

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	Version	Comment
2021-02-04	1.0	Report completed incl. internal review

Table of Contents

1.	Summary	1
2.	Evolution of data access	2
3.	Geographical distribution of data access	2
4.	Data access by category AR	4
5.	Review of user queries	4
6.	News and activities.....	4

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, 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 %.

2. 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.

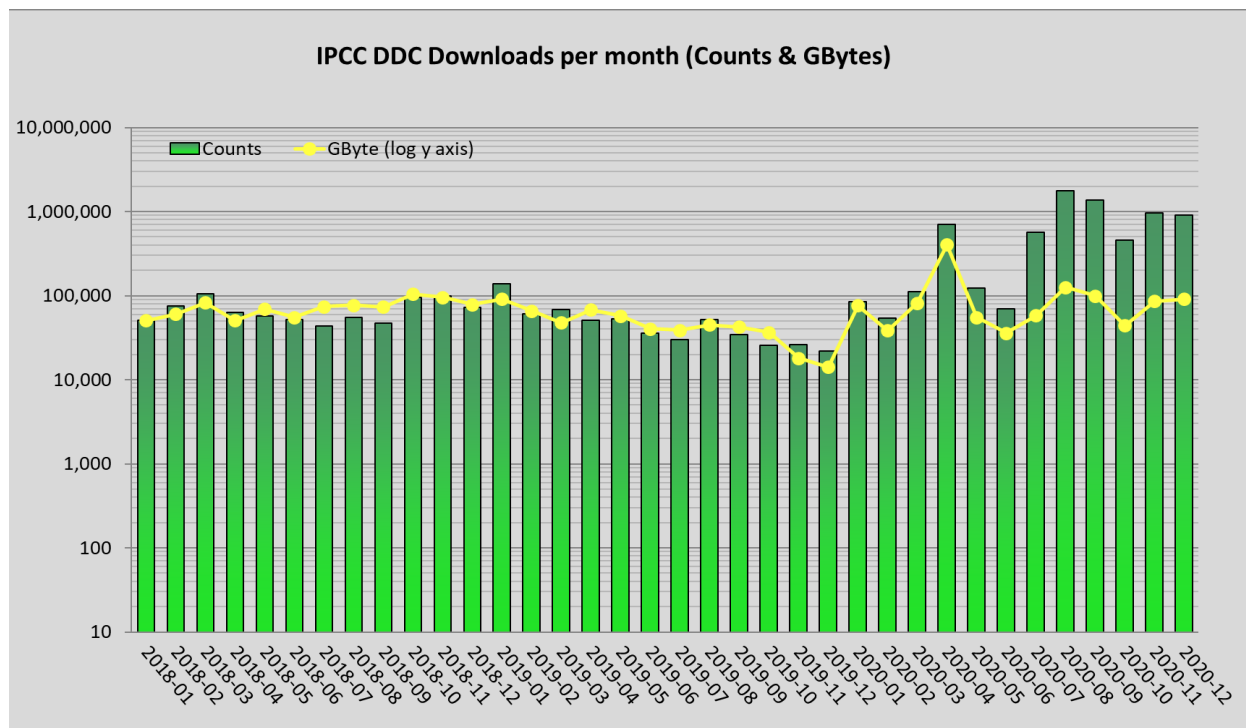


Figure 1: Total data download counts and volumes per months over the last three years in GBytes from the IPCC DDC reference archive.

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.

3. 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.

