25 years IPCC Data Distribution Centre at DKRZ and the Reference Data Archive for CMIP data

Martina Stockhause[1], Michael Lautenschlager[1]

[1]German Climate Computing Center (DKRZ), Hamburg, 20146, Germany

5   *Correspondence to*: Martina Stockhause (stockhause@dkrz.de)

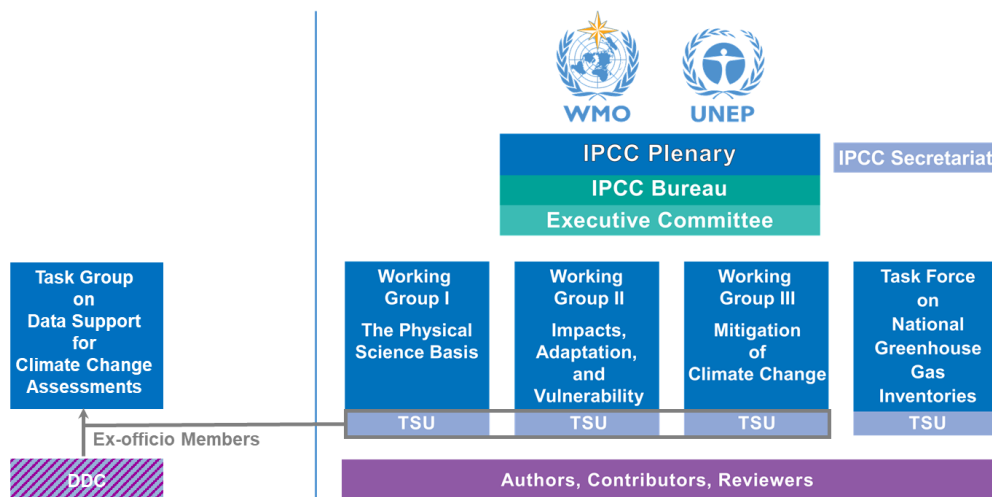
Abstract. The Data Distribution Centre (DDC) of the Intergovernmental Panel on Climate Change (IPCC) celebrates its 25th anniversary in 2022. DKRZ is the last remaining founding member among the DDC Partners. The continuous developments in data standardization, data management best practices and infrastructure developments are reflected in the DDC's operations. Examples for such milestones are the NetCDF/CF data standard, the DataCite data DOIs enabling data citation, the data

10   preservation and stewardship standards of the World Data System (WDS), or the IPCC FAIR Guidelines for the current Sixth IPCC Assessment Report (AR6). The FAIR Guidelines significantly changed the role of the DDC Partner DKRZ from an independent partner for long-term data preservation into an active partner involved in IPCC's Sixth Assessment cycle. As a result, the DDC has become more exposed and visible, posing a challenge and an opportunity to operationalize IPCC's FAIR Guidelines and long-term preservation approaches. Combined with a long-term DDC strategy, the DDC can be positioned in

15   the developing FAIR digital object research landscape as well as in the political and societal realm of climate change.

1   History of TG-Data and the IPCC DDC

The current Data Distribution Centre (DDC, https://ipcc-data.org/, last access: 10 March 2022) of the Intergovernmental Panel on Climate Change (IPCC) is jointly managed by four partners: the German Climate Computing Center (DKRZ, Germany), the Center for International Earth Science Information Network (CIESIN, USA), the Spanish Research Council (CSIC, Spain),

20   and MetadataWorks (UK). DKRZ is the only remaining founding partner, who has been operating the DDC for 25 years. The DDC Partners signed a Memorandum of Understanding defining the DDC's operations (Xing, 2021). The DDC is overseen by the Task Group on Data Support for Climate Change Assessments (TG-Data, https://ipcc.ch/data, last access: 10 March 2022), which is a non-permanent part of IPCC's structure (Fig. 1).

25   **Figure 1: IPCC Structure (from IPCC webpage)**

TG-Data member experts are complemented by ex-officio Members representing the DDC Partners and the Technical Support Units (TSU) of the three IPCC Working Groups (WG). The core role of the DDC is the support for authors and data users in data and scenarios underpinning IPCC outputs (see DDC Guidance, IPCC, 2018a), especially those in developing countries.

30 The DDC was established at the Thirteenth Session of the IPCC (IPCC-13) on 22 and 25-28 September 1997 in the Maldives (IPCC, 1997). Deutsches Klimarechenzentrum (DKRZ) in Germany and the Climatic Research Unit (CRU) in the United Kingdom were selected as a shared DDC operation. The initiative started at the IPCC Workshop on Regional Climate Change Projections for Impact Assessment (London 24-26 September 1996), on which requirements for data availability, data standardization and data quality for the IPCC Assessment process were formulated together with the needs for guidance materials for consistent data usage and capacity building. The workshop led to the establishment of the Task Group on Climate

35 Scenarios for Impact Assessment (TGCIA) expert group by the IPCC Bureau on its Eleventh Session (7-8 November 1996). TGCIA formulated the recommendation for the establishment of a DDC and sought Bureau approval in May 1997. The IPCC Bureau requested governments on 22 July 1997 to nominate institutions to act as DDC and to provide the necessary financial support to establish and maintain the DDC function. This led to the DDC establishment on IPCC-13.

The Task Group TGCIA was renamed to Task Group on Data and Scenario Support for Impact and Climate Analysis (TGICA)

40 in 2003. The German DDC operations were taken over by the Model and Data Group of the Max Planck Institute for Meteorology (MaD/MPI-M) between 2000 and 2009. Since 2010, the DDC has been operated by DKRZ, again.

As a result of an IPCC Expert Meeting on the Future of TGICA (IPCC, 2016), TGICA got under review of the IPCC Ad-hoc Task Force TGICA from 2016 till the reestablishment of the Task Group as Task Group on Data Support for Climate Change Assessments (TG-Data) on the 47th IPCC Plenary session in Paris, 13-16 March 2018 (IPCC, 2018a). The Terms of Reference

45 for TG-Data as well as the Guidance for the DDC were renewed. The focus of the DDC was narrowed and refocused on data support tasks.

