Report 2021 of the DDC at DKRZ

Document ownership and history	
Owner	IPCC DDC at DKRZ (ipcc.wdc-climate.de)
Location	DDC_report_DKRZ_2021.docx
Author team:	M. Stockhause
Version	1.0
Date	2021-01-26

Version history

Date	Version	Comment
2021-01-26	1.0	Final version
2021-01-24	0.1	Draft for internal review

Table of Contents

1. 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, ≤ 10 GBytes each for FAR, SAR, and TAR, and 35 TBytes for SR1.5 data.

From mid 2020 to January 2021, massive downloads of a single dataset from a single European ESGF user occurred. These technical downloads were carefully eliminated from the statistics. IPCC DDC users downloaded ca. 420 TBytes and 820 000 datasets with mean monthly downloads of 68 000 datasets/month and 35 TBytes/month. The total download volume in 2021 decreased by ca 50 % compared to 2020, getting back the data download level of 2019. Data downloads were dominated by AR5 downloads via the ESGF with a share of > 90 %. Downloads for AR5 data in 2021 were about half of the downloads in 2020 in volume. Download volumes show increases from 2020 to 2021 for the other ARs of: ca. 18% for SAR, ca. 86% for TAR and ca. 32% for AR4. DDC users requested no data for selected areas on storage media in 2021.

75 % of the downloads were from Asian (51 %) and North American users (24 %). The remaining quarter of the download counts in 2021 are divided between 16 % European, 8 % African, 1 % Australian and less than half a percent of South American users.

2. Evolution of data access

The user downloads from the DDC reference archive in 2020 showed a significant increase of the number of downloaded datasets, which were not reflected in the download volumes. Data analysis identified a single user of the Earth System Grid Federation (ESGF; or more precisely a single IP) downloading a single dataset in an extremely high frequency from the DDC's ESGF data node. These downloads started in July 2020, peaked in August 2020 with 1.7 million dataset downloads and stopped in mid January 2021 after contacting the institution. These technical downloads were carefully eliminated from all DDC statistics.

The total downloads in 2021 of 420 TBytes are half the downloads in 2020 based on the corrected numbers (**Figure 1**). The volume was downloaded in 820 000 individual datasets. The high download volume in 2020 is mainly caused by the high downloads in spring 2020. The mean monthly downloads in 2021 were ca. 68 000 datasets/month and 35 TBytes/month, which is about 57% of the monthly mean dataset downloads and about half the monthly mean download volume in 2020. Apart from the download peak in 04/2020, for which the reason is unknown, there was no general trend in the monthly downloads observed in the years 2020 and 2021.

