

**FIFTY-THIRD (bis) SESSION OF THE IPCC
Electronic Session, 22 – 26 March 2021**

IPCC-LIII(bis)/INF. 6, Rev.3¹
(24.III.2021)
Agenda Item: 4.6
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)

¹ The revised version is being submitted to clarify under DDC-related matters that “DKRZ has a core funding of about 0.5 FTE. So far, funding to continue the large-scale data archival of CMIP6 has not been confirmed” and to include “Country or Region” in the table in Annex 5, in accordance with comments and requests submitted by governments.

PROGRESS REPORTS

Task Group on Data Support for Climate Change Assessments

1. Introduction

The Task Group on Data Support for Climate Change Assessments (TG-Data) presented a progress at the 58th Session of the IPCC Bureau (BUR-58, Electronic Session, 19 – 20 May 2020). The report contained an update on the TG-Data membership and salient activities of the Task Group and its Subgroups.

The TG-Data presented another progress report at the 59th Session of the IPCC Bureau (BUR-59, Online Session, 13 – 14 October 2020). The report covered certain procedural matters, reports of TG-Data Subgroups and budgetary implications of some TG-Data and Data Distribution Centre (DDC) activities. Some of these matters are summarized below.

The TG-Data submitted another progress report for the 60th Session of the IPCC Bureau (BUR-60, Electronic Session, 16 – 18 March 2021). The report presented discussions and outcomes of the TG-Data virtual meetings which were held on 6th and 19th October 2020, and 11th November 2020, TG-Data recommendations on IPCC Sixth Assessment Report (AR6) licensing and 2020 statistics on DDC usage. Some of these issues are presented in subsequent subsections of this report.

2. Salient TG-Data Activities

2.1 Meetings

- a) TG-Data held three teleconferences: Fourth Teleconference on 27th May 2020; Fifth Teleconference on 4th September 2020; and Sixth Teleconference on 10th February 2021.
- b) The Task Group held virtual meetings on 6th and 19th October 2020, and 11th November 2020. These meetings featured three plenary sessions and Sub-group meetings. The objectives for each one of the plenary sessions were defined as:
 - Plenary meeting #1 - Current state and objectives
 - Plenary meeting #2 - Discussion on subgroups contentious issues
 - Plenary meeting #3 - Adoption of resolutions
- c) Sub-group meetings were organized to address the following main issues:
 - Dataset prioritization Sub-group
 - i) Provide DDC a list of priority datasets to archive
 - FAIR Sub-group
 - i) Unified guidance document for all Working Groups (WGs) + Working Group (WG) specific appendices / details
 - ii) Trial run of the final data archiving at CEDA to estimate workload and potential obstacles (resources, etc)
 - Outreach & Webpages Subgroup
 - i) Migration of TG-Data to new web site - roadmap, list of tasks
 - ii) Training workshop - topics (Atlas, etc), budget, format
 - Partnerships Subgroup
 - i) Clarify expectations to become part of DDC

- DDC Managers Subgroup
 - i) Evaluate resources for input/source dataset archival from dataset prioritization
 - ii) Evaluate resources for final data archival based on trial run
 - iii) Identify short- and mid-term roadblocks and mitigation strategies
 - iv) Report on DDC usage statistics

- d) Some issues which were raised during the virtual meetings, and outcomes included:
 - Advancing the process towards the implementation of the FAIR² Guidelines across the three IPCC Working Groups. The near-final draft of the guidance document was circulated to TG-Data and Ex-Officio members for their final endorsement.

 - Regarding data prioritization:
 - ✓ A data archival prioritization was agreed, with the highest priority for final datasets, second priority for input datasets, and third for intermediate datasets.
 - ✓ For the input datasets of Working Group I (WGI), the prioritization for the archival at DKRZ³ was agreed as 1. CMIP6⁴ datasets, 2. datasets specified by the IPCC authors, and 3. CORDEX⁵ datasets.
 - ✓ The CMIP6 dataset list details were under discussion and have been finalized after the face-to-face meeting.
 - ✓ A timetable was developed for the transfer of data from the Technical Support Unit (TSU) to the DDC Partners to give an overview of the expected flow of information and to enable the DDC Partners to schedule their work accordingly; and
 - ✓ Trial runs were carried out to determine the feasibility of archiving data underlying all the targeted figures. A decision would subsequently be taken on the appropriate number of figures or determine the need to mobilize additional resources to support the work of archiving data underlying all the targeted figures at CEDA⁶.
 - ✓ The proposal for CIESIN⁷ to curate the INFORM datasets was endorsed by TG-Data and therefore considered an official request by the Task Group.

 - On matters related to the FAIR principles:
 - ✓ **Archival and documentation of data and code from the WGI report - update from the WGI TSU**
 - *Data and Code Archival*
 - In collaboration with the CEDA (the UK member of the DDC), the WGI TSU has now finalized the procedures and related software tools for the archival of data plotted in WGI report figures, as well as related scripts. The WGI TSU has kept the WGII and III TSUs informed of progress throughout the development phase. A webinar was held on February 8th 2021 (see below links), outlining the steps that those creating IPCC figures are required to perform in order for the TSU and DDC to archive the related data and code. The process is similar for both data and code, involving 1) Creation of the data files and code. 2) Upload of these files to the DMS (or alternatively GitHub in the case of code). 3) Provision of accompanying metadata using a custom-built online database system created by the WGI TSU - Figure Manager. Following these steps, the TSU and CEDA then undertake a series of checks on the data, code and metadata provided, as a form of quality control. Data and code from the WGI report will be archived from now until the end of September on a rolling basis, as the data/script files and corresponding metadata are made available to the WGI TSU.

² findable, accessible, interoperable, reusable

³ Deutsches Klimarechenzentrum, Germany

⁴ Coupled Model Intercomparison Project Phase6

⁵ Coordinated Regional Downscaling Experiment

⁶ Centre for Environmental Data Analysis, UK

⁷ The Center for International Earth Science Information Network, USA

- Webinar video recording:
https://drive.google.com/file/d/1yWlyPuyV_4Vn7wR_n7ROU-WlgtnuAhmu/view?usp=sharing
- Webinar Slides:
https://drive.google.com/file/d/1knThtsPD_vnLITYEg_Ktfr3YTKcFOAcn/view?usp=sharing
- Input Data Tables
 - As a further contribution to the IPCC's efforts in implementing FAIR principles for data, the WGI TSU have designed tables to document the input datasets used in the WGI report. These 'input data tables' provide information on a figure-by-figure basis, including direct links to the data files, and citations for this data, as well as for related publications. They will be provided as chapter Supplementary Material (online PDFs). For efficiency in preparing metadata on the model data from MIPs used in figures, the TSU has also designed a text file through which detailed information on CMIP6, CMIP6 and CORDEX data can be provided, including the specific version used. In addition to using these 'model metadata files' to fill in information on model data in the input data tables, they will also be used by DKRZ (the German member of the DDC) to undertake the archival of CMIP6 datasets used in the WGI report. A webinar was held on February 12th 2021, reviewing material previously presented on this topic by the WGI TSU in summer 2020.
- On DDC-related matters:
 - ✓ CEDA have submitted a bid to an open competition to be re-appointed as the supplier of the UK component of the DDC.
 - ✓ CIESIN is funded up to July 2021 with a possibility for the funding to be extended afterwards.
 - ✓ DKRZ has a core funding of about 0.5 FTE. So far, funding to continue the large-scale data archival of CMIP6 has not been confirmed.
 - ✓ Clarify the scope of one of the service gaps (i.e. the requirement for a service to provide support for regional data) is needed. There needs to be clarity on how that fits within the existing DDC Memorandum of Understanding in order to make it easy for people who are engaged in discussing with funders to work out what can be done as an additional contribution.
 - ✓ DDCs prepared reports on usage statistics on data downloads (Appendix 4) and the DDC web page (Appendix 5). These reports discuss the time evolution and geographical origin of page views and data downloads.
- On partnerships issues:
 - ✓ There were conversations with the Australian and Japanese government representatives to address how these governments could lend support for activities related to regional data;
 - ✓ The Partnerships Sub-group identified a need to develop a guidance document to cover the logistics of contracting, such as the nature and duration of partnerships. There are uncertainties pertaining to the status of potential new partners (i.e. if they will be directly linked to TG-Data or to the IPCC or to the existing DDCs).
- Licensing of IPCC products:
 - ✓ It was noted that the CMIP6 data license is potentially incompatible with the license of the IPCC report. CMIP6 model outputs have a Share-Alike license, meaning that derived products should be licensed under the same license. This could potentially conflict with the IPCC's own license for its reports. Therefore, some legal issues need to be addressed, which however go beyond the remit of the Sub-groups. Consequently it was recommended that TG-Data informs the IPCC Bureau about the possible incompatibility between the current license of the IPCC products and the implementation of FAIR

principles, and request the Bureau for advice on how to move forward in considering whether or not a change in license is warranted and how to deal with the legal aspects. More elaborate TG-Data recommendations on IPCC licensing issues are presented below.

2.2 TG-Data Activities with Budgetary Implications

At the 52nd Session of the IPCC, the TG-Data presented its workplan (Document [IPCC-LII/Doc. 8, Rev. 1](#)). The Panel approved the TG-Data workplan for the period 2019-2023 (Decision [IPCC-LII-5](#)). Furthermore, the Panel requested the TG-Data Co-Chairs to seek guidance from the Secretariat in mobilizing resources, including in-kind contributions, for their activities and for the Data Distribution Centre (DDC), and to provide the Secretariat with cost estimates for the activities outlined in their workplan (paragraph 17 of Decision IPCC-LII-9).

TG-Data presented its preliminary budget at the 59th Session of the IPCC Bureau (BUR-59) showing two activities which would likely require human resources and funding. One is holding outreach activities (training workshops on access to IPCC-curated data), and the other is ensuring IPCC data follows FAIR principles. It was indicated that the estimates presented at the 50th Session of the IPCC Bureau (BUR-50) were preliminary and would be refined during the virtual meeting.

With regards to the outreach training activities during the virtual meeting it was decided that after the WGI report is released (after September) there will be training activities related to dissemination of the tool and resources associated with the WGI Interactive Atlas Chapter. The authors of the chapter will oversee the preparation of the material and later organization of the training sessions with guidance and aid from TG-Data. The training sessions will be organized in developing countries in 5 regions (South America / Central America / Africa / Asia I / Asia II) for which there are representatives among IPCC Chapter authors. It was also decided that due to the SARS-CoV-2 pandemics the training sessions will be held online. The resources associated with the preparation of the training session are in-kind support estimated at 1 Full-time equivalent (FTE) from the IPCC Chapter authors, logistic support from WGI Technical Support Unit (TSU), and expenses for documentation translation, copy-editing and graphic design. For years 2022 and 2023 the format for outreach activities will be decided according to the evolution of the pandemic.

With IPCC final data to become publicly available, it would be worth improving the cataloguing capability of the DDC website and modernize its overall design. This would be part of the “Keep TG-Data and DDC web pages up to date” item of the work plan. The costs for these web site improvements are unclear and TG-Data will need support from the Secretariat to navigate these questions.