2 The Reference Data Archive at the DDC at DKRZ

DKRZ is the DDC Partner responsible for the long-term preservation of the global climate model data, which is strongly related to Working Group I (WGI) of the IPCC and the research project Coupled Model Intercomparison Project (CMIP) of

50 the World Climate Research Programme (WCRP). Starting with the Second Assessment Report (SAR), core variables for the characterization of the state of the Earth System (Tab. 1) from the model applications for future climate projections were long-term archived at DKRZ, building the Reference Data Archive for the global climate model data underpinning IPCC's ARs. During the long-term archival, the data are stored on tape and the metadata are enriched to provide sufficient information for various downstream users without specific knowledge of climate model applications. The size of the Reference Data Archives

55 for the different ARs increased from around 10 GBytes and 400 datasets for SAR and TAR via ca. 1 TByte and 1 500 datasets for AR4 to 1.7 PBytes and 910 000 datasets for AR5 (Fig. 2). The reasons are an increased number of archived variables per model run, an increased number of models participating in CMIP, and the change in archival strategy from monthly to daily and sub-daily data. Download statistics show the long-term interest of users in the DDC Reference Data (Fig. 3).

As DDC Partner, DKRZ committed to keep its DDC data holdings accessible and reusable on the long-term, which involves

60 cyclic renewal of hardware, continuous maintenance of software, and metadata and data curation. New generations of hardware (tape system) require the copying of the DDC data holdings on new cartridges. Software updates for data discovery, access and exchange are required to comply to new standards and interfaces in order to enhance the user experience. An important metadata curation measure was the introduction of the Climate and Forecast (CF) Standard Names (https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html, last access: 10 March 2022)

65 and the mapping of proprietary model variable names to this new community standard in the DDC metadata. The data format of the datasets in the DDC reflects the change in the community's standard data format from the WMO GRIB1

(https://community.wmo.int/activity-areas/wis/grib-edition-1, last access: 10 March 2022) standard used for SAR and TAR towards NetCDF/CF (https://cfconventions.org/Data/cf-conventions/cf-conventions-1.9/cf-conventions.html, last access: 10 March 2022) used for AR5 with the provision of the datasets in both formats for AR4. In support of the IPCC users located in

70   developing countries with low internet bandwidths, the DDC introduced the service of sending DVD and USB devices with preselected variables for seven regions per airmail. Data of the FAR were added to the DDC in 2008 in the original formats, in which they were rescued. As it was not possible for all datasets to identify whether these data were available at the time of FAR preparation, only a subset of the available datasets was archived in the DDC. Because of the low level of standardization, these datasets are difficult to use.

75

air_pressure_at_sea_level
air_temperature
convective_precipitation_flux
dew_point_temperature
geopotential_height
global_average_thermosteric_sea_level_change
land_area_fraction
land_ice_area_fraction
large_scale_precipitation_flux
precipitation_flux
relative_humidity
sea_ice_amount
snowfall_amount
soil_moisture_content
specific_humidity
surface_altitude
surface_net_downward_shortwave_flux
surface_sensible_heat_flux
surface_downwelling_shortwave_flux_in_air
surface_snow_area_fraction
surface_snow_melt_flux
surface_snow_thickness
surface_temperature
water_evaporation_rate
wind_speed
x_wind
y_wind

Table 1: Core variables of the Reference Data Archive in CF Standard Name convention.

DKRZ adjusted to new data management best practices evolving over time. The DKRZ long-term data archive was approved in 2003 as WDC Climate (WDCC) in the ICSU World Data Center system. DKRZ became a Regular Member of the ISC

80   World Data System (WDS, https://www.worlddatasystem.org/, last access: 10 March 2022) in 2008, the year of WDS's establishment. Therefore, the DDC Partner DKRZ complies with the common research repository standards adopted by the WDS, currently the CoreTrustSeal (https://www.coretrustseal.org/, last access: 10 March 2022). With the foundation of DataCite in 2009, registering DOIs for data in order to make them citable became a community standard, which was taken up by the DDC Partner DKRZ for the AR5 Reference Data Archive published in 2013 and 2014. The long-term archival of AR5

85   provided further major changes in the workflow due to the extremely high data volume and several changes in the CMIP5 (https://pcmdi.llnl.gov/mips/cmip5/, last access: 10 March 2022; Taylor et al., 2012) data infrastructure:

- The data were disseminated by the newly developed federated and decentral infrastructure of the Earth System Grid Federation (ESGF, https://esgf.llnl.gov, last access: 10 March 2022);

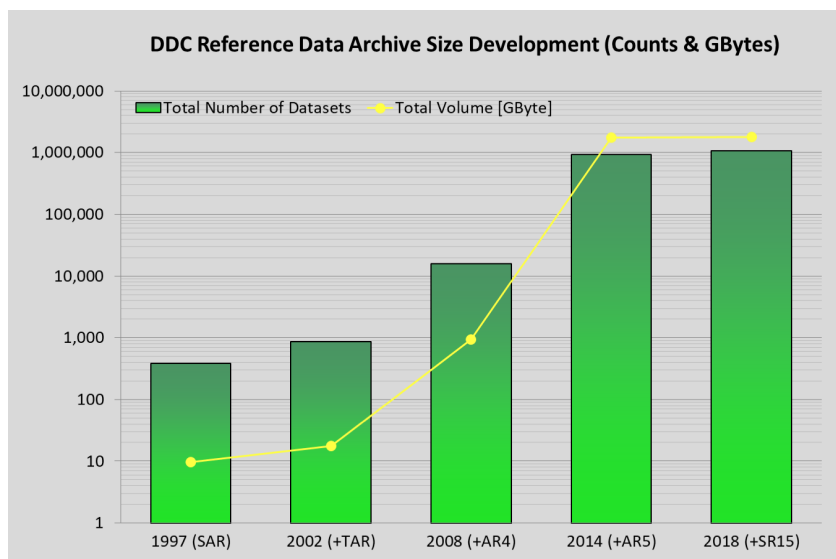



Figure 2: Size development of the DDC Reference Data Archive for the global climate model data.

