

FIFTY-SEVENTH SESSION OF THE IPCC
Geneva, Switzerland, 27 – 30 September 2022

IPCC-LVII/INF. 7
(25.VIII.2022)
Agenda Item: 7.6
ENGLISH ONLY

PROGRESS REPORTS

Task Group on Data Support for Climate Change Assessments

(Prepared by the Co-Chairs of the Task Group on Data Support for Climate Change Assessments)

(Submitted by the Secretary of the IPCC)

PROGRESS REPORTS

Task Group on Data Support for Climate Change Assessments

This report describes the activities undertaken by the Task Group on Data Support for Climate Change Assessments (TG-Data) since the last update presented during the 53rd (bis) Session of the IPCC in March 2021.

1. Salient activities since last report

Meetings

- Seventh TG-Data teleconference was held on 1st June 2021.
- Third (virtual) annual face to face (F2F) TG-Data meeting was held on 22nd October 2021.
- Eighth TG-Data teleconference was held on 5th April 2022.
- Ninth TG-Data teleconference was held on 27th July 2022.
- Monthly DDC managers meetings

DDC membership

The United Kingdom's Department for Business, Energy & Industrial Strategy has selected MetaDataWorks (MDW) through an open bidding competition as the purveyor of the UK's contribution to the Data Distribution Centre (DDC). MDW replaces the Center for Environmental Data Analysis (CEDA) who previously played this role, but CEDA remains involved in the data curation process through an agreement with Working Group I (WGI) Technical Support Unit (TSU) and informal discussions.

MDW was officially included in the DDC through a motion passed during the virtual F2F meeting on 22 October 2021. A new version of the Memorandum of Understanding (MoU) for Operation of the IPCC Data Distribution Centre was presented and adopted in the F2F meeting. The MoU was then signed by the institutional heads of all current DDC partners and linked at the new DDC website (<https://www.ipcc-data.org/about>). The MoU was also published at Zenodo with a DOI: <https://doi.org/10.5281/zenodo.5914483>. Outreach activities

TG-Data together with Working Group I (WGI) have prepared the first formal outreach activity organized for the Sixth Assessment Report (AR6). This first package of outreach activities is linked to the outcomes of the [WGI report](#) that was released in August 2021. The objectives of these activities are:

- Present the main results of the WGI report.
- Engage with the regional research practitioner community over specific regional domains.
- Present the data availability with special emphasis on the Interactive Atlas (IA) providing regional information and synthesis (from the Technical Summary and the regional chapters 10-11-12-Atlas Chapters).

Considering the current pandemic context, the training workshops are being organized in a virtual mode in the form of regional webinars. A package of introductory material (videos) building on previous events (in particular the [COP-26 IPCC regional events](#) and the [official launch of the Interactive Atlas](#)) is provided in advance as context information to set the ground for the webinar. The webinars are intended for practitioners and the research community (but also open to consultants and public sector representatives) and focus on data aspects using the facilities of the IA. These events are envisioned as interactive sessions and require registration, although they could be also followed by streaming.

As a first package, three regional events are organized in Europe, Central and South America and Africa. This is the webpage where the webinars are announced: <https://www.ipcc.ch/event/interactive-atlas-regional-webinars/>

The [European Webinar](#) took place on Friday, 18 March 2022, 11:00 - 12:00 UTC, attracting over 600 registrants. A total of 350 Europeans were invited to the Zoom meeting, and the rest were redirected to a live streaming channel. Around 150 persons attended the webinar, with four Breakout Groups, two in English (~100 and ~20); one in Spanish (with ~30) and one in French (with ~10). The [recorded video](#) has been viewed ~600 times.

The [Central and South America Webinar](#) took place on Friday, 29 April 2022, 13:00 - 14:30 UTC, attracting 690 registrants. A total of 135 persons attended the webinar, with 6 Breakout Groups, one in English, four in Spanish and one in Portuguese.

Webinars for other regions are in preparation, see [Next steps](#).

A [presentation](#) about the new DDC catalog by Adam Milward and Sabine Undorf (MDW) entitled "Climate Change Data for Everyone: Building the Catalogue for the Intergovernmental Panel on Climate Change Data Distribution Centre" was given on 17 December 2021 at the AGU 2021 Fall meeting in the session "N53A: Connecting Disciplines and Data in Earth and Environmental Synthesis Research: Enabling International and Interdisciplinary Data Discovery, Integration, and Reuse".