The refined budget is as shown in the table below.

| Activity | 2021 | 2022 | 2023 |
|--|-------------|-------------|-------------|
| Outreach – Atlas training preparation (in-kind) | 1FTE | TBD | TBD |
| Outreach – Atlas training documentation editing, translations | 3k | 5k | |
| FAIR data (in-kind from current and new partnerships) | 6 FTE | 7 FTE | 5 FTE |
| DDC data catalog | 20k | 15k | |
| DDC design update | 25k | 25k | |

Note that the resources identified for FAIR data aggregate resources required for multiple data-related tasks described in the following table. The activities aim to fill gaps in the current DDC capacity to support the IPCC. It defines quantified blocks of work which could be done by making more resource available to existing partners or through inclusion of new partners, as well as the potential impacts of this work. It includes a rough estimate of the effort required across DDCs and requirements in terms of the participation of partner institutions. In this sense, the budget above could be met with in-kind contribution from partner organizations.

| Title & description | Impact | Partnership (P) vs. Funding (F) | Effort |
|---|---|---|---|
| <p>G1 Developing an API to DDC data A single Application Programming Interface across all the DDC would allow user applications to browse DDC data from a unified interface. The API would follow existing standards to ensure interoperability with other data providers.</p> | <p>Enhanced visibility and accessibility of data. It will be easier for users to find and access resources.</p> | <p>P & D: Requires work in the existing archives and in a new user interface.</p> | <p>3 months effort for scoping study + implementation (TBD). CIESIN: 25-50% FTE</p> |
| <p>G2 Discussion group on environmental data informing global policy Discussion groups of organizations such as RDA, WDS or CODATA bring together technical expertise from many institutions tackling related aspects of management of scientific data. The objective would be to bring together people from organizations such as IPCC and IPBES to discuss the specific issues around implementing FAIR data in the context of supporting scientific assessments informing policy.</p> | <p>Greater clarity about the challenges and create opportunities to learn from best practices in other areas.</p> | <p>P & F: Staff time to organise meetings, moderate communications among members, manage documents etc.</p> | <p>20% FTE</p> |
| <p>G3 Engagement in RDA and WDS technical groups There are existing discussion groups in the RDA and WDS addressing issues of relevance to the DDC. The DDC could benefit by raising issues in these venues. The partner institutions have, in some cases, representatives -- but the DDC managers lack the capacity to engage with the process.</p> | <p>International expertise would feed into ongoing decisions on services and policy.</p> | <p>F: This task needs someone who has a detailed working knowledge of the DDC.</p> | <p>20% FTE (1 day/week) CIESIN: Covered</p> |
| <p>G4 Global data access Work to improve data access for users from countries outside the Annex 1 list of the UNFCCC. More generally, work to provide and enhance access to countries which lack the infrastructure and resources of the current DDC partners. The systems and processes provided at existing partner institutions are guided by the needs and expectations of users in the UK, Germany and US.</p> | <p>Improved accessibility of data for users in developing countries.</p> | <p>P: There would be considerable benefit in having this activity led from an institution which is firmly set within the target user group.</p> | <p>1 FTE upwards SEDAC: 25%FTE</p> |

| | | | |
|---|--|--|---|
| <p>G5 Sustaining partners The services provided by the existing partners provide the core of the DDC. There are two aspects of this: (1) the knowledge and experience which has been accumulated over years and (2) the ability to provide the service in the future. It is well recognised that data management requires long term engagement, but the specifics of national funding often result in short term funding decisions. The route to addressing the problem is unclear: enhanced visibility will help.</p> | Improved stability of DDC services. | F | ~1FTE per partner CIESIN: Covered by 5-year SEDAC contract |
| <p>G6 Exhaustive IPCC data archival DDC archive datasets according to priorities set by the TG-Data prioritization subgroup. It is unlikely that current DDC resources will suffice to archive all key datasets used in all three WGs. Furthermore, the diversity in datasets (data volume, metadata quality, legal agreements, provider collaboration) makes it extremely difficult to reliably estimate the effort and resources required to archive the full IPCC dataset collection. Ideally, DDC would have access to flexible, scalable resources to adapt to this uncertainty.</p> | Larger set of datasets available through DDC and its partners, meaning assessed data is securely archived in long-term archives. | P & F: Data providers should meet CoreTrustSeal criteria to avoid duplicating datasets. Implies coordination with DDC to agree on common API for access, citation, catalog, etc. | 80% FTE for for intermediate data generated by authors. 60% FTE for 3 rd party data. CIESIN: 1-2 FTE |
| <p>G7 Regional to local data and data services There is a growing request for regional to local data, which are currently not archived in the DDC. Apart from local data also regional data services are required. A network of partner repositories in the regions could close the gap in available data, improve data access and data support/information for developing countries</p> | Regional data and information available through DDC and network of partners | P & F: Requires coordinating this effort but can make data reuse easier for developing countries. Overlaps with G6 and G4. | 6 months for consultations and scoping work CIESIN: 25% FTE to coordinate with local/regional networks. |

3. TG-Data Recommendations on IPCC AR6 Data Licensing

A sub-group of some TG-Data and Ex-Officio members prepared the following:

a) Introduction

Per its terms of reference, one of the purposes of Task Group on Data Support for Climate Change Assessments (TG-Data) is to “Facilitate in cooperation with the Data Distribution Centre (DDC) the availability and use of climate change related data resulting from the activities of the IPCC in accordance with the mandate of the IPCC.” As the IPCC and many of the organizations responsible for data products which underpin IPCC assessments are moving towards clearer formulations of license policies, TG-Data has a number of recommendations to make to ensure the availability and use of IPCC products, and help clarify the rights and obligations of both users and contributors.

Our understanding is that, in line with long-standing policy, the IPCC Sixth Assessment Report (AR6) will be published under a license prohibiting commercial use and the creation of derivative products, unless given explicit permission. If the same license applied to data products published by the IPCCDDC, it would severely limit the role of the DDC and render the IPCC data products useless for most users. A license allowing reuse and the creation of derivatives is essential for the pursuit of research and the reuse of IPCC products for national assessments, adaptation and mitigation policies.

A second issue is that a strict interpretation of the licenses of the climate projections from the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project 6 (CMIP6) project would require IPCC CMIP6 derived products to be licensed under a Creative Commons (CC) Attribution Share-Alike license. The Share-Alike clause gives freedom to use and manipulate the data, but imposes the same license to all down-stream products: “If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.” This would require CMIP6 derivative products created by AR6 authors to be published under a Share-Alike license, which is incompatible with the IPCC AR6 license. More generally, Share-Alike clauses restrict mixing and commercial use, which again would severely limit the scope and usability of IPCC data products distributed by the DDC. The IPCC should consider asking CMIP6 data providers an exemption from the Share-Alike clause for all derivative products created in the course of AR6.

Navigating the license arrangements to establish a clear set of agreements between data providers, publishers and users requires both an agreement on the policies and objectives and a signed acceptance of any licensing exemptions which are needed to implement the policies and objectives. This document presents recommendations from the IPCC Task Group on Data Support for Climate Change Assessments (TG-Data) on policies and objectives for the licensing arrangements in AR6.

b) TG-Data Recommendations

- i. The IPCC DDC should adopt the license policy specified in Appendix 3. Under this policy, the IPCC Data Distribution Centre will aim to publish data products, where this does not infringe the interests of relevant license holders, under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. This will enable re-use of the data both in research and in adaptation and policy support arenas. Additionally, use of IPCC DDC data will be clarified and enhanced through Terms of Use incorporated into the IPCC DDC website.
- ii. The data downloads from the AR6 Working Group I (WGI) Interactive Atlas will be part of the IPCC DDC and should fall under the CC BY 4.0 license.
- iii. The IPCC should request an exemption from the "ShareAlike" clause of the CMIP6 data license to enable publication of specific and limited data products, as explained in point 5 and 6 below, in the IPCC Data Distribution Centre with the CC BY 4.0 license.

- iv. The CC BY 4.0 license should apply to the following data products in the IPCC DDC
 - ✓ Figure and table data from AR6: the highly aggregated data sets prepared by authors to generate figures or populate tables in the report. For example, time series of global or regional annual means, with indications of uncertainty derived from ensemble spread and other sources.
 - ✓ The data products which can be downloaded through the AR6 WG1 Interactive Atlas, which will include near-surface temperature and precipitation fields for historical simulations and a number of future scenarios.

- v. The IPCC DDC will also hold a copy of a larger portion of the CMIP6 archive in order to ensure that the data remains available in the longer term. This data will be distributed under the CC BY-SA 4.0 and CC BY-NC-SA 4.0 licenses exactly as set by the providers. Intermediary data collections prepared by IPCC authors using CMIP6 data and archived to provide transparency regarding the production of IPCC figures and to enable re-use in policy support and adaptation planning will, if published, carry the CC BY-[NC-]SA 4.0 license.

c) Motivation

Licensing of Content in IPCC Reports and the Data Distribution Centre

The IPCC Assessment Report is based on an assessment of scientific literature. The underlying datasets are used extensively in the assessment process to corroborate, illustrate and communicate assessment conclusions. The tables and figures of the Assessment Report become important resources for activities associated with the implementation of global policy decisions taken on the basis of the report, activities which are becoming ever more widespread. The recommendations made here are intended to be taken together with work on the implementation of FAIR Principles (refs 6, 7).

Licensing, with a maximum of clarity and transparency, of IPCC material is essential to facilitate appropriate use of this material to address pressing climate change challenges, while protecting the rights of data providers. Open access and transparency of scientific data have been widely recognized as the “bedrock” of modern scientific progress (Science International, 2015 [4]). Ensuring that everyone has the right “to share in scientific advancement and its benefits” and “to the protection of the moral and material interests resulting from any scientific, literary or artistic production” is enshrined in Article 27 of the Universal Declaration of Human Rights (1948) [5].

The IPCC restricts the rights of users to manipulate content taken from its web domain (ipcc.ch) and published reports. This is standard practice for printed publications in order to prevent distortion and manipulation of content by others. This ensures that content subject to the rigorous IPCC approval process or bearing the IPCC logo is fully protected.

The shift towards digital publication of IPCC reports blurs the division between report content and data products held in the IPCC DDC. The DDC plays a supporting role in the assessment process (Appendix 3. and ref [8]) and is overseen by the TG-Data. Fulfilment of this role includes a commitment to making data accessible to promote transparency in the assessment process. A restrictive license as used for the publication of assessment reports would severely limit the role of the DDC and render the data products useless for most users. It is, therefore, essential to maintain a clear distinction between the content of the DDC and the published IPCC reports.

There are many complexities associated with data licensing, especially when applied to an activity on the scale of the IPCC assessment reports, which combine international collaborative efforts involving thousands of scientists and rigorous review and approval process involving all of its member states. Firstly, there is the problem of communicating license requirements clearly to institutions around the globe, which operate in a variety of languages and legal jurisdictions. Communication can be greatly simplified by using well known standardised license conditions, so that the license requirements can be communicated just by naming the license. This can be achieved by selecting licenses from the Creative Commons portfolio of standard licenses, some

of which are listed in Appendix 1. Further background on the topic is given in Appendix 2. TG-Data recommends that the IPCC adopt a Creative Commons license for its next assessment reports.

Whether or not the IPCC adopts a Creative Commons license for AR6, the adoption of a ShareAlike license by CMIP6 means that it is now necessary for IPCC to request an exemption from the CMIP6 license terms to enable inclusion of tables and figures derived from CMIP6 data in the report.

The DDC has previously dealt with licensing issues on a case by case basis. In order to deal consistently with licensing in the future, and to clarify the reasoning behind licensing decisions, the DDC is now adopting the policy set out in Appendix 3. This policy both protects the rights of data providers and facilitates re-use of data.

The proposed DDC policy recognises that many organisations seeking to respond to IPCC assessments of risks and likely outcomes rely on guidance from commercial services, and those services should have access to the data underlying IPCC assessments. The policy also recognises that a "ShareAlike" (also known as "copyleft") licensing constraint can act as a barrier to commercial users. This policy therefore seeks to avoid such constraints, enable reuse and promote iterative improvements, a widely used approach in the publishing of software and research data. Some scientific journals refuse to publish data licensed under ShareAlike licenses and such licenses could restrict further use of these data in interdisciplinary research and applications that build on the IPCC assessment reports.

Implementing the proposed DDC policy on final assessment data, which applies to both figure data and table data in AR6 (see section 5 below), will require exemptions from the CMIP6 data providers. Given the highly aggregated nature of this data, and based on discussions with WCRP, we do not expect objections from these data providers, so long as the nature of the requested exemption is clearly presented.

The Status of the AR6 WGI Interactive Atlas

In addition to the final assessment data, the AR6 WGI Interactive Atlas will provide users with visualisations of selected scenario data and access to underlying data. The visualisations will be considered as part of AR6, and hence fall under the license covering that report. The underlying scenario data will also be available. Users need to manipulate this data, so it is appropriate that it should come from the DDC rather than be considered as part of the report. This data, which comprises a very limited and specific subset of the CMIP6 scenario data, is of particularly high importance for people attempting to deal with IPCC assessment outcomes. For this reason, we will seek to obtain permission to publish under the CC BY 4.0 license, dropping the "ShareAlike" clause from the CMIP6 license.

Procedures for Dealing with Data Provider Licenses

The "Share Alike" clause of the CMIP6 data license prevents the use of that data to generate material for distribution under the licenses recommended for AR6 and the IPCC DDC above, unless exemption is gained from the data providers. There are multiple copyright owners of the data in the CMIP6 archive, as each participating modeling centre retains ownership of their contributions, but the form of the data policy has been set by a consensus decision moderated by the WCRP Working Group on Coupled Models (WGCM).

d) The Issue of Jurisdiction in the IPCC Context

Licenses need to have maximum clarity within any jurisdiction where use of material is expected. In the case of IPCC material, usages may have multiple jurisdictional implications. The problem of porting licenses from one jurisdiction to another, i.e., translating and adapting the language, rights, and constraints to reflect local laws, can be quite difficult. This problem can be avoided by using the Creative Commons family of licenses, which have been designed to provide legal

interoperability across multiple jurisdictions. Further information on licenses is provided in Appendix 1.

e) CMIP6 Data in the IPCC Sixth Assessment Report

The AR6 WGI Interactive Atlas

The WGI Interactive Atlas displays data extracted from CMIP6 and CORDEX archives. The CMIP6 data is published under the Creative Commons Attribution-ShareAlike 4.0 license. As noted above, this license imposes constraints which will significantly restrict many potential commercial users. This is currently at odds with the full implementation of the FAIR data principles, a core attribute of the Interactive Atlas to be a fully transparent and accessible product for IPCC users.

