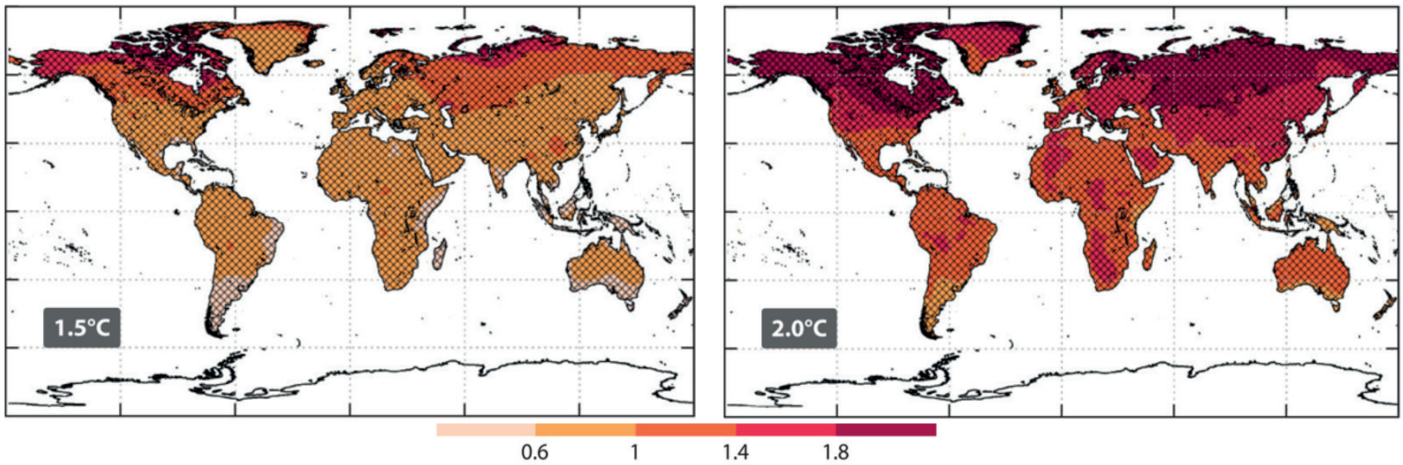
Data Management Stories at DKRZ

Issue 6, August 2019



Annual mean 2 m temperature differences between the 1.5 °C warmer climate (left; 2106-2115) and the 2.0 °C warmer climate (right; 2106-2115), respectively, and the current decade climate (2006-2015).

Data Management for HAPPI at DKRZ

The HAPPI-MIP initiative proposes three core experiments (Mitchell et al., 2017) representing three different climate warming equilibria:

- current decade climate (2006-2015)
- future climate warmer by 1.5 °C than at pre-industrial (1861-1880) conditions (2106-2115)
- future climate warmer by 2.0 °C than at pre-industrial conditions (2106-2115)

German Climate Computing Center (DKRZ)

The mission of DKRZ is to provide high performance computing (HPC) platforms, sophisticated and high capacity data management and services for premium climate science.

Besides providing HPC services, DKRZ supports projects in all aspects relevant to data management. This includes preparation, quality assessment, distribution, and long-term archiving of data.

www.dkrz.de

The German contribution to HAPPI-MIP is the HAPPI-DE project (Schleussner et al., 2018). In that framework, DKRZ contributed to the HAPPI-MIP database with global model simulation results based on the Atmospheric General Circulation Model (AGCM) ECHAM6.3-LR.

DKRZ manages the long-term archiving (LTA) and the publication of HAPPI-related data at CERA-WDCC (CERA: Climate and Environmental data Retrieval and Archiving; WDCC: World Data Center for Climate). The LTA includes a systematic inspection of the data and the assigment of DataCite Digital Object Identifiers (DOIs) making the data citable to a high scientific standard.

HAPPI: Half a degree Additional warming, Prognosis and Projected Impacts

The Conference Of the Parties of the United Nations Framework Conventions on Climate Change recommended in its Paris Agreement of 2015 to hold the increase in the global temperature to well below 2 °C above pre-industrial levels recognizing that this would significantly reduce the risks and impacts of climate change.

It invited the Intergovernmental Panel on Climate Change (IPCC) to prepare a Special Report (SR1.5) on the impacts of global warming of 1.5 °C above pre-industrial levels and related greenhouse gas emission pathways in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.

The international HAPPI – Model Intercomparison Project (HAPPI-MIP) initiative was launched in response to the announcement by the IPCC of such a report (SR1.5).