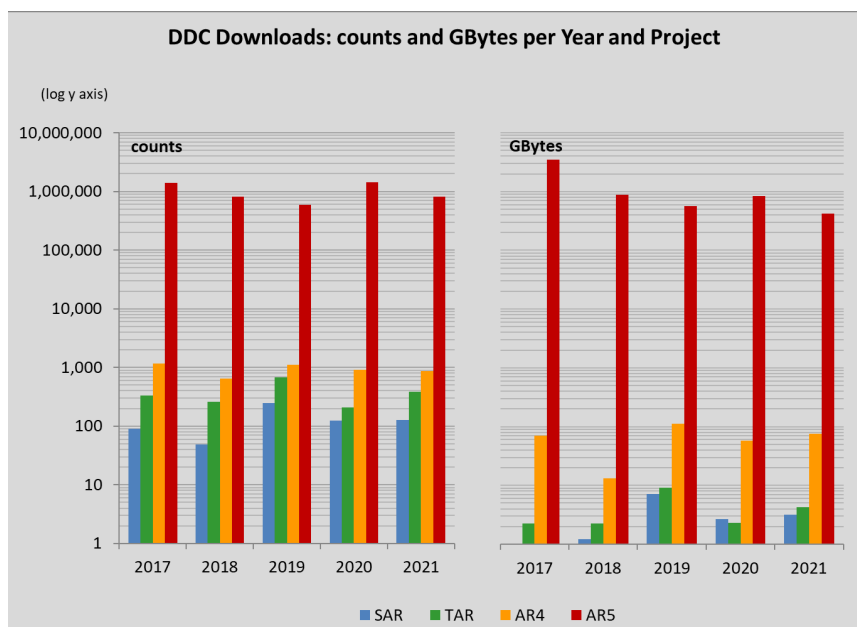



Figure 3: Downloads in number of datasets [counts] and volume [GByte] from the DDC Reference Data Archive over the last 5 years per Assessment Report (AR6 Reference Data is not yet available).

- detailed model and experiment documentations were gathered from the CMIP5 participants by the Earth System Documentation (ES-DOC, https://es-doc.org, last access: 10 March 2022);

- a three level quality control procedure (CMIP5 QC, https://cmip5qc.wdc-climate.de, last access: 10 March 2022) was applied to ensure basic data quality, the consistency of metadata, and metadata conformance with community standards like NetCDF/CF and project standards like the Data Reference Syntax (DRS). Passing the three quality control levels was the prerequisite for the integration of the data in the IPCC AR5 Data Reference Archive (Stockhause et al., 2012; Fig. 4).

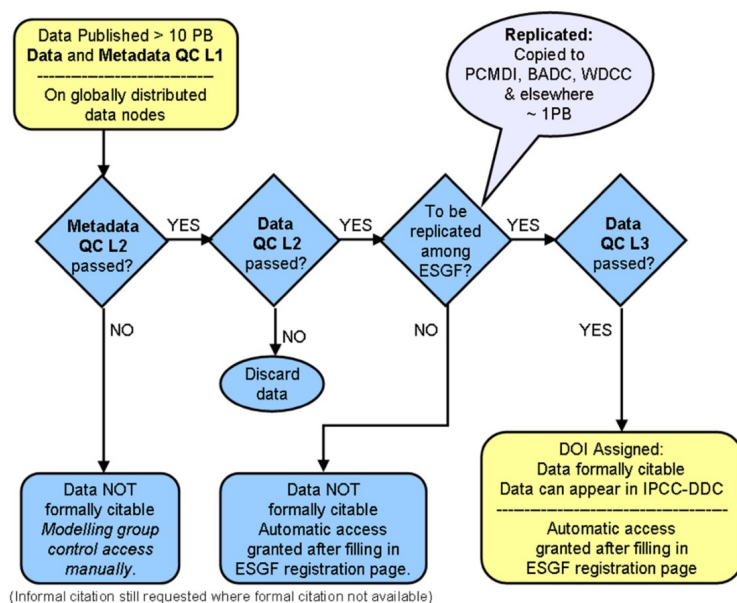



Figure 4: Schematic workflow of the three-level quality control procedure of CMIP5 data

The size of the CMIP5 data archive required a high level of automation for metadata and data ingest as well as for the quality
105 control checks. New interfaces to the infrastructure components ESGF, ES-DOC and DataCite had to be developed for
insertion of use and discovery metadata and data DOI registration. The long-term archived DDC AR5 data were made
accessible through the ESGF. ETH Zurich collected a CMIP5 data subset in support of the IPCC authors in an alternate data
structure. Due to the difficulty to relate the individual datasets back to the CMIP5 datasets, the data archive of the ETH Zurich
was added as additional IPCC Working Group I AR5 snapshot to the DDC AR5 Reference Data Archive. Discussions with
110 the ETH Zurich provided valuable input for the IPCC FAIR Guidelines concept for AR6 and the long-term archival of the
CMIP6 data underpinning the AR6.

3 AR6 and the IPCC FAIR Guidelines

TGICA was under review of the IPCC from the start of the Sixth Assessment Cycle in January 2016 until the re-established
TG-Data held its First Meeting in November 2019. This was less than a year prior to the original WGI literature and data cut-
115 off date on 30 September 2020, which was postponed to 31 January 2021 due to the Corona pandemic. The lack of the
coordinating task group was a difficult situation for the formulation and implementation of the FAIR Guidelines into the Sixth
Assessment Report (AR6).