Report Tables and Embedded Data

The Assessment Reports include many tables and figures which illustrate key conclusions. These typically contain results obtained through a synthesis of data from many different sources. For instance, a figure may show a time series based on the multi-model ensemble mean of global or regional spatial means of CMIP6 data. In some cases, tabulated summary data is included in the report text. Digital images, such as GeoTIFF maps, may also contain embedded data values which can be considered as derived data products. The CMIP6 ShareAlike clause applies to such content and, by its nature, consequently applies to the whole assessment report.

Curated Figure Data

In order to make such data available to readers of the 6th Assessment Report, the final time series for many figures will be published by the IPCC Data Distribution Centre as a new dataset. These datasets will be highly aggregated products, representing a synthesis of many input datasets. In past reporting cycles such datasets, when they were made available (there was no systematic approach to preserving them) were done so under terms of use specified by the authors, usually without any Sharealike restriction. The Sharealike clause of the CMIP6 data imposes a new restriction on how such products can be disseminated. This material is not part of the published IPCC Assessment Report, but accessibility and transparency of the process of assessment will be impaired if this data is licensed with the Sharealike clause. Given the highly aggregated nature of the data, we do not expect that there was any intention on the part of the CMIP6 data providers for restriction. Many users of CMP6 data are likely to neglect the license constraint from ignorance. The IPCC DDC will, of course, only relax the license constraint if it is authorised by the providers.

Intermediate data products prepared by authors

Intermediate data products may be produced when authors are synthesising very large volumes of data. In some cases, in order to maximise the benefit obtained from these intermediate datasets and to promote transparency, these intermediate datasets may, in some cases, be archived by the IPCC DDC.

Appendix 1: Licenses and Rights

Licensing of Material on the IPCC Web Pages

"Unless otherwise stated, the information available on this website, including text, logos, graphics, maps, images and electronic downloads is the property of the IPCC and is protected by intellectual property laws.

You may freely download and copy the material contained on this website for your personal, non-commercial use, without any right to resell or redistribute it or to compile or create derivative works there from, subject to more specific restrictions that may apply to specific materials.

Reproduction of limited number of figures or short excerpts of IPCC material is authorized free of charge and without formal written permission provided that the original source is properly acknowledged, with mention of the complete name of the report, the publisher and the numbering of the page(s) or the figure(s). Permission can only be granted to use the material exactly as it is in the report. Please be aware that figures cannot be altered in any way, including the full legend. For media use it is sufficient to cite the source while using the original graphic or figure. In line with established Internet usage, any external website may provide a hyperlink to the IPCC website or to any of its pages without requesting permission.

For any other use, permission is required. To obtain permission, please email a scanned copy of a letter on official letterhead addressed to the Secretary of the IPCC with all relevant details. All communications by mail should be addressed to:

IPCC Secretariat
World Meteorological Organization
7bis Avenue de la Paix, P.O. Box No. 2300
CH-1211 Geneva 2,
Switzerland

All e-mail communications should be addressed to ipcc-sec@wmo.int.
(<https://www.ipcc.ch/copyright/> accessed 25th January, 2021)

Sui generis database rights

The Creative Commons licenses refer to "sui generis database rights". This refers to an aspect of copyright law which, through the 1995 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), applies in all 165 WTO member states.

The European Union interpretation of the sui generis rights states: "it should be laid down that the maker of a database or rightholder may not prevent a lawful user of the database from extracting and re-utilizing insubstantial parts; whereas, however, that user may not unreasonably prejudice either the legitimate interests of the holder of the sui generis right or the holder of copyright or a related right in respect of the works or subject matter contained in the database;" (DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, clause 49).

The Creative Commons licenses also refer to the obligations of users who incorporate "substantial" portions of the licensed data into their own material. From the above quote it appears that "substantial" should be interpreted, at least in the European jurisdiction, in terms of the interests of the copyright holder

Creative Commons Licenses

No Rights Reserved (CC0)

This license is essentially opting out of all license restrictions. CC0 enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those

interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

<https://creativecommons.org/licenses/by-sa/4.0/>

This license allows data to be used for commercial purposes, but imposes the restriction that any derived products must be distributed with the same license. This condition effectively prevents many commercial applications, so there is an indirect restriction on the way in which the data can be used.

For instance, if a consultant produces some guidance on flood risk, with tables of risk categories derived from climate projections published under the CC BY-SA license, then they would be obliged to publish those tables under the same license. Their customers would also have to apply the same license if they develop value-added products or desire to redistribute material. In US law, the doctrine of Fair Use provides protection from such restrictions if the published tables can reasonably be considered as a new creation, but this does not necessarily apply outside the US. The UK equivalent of the US Fair Use doctrine, for example, is the Fair Dealing doctrine which does not apply to any commercial activity.

Attribution-NonCommercial-NoDerivs 4.0 International (CC BY-NC-ND 4.0)

This license allows the user to copy and redistribute material, but prohibits any commercial use. It also prohibits the distribution of adapted material (the "noDerivs" clause), which, in the Creative Commons terminology, refers to material which is "translated, altered, arranged, transformed, or otherwise modified".

Transformation of data is a common element of data usage, so the prohibition on adapted material would be a substantial barrier to downstream use of the data.

Attribution 4.0 International (CC BY 4.0)

<https://creativecommons.org/licenses/by/4.0/>

The Creative Commons attribution license allows unrestricted use. The primary requirements are that users should provide proper attribution and also, when distributing modified products, provide clear information about the modifications that have been made. These last two restrictions may be considered important for IPCC material, as compared to the complete freedom given by CC0.

There are also specific attribution requirements, including a requirement to link back to source data. In the case of datasets produced by the IPCC during AR6, users of IPCC data will be required to link back to a DOI landing page that will provide full details of primary datasets.

Review

Creative Commons (CC) licenses implicitly allow the licensor to waive certain clauses. In particular, the "3b. ShareAlike" clause of the CC BY-SA license is problematic for the IPCC WGI Interactive Atlas and for the distribution of data related to figures in the report, and the IPCC may wish to consider requesting a waiver of this condition for products used in the Atlas (see below). The license holders who could grant such a waiver are the CMIP6 modeling groups.

Appendix 2: Background on licensing issues

Publication in journal with CC-BY license

Publication in the Earth System Science Data (ESSD) journal. ESSD will not accept papers containing "ShareAlike" data.

Categories of Licenses

There is a good discussion of 3 main categories of license in a blog article by Matt Mecoli [1], looking specifically at the GNU General Public License, which is a Sharealike/copyleft license, the MIT license and the Apache license. Mecoli notes that the Sharealike license "can add a lot more legal and technical complexity" for potential commercial users and that many such users simply avoid using such products.

David Wiley [2] asserts that the Sharealike clause "makes creators of derivative works second-class citizens".

A more detailed analysis of the legal issues and licensing constraints is provided by a publication of the Research Data Alliance (RDA) and the International Science Council (ISC) Committee on Data (CODATA), *Legal Interoperability of Research Data: Principles and Implementation Guidelines* [3]

Appendix 3: IPCC DDC Data Licensing Policy

About this Policy

This licensing policy has been created to support the execution of the core functions of the DDC set out at the 47th Session of the IPCC, Paris, 2018 and listed at the end of this appendix.

This licensing policy is intended to protect the interests of the copyright holders of data which is distributed through the DDC while giving data users clarity about their rights and responsibilities when they use data obtained from the DDC.

In the interests of presenting clarity and transparency, while protecting the rights of data providers, the IPCC DDC adopts the Creative Commons Attribution License (CC-BY 4.0) as its preferred data license. As explained below, this will not apply to all data served through the IPCC DDC, but will be used for derived data products generated in IPCC assessments where possible.

The Material to be Licensed

Input Data: Scientific Data and Scenarios

Scientific data and scenarios which play an important role in the IPCC assessment reports will, if necessary, be held in the DDC to ensure transparency, traceability, accessibility and stability. The DDC may work also in collaboration with an external archive or network to underwrite the stability and quality of long-term data preservation complying to international repository standards, such as the CoreTrustSeal. In both such cases, the data may be a direct copy of the original source data.

Final Assessment Data

Final Assessment Data refers to data which is explicitly presented in data tables in the report or graphically displayed (e.g. as a line graph or a coloured spatial map). Such datasets are compiled from multiple input datasets. Many downstream users of the data need access to this data and wish to manipulate the data to combine it with their own datasets and models. The DDC aims to archive and preserve Final Assessment Data as part of the implementation of the TG-DATA FAIR Data Policy [7,6], ensuring full documentation and attribution of the input data products and accessibility of the final assessment data.

Intermediate Assessment Data

In addition to the Final Assessment Data, the DDC may also hold datasets produced by authors as an intermediate step in the generation of Final Assessment Data. These intermediate datasets may be of high value both in terms of ensuring transparency around the assessment process and as a resource for users trying to implement policies based on the IPCC work. However, they are also likely to be much closer in structure and content to the source data products than the Final Assessment Data.

Selected Assessment Data

The WG1 Interactive Atlas is a new component of the Assessment Report. It will allow users to browse scenario data through an interactive graphical user interface. The images which can be accessed from the Atlas will be covered by the license applied to the assessment report which will restrict usage, except where permission is obtained, and prohibit manipulation of the images. The Atlas will also provide access to the data being displayed, which takes the form of multiyear averages of data from the CMIP6 and CORDEX archives. The data will be provided from the IPCC DDC, and covered by a license agreed between the DDC and the copyright holders of the input data.

Mapping Data to Roles

| Mapping of DDC roles (column 1) to categories of data listed in this section. | | | | |
|---|---|------------------------|------------------------------|-------------------------------|
| Role (see App. 1) | Input Data: Scientific Data and Scenarios | Final Assessment Table | Intermediate Assessment Data | Selected Assessment Data |
| 2.1 | Yes, e.g. CMIP | | | |
| 2.2 | | Yes | Partial | Yes: data behind atlas images |
| 2.3 | Yes, e.g. CMIP | | | |
| 2.4 | | Yes | Yes | |

Drivers for the Licensing Policy

The licensing policy should ensure that the data is accessible to both IPCC authors and external users. To this end, it should not place undue restrictions on the users. The license also needs to protect the legitimate interests of the owners of data.

The policy recognises that full transparency of the IPCC assessment process depends on commercial use of the final assessment data associated with the reports. Many organisations rely on commercial services and consultancy for expert guidance. We also recognise that the "ShareAlike" constraint presents a barrier to many commercial services. It is not an insurmountable barrier in general, but may significantly restrict the flow of information and could result in people turning to alternative, less restrictive, data sources.

The intention of the "ShareAlike" clause is to protect the data providers from loss of due accreditation which could follow republication of key parts of their data by other providers. In alignment with CC license provisions, providers may be assured that accreditation is required as to the data authors and providers.

The Licensing Policy

- i. The Creative Commons Attribution (CC BY 4.0) license provisions shall apply for data made available on the DDC website, except where otherwise noted, in particular where provisions of licenses of input data apply and when no exemptions can be obtained.
- ii. Input Data - Scientific Data and Scenarios: Data in this category, such as the CMIP data held at the DDC Partner DKRZ, is no different to the data originally published by the data providers. In this case the license conditions imposed by the providers will apply. See item 5 for exceptions.
- iii. Final Assessment Data: Data in this category will be published under the CC BY 4.0 license, subject to restrictions from input data licenses and where no exemptions are available. The DDC will prioritise working with TG-DATA and the IPCC to obtain exemptions from any restrictions imposed by input data licenses.
- iv. Intermediate Assessment Data: data in this category will be published with licenses that comply with all the licenses of input datasets. See item 5 for exceptions.

v. For some limited and specific subsets of scenario data, such as the data which is visualised through the AR6 WGI Interactive Atlas, or for intermediate assessment data of high value, there may be grounds for requesting an exemption in order to permit publication under the CC BY 4.0 license. In AR6, this is the case for the IPCC WGI Interactive Atlas.

DDC Roles and Responsibilities

Taken from "Guidance for the core functions of the IPCC Data Distribution Centre (DDC)" (https://www.ipcc.ch/site/assets/uploads/2018/12/Guidance_DDC.pdf).

2.1 Archive and provide transparency, traceability, and stability of data and scenarios used by the IPCC in its reports, available at the DDC or elsewhere.

2.2 Archive and provide transparency, traceability, and stability of data and scenarios underpinning key figures and tables, and headline statements in the IPCC reports.

2.3 Collaborate as appropriate with data centers that hold data or provide functions relevant to the IPCC in a transparent manner, under the guidance of the TG-DATA and the Working Group Bureaus (WGBs), to provide information via IPCC websites relevant to data and scenarios.

2.4 Curate new data sets unavailable elsewhere and link to external data sets of relevance.

2.5 Improve accessibility to Data Distribution Centre materials for supporting IPCC authors and external users, especially in developing countries.