Outreach activities for **Working Group III (WGIII) report** are under planning. The objectives of these activities are to:

- Engage with the regional research practitioner community over specific regional domains.
- Present the main results of the WGIII report.
- Present the data availability with special emphasis on the Scenario Explorer developed by IIASA for IPCC WGIII.

Based on the experience of WGI outreach activities, it is expected to hold the training workshops for 2022 (second semester around November) in a virtual mode. The workshops are expected to last no more than 2 hours, with some general introductions and presentation of general data relevant information about WGIII-AR6 and hands-on activities. Although the workshops will not be region specific but global in spirit, it is expected that due to time zone differences the workshops are deployed in a series of events for different regions of the world. It is expected to have a mixed audience between practitioners and researchers. It is expected that the audience will have some basic technical skills.

FAIR Guidelines

A document describing recommended FAIR¹ practices for IPCC authors has been completed and adopted by TG-Data members at the 8th teleconference on 5 April 2022.

This document introduces the implementation of FAIR data principles into the IPCC process and reflects the four elements of FAIR to find the data, produce and reproduce figures and, finally, to document the provenance for reusability. It presents standard (basic) measures that are recommended for all digital data that is assessed, intermediate measures that achieve reproducibility of assessed digital information, for example through the use of collaborative platforms for figure development, and also full (advanced) measures to achieve reusability of digital products with complete provenance documentation. The document has been copy-edited by the IPCC Secretariat and published on Zenodo: [The implementation of FAIR data principles in the IPCC AR6 assessment process \(V1\)](#).

¹ Findable, accessible, interoperable and reusable

Licensing guidelines

A second document presented at the 9th teleconference on 27 July 2022 outlines TG-Data recommendations regarding data licensing. Licenses describe the terms under which data can be used and shared. Some licenses restrict usage to non-commercial purposes, require derivative works to be published under the same license as the original material, prohibit redistribution, etc. TG-Data licensing guidelines recommend publishing IPCC final datasets under the CC BY 4.0 license, which gives data users the permission to share and modify the data, as long as the original authors are given credit and changes to the original data are indicated. The document will be formally adopted and published over the coming weeks.

Data curation activities

IPCC Data Distribution Centres (DDC) at DKRZ² (Germany), CIESIN³ (USA), MetaDataWorks & CEDA⁴ (UK) and CSIC-IFCA⁵ (Spain) are working with WG TSUs to curate input, intermediate and final datasets that are used in the assessment report (AR). These datasets are archived across the different DDCs, and cataloged on the new DDC website developed by MetaDataWorks (<https://www.ipcc-data.org/>).

The DDC data catalog is built on a new metadata specification (schema) clearly defining the expected fields and their definition. See [documentation](#) for more details. This specification is the result of cross-WG consultations led by MDW with TSUs and DDCs.

Recently, IPCC authors and contributors have made relevant datasets, code libraries and publications available through an online data curation service called Zenodo. An [IPCC AR6](#) Zenodo community was created to collect AR6 relevant datasets published on the platform. As of July 2022, the community links to five AR6 datasets.

Final data

Final data sets refer to data directly underlying graphics included in the AR. As of 26th July 2022, the DDC catalog has the following datasets. Additional datasets are expected within the next few months. (See Table 1).

Table 1 Status of final dataset curation across AR6 WGs as of August 2nd, 2022. Source: Jyoti Rogers, MetaDataWorks.

Group	Live	Upcoming
AR6 WGI	21	1
AR6 WGII	14	14
AR6 WGII related SROCC	6	0
AR6 WGIII	29	12
AR6 SYR	0	5

² Deutsches Klimarechenzentrum

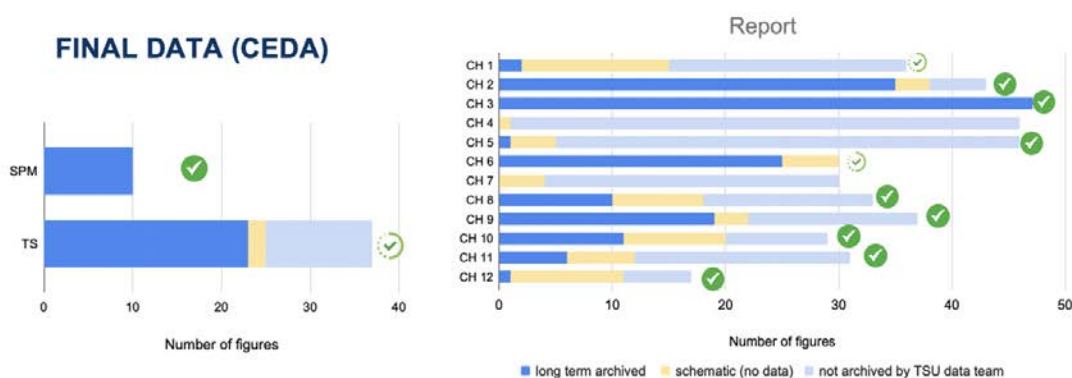
³ Center for International Earth Science Information Network

