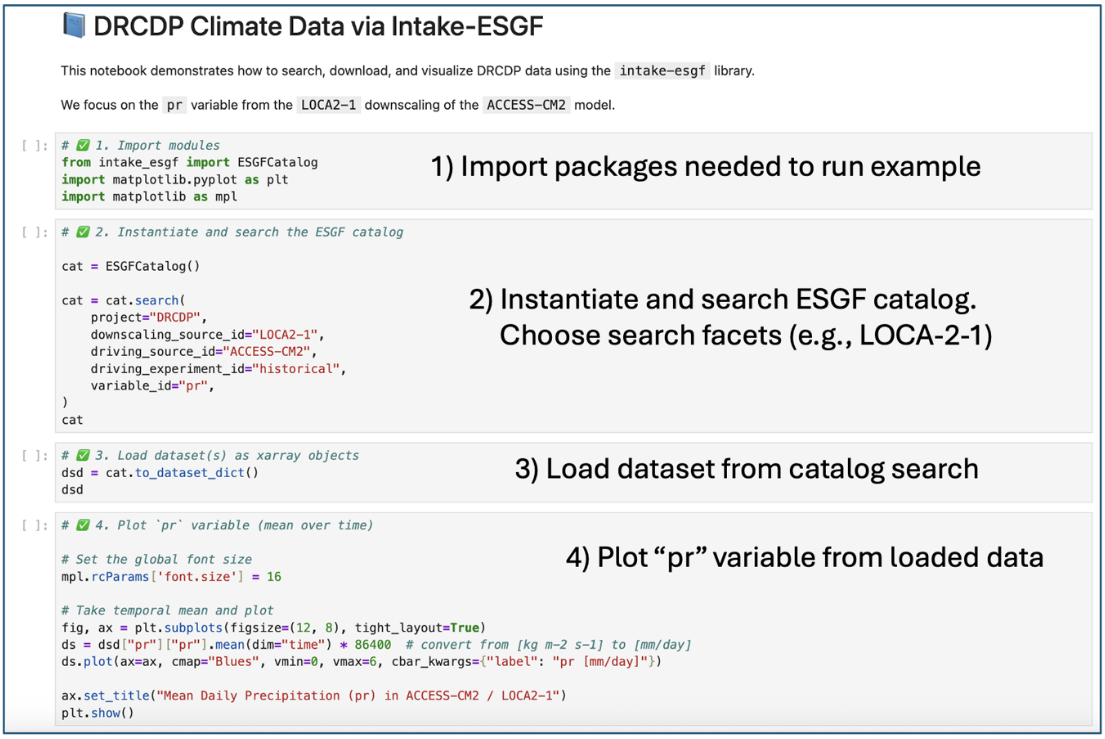
left panel of a typical MetaGrid search session with the DRCDP project selected, highlighting the available filtering options for querying specific datasets.

3.2. Tools for data discovery and analysis: intake-esgf

As a complement to the MetaGrid web portal, programmatic, scalable access to ESGF-hosted datasets is available via the intake-esgf package, a powerful Python-based interface. This tool enables users to query ESGF catalogs, inspect metadata, and stream datasets directly into memory using familiar data structures such as xarray and dask. By supporting scriptable workflows, intake-esgf facilitates reproducible research and enables power users to automate discovery, analysis, and visualization of large, high-resolution datasets.

Figure 3 shows how just a few lines of code in a Jupyter notebook environment can search for DRCDP datasets, load them, and generate a quick visualization, lowering barriers for both new and advanced users. The example highlights the LOCA2-1 downscaled ACCESS-CM2 daily precipitation simulation, while figure 4 shows the resulting high-resolution map of long-term mean precipitation (pr (mm d^{-1})) for the selected region and period. This demonstrates how intake-esgf unlocks rich scientific insights from terabytes of ESGF-hosted output with minimal setup.

To support users at all skill levels, ESGF2-US maintains extensive documentation and example notebooks (https://intake-esgf.readthedocs.io/en/latest/) demonstrating common tasks such as dataset discovery, download, and analysis. Additional resources and community support are available on GitHub (https://github.com/esgf2-us/intake-esgf).