2.6 Contribute to a sustainable structure established and approved by the IPCC to provide observed and model data and information relevant at regional scales.

References

[1] Matt Mecoli: <https://towardsdatascience.com/a-data-scientists-guide-to-open-source-licensing-c70d5fe42079>

[2] David Wiley, <https://opencontent.org/blog/archives/347>

[3] RDA & CODATA, <https://doi.org/10.5281/zenodo.162241>

[4] Science International (2015): Open Data in a Big Data World. Paris: International Council for Science (ICSU), International Social Science Council (ISSC), The World Academy of Sciences (TWAS), InterAcademy Partnership (IAP). https://council.science/wp-content/uploads/2017/04/open-data-in-big-data-world_long.pdf

[5] Universal Declaration of Human Rights (1948), United Nations General Assembly resolution 217 A, Paris, 10 December 1948. www.un.org/en/universal-declaration-human-rights/index.html

Article 27:

(1) Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.

(2) Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.

<https://www.un.org/en/universal-declaration-human-rights/index.html>

[6] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

[7] AR6 WG1 TSU/TG-DATA paper on Fair Data

[8] "Guidance for the core functions of the IPCC Data Distribution Centre (DDC)" (https://www.ipcc.ch/site/assets/uploads/2018/12/Guidance_DDC.pdf): "The purpose of the IPCC Data Distribution Centre (DDC) is to archive and provide transparency, traceability, and stability of data and scenarios that are relevant in the context of the IPCC."

Glossary

CMIP6: Coupled Model Intercomparison Project Phase 6 (<https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6>)

WCRP: World Climate Research Programme (www.wcrp-climate.org)

WGCM: Working Group on Coupled Modelling (<https://www.wcrp-climate.org/wgcm-overview>) A WCRP working group.

Share-Alike or copyleft: A licensing approach which obliges users of a product to implement an equivalent license on all derived products. Sometimes referred to as a "viral" license. The advocates of this approach see it as a way of preventing others from imposing restrictive licensing conditions, critics see it as a disproportionate restriction on the freedom of users to decide how they deal with their own products.

Appendix 4: DKRZ Usage

Report 2020 of the DDC at DKRZ

| Document ownership and history | | |
|--------------------------------|---|--|
| Owner | IPCC DDC at DKRZ (ipcc.wdc-climate.de) | |
| Location | DDC_report_DKRZ_2020.docx | |
| Author team: | M. Stockhause | |
| Version | 1.0 | |
| Date | 2021-02-04 | |
| Version history | | |
| Date | Versio | Comment |
| 2021-02-04 | 1.0 | Report completed incl. internal review |

a) Summary

The total AR5 data volume provided by IPCC DDC hosted at DKRZ is 1.7 PBytes: 1.6 PBytes in the DDC AR5 Reference Archive, 100 TBytes in the IPCC AR5 WG1 Archive, ca. 1 TBytes for AR4, less than 10 GBytes each for FAR, SAR, and TAR, and 35 TBytes for SR1.5 data.

In 2020, massive downloads of a single dataset from a single European ESGF user occurred. The statistics were corrected by excluding the estimated effect of these possibly technical downloads. IPCC DDC users downloaded ca. 0.8 PBytes and 1.4 million datasets with mean monthly downloads of 120 000 datasets/month and 70 TBytes/month. The total download volume in 2020 increased by ca 50 % compared to 2019. Data downloads were dominated by AR5 downloads via the ESGF with a share of ca. 93 %. Downloads for AR5 data increased by 140 % in download numbers and 50 % in download volume compared to 2019. The other AR download volumes decreased by -49 % for AR4, -75 % for TAR and -62 % for SAR. DDC users requested no data for selected areas on storage media in 2020.

The majority of downloads were from Asian users with 85 %. Downloads from European and North American users were similar with 6.5 % and 6.7 %, respectively. Users from other continents have shares of ≤ 1 % of the total download counts: Africa: 1.0 %, Oceania: 0.4 %, South America: 0.8 %.

b) Evolution of data access

The user downloads from the DDC reference archive show a significant increase of the number of downloaded datasets, which is not reflected in the download volumes (**Figure 1**). Data analysis identified a single ESGF user (or more precisely from a single IP via the DDC's ESGF node) is responsible for this by a repeated download of a single dataset with extremely high frequency. These downloads started in July 2020 and peaked in August 2020 with 1.7 million dataset downloads. This single user dominates the download statistics in the second half of 2020. Therefore, the influence of this possibly technical downloads on the different DDC statistics is discussed in this report.

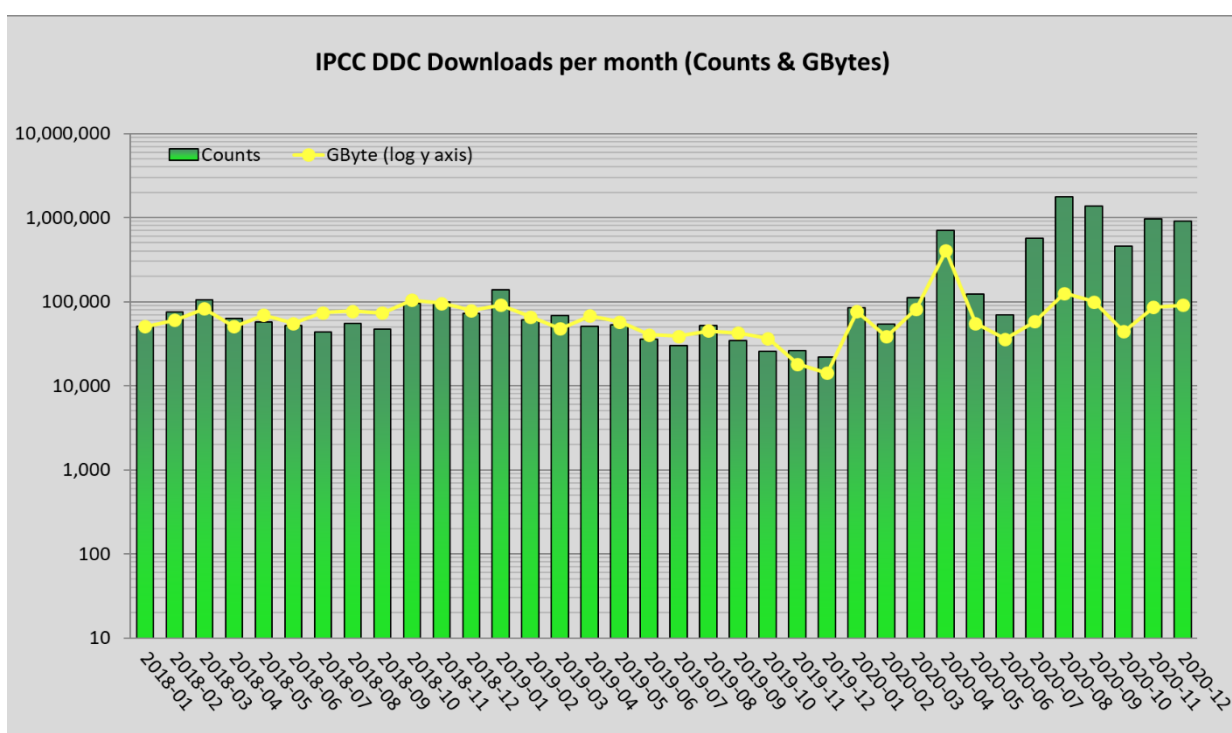


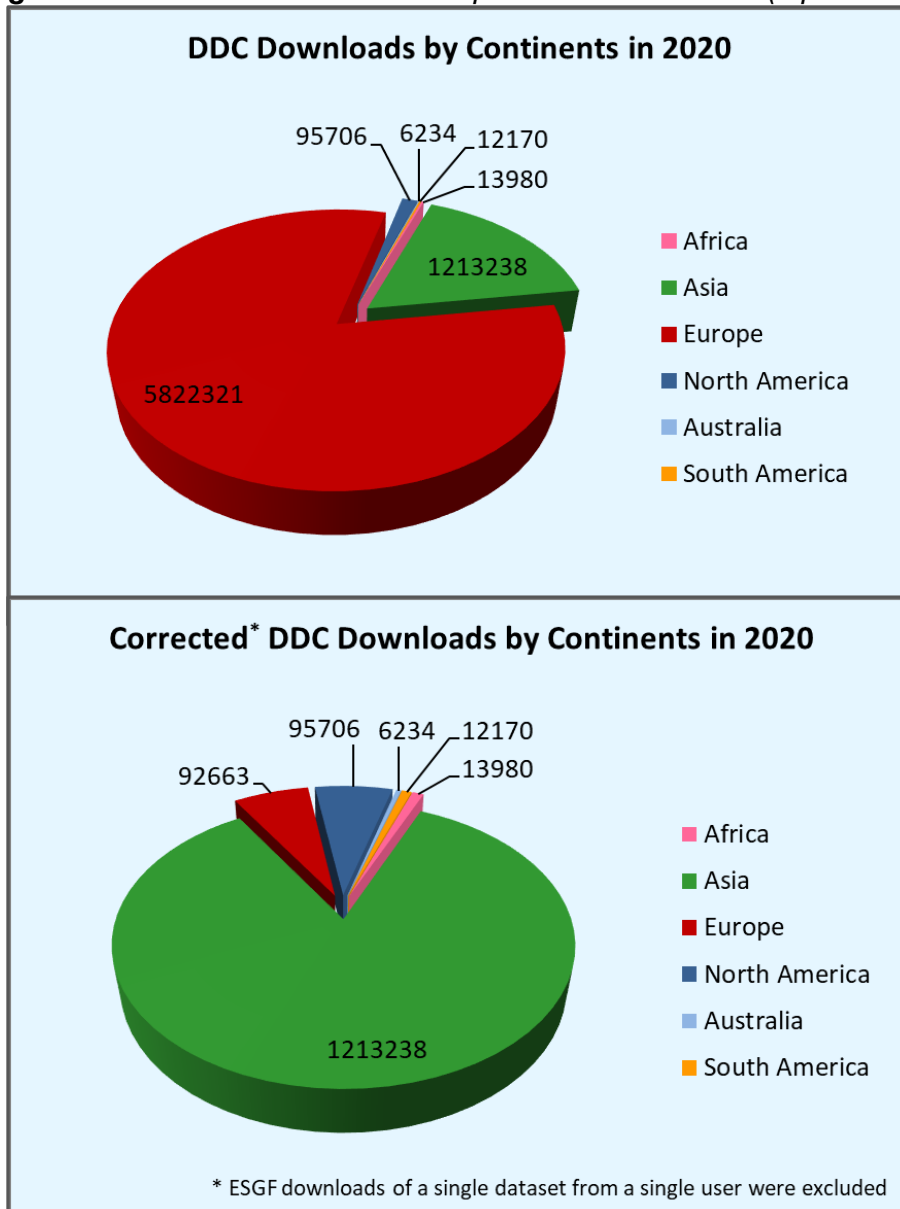
Figure 1: Total data download counts and volumes per months over the last three years in GBytes from the IPCC DDC reference archive.

The total downloads in 2020 of 1.2 PBytes doubled compared to 2019, excluding the influence of the single dataset downloads this increase is estimated as +50 %. The mean monthly downloads in 2020 were ca. 600 000 datasets/month and 100 TBytes/month. The corrected values are estimated as a total download of 0.8 PBytes and 1.4 million datasets with mean monthly downloads of 120 000 datasets/month and 70 TBytes/month. Downloads peaked in 04/2020, the reason is unknown. A plausible explanation might be an increased replication activity of external repositories. No general trend was observed in the corrected download data for year 2020.

c) Geographical distribution of data access

For the IPCC DDC AR5 data, direct data access at the DKRZ and access via ESGF (Earth System Grid Federation) are supported. Downloads from ESGF dominate the statistics with 98% or 93% after correction with downloads from the single European ESGF user. The ESGF file downloads in 2020 were merged with the DDC continental download information. As the influence of this massive downloads from the single European ESGF user have a huge influence on the continental distribution, both distributions are shown in **Figure 2**. As an information on the number of active users is not available for the dominating ESGF downloads, no reliable information on active users can be provided for 2020.

Figure 2: Downloads counts of users per continent for 2020 (top: uncorrected values;



bottom: values corrected by exclusion of single European ESGF user downloads).

The download counts per continent show that 81 % of the total downloads were from European users (**Figure 2** top). This is reduced to a more reliable share of 6.5 % by exclusion of downloads from the single European ESGF user (**Figure 2** bottom). Thus, the following discussion will focus on the corrected values. As in the previous years, the majority of downloads are from Asian users with 85 %. Downloads from European and North American users are similar with 6.5 % and 6.7 %, respectively. Downloads from the other continents are ≤ 1 % of the total download counts: Africa: 1.0 %, Oceania: 0.4 %, South America: 0.8 %. Downloads from Africa, Asia and South America, which can be roughly regarded as developing and economy-in-transition countries, add up to a 84 % share of the total download counts. The continental shares for direct DDC downloads and downloads via the ESGF are different: The dominance of Asian user downloads is more pronounced among ESGF users with 86 % than among DDC users with 61 %. Accordingly, direct DDC downloads have a higher share of South American (12 %) and of European (16 %) user downloads.