The idea for the FAIR Guidelines was born during the IPCC Expert Meeting on the Future of TGICA in January 2016. The
aim of the FAIR Guidelines is the enhancement of the transparency of the IPCC AR6 and thus to contribute to IPCC's integrity.
120 The FAIR Guidelines implement the established data management principles of FAIR (Findable, Accessible, Interoperable,
Reusable; Wilkinson et al., 2016) for data and TRUST (Transparency, Responsibility, User Focus, Sustainability, Technology;
Lin et al., 2020) for repository operations into the Sixth Assessment cycle. The FAIR data principles describe requirements
for datasets to become an integral part of the research environment. The TRUST principles for repositories and its
implementation in the CoreTrustSeal (CTS) complement these essential data properties by best practices for repository
125 operations in long-term data preservation and data stewardship. The CTS certification is part of the Regular Membership
application and approval procedure of the ISC World Data System.

The concept development for the FAIR Guidelines started at the First IPCC AR6 Data Workshop in Hamburg, Germany, 19-20 September 2017 (Stockhause et al., 2017) and continued at the second virtual meeting on 20 February 2018. In collaboration with the WDS, which started at the Data Repository Day 2018 (WDS, 2018), the FAIR Guidelines concept was formulated in

130 Stockhause et al. (2019). This concept was discussed with IPCC authors of WGI and WGII on the IPCC Expert Meeting on Assessing Climate Information for Regions in Trieste, 16-18 May 2018 (IPCC, 2018b). Finally, the implementation of the FAIR Guidelines into tools supporting the authors was the topic of a WGI Training on Data and Software Development in Oberpfaffenhofen, Germany, 6-7 June 2019 (IPCC, 2019). The FAIR Guidelines were approved by TG-Data on its first meeting in Montreal, Canada, 6-8 November 2019.

135 The IPCC FAIR Guidelines (Pirani et al., 2022) can be described along three aspects:

•   Traceability of key statements and of figure and table creation: Information on input datasets like CMIP6 (Eyring et al., 2016), final data displayed in figure graphics, and the analysis script applied to produce the figures are collected from the authors by the WGI TSU. This information is provided for every figure in a special section of the Supplementary Materials for each chapter. Moreover, bidirectional references between the digital AR6, final datasets and input datasets will interlink

140 the different AR6 products (Fig. 5).

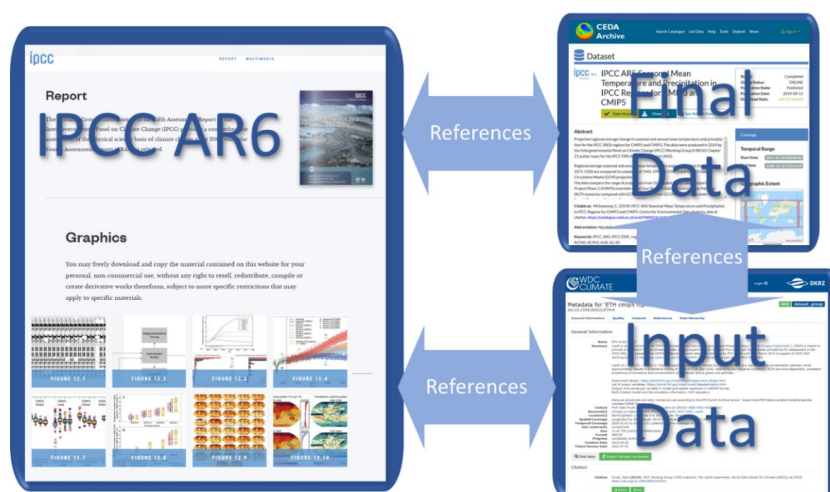



Figure 5: Schematic vision of the bi-directional references between report, input data and final datasets in IPCC AR6 (screenshots from IPCC, CEDA and DKRZ webpages).

•   Credit for input data: Input datasets used by the authors are cited in the AR6 in compliance with Good Scientific

145 Practices (DFG, 2019). In case of CMIP6 data, data citation is required by the Creative Commons licenses (CC, https://creativecommons.org/, last access: 10 March 2022), CC BY-SA and CC BY-NC-SA, under which CMIP6 data were published. CMIP6 data are cited in a summarized form in Appendix II of the WGI AR6 (https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/, last access: 10 March 2022; IPCC, 2021), the provenance metadata of the IPCC WGI Interactive Atlas (https://interactive-atlas.ipcc.ch/, last access: 10 March 2022), and for each figure in the Supplementary

150 Materials.

•   Long-term preservation of input data, scripts, and final data: The information, scripts and final datasets collected by the WGI TSU are transferred to the designated repository for long-term preservation. In case of CMIP6, dataset lists are compiled by the TSU and provided for the DDC Partner DKRZ. Based on these lists, datasets are replicated, use metadata are accessed from the ESGF, and further documentations from the Citation Service (Stockhause and Lautenschlager, 2017) and

155 ES-DOC (Pascoe et al., 2020) are added. The long-term archival workflow is depicted in Figure 6.

The implementation of the FAIR Guidelines required a close cooperation between WGI TSU and the DDC Partners.