📘 DRCDP Climate Data via Intake-ESGF

This notebook demonstrates how to search, download, and visualize DRCDP data using the `intake-esgf` library.

We focus on the `pr` variable from the `LOCA2-1` downscaling of the `ACCESS-CM2` model.

```
[ ]: # ✅ 1. Import modules
from intake_esgf import ESGFCatalog
import matplotlib.pyplot as plt
import matplotlib as mpl
```

1) Import packages needed to run example

```
[ ]: # ✅ 2. Instantiate and search the ESGF catalog

cat = ESGFCatalog()

cat = cat.search(
    project="DRCDP",
    downscaling_source_id="LOCA2-1",
    driving_source_id="ACCESS-CM2",
    driving_experiment_id="historical",
    variable_id="pr",
)
cat
```

2) Instantiate and search ESGF catalog. Choose search facets (e.g., LOCA-2-1)

```
[ ]: # ✅ 3. Load dataset(s) as xarray objects
dsd = cat.to_dataset_dict()
dsd
```

3) Load dataset from catalog search

```
[ ]: # ✅ 4. Plot `pr` variable (mean over time)

# Set the global font size
mpl.rcParams['font.size'] = 16

# Take temporal mean and plot
fig, ax = plt.subplots(figsize=(12, 8), tight_layout=True)
ds = dsd["pr"]["pr"].mean(dim="time") * 86400  # convert from [kg m-2 s-1] to [mm/day]
ds.plot(ax=ax, cmap="Blues", vmin=0, vmax=6, cbar_kwargs={"label": "pr [mm/day]"})

ax.set_title("Mean Daily Precipitation (pr) in ACCESS-CM2 / LOCA2-1")
plt.show()
```

4) Plot "pr" variable from loaded data

Figure 3. A Jupyter notebook showing the process of using intake-esgf to search, load, and visualize downscaled Earth system data residing in ESGF. With just a few lines of code, a user can search ESGF for available DRCDP data, load the file to their own machine, and create a quick plot to visualize the data. The produced map is shown in figure 4. The intake-esgf library is available on GitHub at https://github.com/esgf2-us/intake-esgf, with documentation available at https://intake-esgf.readthedocs.io/en/latest/. The version of intake-esgf used in this work corresponds to release v2025.10.22 (commit 29dcc56d258f55e2c795f7bd5f90ee2680249263) on the *main* branch.

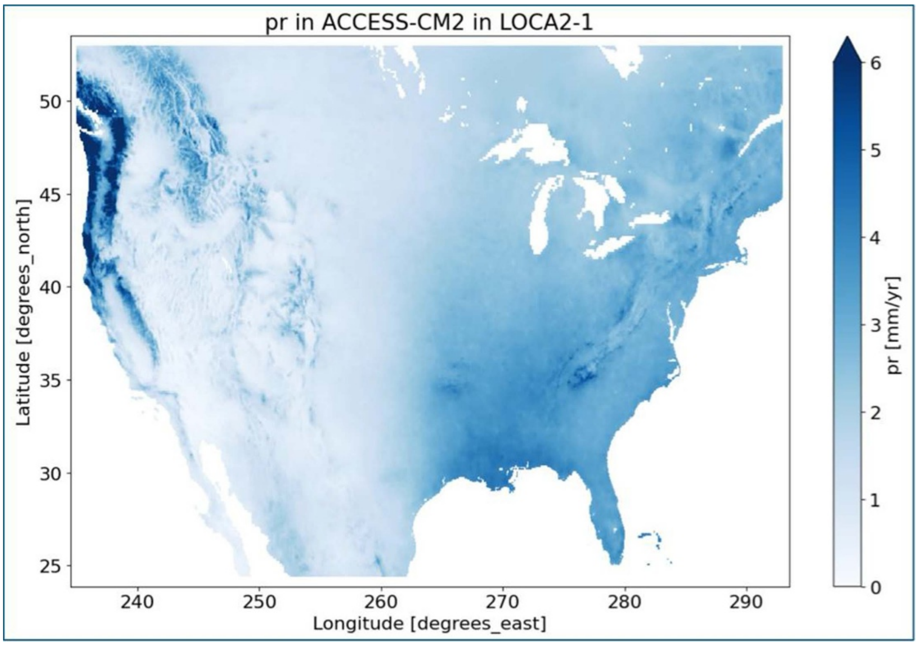


Figure 4. Map showing the LOCA2-1 version of the downscaled ACCESS-CM2 model simulation of long term mean precipitation (pr (mm d^{-1})). This is the plot that is produced using the Jupyter notebook code from figure 3.