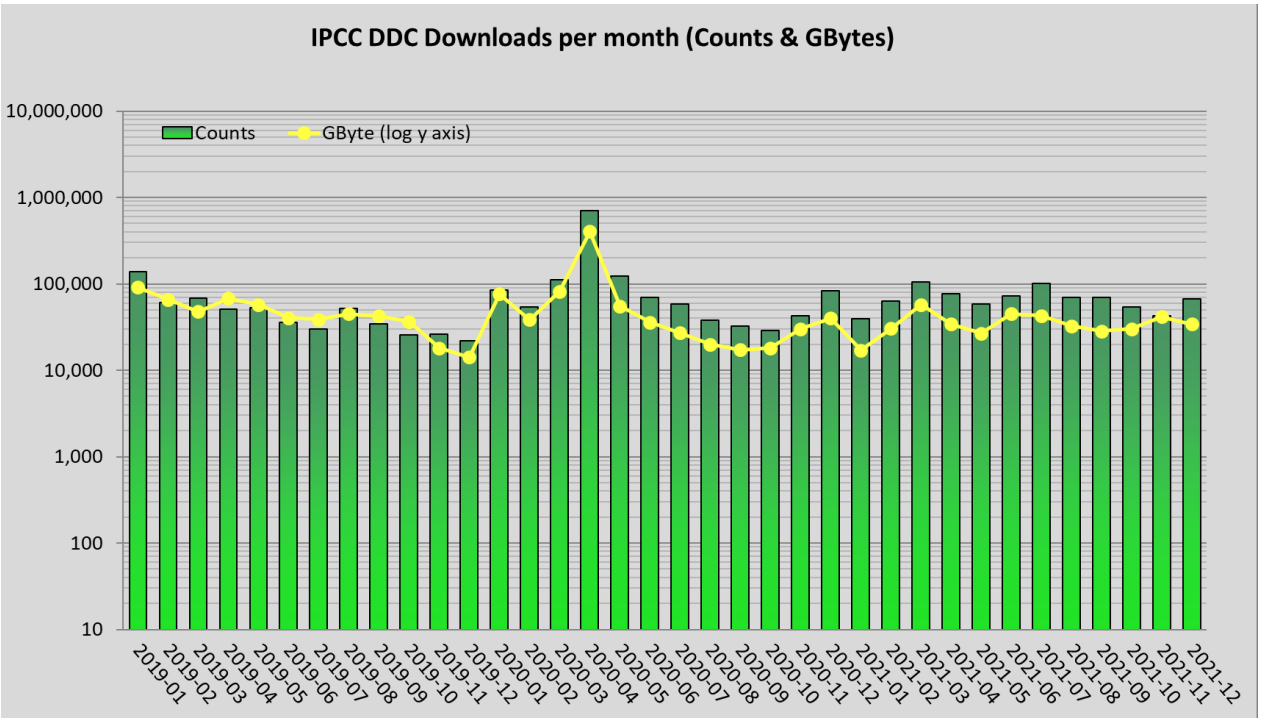

Figure 1: *Total data download counts and volumes per months over the last three years in GBytes from the IPCC DDC reference archive.*

3. Geographical distribution of data access

For the IPCC DDC AR5 data, direct data access at the DKRZ and access via ESGF are supported. Downloads from ESGF dominate the statistics with > 90 % of the downloads. The ESGF file downloads in 2021 were merged with the DDC continental download information (**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 2021.

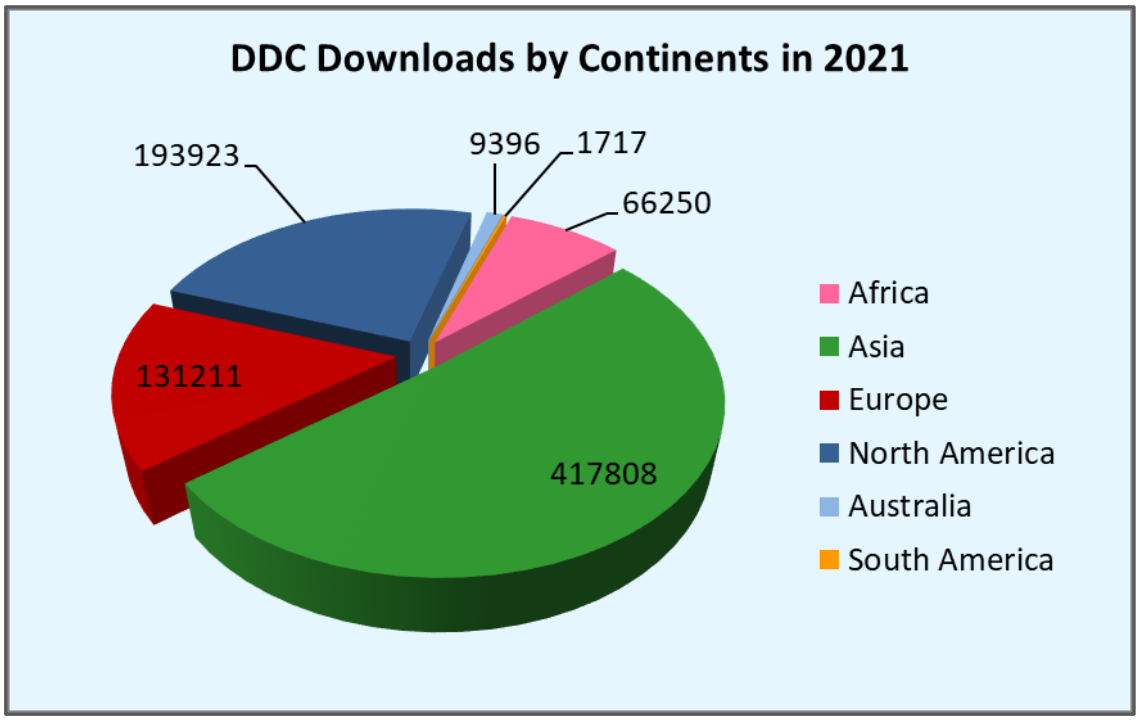

***Figure 2**: Number of dataset downloads (download counts) per continent for 2021.*

The download counts per continent show that 51 % of the total downloads were from Asian and 24 % from North American users (**Figure 2**). The remaining quarter of the download counts in 2021 are divided between 16 % European, 8 % African, 1 % Australian and less than half a percent of South American users. Downloads from Africa, Asia and South America, which can be roughly regarded as developing and economy-in-transition countries, add up to a 59 % share of the total download counts. The percentage of North American users among the ESGF portal users is significantly higher than among DDC portal users and the percentage of European users has the opposite trend.

3.1 Data on storage media

DDC users requested no data for selected areas on storage media in 2021.

4. Data access by category AR

The monthly download rates in 2021 from the IPCC DDC Reference Data Archive were dominated by AR5 downloads as in the previous years (**Figure 3**; online monthly download statistics[1]). The majority of AR5 data (> 90 %) were downloaded via DKRZ's DDC ESGF data node. The corrected numbers show a total download volume of ca. 420 TByte for AR5 and 2021, which is about half of that in 2020 and about 74 % of that in 2019. The general trend is a decrease of the downloaded

[1] 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

data volume for AR5 over the past five years with the exception of an intermediate increase observed in 2020. The individual 820 000 AR5 file downloads in 2021 are 43% less than the number in 2020 but 40% higher than in 2019. This indicates that the average size of a downloaded AR5 file in 2021 was slightly larger than in 2020.