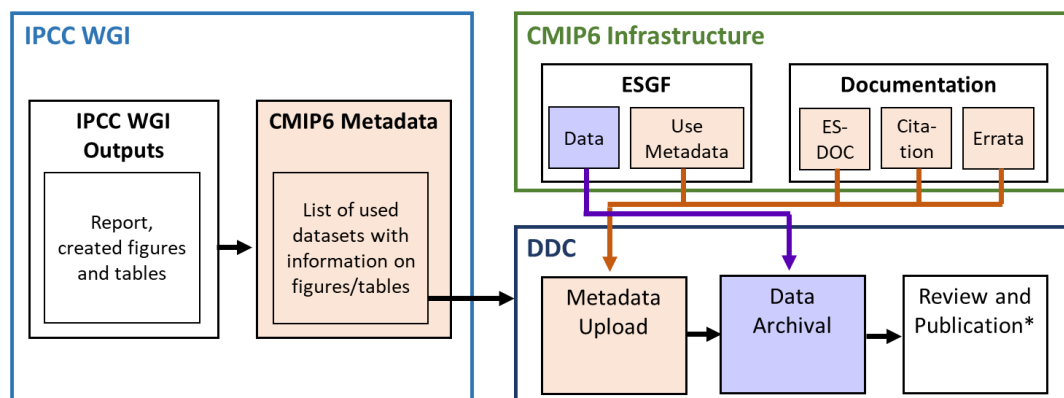



Figure 6: CMIP6 input data archival workflow to build the DDC AR6 Reference Data Archive

4   Changed role of the DDC Partner DKRZ in AR6

160   The implementation of the FAIR Guidelines changed the role of the DDC Partner DKRZ from a long-term archive, operating mostly independent of WGI and the assessment cycle, into an active partner with an enhanced role within the Sixths Assessment cycle. The closest cooperation was required with the WGI TSU on the formulation and implementation of the FAIR Guidelines. Thus, DDC Managers participated in the IPCC Expert Meeting on Assessing Climate Information for Regions in May 2018 and jointly organized the WGI Training on Data and Software Development in June 2019 together with

165   the WGI TSU. Advice based on the DDC's long experience in data management were provided for the necessary information on data usage required from the authors, best practices in data citation and the definition of machine-actionable interfaces. The DDC Manager at DKRZ joined the WGI AR6 authors as contributing author and reviewed the First and Second Order Drafts of the report to provide expert advice on data management aspects.

This active role of the DDC in AR6 increased the DDC's visibility and resulted in requests for further support of the IPCC

170   authors during the preparation of the AR6. The DDC Partner DKRZ and former DDC Partner CEDA provided Virtual Workspaces (Stockhause, 2020; Fig. 7) for the authors co-funded by the EU project Infrastructure for the European Network for the Earth System Modelling (IS-ENES, http://is.enes.org, last access: 10 March 2022). These collaboration platforms provided storage and compute resources for the chapter author groups together with access to requested core datasets and common software packages. Moreover, DKRZ supported the ESMValtool (https://www.esmvaltool.org/, last access: 10 March

175   2022; Eyring et al., 2020) development in technical aspects and hosts the webpage with CMIP evaluation results (https://cmip-esmvaltool.dkrz.de/, last access: 10 March 2022). On the national level, the DDC Manager at DKRZ joined the authors' subgroup of the German IPCC Coordination Office (https://www.de-ipcc.de/, last access: 10 March 2022).

For the provision of sustainable DDC services at DKRZ, it is essential to collaborate with a wide range of external partners:

-        General Infrastructure Partners:

180   o   National: Nationale Forschungsdateninfrastruktur (NFDI, https://www.nfdi.de/, last access: 10 March 2022)

o   European: European Open Science Cloud (EOSC, https://eosc.eu/, last access: 10 March 2022)

o   International: W3C, ISO, Research Data Alliance (RDA), WDS, CODATA, DataCite

-        Geophysical Infrastructure Partners:

o   Open Geospatial Consortium (OGC)

185   o   Coalition for Publishing Data in the Earth and Space Sciences (COPDESS, https://copdess.org, last access: 10 March 2022)

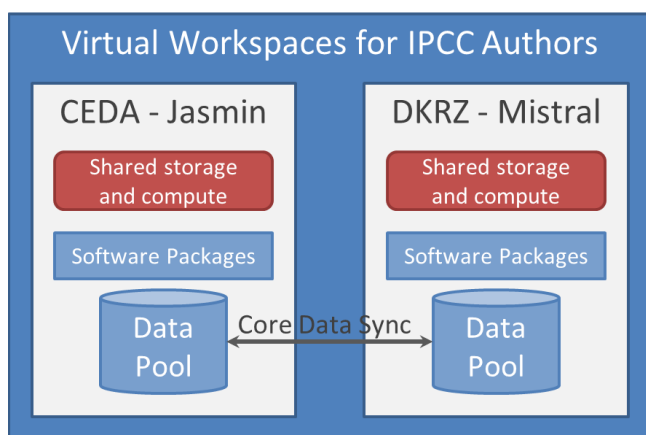
        o   NFDI Consortium for Earth System Sciences (NFDI₄Earth, https://www.nfdi4earth.de/, last access: 10 March 2022)

    -    Climate Infrastructure Partners:

190   o   European Network for Earth System Modelling Research Infrastructure (ENES-RI) Partners

        o   ESGF

        o   Copernicus Climate Data Store (CDS, https://cds.climate.copernicus.eu/, last access: 10 March 2022)

        o   WIP and Climate Data Node Operation Team (CDNOT)

195   **Figure 7: Virtual Workspaces provided by CEDA and DKRZ for IPCC AR6 authors (co-funded by IS-ENES).**

The implementation of the FAIR Guidelines introduced questions to be solved with the IPCC Bureau like license issues. IPCC had to ask the CMIP6 participants through the Working Group on Coupled Modelling (WGCM) for an exemption of the CMIP6 data licenses, allowing the DDC Partners to publish the derived final datasets underpinning figures and intermediate and final datasets underpinning the Interactive Atlas under the CC BY license. As representatives of TG-Data, DDC Partner

200   DKRZ and former DDC Partner CEDA represented the technical requirements of IPCC in the WGCM Infrastructure Panel (WIP, https://www.wcrp-climate.org/wgcm-cmip/wip, last access: 10 March 2022) and contributed data aspects to the IPCC Informal Group on Publications.

Independent of the FAIR Guidelines, the DDC Partners intensified their collaboration. The new UK DDC Partner MetadataWorks set up a joint DDC catalogue to improve the discovery of DDC data holdings. The DKRZ's DDC Manager

205   contributed to the development of the DDC's profile of the Data Catalog Vocabulary standard (W3C DCAT, https://www.w3.org/TR/vocab-dcat-3/, last access: 10 March 2022) and provided the metadata of its Reference Data Archive in December 2021. A central DDC help desk was set up to coordinate the DDC user support. The revision of the DDC webpages is ongoing aiming to retire outdated pages and to refocus on IPCC-related data according to the renewed DDC Guidance.