4. Discussion

4.1. ESGF2-US as a centralized resource

While many products already have dedicated access points, ESGF2-US provides added value via a centralized, standardized, and federated platform that integrates the diverse data assets into a common ecosystem. This integration improves discoverability, ensures metadata consistency, enhances interoperability across tools and workflows, and ultimately accelerates model evaluation and research. These features are especially important for enabling cross-model and cross-dataset analysis, directly benefiting tools like the Program for Climate Model Diagnosis and Intercomparison (PCMDI) Metrics Package (PMP) (Lee *et al* 2024) and the International Land Model Benchmarking system (ILAMB) (Collier *et al* 2018), which support national and international assessments, or the upcoming Rapid Evaluation Framework which is designed to provide immediate quantitative feedback that allows model developers and scientists to quickly identify model biases and performance issues (Hoffman *et al* 2025).

Standardization is key to ESGF2-US's role as a centralized resource. By requiring adherence to CF conventions and project-specific data specifications (both delivered through CMOR formatting), the platform ensures that DRCDPs are consistently structured, semantically rich, and interoperable across tools and domains. These standards, along with clear directory hierarchies and file naming conventions, make datasets easier to discover, track, and integrate into workflows such as those used by PMP and ILAMB. The successful integration of LOCA2-1, LOCA2-0, and STAR-ESDM-v1 into ESGF2-US exemplifies how these standards enable efficient publication and enhance the platform's value as an interoperable, centralized resource.

Because ESGF operates as a federated system, datasets hosted at ESGF2-US can potentially be discovered and accessed through the broader ESGF infrastructure. This US-centric effort does not preclude global engagement; in fact, if international partners express interest, DRCDPs hosted on ESGF2-US could be replicated and served at other federation nodes, enabling broader access and redundancy across regions, and would likely involve coordination with other communities, such as CORDEX.

4.2. Challenges, community progress, and incentives for standardizing data

Despite the clear benefits of standardized data practices, significant barriers remain, chief among them the time and technical investment required to conform to evolving conventions. For many data producers, aligning with ESGF standards can be a non-trivial effort. To address these challenges, efforts have been made to improve metadata interoperability and streamline workflows for downscaled products. These improvements have supported the successful standardization and publication of datasets such as LOCA2-1, LOCA2-0, and STAR-ESDM-v1, demonstrating what is possible when technical guidance, shared tooling, and clear conventions are available.

Demonstrating the ability to CMORize a data file (figure 1) and to access datasets programmatically using intake-esgf (figure 2) is intended to give the community confidence that preparing DRCDP data for ESGF ingestion, and later discovering and accessing it, is a manageable and well-supported process. Continued community efforts to develop shared tools, improve guidance, and automate key processes are steadily lowering the barriers, making it increasingly feasible for both data producers and users to engage with standardized DRCDP data through a more accessible and interoperable ESGF ecosystem.

4.3. Future data ingestion priorities

Building on the workflows and standards established for LOCA2-1, LOCA2-0, and STAR-ESDM, ESGF2-US anticipates steadily broadening its archive of high-resolution DRCDPs. In addition to the statistically downscaled LOCA2 and STAR-ESDM data already available, forthcoming candidates include Ensemble Generalized Analog Regression Downscaling (Gutmann *et al* 2022) and the NASA Earth-Exchange suites NEX-GDDP and NEX-DCP30 (Thrasher *et al* 2012, 2024). Dynamically downscaled products under consideration include the EPA dynamically downscaled ensemble, version 2 (EDDE v2) (Spero *et al* 2025), the Climate Risk and Resilience Portal dataset (ANL 2023), the Western United States dynamically downscaled dataset (Rahimi *et al* 2024), and the North American CORDEX (Bukovsky and Mearns 2020). Most of these datasets are profiled in the forthcoming BAMS article by Ullrich *et al* (2025), which lays out a coordinated national strategy for DRCDPs. Looking ahead, ESGF2-US also expects to evaluate and ingest emerging products generated with artificial intelligence and machine learning (AI/ML) techniques as the downscaling landscape continues to evolve. These future priorities will further test and refine the workflows, standards, and collaborative practices that have already proven effective for the initial datasets.