Trends for the continental downloads from 2019 to 2020 are difficult to estimate, as the ESGF download statistics were not available for 2019. As the total download volume for 2020 increased by 50 % in volume and by 140 % in download counts, downloads from all continents are likely to have increased.

Data on storage media

DDC users requested no data for selected areas on storage media in 2020.

d) Data access by category AR

The monthly download rates in 2020 from the IPCC DDC Reference Data Archive were dominated by AR5 downloads as in the previous years (**Figure 3**; online monthly download statistics⁸). The majority of AR5 data were downloaded via DKRZ's DDC ESGF (Earth System Grid Federation) data node. AR5 downloads show an increase by a factor of 11 in dataset numbers and 110 % in volume. Excluding the effect of the single dataset download from a single user significantly reduces this increase to 140 % in dataset numbers and to a 50 % increase in data volume. The data downloads for previous ARs show decreases from 2019 to 2020 of -18 % for AR4, -70 % for TAR and -50 % for SAR in dataset numbers and of -49 % for AR4, -75 % for TAR and -62 % for SAR in data volumes.

e) Review of user queries

User requests are directed to the DDC and for AR5 data partly to the ESGF support. A separation of user requests on IPCC DDC issues is not possible.

In parallel to the regular user support channels, additional requests were sent to individuals at the modelling centers or at the data centers.

f) News and activities

As a further DDC activity, a Virtual Workspace was offered as collaboration platform in support of the IPCC AR6 authors.

⁸ Online monthly download statistics are available at:

https://cera-www.dkrz.de/WDCC/ui/cerasearch/statistics?type=downloads_by_domain&domain=IPCC-DDC

https://cera-www.dkrz.de/WDCC/ui/cerasearch/statistics?type=downloads_by_domain&domain=IPCC-DDC_AR5

https://cera-www.dkrz.de/WDCC/ui/cerasearch/statistics?type=downloads_by_domain&domain=IPCC-DDC_AR4

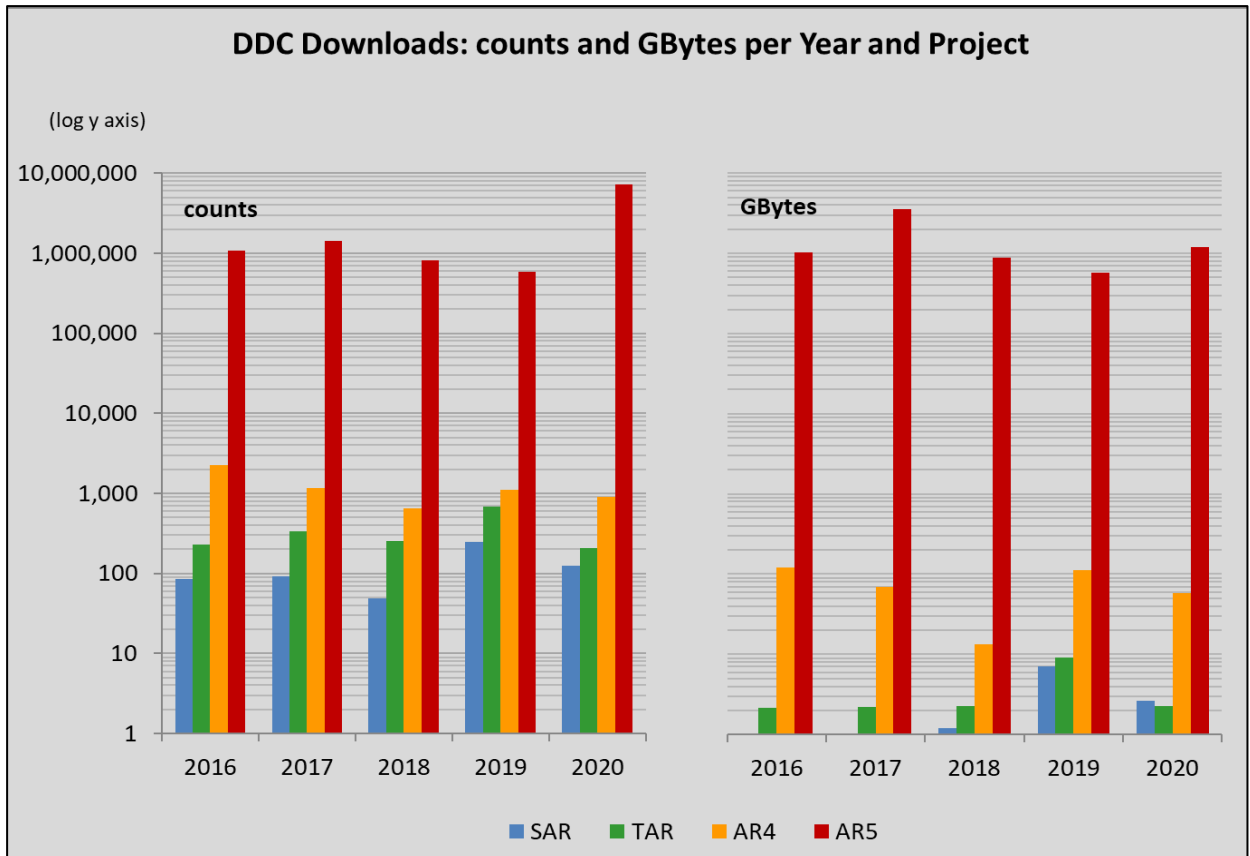


Figure 3: Total annual data download counts (left) and volumes in GBytes (right) over the last five years for the different DDC reference archives (without FAR and SR1.5).

Appendix 5: CEDA Usage

2020 Usage Report for the IPCC DDC

| Document ownership and history | | |
|--------------------------------|-------------------|---|
| Owner | CEDA | |
| Location | 2020DDCUsage.docx | |
| Author(s) | C. Pascoe | |
| Version | 1.0 | |
| Date | 2021-02-12 | |
| Version history | | |
| Date | Version | Comment |
| 2021-02-11 | 0.1 | Initial Report for DDC managers. This includes the separate SEDAC statistics but does not yet include browser statistics or the country by country breakdown. |
| 2021-02-12 | 1.0 | Completed outstanding sections |

a) Summary

This report provides details of how the DDC web site, ipcc-data.org, was used in the 12 months from January 1st 2020 to December 31st 2020. The information presented here is generated using the Google Analytics service; providing a range of different statistics that cover both temporal and spatial trends. Only non-bounce sessions in which users have interacted with the website have been used in this analysis.

The annual DDC page views for 2020 show a 21% decrease in usage compared to 2019 but are nevertheless comparable to the number page views received in previous years. 2020 began with page views at the upper end of the range and ended at the lower end of the range. The annual total for 2020 is above that seen in 2017 and 2018.

The Home page remains the most commonly viewed page on the DDC. In 2020 Europe was the most active region using DDC resources.

b) Evolution of User Sessions

The number of DDC page views recorded in 2020 are about 21% less than those received in 2019. The decrease in usage began in mid-March 2020 and the negative trend (relative to previous years) continued until August 2020 whereupon the pattern of use began to show a similar behaviour to previous years albeit from a lower baseline.

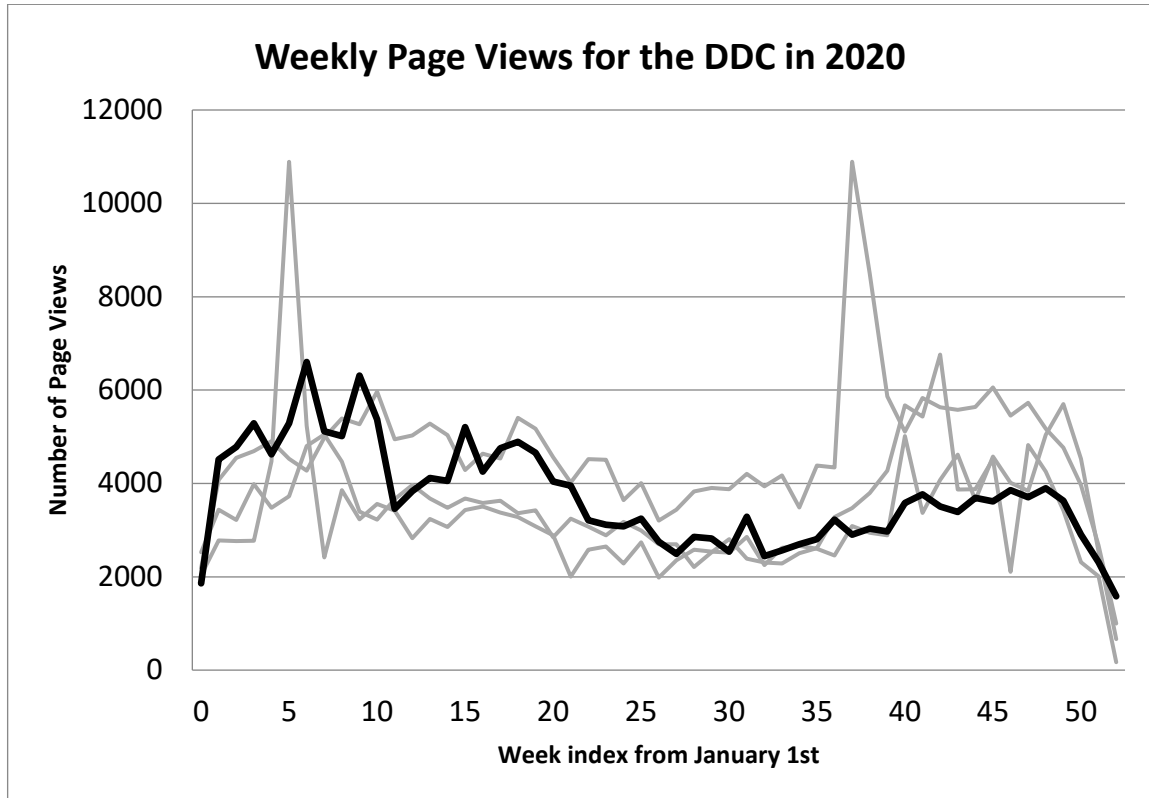


Figure 1: Time series of weekly DDC page views for non-bounce sessions. The time axis is shown in weeks from January 1st. The thick black line shows the page views in the period from January 1st to December 31st 2020, thin grey lines show the page views in previous years (2017, 2018 and 2019).

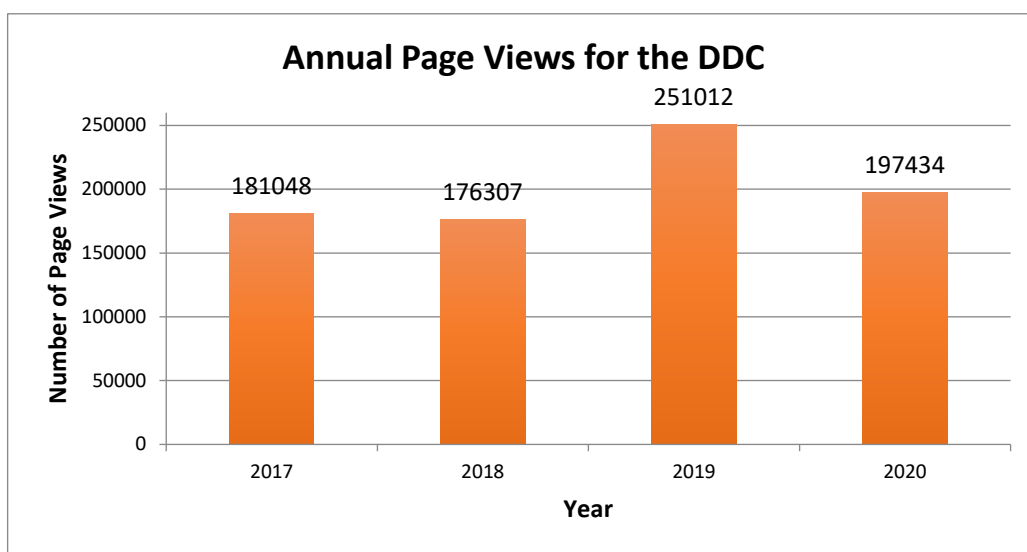


Figure 2: Annual page views for non-bounce sessions from 2017 – 2020.

c) Access Patterns by Page on the DDC

This page view analysis examines the 50 most visited URLs on the IPCC-DDC.