5   Role of CMIP6 infrastructure and other projects for the AR6 cycle

210   All IPCC Assessment cycles relied on the latest research on climate change projections provided by the WCRP project CMIP. The core of the authors' work is the assessment of the latest peer-reviewed literature. CMIP data were used in the peer-reviewed literature as well as for the creation of IPCC report figures. With the introduction of the FAIR Guidelines, the dependency on CMIP-related literature and CMIP data were complemented by the dependency on CMIP6 infrastructure components (Petrie et al., 2021) and further DDC support activities (see section 4). For the CMIP6 data and infrastructure coordination, the WIP

215   was set up, which endorsed concepts for infrastructure contributions and project data standards. The standardization of CMIP6 data is important for the reusability of the data. This includes compliance to the NetCDF/CF standard and specific file name

conventions, a uniform directory structure, and the collection and dissemination of the CMIP6 Controlled Vocabularies (CMIP6-CVs, https://github.com/WCRP-CMIP/CMIP6_CVs, last access: 10 March 2022; Taylor et al., 2018)

The necessary infrastructure components include the data infrastructure Earth System Grid Federation (ESGF; Williams et al.,
220 2016; Cinquini et al., 2014), which disseminates the data and provides use metadata and references to further information like data citation through its index. The CMIP6 Citation Service (http://cmip6cite.wdc-climate.de, last access: 10 March 2022; Stockhause and Lautenschlager, 2017) contributes data references to the long-term archival concept (Stockhause et al., 2015) and thus links the data infrastructure to the long-term data preservation of the CMIP6 data subset in the DDC AR6 Reference Data Archive.

225 Apart from these necessary infrastructure components, ES-DOC provides detailed information on models, experiments and errata for further metadata enrichment in the DDC Reference Data Archive. While the Citation Service succeeded in having full data citation coverage for all datasets published in the ESGF on the literature and data cut-off date for WGI AR6 on 31 January 2021, the coverages of ES-DOC model descriptions and errata information are disappointingly low. That means that the Citation Service and DDC mostly rely on the brief model descriptions contributed by the CMIP6 participants during the
230 registration process, which are available in the CMIP6-CVs.

The agreed high data standards and the high volume of the contributed data require thorough quality checks by the participants to ensure compliance and high quality of the CMIP6 data. The conformance with the NetCDF/CF and the additional project metadata rules is automatically checked during ESGF data publication.

6   Conclusion and Perspectives

235 The role of the DDC has changed over the 25 years of its existence in order to adjust to evolving data management standards and altered requirements from IPCC WGI. Furthermore, the tasks of the DDC Partner DKRZ have been adjusted to developments in the CMIP6 infrastructure, which provides the data and documentation for the DDC's Reference Data Archive. DDC's data holdings provide valuable ancillary information for IPCC Assessment Reports. AR6 stands for a major change: On the one hand, the role of the DDC turned from an independent long-term data archive into an active service provider for
240 the IPCC. On the other hand, the IPCC FAIR Guidelines significantly enhanced the transparency of AR6 key findings. Their implementation posed a challenge to all partners: WGI TSU, authors and the DDC Partners. Data usage documentation in AR6 und the long-term archival of related input and final datasets enable the traceability of results and the reuse of datasets but it is still incomplete esp. in the long-term preservation of input datasets. The DDC Partners identified this as part of their gap analysis in 2020:

245 1.   Exhaustive IPCC data archival,

2.   Improved global data access (e.g. compute service for reduction of data transfer volume to support DDC users in developing countries and support for users from various domains),

3.   Data Discovery,

4.   Machine-accessible DDC data,

250 5.   Regional to local data and data services,

6.   Sustaining DDC Partners,

7.   Collaboration with data infrastructure networks, e.g. RDA, WDS or CODATA,

8.   Collaboration with cognate data providers, e.g. IPBES.

Some gaps are easy to target like the collaboration aspects (gaps 7 and 8) and the data discovery issue was solved with the
255 joint DDC catalogue, but the remaining gaps require funding and are related to the missing long-term strategy for the DDC (gap 6).

The first review of the FAIR Guidelines by TG-Data has started, in order to derive recommendations for the AR7 data management and to perpetuate the introduced FAIR data concept. Aspects under discussion include the review of the content of the FAIR Guidelines and their implementations, a better integration of the FAIR Guidelines into the IPCC assessment
260 process and the formulation of a long-term strategy. For operationalizing the FAIR Guidelines' implementation, the integration of recommendations and standards from various infrastructure expert groups is required in order to unite best practices in data management with Good Scientific Practice. Thus, the DDC can be positioned in the developing FAIR digital object research landscape as described in the concept of a FAIR Digital Object Framework by De Smedt et al. (2020).

The IPCC FAIR Guidelines' implementation in the AR6 and in the IPCC WGI Interactive Atlas or even the input4MIPs project
265 publication (Durack et al., 2018) contribute practice examples to the discussions on the integration of data into publications and climate services. Furthermore, the user focus need to be broadened to include the requirements from the impact user community, policy advisors, economics and society into DDC service developments. The importance of data has reached the policymakers. The German Minister of Education and Research, Mrs. Stark-Watzinger, as representative of the German government explicitly mentioned Germany's involvement in the IPCC DDC and the importance of data for the IPCC process
270 at the opening of the 55th Session of the IPCC and 12th Session of WGII on 10 February 2022 (IPCC, 2022).

The DDC lacks a long-term funding strategy. The current DDC Partner funding is provided by few IPCC member states, who request sharing the burden with other IPCC members. Thus, DDC services are periodically at risk of discontinuation. Consolidation of DDC funding is required to ensure sustainable DDC services for the DDC users and for the IPCC WGs and their authors.