As ESGF2-US expands to ingest new DRCDPs, attention will also be given to supporting variables that currently lack formal CMOR standards, such as convective available potential energy which serves as a proxy for severe weather potential. Addressing these variables will involve community collaboration to develop provisional naming conventions and metadata schemas, ensuring that future datasets with novel or specialized variables can be integrated without sacrificing discoverability or interoperability.

4.4. Community engagement and responsiveness

Central to ESGF2-US is a commitment to community engagement and iterative refinement based on user feedback. The project will actively gather input through a variety of channels, including periodic user surveys, public issue tracking and feature requests on GitHub, and dedicated workshops or webinars focused on user experience and technical needs. By systematically analyzing this feedback, the ESGF2-US team can identify common challenges, prioritize enhancements, and adapt the platform to better serve diverse stakeholder groups. This responsive approach not only fosters transparency, trust, and inclusivity, but also lowers barriers to standardization and contribution, ensuring that ESGF2-US evolves in step with the changing priorities and workflows of the Earth system data user community. Together, these efforts position ESGF2-US as an evolving, community-driven platform capable of supporting the next generation of DRCDPs.

5. Summary and conclusions

This paper has documented the successful integration of DRCDPs into the ESGF2-US framework, marking a key milestone in making high-resolution, decision-relevant information broadly accessible, standardized, and interoperable. Currently, flagship statistically downscaled datasets such as LOCA2-1, LOCA2-0, and STAR-ESDM are hosted and discoverable alongside global model and observational data, with plans underway to incorporate additional statistical, dynamical, and emerging AI/ML-based products in future iterations. By adopting community-driven conventions like CF metadata standards and CMORization, ESGF2-US provides a foundation for interoperability, enabling seamless integration across models, tools, and users. User access is further enhanced through the complementary platforms MetaGrid and intake-esgf, which support diverse workflows ranging from exploratory analysis to scalable, reproducible pipelines.

While challenges remain, including cyberinfrastructure demands, community awareness, and adoption barriers, targeted strategies centered on standardization, outreach, and responsive feedback are already addressing these issues. Looking ahead, ESGF2-US aims not only to expand its data holdings but also to serve as a catalyst for broader adoption and innovative applications of DRCDPs. By fostering a more connected, transparent, and user-driven ecosystem, ESGF2-US is helping to ensure that high-quality, decision-relevant data are readily available to inform science, policy, and societal resilience in the face of environmental change.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://esgf-node.ornl.gov/search/DRCDP (Earth System Grid Federation (ESGF) 2025).

Acknowledgments

This research was supported by the Earth System Grid Federation 2—US (ESGF2-US) project, which is sponsored by the Data Management Program of the Earth and Environmental Systems Sciences Division (EESSD) of the Office of Biological and Environmental Research (BER) in the US Department of Energy (DOE) Office of Science. The work of S K A, P J D, P J G, K E T, and P U at PCMDI is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. IM Release No. LLNL-JRNL-2010566. The material from the Lawrence Berkeley and Lawrence Livermore National Laboratories was funded by a contract from the Environmental Security Technology Certification Program (ESTCP) under Project NH24-8916. This manuscript has been authored at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US DOE. The US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or to allow others to do so, for US government purposes. By accepting this article for publication, the publisher acknowledges this license. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Author contributions

Elias C Massoud 0000-0002-1772-5361
Conceptualization (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Software (equal), Supervision (equal), Visualization (equal), Writing – original draft (equal), Writing – review & editing (equal)