⁴ Centre for Environmental Data Analysis

⁵ Consejo Superior de Investigaciones Científicas - Instituto de Física de Cantabria

WGI TSU is in the process of working with CEDA to assemble, quality control and curate and catalog final data from chapter and Technical Summary Figures. The current status is summarized in Figure 1 where dark blue shows what is being curated, yellow shows figures that are not data-based (e.g. conceptual figures or schematics) and light blue is where data is currently not available to the TSU. Efforts are underway to mobilize data that is currently not available.

Figure 1 Status of curation of WGI final data by WGI TSU and CEDA as of 25 July 2022. Source: Lina Sitz, WGI TSU



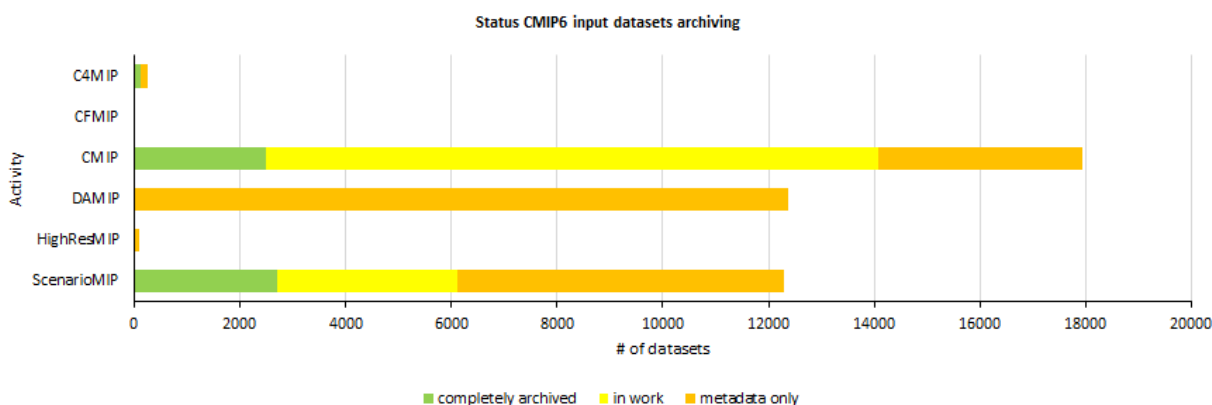
Input data

Input data refer to source material used by IPCC authors. The focus of DDC is to archive data relied on by IPCC that are not already curated in other trusted repositories. The DDC catalog has been populated with input datasets from previous ARs, 1,646 hosted at DKRZ and 9 hosted at CIESIN.

For AR6, CIESIN has now archived the [INFORM Global Risk Index](#) dataset used in WGII. Note that WGII also uses the [World Risk Index - Vulnerability Component 2019](#), which has been published by its authors on the Zenodo platform and will be included in the DDC catalog. Similarly, a link to the [AR6 Scenarios Database](#) has been published to Zenodo, and discussions are underway to preserve a dark snapshot of the version used in the AR6 report at the DDC.

The archival of CMIP6⁶ data at DKRZ is ongoing (see Figure 2 for the status) but is hindered by a number of challenges: incomplete list of data from authors, missing usage information, errors in dataset identifiers, reference to dataset versions that have been “unpublished” and replaced by newer versions, among other issues. TG-Data will make recommendations to improve the process for AR7.