275 **Code/Data availability**

No code nor data was created for this study.

Author contribution

MS wrote the manuscript draft and ML reviewed and contributed to the manuscript.

Competing interests

280 The authors declare that they have no conflict of interest.

Acknowledgements

We thank Tim Carter (former co-chair of IPCC TGICA, Finnish Environment Institute SYKE), who generously answered questions and shared unpublished materials about the early days of the IPCC DDC and TGCIA; Anna Pirani (head of IPCC WGI TSU) for the coordination of the formulation and FAIR Guidelines and their implementation in the AR6 WGI; and the
285 DDC Partners, the WG TSUs esp. the WGI TSU colleagues, and the TG-Data members for their contributions to the IPCC FAIR Guidelines.

References

De Smedt, Koenraad, Koureas, Dimitris, and Wittenburg, Peter: FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications 2020, 8(2), 21, MPDI. https://doi.org/10.3390/publications8020021, 2021.

290   Deutsche Forschungsgemeinschaft: Guidelines for Safeguarding Good Research Practice. Code of Conduct, https://doi.org/10.5281/zenodo.3923602, 2019.

Durack, P. J.,Taylor, K. E.,Eyring, V.,Ames, S. K.,Hoang, T.,Nadeau, D.,Doutriaux, C.,Stockhause, M., and Gleckler, P. J.: Toward standardized data sets for climate model experimentation, Eos, 99, https://doi.org/10.1029/2018EO101751, 2018.

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L.-P.,
295   Carvalhais, N., Cionni, I., Cortesi, N., Crezee, B., Davin, E., Davini, P., Debeire, K., de Mora, L., Deser, C., Docquier, D., Earnshaw, P., Ehbrecht, C., Gier, B. K., Gonzalez-Reviriego, N., Goodman, P., Hagemann, S., Hardiman, S., Hassler, B., Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N., Lejeune, Q., Lembo, V., Lovato, T., Lucarini, V., Massonnet, F., Müller, B., Pandde, A., Pérez-Zanón, N., Phillips, A., Predoi, V., Russell, J., Sellar, A., Serva, F., Stacke, T., Swaminathan, R., Torralba, V., Vegas-Regidor, J., von Hardenberg, J., Weigel, K., and Zimmermann, K.: Earth System Model
300   Evaluation Tool (ESMValTool) v2.0 - an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, Geosci. Model Dev., 13, 3383-3438, https://doi.org/10.5194/gmd-13-3383-2020, 2020.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958,
305   https://doi.org/10.5194/gmd-9-1937-2016, 2016.

IPCC: Report of the thirteenth session of the IPCC, Maldives, 22, 25-28 September 1997. https://www.ipcc.ch/site/assets/uploads/2018/05/thirteenth-session-report.pdf, last access: 10 March 2022, 1997.

IPCC: Expert Meeting on the Future of the Task Group on Data and Scenario Support for Impact and Climate Analysis (TGICA), 26-27 January 2016, Geneva, https://www.ipcc.ch/site/assets/uploads/2018/05/EMR_TGICA_Future.pdf, last
310   access: 10 March 2022, 2016.

IPCC: Guidance for the core functions of the IPCC Data Distribution Centre (DDC). https://www.ipcc.ch/site/assets/uploads/2018/12/Guidance_DDC.pdf, last access: 10 March 2022, 2018a.

IPCC: Report of the forty-seventh session of the IPCC, Paris, France, 13 – 16 March 2018. https://www.ipcc.ch/site/assets/uploads/2018/03/final_report_p47.pdf, last access: 10 March 2022, 2018b.

315   IPCC: Expert Meeting on Assessing Climate Information for Regions. Trieste, Italy, f16-18 May 2018. http://www.ipcc.ch/apps/eventmanager/documents/52/120920180323-INF5EM-Regions.pdf, last access: 10 March 2022, 2018c.

IPCC: Working Group I Training on Data Access and Software Development. Oberpfaffenhofen, Germany, 6-7 June 2019. https://www.ipcc.ch/event/wgi-training-on-data-and-software-development/, last access: 10 March 2022, 2019.

320   IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press, 2021.

IPCC: 55th Session of the IPCC, Opening Ceremony, 14th February 2022, https://www.youtube.com/watch?v=00_0udBrw_w,
325   last access: 10 March 2022, 2022.

Cinquini, Luca, Crichton, Daniel, Mattmann, Chris, Harney, John, Shipman, Galen, Wang, Feiyi, Ananthakrishnan, Rachana, Miller, Neill, Denvil, Sebastian, Morgan, Mark, Pobre, Zed, Bell, Gavin M., Doutriaux, Charles, Drach, Robert, Williams, Dean, Kershaw, Philip, Pascoe, Stephen, Gonzalez, Estanislao, Fiore, Sandro, and Schweitzer, Roland: The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data, Future Generation Computer Systems, Volume
330   36, 2014, Pages 400-417, ISSN 0167-739X, https://doi.org/10.1016/j.future.2013.07.002, 2014.

Lin, Dawei, Crabtree, Jonathan, Dillo, Ingrid, Downs, Robert R., Edmunds, Rorie, Giaretta, David, De Giusti, Marisa, L'Hours, Hervé, Hugo, Wim, Jenkyns, Reyna, Khodiyar, Varsha, Martone, Maryann E., Mokrane, Mustapha, Navale, Vivek, Petters,

Jonathan, Sierman, Barbara, Sokolova, Dina V., Stockhause, Martina, and Westbrook, John: The TRUST Principles for digital repositories. Sci Data 7, 144. https://doi.org/10.1038/s41597-020-0486-7, 2020.

335 Pascoe, C., Lawrence, B. N., Guilyardi, E., Juckes, M., and Taylor, K. E.: Documenting numerical experiments in support of the Coupled Model Intercomparison Project Phase 6 (CMIP6), Geosci. Model Dev., 13, 2149–2167, https://doi.org/10.5194/gmd-13-2149-2020, 2020.