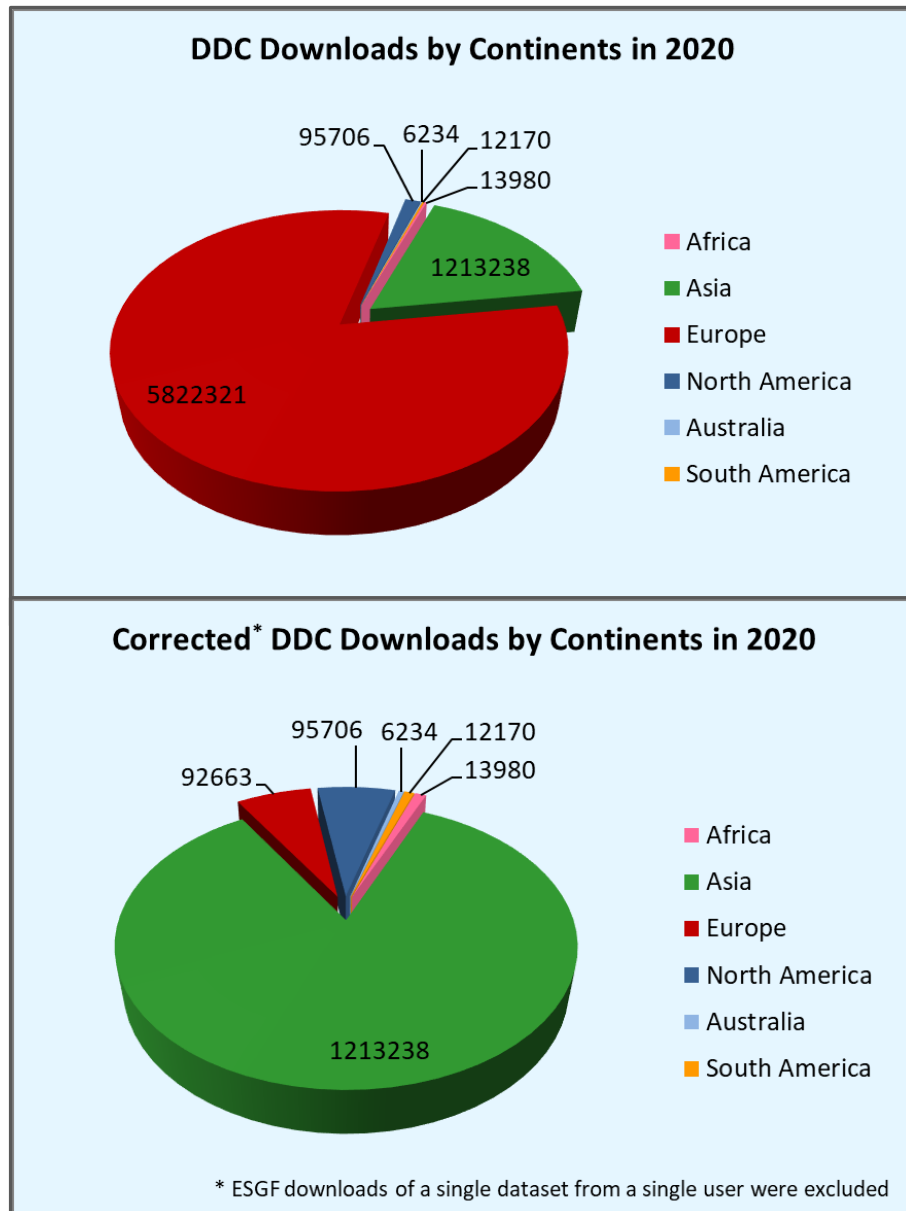


Figure 2: Downloads counts of users per continent for 2020 (top: uncorrected values; bottom: values corrected by exclusion of single European ESGF user downloads).

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.

3.1 Data on storage media

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

4. 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 statistics¹). 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.

5. 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.

6. News and activities

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

¹ Online monthly download statistics are available at:

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

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

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

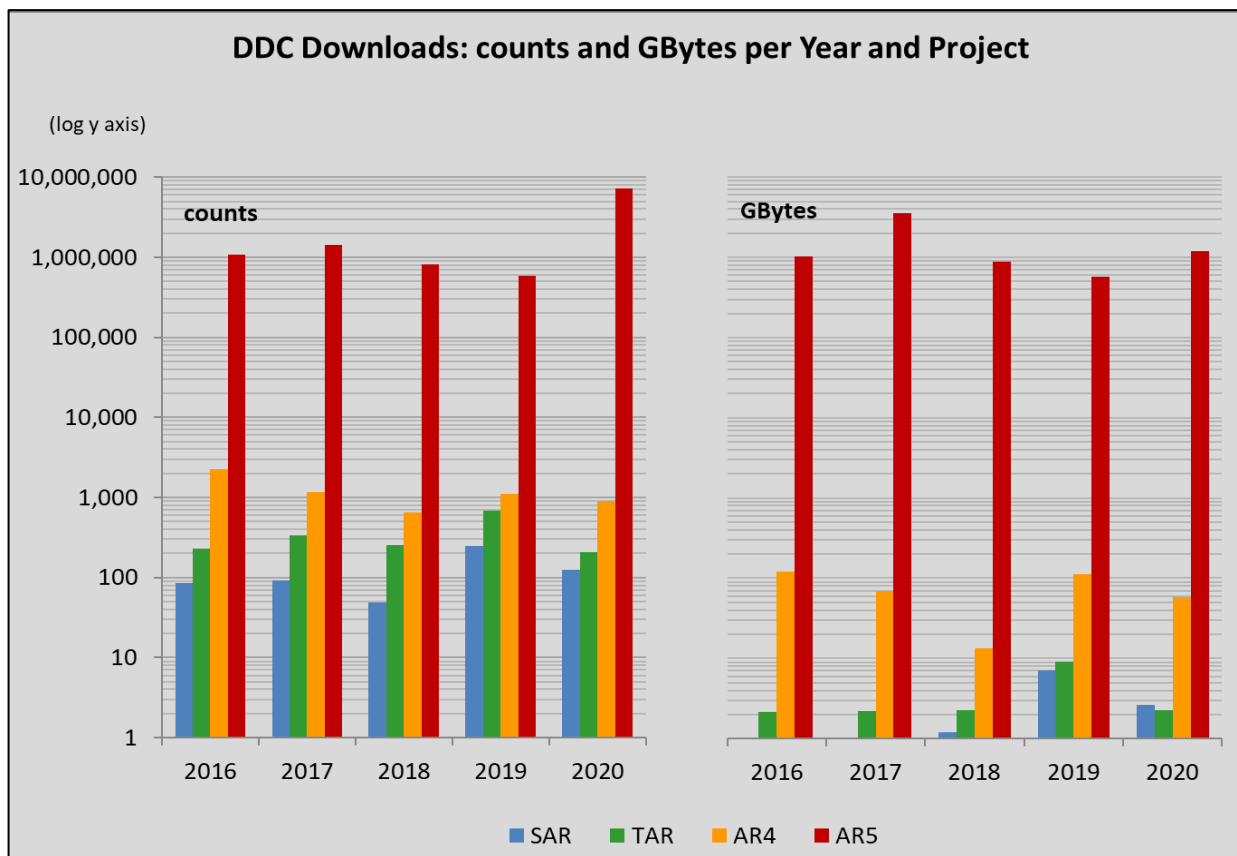


Figure 3: Total annual data download counts (left) and volumes in GBytes (right) over the last five years for the different DDC reference archives (without FAR and SR1.5).