Figure 2 Status of CMIP6 input dataset curation as of August 1st, 2022. Source: Beate Krüss/Martina Stockhause, DKRZ



⁶ Coupled Model Intercomparison Project Phase 6

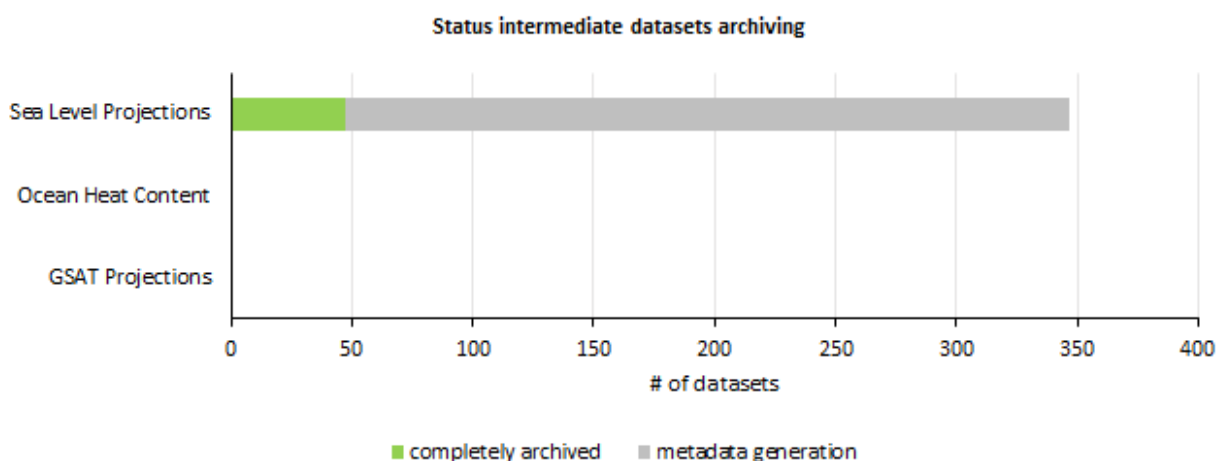
First Assessment Report (FAR) input model data at DKRZ from GISS was processed in June 2021 to create an ASCII version of the original data available in native data format. This curation measure eases data reuse by DDC users. Two datasets and 4 GBytes of data were added to the FAR Reference Data Archive, which now has a total volume of 10 GBytes.

Intermediate data

Intermediate data refer to processed data that is generated in the production of the IPCC report. A new workflow has been developed by WGI TSU and DKRZ for the curation of intermediate data. Work is underway to archive intermediate data assessed in WGI for constrained sea level, ocean heat content and global surface temperature projections (see Figure 3 for the status).

Intermediate datasets used in the production of the WGI Interactive Atlas are curated by CSIC-IFCA DDC.

Figure 3 Status of intermediate dataset curation as of August 1st, 2022. Source: Beate Krüss/Martina Stockhause, DKRZ



Code archival

WGI TSU is working on the collection and archival of code libraries developed by authors for the preparation of WGI report figures. The libraries are assembled on the [WGI GitHub repository](#) and published after author and TSU quality control checks, and on completion of documentation to guide users on the resources available. The status is shown in Figure 4 from the assembly of code to publication and the assignment of digital object identifiers (DOIs) via Zenodo.

Figure 4 Status of code archival of scripts underpinning WGI figures. Source: Diego Cammarano, WGI TSU

Chapter	Code on IPCC WGI Github	Dedicated Figure Github Repository	DOIs
1	⊗	⊗	⊗
2	✓	✓	✓
3	✓	✓	✓
4	🕒	🕒	🕒
5	⊗	⊗	⊗
6	✓	✓	🕒
7	✓	🕒	🕒
8	✓	✓	🕒
9	✓	✓	🕒
10	✓	✓	🕒
11	✓	🕒	🕒
12	✓	🕒	🕒
Atlas	✓	✓	✓
TS	🕒	🕒	🕒
SPM	⊗	⊗	⊗

2. Lessons learned on FAIR implementation for WGI

WGI TSU and DDC team members are in the process of taking stock of the work done so far to implement FAIR principles as part of the WGI AR6. The following is a summary of the key challenges encountered that need to be addressed to develop improved guidance for AR7:

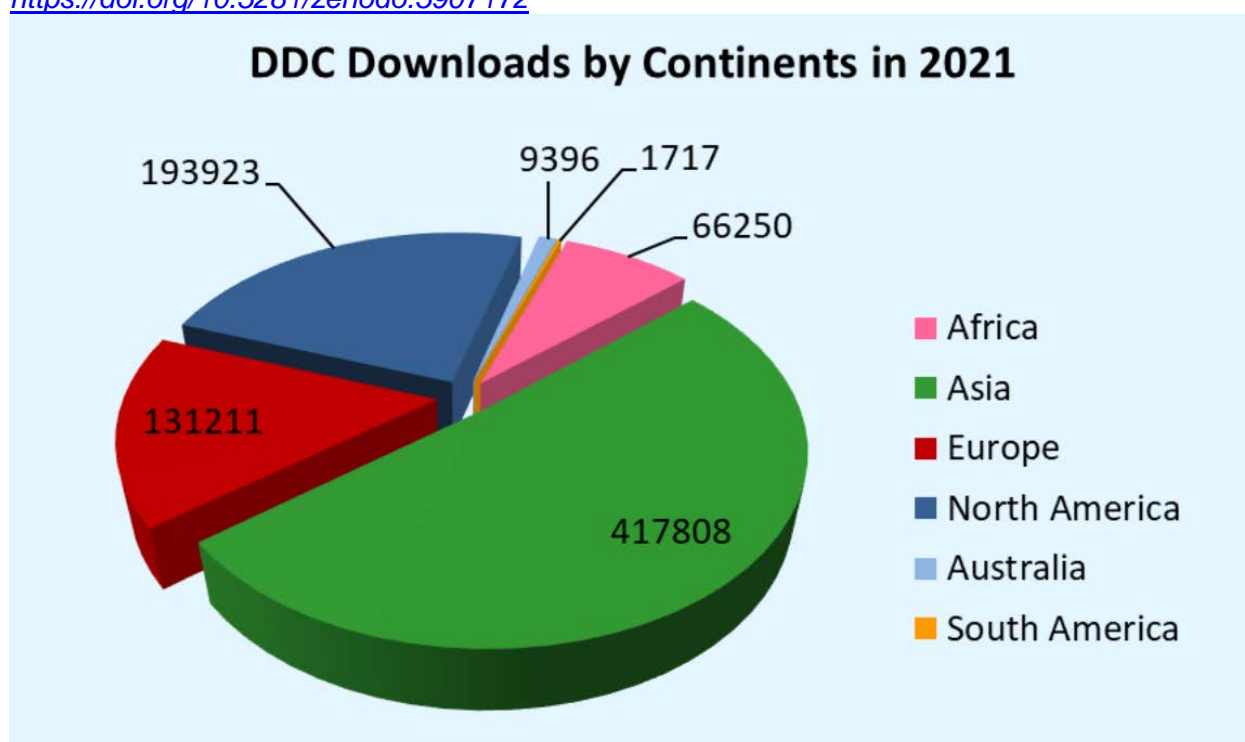
1. Novelty of the implementation of FAIR data principles - new roles, workflows for TSU and DDC.
2. Challenges faced by authors of contributing to a process put in place during the assessment.
3. Missing datasets, code and metadata.
4. Quality control process.
5. Ensuring that licensing constraints are met.
6. Challenges for established curation practices.
7. Interconnecting archived/curated data and code archived to the report.
8. Implementation of IPCC Error Protocol for data and code.
9. Time pressure towards the end of assessment.
10. Awareness building and guidance on what resources are available and for their proper use.

Lessons learned from the AR6 experience and recommendations for AR7 are being developed by the TSU and DDC representatives (Pirani et al., 2022, <https://zenodo.org/record/6992173>). To facilitate the implementation of FAIR principles, these need to be adopted as a core practice within the assessment process from the start of the AR7. Authors need to be provided with clear instructions on mandatory requirements, templates accompanied by guidance and online tools to assemble information, with a clear timeline. Dedicated TSU staff is needed to support authors and work with the DDC. Dedicated DDC resources are needed to staff this work for the duration of the cycle. Attention to resources for the end of the cycle timeline is needed since a substantial part of the archival and curation work will be completed, including fully interconnecting the code and data products with the report itself. Oversight and guidance is needed throughout the process. TG-Data has a role as a persistent group here, as well as Working Group Bureau members.

3. Data Access

In 2021, 820,000 datasets adding up to 420 Tbytes were downloaded from the DKRZ archive. Ninety percent of these downloads were made through the Earth System Grid Federation (ESGF) access point. The geographic distribution of downloads per continent is shown in Figure 5. The annual DDC at DKRZ report provides additional information, including time series of downloads and a breakdown of downloads per AR cycle.

Figure 5 Number of dataset downloads from DKRZ long-term archive (download counts) per continent for 2021. Source: Stockhause, Martina (2022). Report 2021 of the DDC at DKRZ. Available at <https://doi.org/10.5281/zenodo.5907172>



4. Website

The DDC website (<https://www.ipcc-data.org/>) has undergone a complete overhaul, and the new version has been live since November 1st and has attracted 4,200 visitors up to 31 March 2022. The previous website has been archived and its content is still available. The site content has been reviewed by TG-Data members and adopted during the 2021 F2F virtual meeting.