The IPCC-DDC home page remains the most commonly viewed page on the DDC, receiving 15% of page views in 2020 (Figure 3). Other popular pages are those that explain some of the basic science behind climate change such as: Understanding the observational record; Projected emissions and concentrations of carbon dioxide. It is clear that users are interested in using the DDC to access data and information about the IPCC 5th Assessment Report (AR5).

Figure 4 shows the number of unique page views and number of entrances for the most used pages on the DDC during 2020. Users tend to arrive at pages where the number of entrances are about equal to the number of page views and navigate to pages where the number of entrances are much less than the number of page views.

Most users arriving at the DDC continue to do so via the Home page. It is reassuring to know that the first DDC page users generally see is our home page because this is where we provide information about the purpose of the site and an overview of the content that we provide. Information pages about the scenario process for AR5 and the description of RCPs are also places of entry to the DDC. 2020 showed that users are also arriving at the DDC via the Glossary.

Most users visiting the guidance pages navigate there from another page on the DDC, this fits with the behaviour of a person who, having found some interesting data, decides to use our guidance documents to learn how to use it.

The pageview statistics for pages hosted by the Socioeconomic Data and Applications Centre (SEDAC) is shown in Table 2 and Figure 5.

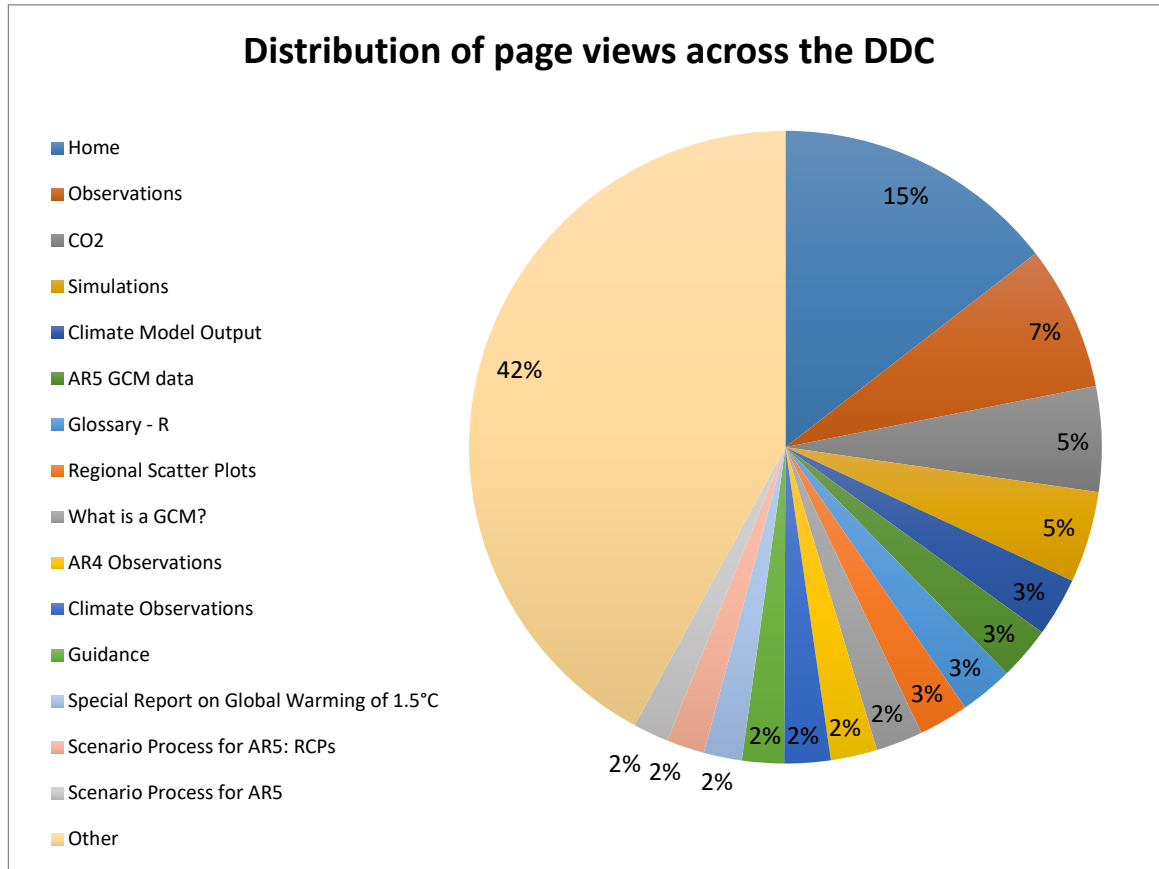


Figure 3: Distribution of page views for the most visited pages on the IPCC-DDC during 2020.

| Page on the DDC | URL | Number of Views | % |
|---|--|-----------------|------|
| Home | www.ipcc-data.org/index.html | 28,159 | 14.3 |
| Observations | www.ipcc-data.org/observ/index.html | 14,336 | 7.3 |
| CO2 | www.ipcc-data.org/observ/ddc_co2.html | 10,430 | 5.3 |
| Simulations | www.ipcc-data.org/sim/index.html | 9,057 | 4.6 |
| Climate Model Output | www.ipcc-data.org/sim/gcm_monthly/index.html | 5,839 | 3.0 |
| AR5 GCM data | www.ipcc-data.org/sim/gcm_monthly/AR5/index.html | 5,272 | 2.7 |
| Glossary - R | www.ipcc-data.org/guidelines/pages/glossary/glossary_r.html | 5,263 | 2.7 |
| Regional Scatter Plots | www.ipcc-data.org/syn/tar_scatter/index.html | 4,949 | 2.5 |
| What is a GCM? | www.ipcc-data.org/guidelines/pages/gcm_guide.html | 4,650 | 2.4 |
| AR4 Observations | www.ipcc-data.org/observ/clim/ar4_global.html | 4,610 | 2.3 |
| Climate Observations | www.ipcc-data.org/observ/ddc_envdata.html | 4,577 | 2.3 |
| Guidance | www.ipcc-data.org/guidelines/index.html | 4,158 | 2.1 |
| Special Report on Global Warming of 1.5°C | www.ipcc-data.org/sim/gcm_monthly/SR15/index.html | 3,780 | 1.9 |
| Scenario Process for AR5: RCPs | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/RCPs.html | 3,766 | 1.9 |
| Scenario Process for AR5 | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/index.html | 3,526 | 1.8 |

Distribution of page views for the most visited pages on the IPCC-DDC during 2020. This table contains the same information as the pie chart in figure 3 but with the page URLs shown.

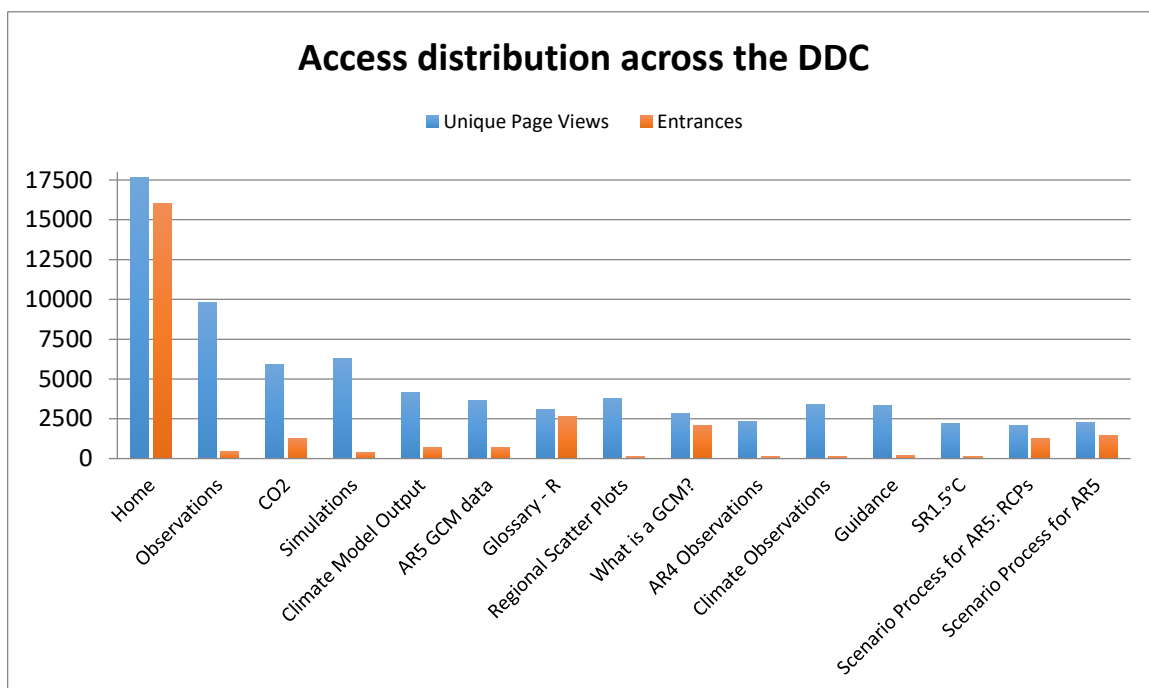


Figure 4: The number of unique views per page (blue) and the number of entrances per page (red) during 2020. Users tend to arrive at pages where the number of entrances are about equal to the number of page views and navigate to pages where the number of entrances are much less than the number of page views. Note that the “unique page views” statistic lists all pages that are visited in a session, it does not take into account the number of times a page is visited during that session.

| SEDAC Page | SEDAC Page on the DDC | Page views | Unique Page views | Entrances |
|--|--|------------|-------------------|-----------|
| Scenario Process for AR5: RCPs | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/RCPs.html | 3,766 | 2,089 | 1,292 |
| Scenario Process for AR5 | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/index.html | 3,526 | 2,256 | 1,468 |
| IPCC AR5 Observed Climate Change Impacts | sedac.ciesin.columbia.edu/ddc/observed_ar5/index.html | 2,839 | 1,972 | 1,008 |
| SRES Emissions Scenarios | sedac.ciesin.columbia.edu/ddc/sres/index.html | 2,371 | 1,297 | 796 |
| Socio-Economic Baseline Dataset | sedac.ciesin.columbia.edu/ddc/baseline/index.html | 1,980 | 1,251 | 643 |
| Socio-Economic Data and Scenarios | sedac.ciesin.columbia.edu/ddc/index.html | 1,500 | 879 | 722 |
| IPCC AR5 Observed Climate Change Impacts | sedac.ciesin.columbia.edu/ddc/observed_ar5/ar5_2.html | 1,472 | 817 | 20 |
| Scenario Process for AR5: The IPCC and Scenario Development | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/ipcc_scenarios.html | 1,067 | 792 | 95 |
| Scenario Process for AR5: Parallel Phase: Climate Modeling Activities | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/parallel_climate_modeling.html | 931 | 709 | 27 |
| IPCC IS92 Scenarios | sedac.ciesin.columbia.edu/ddc/is92/index.html | 911 | 568 | 235 |
| IPCC AR4 Observed Climate Change Impacts | sedac.ciesin.columbia.edu/ddc/observed/index.html | 897 | 694 | 120 |
| Scenario Process for AR5: Parallel Phase: New Narratives and Scenarios | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/parallel_nat_scen.html | 761 | 512 | 60 |

| | | | | |
|--|---|-----|-----|----|
| Scenario Process for AR5: Scenarios Background Information | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/scenario_background.html | 517 | 396 | 34 |
| Scenario Process for AR5: Scenario Process Overview | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/scenario_overview.html | 454 | 355 | 29 |
| Scenario Process for AR5: References and Resources | sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/reference_resource.html | 308 | 226 | 28 |

Distribution of page views for the 15 most visited DDC pages hosted by the Socioeconomic Data and Applications Centre (SEDAC).