Sasha K Ames 0000-0001-9981-2997
Conceptualization (equal), Formal analysis (equal), Investigation (equal), Resources (equal), Writing – review & editing (equal)

Nathan Collier
Conceptualization (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Software (equal), Validation (equal), Writing – review & editing (equal)

Paul J Durack 0000-0003-2835-1438
Conceptualization (equal), Formal analysis (equal), Investigation (equal), Software (equal), Validation (equal), Writing – review & editing (equal)

Daniel Feldman
Conceptualization (equal), Formal analysis (equal), Writing – review & editing (equal)

Peter J Gleckler
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (lead), Resources (equal), Software (equal), Validation (equal), Writing – review & editing (equal)

Forrest M Hoffman
Conceptualization (equal), Funding acquisition (equal), Supervision (equal), Writing – review & editing (equal)

Jitendra Kumar
Conceptualization (equal), Formal analysis (equal), Writing – review & editing (equal)

Fred Lipschultz 0000-0003-1634-6754
Conceptualization (equal), Formal analysis (equal), Supervision (equal), Writing – review & editing (equal)

Karl E Taylor 0000-0002-6491-2135
Conceptualization (equal), Methodology (equal), Writing – review & editing (equal)

Paul Ullrich
Conceptualization (equal), Formal analysis (equal), Investigation (equal), Supervision (equal), Validation (equal), Writing – review & editing (equal)

References

Argonne National Laboratory 2023 Climate risk and resilience portal (ClimRR) (available at: https://disgeoportal.egs.anl.gov/ClimRR/)
Bernholdt D *et al* 2007 The Earth system grid: supporting the next generation of climate modeling research (arXiv: 0712.2262)
Bukovsky M S and Mearns L O 2020 Regional climate change projections from NA-CORDEX and their relation to climate sensitivity *Clim. Change* **162** 645–65
CESSDA 2021 Data management expert guide: file naming and folder structure (Consortium of European Social Science Data Archives) (available at: https://dmeg.cessda.eu/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure)
Cimadevilla E, Lawrence B N and Cofiño A S 2025 The Earth system grid federation (ESGF) virtual aggregation (CMIP6 v20240125) *Geosci. Model Dev.* **18** 2461–78
Collier N, Hoffman F M, Lawrence D M, Keppel-Aleks G, Koven C D, Riley W J and Randerson J T 2018 The international land model benchmarking (ILAMB) system: design, theory, and implementation *J. Adv. Model. Earth Syst.* **10** 2731–54
Durack P J *et al* 2025 The coupled model intercomparison project (CMIP): reviewing project history, evolution, infrastructure and implementation *EGUsphere Preprint* (https://doi.org/10.5194/egusphere-2024-3729)
Earth System Grid Federation (ESGF) 2025 *DRCDP data archive* (Oak Ridge National Laboratory Distributed Active Archive Center) (available at: https://esgf-node.ornl.gov/search/DRCDP)
Eaton B *et al* 2006 NetCDF climate and forecast (CF) metadata conventions version 1.11 Zenodo https://doi.org/10.5281/zenodo.14274887
Gleckler P *et al* 2024 Obs4MIPs data specifications version 2.5 (ODS2.5) (2.5) Zenodo https://doi.org/10.5281/zenodo.11500474
Gutmann E D, Hamman J J, Clark M P, Eidhammer T, Wood A W and Arnold J R 2022 En-GARD: a statistical downscaling framework to produce and test large ensembles of climate projections *J. Hydrometeorol.* **23** 1545–61
Harvard Biomedical Data Management 2023 File naming conventions (Harvard Medical School) (available at: https://datamanagement.hms.harvard.edu/plan-design/file-naming-conventions)