Petrie, R., Denvil, S., Ames, S., Levavasseur, G., Fiore, S., Allen, C., Antonio, F., Berger, K., Bretonnière, P.-A., Cinquini, L., Dart, E., Dwarakanath, P., Druken, K., Evans, B., Franchistéguy, L., Gardoll, S., Gerbier, E., Greenslade, M., Hassell, D.,

340 Iwi, A., Juckes, M., Kindermann, S., Lacinski, L., Mirto, M., Nasser, A. B., Nassisi, P., Nienhouse, E., Nikonov, S., Nuzzo, A., Richards, C., Ridzwan, S., Rixen, M., Serradell, K., Snow, K., Stephens, A., Stockhause, M., Vahlenkamp, H., and Wagner, R.: Coordinating an operational data distribution network for CMIP6 data, Geosci. Model Dev., 14, 629–644, https://doi.org/10.5194/gmd-14-629-2021, 2021.

Pirani, Anna, Alegria, Andres, Al Khourdajie, Alaa, Gunawan, Wuwan, Gutiérrez, José Manuel, Holsman, Kirstin, Huard,

345 David, Juckes, Martin Kawamiya, Michio, Klutse, Nana, Krey, Volker, Matthews, Robin, Pascoe, Charlotte, van der Shrier, Gerard, Spinuso, Alessandro, Stockhause, Martina, and Xing, Xiaoshi: The implementation of FAIR data principles in the IPCC assessment process, final version, https://docs.google.com/document/d/1y42IqWx9XtbSpRC0xT7F49-lTmsHPgmXB7ytHZompZM, last access: 10 March 2022.

Stockhause, M.: IPCC Virtual Workspace at DKRZ, http://bit.ly/IPCC_DKRZ_Virtual_Workspace, last access: 10 March

350 2022, 2020.

Stockhause, M., Höck, H., Toussaint, F., and Lautenschlager, M.: Quality assessment concept of the World Data Center for Climate and its application to CMIP5 data, Geosci. Model Dev., 5, 1023-1032, 2012, https://10.5194/gmd-5-1023-2012, 2012.

Stockhause, M., Juckes, M., Chen, R., Moufouma Okia, W., Pirani, A., Waterfield, T., Xing, X. and Edmunds, R.: Data Distribution Centre Support for the IPCC Sixth Assessment. Data Science Journal, 18(1), p.20. http://doi.org/10.5334/dsj-

355 2019-020, 2019.

Stockhause, Martina, Juckes, Martin, Pirani, Anna, Poloczanska, Elvira, and Waterfield, Tim: First IPCC AR6 Data Workshop Minutes. Zenodo. https://doi.org/10.5281/zenodo.1036460, 2017.

Stockhause, M. and Lautenschlager, M.: CMIP6 Data Citation of Evolving Data. Data Science Journal, 16, p.30, http://doi.org/10.5334/dsj-2017-030, 2017.

360 Stockhause, M., Toussaint, F., and Lautenschlager, M.: CMIP6 Data Citation and Long-Term Archival. Zenodo. https://doi.org/10.5281/zenodo.35178, 2015.

Taylor, Karl E., Juckes, Martin, Balaji, V., Cinquini, Luca, Denvil, Sébastien, Durack, Paul J., Elkington, Mark, Guilyardi, Eric, Kharin, Slava, Lautenschlager, Michael, Lawrence, Bryan, Nadeau, Denis, and Stockhause, Martina: CMIP6 Global Attributes, DRS, Filenames, Directory Structure, and CV's. 10 September 2018 (v6.2.7). https://goo.gl/v1drZl, last access: 10

365 March 2022, 2018.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93(4), 485-498. https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

World Data System: Data Repositories Day 2018, Gaborone, Botswana, 9 November 2018. https://www.worlddatasystem.org/community/wds-members-forum/data-repositories-day-2018, last access: 10 March 2022,

370 2018.

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, Boiten, Jan-Willem, da Silva Santos, Luiz Bonino, Bourne, Philip E., Bouwman, Jildau, Brookes, Anthony J., Clark, Tim, Crosas, Mercè, Dillo, Ingrid, Dumon, Olivier, Edmunds, Scott, Evelo, Chris T., Finkers, Richard, Gonzalez-Beltran, Alejandra, Gray, Alasdair J.G., Groth, Paul, Goble, Carole, Grethe, Jeffrey S., Heringa, Jaap, 't Hoen, Peter A.C, Hooft, Rob, Kuhn, Tobias,

375 Kok, Ruben., Kok, Joost, Lusher, Scott J., Martone, Maryann E., Mons, Albert, Packer, Abel L., Persson, Bengt, Rocca-Serra,

Philippe, Roos, Marco, van Schaik, Rene, Sansone, Susanna-Assunta, Schultes, Erik, Sengstag, Thierry, Slater, Ted, Strawn, George, Swertz, Morris A., Thompson, Mark, van der Lei, Johan, van Mulligen, Erik, Velterop, Jan, Waagmeester, Andra, Wittenburg, Peter, Wolstencroft, Katherine, Zhao, Jun, and Mons, Barend: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18, 2016

380   Williams, D. N., Balaji, V., Cinquini, L., Denvil, S., Duffy, D., Evans, B., Ferraro, R., Hansen, R., Lautenschlager, M., and Trenham, C.: A Global Repository for Planet-Sized Experiments and Observations, Bulletin of the American Meteorological Society, 97(5), 803-816. https://doi.org/10.1175/BAMS-D-15-00132.1, 2016.

Xing, Xiaoshi, Stockhause, Martina, Gutiérrez Llorente, José Manuel, and Irwin, Courtney: Memorandum of Understanding (MoU) for Operation of the IPCC Data Distribution Centre. Zenodo. https://doi.org/10.5281/zenodo.5914482, 2021.