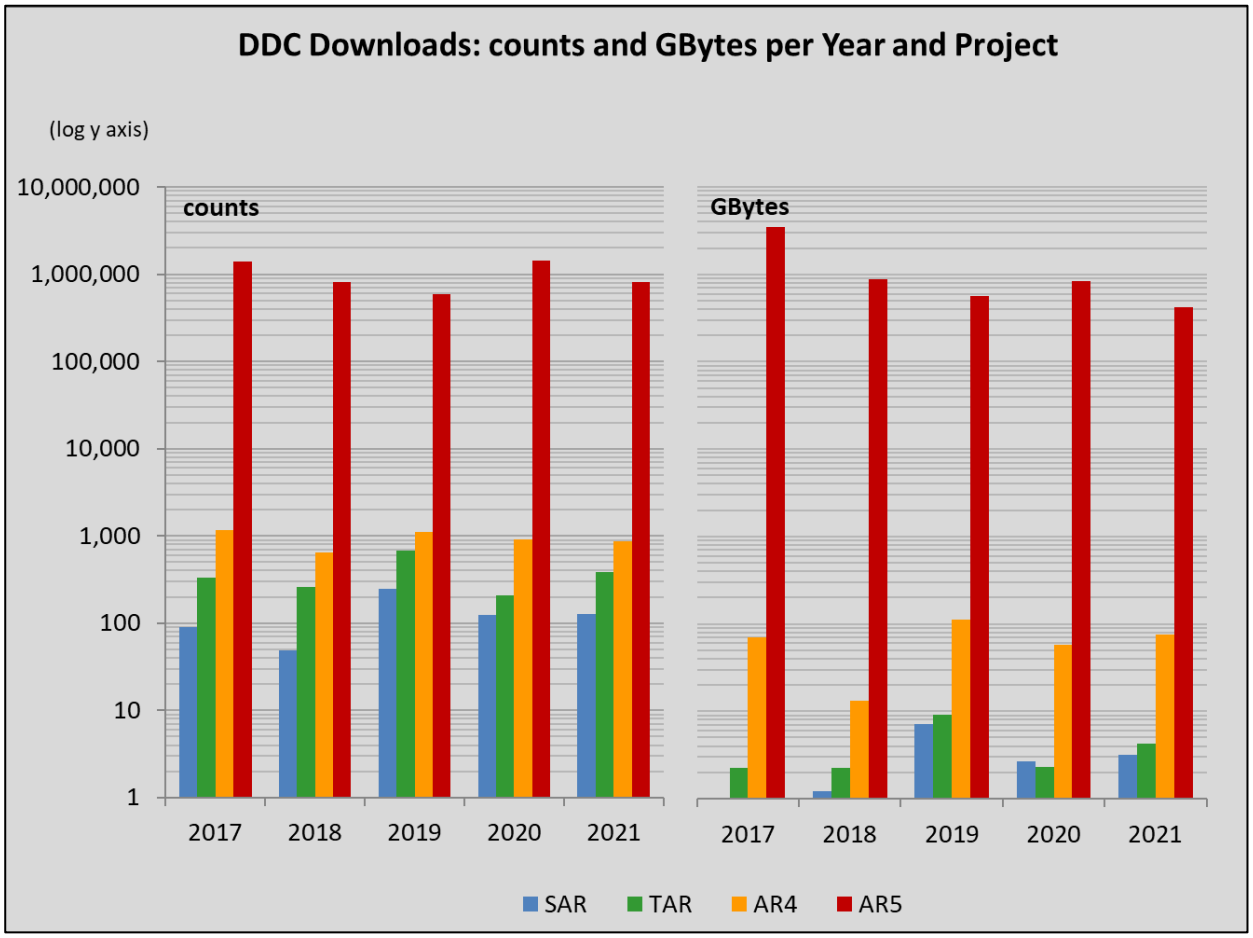

Figure 3: Total annual data download counts (left) and volumes in GBytes (right) over the last five years for the different DDC reference archives (without FAR and SR1.5).

The data download numbers for previous ARs show no significant change in download numbers for SAR and AR4, and an increase of ca. 85% for TAR from 2020 to 2021. Download volumes show increases from 2020 to 2021 of ca. 18% for SAR, of ca. 86% for TAR and of ca. 32% for AR4. Because of the overall low download volume for SAR, TAR and AR4 this increase has no influence on the trend in the total DDC data downloads.

5. Review of user queries

User requests are directed to the DDC Partner DKRZ and for AR5 data partly to the ESGF support. A separation of user requests on IPCC DDC issues is not possible. The joint DDC support infrastructure was set up end of 2021. No requests were forwarded to DKRZ from there.

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

6. News and activities

The FAR data of the institution 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 has now a total volume of 10 GBytes.