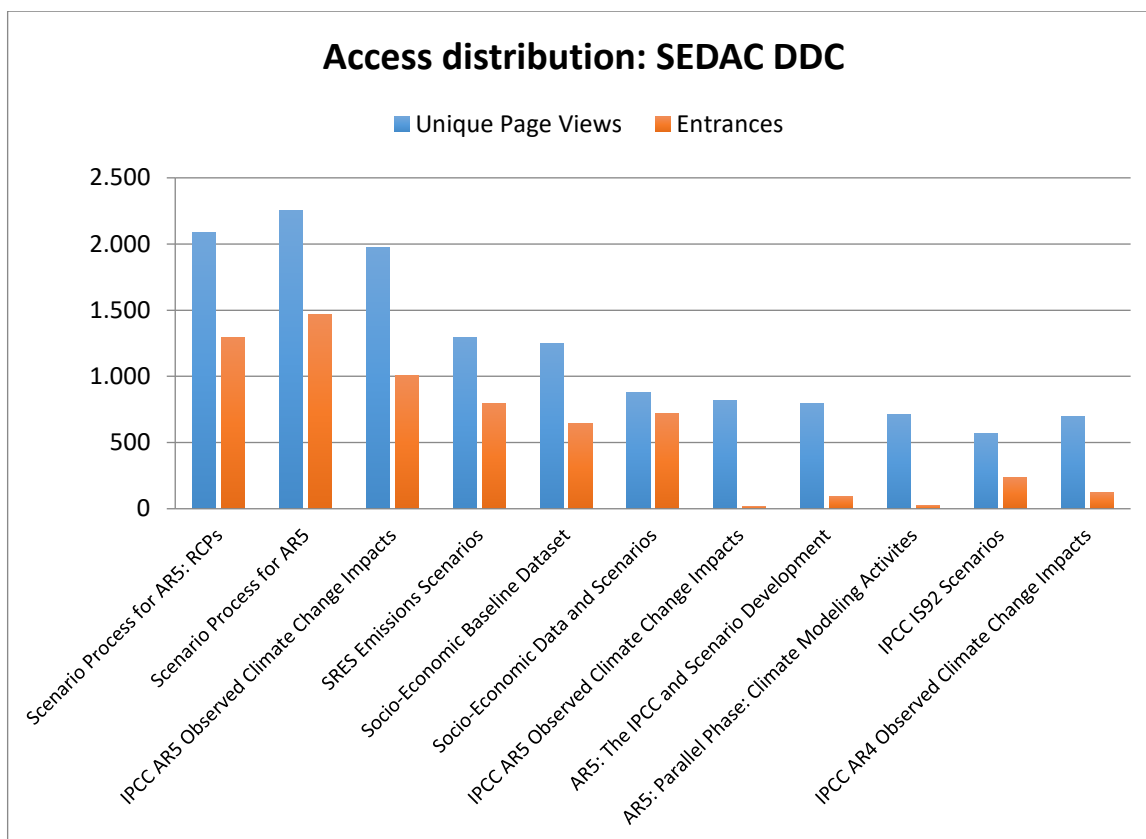


Figure 5: The number of unique views per page (blue) and the number of entrances per page (red) for the SEDAC hosted pages of the DDC.

d) Geographical Distribution of Engagement with the DDC

In 2020 Europe was the most active region using DDC resources (Figure 6). Overall there was a 21.7% fall in the number of DDC sessions from 2019 to 2020 however the fall was experienced differently in different continents (Figure 7). The smallest fall in sessions was seen in Central and Southern Americas where the fall was just 10.8%, the largest fall in sessions was seen in North America where the fall was 28.2%.

Figure 8 names the top 17 countries that make use of the DDC. Users in 7 countries account for 52% of all sessions: USA 18%, UK 8%, India 6%, China 6%, Germany 5%, Canada 5%, France 4%. “other” countries not listed in Figure 8 account for 29% of user sessions. A full breakdown by country of the DDC user sessions in 2019 and 2020 is shown in appendix A.

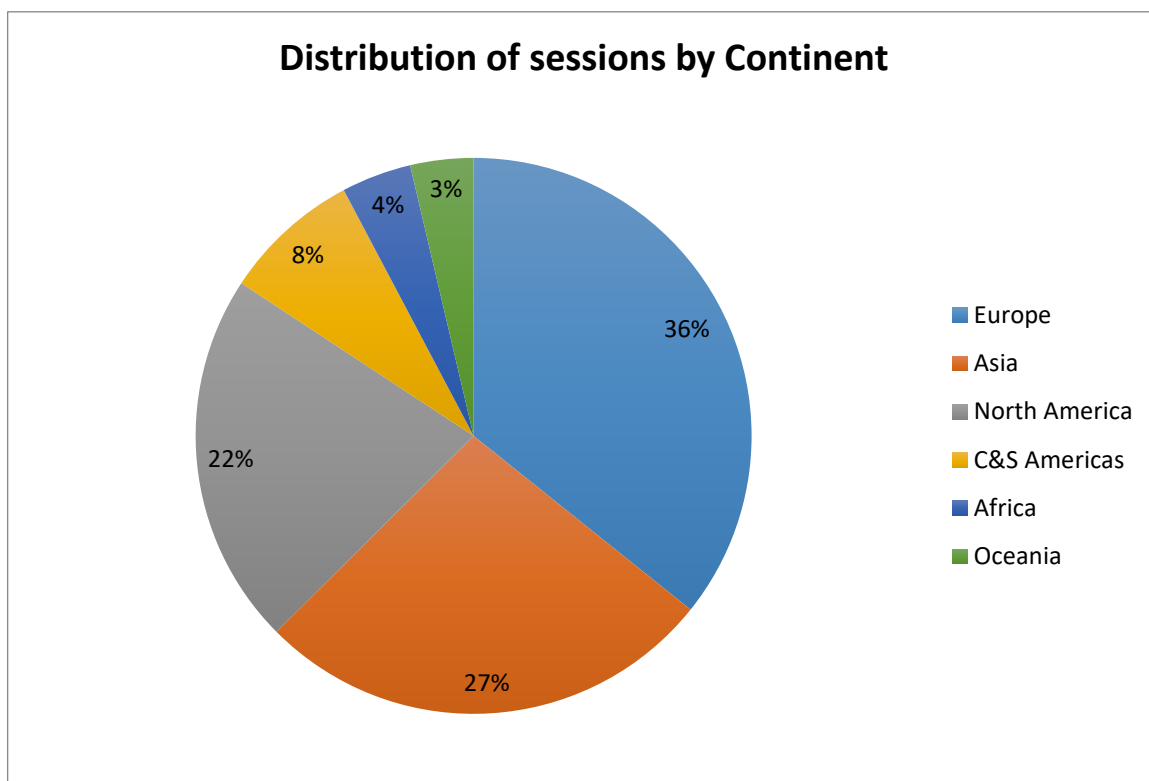


Figure 6: The geographical distribution of IPCC-DDC sessions by continent for 2020.

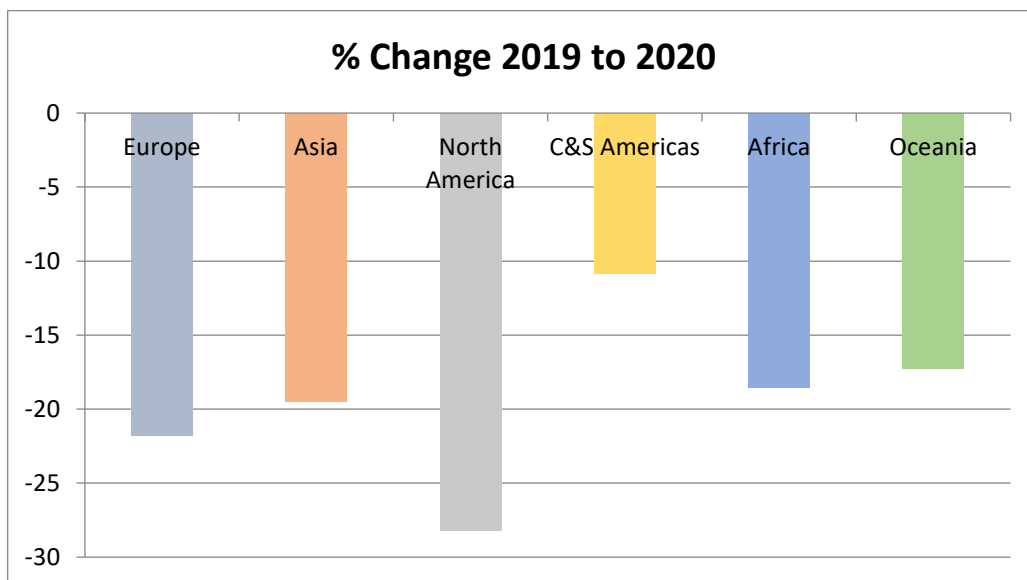


Figure 7: The percentage change in session numbers from 2019 to 2020 by continent.

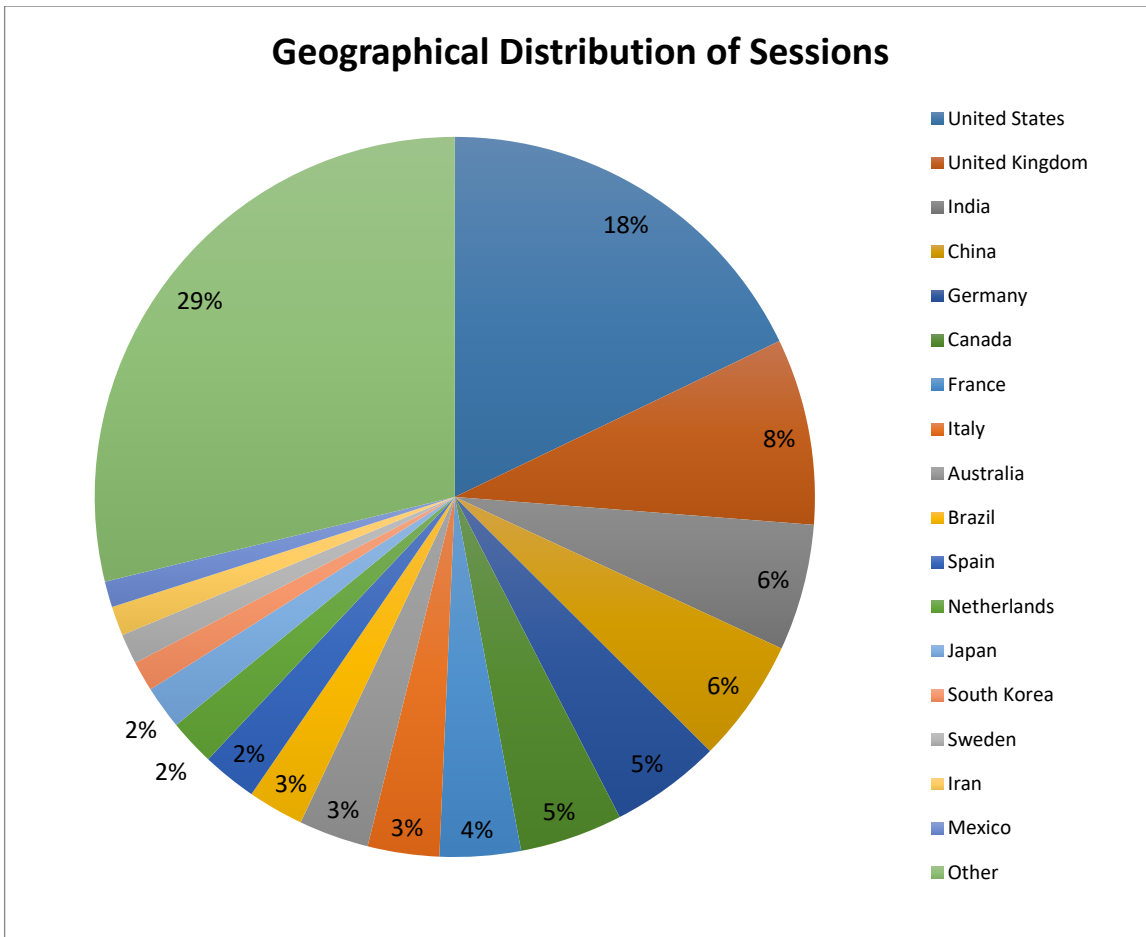


Figure 8: The geographical distribution of IPCC-DDC user sessions by country for 2020. Sessions initiated by users in Europe, Asia and North America account for 85% of DDC activity (Figure 7) however approximately half of the DDC sessions are initiated by users in just seven countries (Figure 8).

e) Browser Statistics

The IPCC-DDC site is available and accessible to all the major internet browsers and operating systems. English is the most used language setting, accounting for 60% of sessions (Figure 9). Chrome is the most popular browser, accounting for 71% of sessions and Windows is the most popular operating system, accounting for 74% of sessions (Figure 10).

