

EGU24-17356, updated on 16 Oct 2024
<https://doi.org/10.5194/egusphere-egu24-17356>
EGU General Assembly 2024
© Author(s) 2024. This work is distributed under
the Creative Commons Attribution 4.0 License.



How do the CMIP7 infrastructure plans support the implementation of the IPCC data guidelines?

Martina Stockhause¹, Matthew Mizielski², Anna Pirani³, Lina Sitz⁴, Alessandro Spinuso⁵, Mudathir Abdallah⁶, Jesbin Baidya⁶, Paul Durack⁷, and Daniel Ellis⁸

¹German Climate Computing Center (DKRZ), Hamburg, Germany (stockhause@dkrz.de)

²MetOffice, Exeter, UK

³Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Venezia, Italy

⁴Instituto de Física de Cantabria (IFCA), Universidad de Cantabria (UC), Santander, Spain

⁵Royal Netherlands Meteorological Institute (KNMI), Utrecht, Netherlands

⁶Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland

⁷Program for Climate Model Diagnosis & Intercomparison (PCMDI), Lawrence Livermore National Laboratory (LLNL), Livermore, CA, USA

⁸WCRP CMIP International Project Office (CMIP-IPO), Oxford, UK

The Intergovernmental Panel on Climate Change (IPCC) regularly assesses a wide range of research results related to climate change reaching from physical sciences to economic and social sciences to provide policy makers with options for combatting the challenges of climate change. The IPCC authors analyze data across multiple domains and from multiple sources.

The IPCC data guidelines enhance the transparency of IPCC outcomes by ensuring that figure creation is traceable, citing input data and long-term preserving data and software. The related data and metadata requested from the authors includes detailed information on datasets used in every figure, for which standardized machine-accessible and -readable information needs to be supported by the input data providers.

An example of an important input data provider is the Coupled Model Intercomparison Project (CMIP), which has continuously improved its standards and data infrastructure to keep track with the significant increase in the scale of the project over recent phases. CMIP defines a set of standards including vocabularies for controlled metadata fields, e.g. variable and experiment names, along with the data itself and its structure. A set of infrastructure services provide access to data, through the Earth System Grid Federation (ESGF), description of the climate models used and known errata, through ES-DOC services, and data citation information including data usage in derivative data sets and published papers where known.

The contribution will discuss the diverse data-workflows of the IPCC authors and the ways that the CMIP infrastructure supports them. Authors access data from the primary data portals of the ESGF, but also from secondary data portals (Copernicus, Pangeo, Climate4Impact) or local data pools hosted by national institutions. The IPCC authors have faced a number of challenges,

including accessing data citation and model description information together with the data, and in identifying new dataset versions with significant changes. With the IPCC's plan to utilize provenance records in AR7 to gather all information requested by the IPCC data guidelines, machine-readable information accessible through the file headers becomes essential.

The entire IPCC AR7 data workflow needs to be supported by tools: the figure creation by the IPCC authors, the report editing process by the TSU and the curation of the CMIP7 input data subset used and the intermediate and final datasets created by various DDC Partners, including bi-directional references between outputs. Virtual workspaces such as CEDA and DKRZ provided for the authors in AR6, which give access to their data pools and common software packages like the ESMValtool, can support the authors in the preparation of their figures and the provision of the requested documentation and provenance information. A dedicated Figure Manager will play a central role in managing the report figures and supporting the overall data workflow. Ultimately, lowering the burden for the authors, the TSU staff and the DDC Partners. This timely gathered information can then be analyzed and used for a harmonization of dataset version usage across the chapters and reports.