Hayhoe K, Scott-Fleming I, Stoner A and Wuebbles D J 2024 STAR-ESDM: a generalizable approach to generating high-resolution climate projections through signal decomposition *Earth's Future* **12** e2023EF004107
Hoffman F M *et al* 2025 Rapid evaluation framework for the CMIP7 assessment fast track *EGUsphere Preprint* (https://doi.org/10.5194/egusphere-2025-2685)
Kershaw P, Abdulla G, Ames S and Evans B 2020 *ESGF future architecture report* U.S. Department of Energy (https://doi.org/10.2172/1643759)
Lee J *et al* 2024 Systematic and objective evaluation of Earth system models: PCMDI metrics package (PMP) version 3 *Geosci. Model Dev.* **17** 3919–48
Massoud E C *et al* 2024 Hosting downscaled climate data products in ESGF2-US: expectations, requirements, and community standards *Abstract (GC23E-0310) American Geophysical Union Fall Meeting 2024* (*10 December 2024*) https://agu24.ipostersessions.com/?s=18-51-D7-58-2E-E3-CD-AB-64-FE-5B-B6-D9-2F-F4-D2
Mauzey C *et al* 2025 The climate model output rewriter (CMOR), version (3.11.0) Zenodo https://doi.org/10.5281/zenodo.15644808
Nikulin G, Buntemeyer L, Fernández J, McGinnis S and Evans J P 2024 CORDEX-CMIP6 archiving specifications for dynamical downscaling, version 1 Zenodo https://doi.org/10.5281/zenodo.10961069
Pierce D W, Cayan D R, Feldman D R and Risser M D 2023 Future increases in North American extreme precipitation in CMIP6 downscaled with LOCA *J. Hydrometeorol.* **24** 951–75
Potter G L, Carriere L, Hertz J, Bosilovich M, Duffy D, Lee T and Williams D N 2018 Enabling reanalysis research using the collaborative reanalysis technical environment (CREATE) *Bull. Am. Meteorol. Soc.* **99** 677–87
Rahimi S *et al* 2024 An overview of the Western United States dynamically downscaled dataset (WUS-D3 *Geosci. Model Dev.* **17** 2265–86
Spero T L, Bowden J H, Mallard M S, Nolte C G, Willison J, Jalowska A M, Talgo K D, Reynolds L J and Brehme K 2025 EPA dynamically downscaled ensemble (EDDE): datasets for adaptation and resilience planning for environmental endpoints *Bull. Am. Meteorol. Soc.* **106** E1539–57
Taylor K E *et al* 2018 CMIP6 model output metadata requirements, data reference syntax (DRS) and controlled vocabularies (CVs), version 1 *Program for Climate Model Diagnosis and Intercomparison* Zenodo https://zenodo.org/records/12768887
Thrasher B, Khajehei S, Kim J B and Brosnan I 2024 NASA Earth exchange downscaled climate projections 30 arcseconds CMIP6 *Sci. Data* **11** 1346
Thrasher B, Maurer E P, McKellar C and Duffy P B 2012 Technical note: bias correcting climate model simulated daily temperature extremes with quantile mapping *Hydrol. Earth Syst. Sci.* **16** 3309–14
U.S. Department of Energy 2024 Understanding decision-relevant regional climate data products workshop (U.S. Department of Energy) (https://doi.org/10.2172/2474992)
Ullrich P A *et al* 2025 Advancing a coordinated strategy for decision-relevant climate data products for the United States *Bull. Am. Meteorol. Soc.* submitted
Waliser D *et al* 2020 Observations for model intercomparison project (Obs4MIPs): status for CMIP6 *Geosci. Model Dev.* **13** 2945–58
Wilkinson M *et al* 2016 The FAIR guiding principles for scientific data management and stewardship *Sci. Data* **3** 160018
Williams D N *et al* 2009 The Earth system grid: enabling access to multimodel climate simulation data *Bull. Am. Meteorol. Soc.* **90** 195–206
Wootten A M, Massoud E C and Raymond C 2025 "Which projections do I use?" strategies for climate model ensemble subset selection based on regional stakeholder needs *Geophys. Res. Lett.* **52** e2025GL116492
Wootten A M, Massoud E C, Sengupta A, Waliser D and Lee H 2020 The effect of statistical downscaling on the weighting of multi-model ensembles of precipitation *Climate* **8** 138
Wootten A M, Massoud E C, Waliser D E and Lee H 2023 Assessing sensitivities of climate model weighting to multiple methods, variables, and domains in the south-central United States *Earth Syst. Dyn.* **14** 121–45