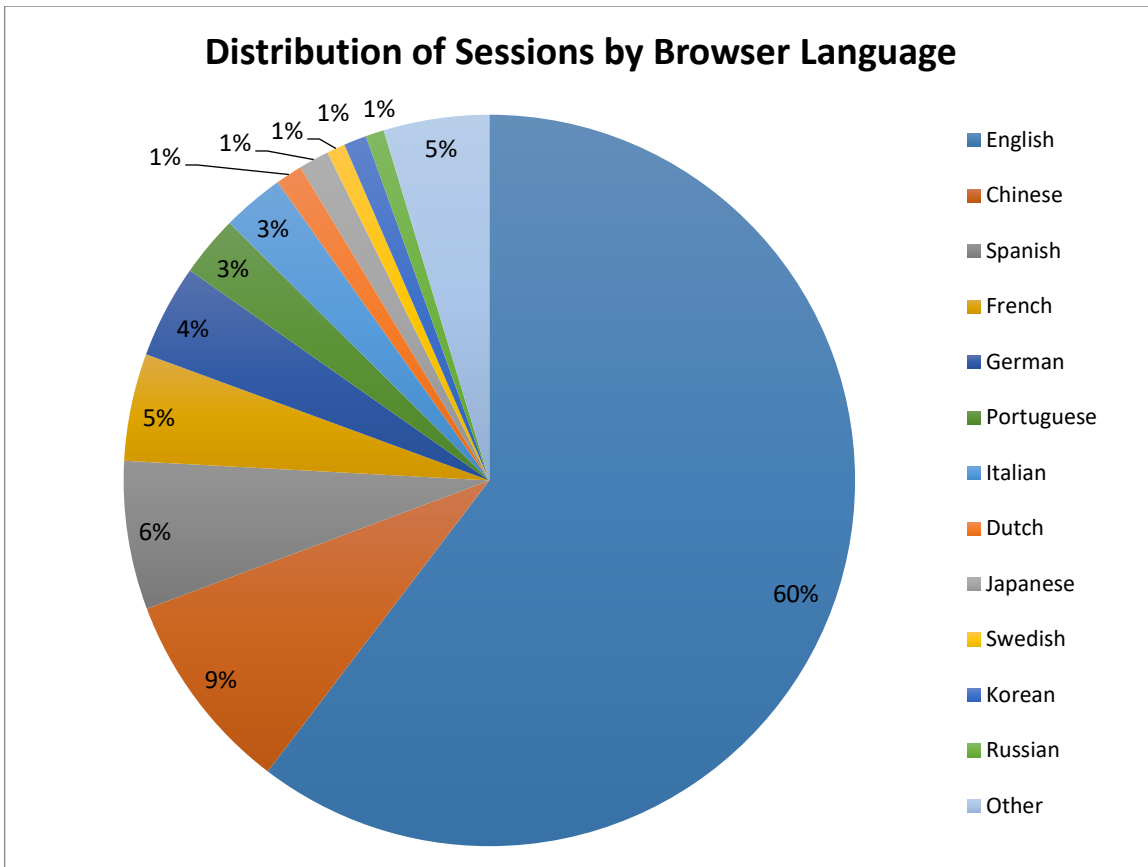


Figure 9: Language settings for DDC user sessions during 2020.

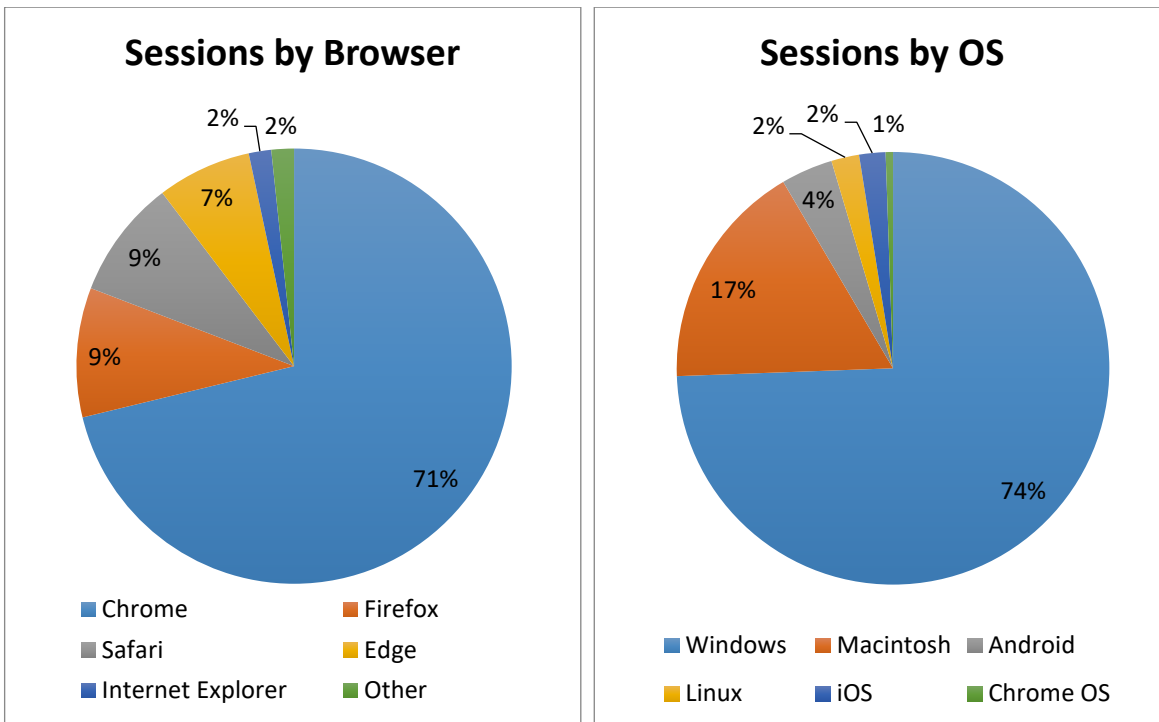


Figure 10: Browser and technology characteristics of DDC user sessions with information about Browsers, Operating Systems that are used to access the DDC during 2019.

f) Appendix

A full breakdown, by country or region, of the DDC user sessions during 2020 and also during 2019 for comparison.

| Country or Region | 2020 Sessions | 2019 Sessions | Country or Region | 2020 Sessions | 2019 Sessions |
|---------------------------|---------------|---------------|----------------------|---------------|---------------|
| United States | 7013 | 9948 | Argentina | 242 | 262 |
| United Kingdom | 3360 | 4495 | South Africa | 242 | 311 |
| China | 2498 | 3570 | Pakistan | 236 | 290 |
| India | 2292 | 3114 | Philippines | 218 | 278 |
| Germany | 1990 | 2669 | Peru | 202 | 205 |
| Canada | 1858 | 2412 | New Zealand | 195 | 296 |
| France | 1477 | 1761 | Ireland | 188 | 226 |
| Italy | 1320 | 1418 | Poland | 187 | 234 |
| Australia | 1272 | 1488 | Finland | 181 | 374 |
| Brazil | 1029 | 1200 | Ecuador | 175 | 144 |
| Spain | 961 | 1346 | Vietnam | 150 | 171 |
| Netherlands | 885 | 1185 | Bangladesh | 138 | 171 |
| Japan | 824 | 1199 | Morocco | 128 | 94 |
| Iran | 688 | 957 | (not set) | 128 | 119 |
| Sweden | 560 | 852 | Nigeria | 123 | 161 |
| South Korea | 552 | 576 | Israel | 111 | 78 |
| Mexico | 482 | 545 | Egypt | 110 | 108 |
| Belgium | 464 | 569 | United Arab Emirates | 105 | 57 |
| Switzerland | 453 | 659 | Hungary | 103 | 113 |
| Indonesia | 440 | 482 | Nepal | 103 | 107 |
| Turkey | 394 | 357 | Kenya | 101 | 143 |
| Taiwan, Province of China | 384 | 425 | Czechia | 98 | 153 |
| Hong Kong, SAR of China | 363 | 314 | Nicaragua | 89 | 15 |
| Denmark | 352 | 434 | Sri Lanka | 88 | 70 |
| Colombia | 346 | 384 | Bolivia | 81 | 48 |
| Portugal | 325 | 343 | Romania | 81 | 82 |
| Greece | 299 | 275 | Ukraine | 67 | 74 |
| Malaysia | 292 | 291 | Uganda | 60 | 56 |
| Russia | 288 | 245 | Ghana | 59 | 57 |
| Chile | 282 | 393 | Costa Rica | 57 | 66 |
| Thailand | 276 | 283 | Algeria | 57 | 53 |
| Singapore | 272 | 250 | Luxembourg | 57 | 60 |
| Norway | 269 | 359 | Iraq | 54 | 53 |
| Ethiopia | 262 | 382 | Saudi Arabia | 52 | 58 |
| Austria | 257 | 342 | Tanzania | 49 | 92 |

| Country or Region | 2020 Sessions | 2019 Sessions | Country or Region | 2020 Sessions | 2019 Sessions |
|-------------------|---------------|---------------|----------------------|---------------|---------------|
| Lithuania | 48 | 37 | Kuwait | 18 | 7 |
| Slovakia | 48 | 40 | Latvia | 18 | 26 |
| Cambodia | 44 | 56 | Cuba | 17 | 6 |
| Jordan | 41 | 54 | Myanmar (Burma) | 17 | 43 |
| Tunisia | 39 | 46 | Namibia | 17 | 10 |
| Serbia | 38 | 25 | Guyana | 16 | 11 |
| Slovenia | 38 | 37 | Georgia | 15 | 15 |
| Zimbabwe | 36 | 48 | Malta | 15 | 13 |
| Burkina Faso | 34 | 17 | Mauritius | 15 | 14 |
| Croatia | 34 | 38 | Belarus | 14 | 15 |
| Bulgaria | 33 | 41 | Cyprus | 14 | 32 |
| Côte d'Ivoire | 33 | 59 | Afghanistan | 13 | 8 |
| Cameroon | 33 | 32 | Mali | 13 | 4 |
| Trinidad & Tobago | 33 | 44 | Malawi | 13 | 8 |
| Benin | 32 | 18 | Azerbaijan | 12 | 8 |
| Panama | 31 | 24 | Laos | 12 | 23 |
| Venezuela | 30 | 14 | Syria | 12 | 26 |
| Madagascar | 29 | 33 | Senegal | 11 | 40 |
| Kazakhstan | 28 | 14 | Bosnia & Herzegovina | 10 | 14 |
| Lebanon | 28 | 28 | Paraguay | 10 | 33 |
| Zambia | 28 | 19 | Yemen | 10 | 3 |
| Guatemala | 27 | 43 | Botswana | 9 | 12 |
| Sudan | 27 | 34 | Honduras | 9 | 15 |
| Uruguay | 27 | 19 | Togo | 9 | 16 |
| Estonia | 25 | 38 | Bahrain | 8 | 7 |
| Fiji | 25 | 16 | Dominican Republic | 8 | 103 |
| Mozambique | 24 | 58 | Moldova | 8 | 7 |
| Bhutan | 23 | 6 | Uzbekistan | 8 | 11 |
| Brunei | 22 | 36 | Monaco | 7 | 3 |
| Jamaica | 22 | 16 | Sierra Leone | 7 | 4 |
| Oman | 22 | 11 | Libya | 6 | 4 |
| Iceland | 21 | 33 | Macao, SAR of China | 6 | 8 |
| Qatar | 21 | 17 | Niger | 6 | 19 |
| Mongolia | 20 | 11 | Kyrgyzstan | 5 | 6 |
| El Salvador | 19 | 9 | Lesotho | 5 | 11 |

| Country or Region | 2020 Sessions | 2019 Sessions | Country or Region | 2020 Sessions | 2019 Sessions |
|--------------------------|---------------|---------------|------------------------|---------------|---------------|
| North Macedonia | 5 | 9 | Guinea | 2 | 1 |
| Palestine | 5 | 10 | Guadeloupe | 2 | 2 |
| Rwanda | 5 | 24 | Jersey | 2 | 1 |
| Tajikistan | 5 | | Chad | 2 | 2 |
| Albania | 4 | 6 | Antigua & Barbuda | 1 | 8 |
| Armenia | 4 | 3 | Aruba | 1 | |
| Cook Islands | 4 | 1 | Curaçao | 1 | 2 |
| Guinea-Bissau | 4 | 1 | Djibouti | 1 | 7 |
| St. Lucia | 4 | | Dominica | 1 | |
| Montenegro | 4 | 3 | Micronesia | 1 | 1 |
| Réunion | 4 | 9 | Gabon | 1 | 1 |
| Suriname | 4 | 4 | Greenland | 1 | |
| Turkmenistan | 4 | | Guam | 1 | |
| Kosovo | 4 | 13 | Liberia | 1 | 2 |
| Barbados | 3 | 7 | St. Martin | 1 | |
| Bahamas | 3 | 1 | Marshall Islands | 1 | |
| Belize | 3 | 5 | Mauritania | 1 | 1 |
| Congo - Kinshasa | 3 | 12 | Maldives | 1 | 2 |
| Central African Republic | 3 | | Papua New Guinea | 1 | 8 |
| Congo - Brazzaville | 3 | | Svalbard & Jan Mayen | 1 | |
| Faroe Islands | 3 | 7 | Eswatini | 1 | 4 |
| Grenada | 3 | 3 | British Virgin Islands | 1 | 1 |
| Gambia | 3 | 2 | U.S. Virgin Islands | 1 | 1 |
| Haiti | 3 | 4 | Vanuatu | | 1 |
| Comoros | 3 | 5 | Burundi | | 1 |
| Liechtenstein | 3 | 6 | Anguilla | | 1 |
| Puerto Rico | 3 | 19 | | | |
| Seychelles | 3 | 1 | | | |
| Somalia | 3 | 3 | | | |
| Sint Maarten | 3 | | | | |
| Timor-Leste | 3 | 2 | | | |
| Andorra | 2 | 1 | | | |
| Angola | 2 | 2 | | | |
| Bermuda | 2 | 2 | | | |
| Gibraltar | 2 | 2 | Total | 40,996 | 52,286 |