www.happimip.org/

HAPPI DE

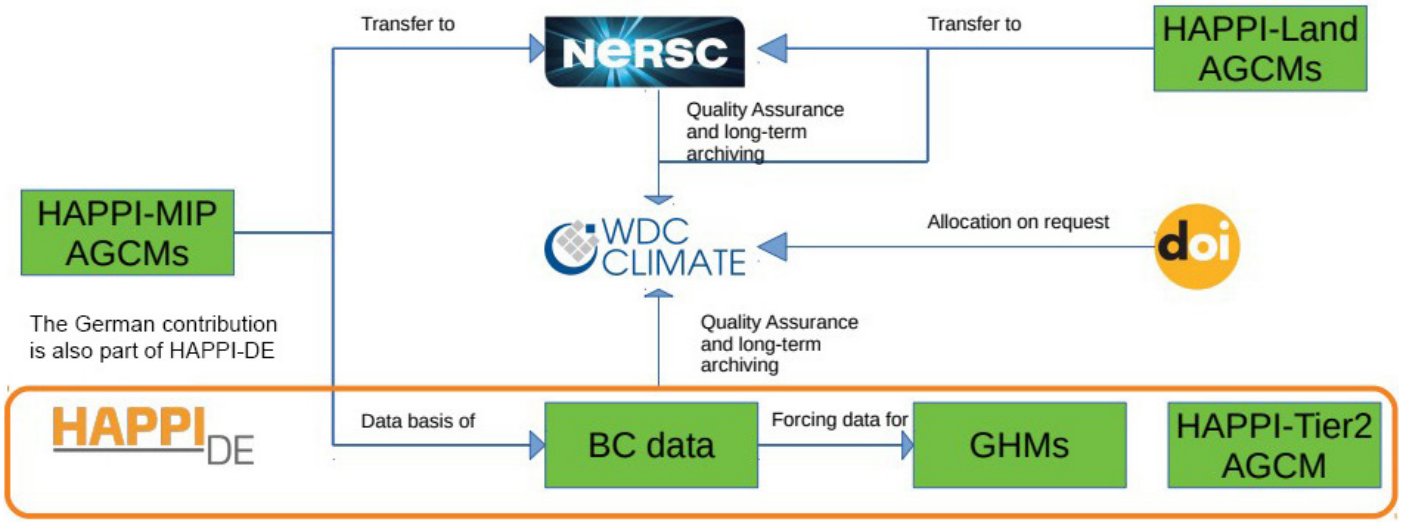
HAPPI-DE is a German contribution to the international HAPPI-MIP initiative. It is a cooperative project of DKRZ, ClimateAnalytics GmbH (CA), the Institute of Physical Geography at Goethe University Frankfurt (GU) and the Climate Service Center Germany (GERICS) funded by the German Federal Ministry of Education and Research (BMBF) from 01.02.2017 to 31.01.2019 under the grant agreement 01LS1613E.

CLIMATE ANALYTICS

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

GERICS Climate Service Center Germany
An institution of Helmholtz-Zentrum Geesthacht

Bundesministerium für Bildung und Forschung



HAPPI long-term archiving workflow

HAPPI long-term archiving in CERA-WDCC

Data providers within the international HAPPI-MIP initiative transferred their standardized model results to a public portal at the National Energy Research Scientific Computing Center (NERSC) operated by the Lawrence Berkeley National Laboratory, California (*https://portal.nersc.gov/c20c/data.html*). The international HAPPI-MIP data comprises model output from Canadian, Swiss, Japanese, Norwegian and German scientific institutes.

As a completion of HAPPI-MIP, some more experiments were designed in HAPPI-DE and at the Eidgenössische Technische Hochschule Zurich. On the one hand, these additional experiments aim to survey the robustness of the HAPPI-MIP results with respect to the variability of the El Niño Southern Oscillation (HAPPI-Tier2) or with respect to different land-use scenarios (HAPPI-Land). On the other hand, climate impact models help to analyze the impacts of climate change across affected sectors and spatial scales more precisely. Global hydrological model (GHM) experiments aim to survey risks for the global freshwater system under global warming. Those GHM simulations were forced by bias-corrected HAPPI-MIP data (BC).

The long-term archiving of the HAPPI data using DKRZ's Core Trust Seal-certified LTA WDCC service was essential to ensure long-term reusability of the data and traceability of results. For this purpose, the HAPPI data was first transferred to the parallel file system at DKRZ and then systematically inspected for compliance with established data standards. For this compliance check DKRZ used its in-house-developed quality assurance tool for climate data. Checking for data quality is one aspect, ensuring that the data are correctly described in DKRZ's CERA database is another aspect of long-term archiving in WDCC. Therefore, the metadata published on *https://cera-www.dkrz.de* were compiled in close cooperation with the data providers – and the data were finally ingested into the LTA WDCC tape archive. Finally, DataCite DOIs, which allow for data citation following best practices, were assigned to the data. One DOI was assigned per model and initiative and/or project (e.g. one DOI assigned for all HAPPI-MIP simulations performed with the AGCM ECHAM6.3-LR; *https://doi.org/10.26050/WDCC/HAPPI-global-ECHAM6.3_v2*).

By August 2019, over 40 Terabyte of data had been archived, organized in more than 185,000 datsets and to which six DOIs had been assigned so far. 48 scientific publications are based on HAPPI-related data now archived in the WDCC at DKRZ. Further DOIs and publications will follow.

Benefits from CERA-WDCC data storage

Scientific data is the basis for published results, as e.g. IPCC's SR1.5. In order to comply with good scientific practices and due to the need for scientific results to be traceable and reproducible, it is becoming common practice that scientific data is made FAIR (Findable, Accessible, Interoperable, Reusable).

The FAIR data guiding principles are comprised of a set of metrics describing a continuum of properties which, when complied with, increase the reusability of a digital data object in a discipline-specific context. Details on the FAIR data guiding principles can be found on the webpages of the GOFAIR-Initiative (*https://www.go-fair.org/fair-principles/*).

DKRZ provides the externally certified LTA service at WDCC and archives the data guided by FAIR principles. It assures that the data is accessible for a period of at least ten years. Furthermore, the data will be assigned a DOI, making the data citable by their users. Thereby the efforts of the data providers can be more adequately acknowledged.

Reference

Mitchell et al. (2017). Half a degree Additional warming, Prognosis and Projected Impacts (HAPPI): Background and experimental design. https://doi.org/10.5194/gmd-10-571-2017

Schleussner et al. (2018). Klimafolgen bei 1.5 °C und 2 °C – Ergebnisse des HAPPI-DE Konsortiums. https://climateanalytics.org/publications/2018/happi-de-15c/

Hirsch et al. (2018). Biogeophysical Impacts of Land-Use change on Climate Extremes in Low-Emission Scenarios: Results from HAPPI-Land. https://doi.org/10.1002/2017EF000744