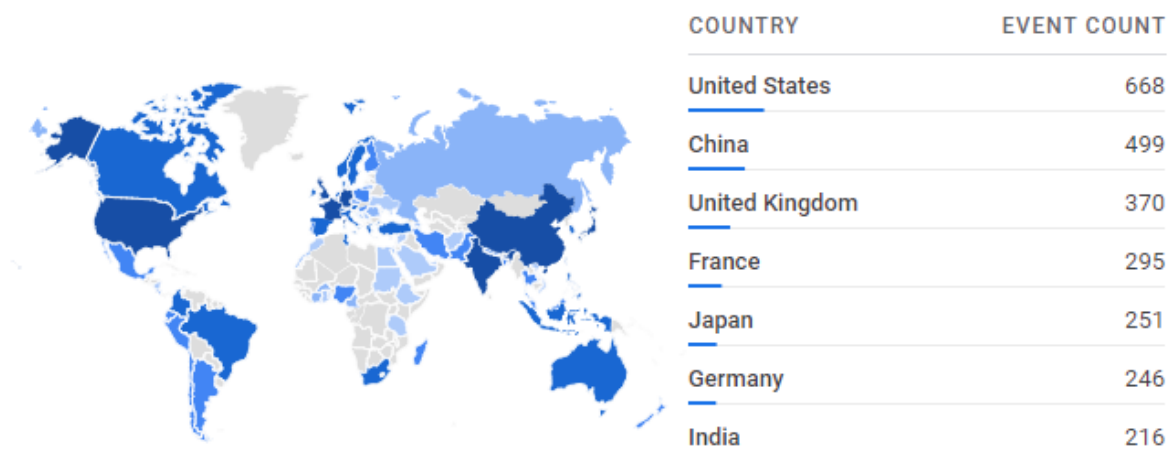
A noticeable spike in traffic has been driven by the publication of WGII's report and associated final datasets on 28 February 2022. A total of 4,200 files were downloaded from the catalog interface (Figure 6), and their geographical distribution is illustrated in Figure 7.

Figure 6 Dataset downloads from DDC website since launch of DDC data catalog (1 November 2021 to 30 June 2022). Source: Jyoti Rogers, MetaDataWorks.



Figure 7 Number of dataset downloads from DDC website per country from 1 November 2021 to 30 June 2022. Source: Jyoti Rogers, MetaDataWorks.

Event count by Country



Since last August, MetaDataWorks responded and resolved 24 service desk tickets (issues or questions from web site visitors).

On 23 March 2022 a new landing page for [AR6 data](#) was added to the web site, describing and linking key input, intermediate and final datasets for each WG.

5. DDC anniversary

The DDC will celebrate its 25th anniversary during IPCC-57 as it was established on IPCC-13 (25-28 September 1997). As the last remaining DDC Partner of the original group, DKRZ has published a review of its work (Stockhause and Lautenschlager, 2022, <https://doi.org/10.5194/gmd-15-6047-2022>).

6. Next steps

- Outreach activities:
 - Hold regional TG Data WGI Interactive Atlas Webinar in Africa (September 13th)
 - Organize other regional TG Data IA Webinars (Oceania, Asia, Middle East, North America)
 - TG Data WGIII Workshops will be organized after COP27 based on the Scenario Explorer developed by IIASA for IPCC WG3.
- Data related activities:
 - Review metadata schema for the DDC catalog based on experience so far.
 - Pursue data curation activities with all TSUs, including SYR.
 - Develop data citation guidelines document for the DDC
- TG Data organization
 - Planning for in-person 2022 face to face TG-Data meeting occurring on October 12-14 in Yokohama.
 - Pursue discussions on the AR6 experience and formulate recommendations for AR7.
- DDC funding:
 - DDC is not centrally funded and informal funding from the US, UK and Spain currently allows partners to deliver core data service for the IPCC. DKRZ received funding from Germany for past cycles but relies on institutional and European project funds for AR6.
 - The future of the DDC is in jeopardy as two of the four partners, DKRZ (Germany) and MetadataWorks (UK), are not funded post-March 2023
 - DDC is creating an options paper that outlines three different options for IPCC data distribution, including the benefits, activities and funding that would be required to support each option
 - DDC is looking to present the options paper to funders to establish a sustainable funding model post March 2023