

Project: **1017**

Project title: **Konsortialdatenprojekt CMIP-DKRZ-Datenpool**

Principal investigator: **Stephan Kindermann**

Report period: **2020-01-01 to 2020-12-31**

Text: maximum of two pages including figures.

Im Rahmen des Konsortialdatenprojektes 1017 wurden im Berichtszeitraum wichtige CMIP6 Datenbestände am DKRZ zur direkten Nutzung der deutschen Klimaforschungsgemeinschaft zusammengeführt. Diese umfassen aktuell (Okt. 2020) ca. 4 PByte, die sich aus ca. 1.5 PByte deutscher CMIP6 Beiträge und 2.5 PByte CMIP6 Replikate externer Modellierungsgruppen zusammenfassen (siehe Abbildung 1). Die zusammengeführten Replikate reflektieren zum Einen die am häufigsten zugriffenen Variablen und CMIP6 Läufe und zum anderen spezifische Anforderungen von deutschen Nutzergruppen, die u.A. über die support Adresse data-pool@dkrz.de am DKRZ angemeldet wurden.

Die zur Verfügung stehenden übrigen 1PByte wurden für die operationellen Abläufe rund um den Datenpool verwendet: temporäre Zwischenspeicherung von Datenkollektionen vor dem Eintrag in den Datenpool (z.B. als Teil der Replikation und zur Qualitätsprüfung sowie Vervollständigung), Bereitstellung von temporären Daten, die zur CMIP6 Datenauswertung angefragt wurden (z.B. CMIP5 Vergleichsdaten). Hiermit wurden alle dem CMIP-DKRZ-Datenpool zugeordneten Ressourcen (5 PByte) im Jahr 2020 genutzt. Aktuell (Oktober 2020) sind 4.6 PByte belegt. Die CMIP6 Datenauswertung durch die Forschungsgemeinschaft ist aktuell noch in vollem Gange, anstehende und zukünftige Anforderungen zur Bereitstellung von CMIP6 Datenkollektionen als Teil des CMIP-DKRZ-Datenpools werden nur erfüllt werden können indem weniger oft genutzte Daten des pools auf tape ausgelagert werden.

Der CMIP-DKRZ-Datenpool bildet zudem wesentliche Grundlage des zukünftigen CMIP6 Langzeitarchives am DKRZ (als Teil des WDCC und IPCC DDC). Der Archivierungsworkflow hierfür wird 2021 beginnen.

Nutzung des Datenpools:

Der Datenpool wurde im Jahr 2020 durch eine grosse Forschungsgemeinschaft genutzt:

- Über die ESGF Schnittstellen wurden mehr als 1.4 PByte von Nutzern aus aller Welt zugriffenen, Zugriffstatistiken sind unter <http://esgf-ui.cmcc.it/esgf-dashboard-ui/cmip6.html> einsehbar (unter Datenknoten den DKRZ CMIP6 Datenknoten names esgf3.dkrz.de auswählen).
- Die direkte Nutzung durch Forscher mit DKRZ account ist schwer zu quantifizieren, da die Datenbestände für jeden DKRZ Nutzer frei zugreifbar sind als Teil des Mistral filesystems. Jedoch belegen die Nachfragen über die support Adressen ein reges Interesse.
- Über die service Aktivitäten des Horizon 2020 Projektes IS-ENES3 wurden zudem internationale Forschergruppen in der Nutzung des Datenpools in Datenauswerteprojekten unterstützt.
- Spezifischer support wurde geleistet zur Unterstützung von IPCC working groups zur zeitnahen Datenauswertung
- Der Datenpool bildet die Grundlage zur Bereitstellung von CMIP6 Modell-Evaluierungsergebnissen auf Basis des ESMValTools (siehe <https://cmip-esmvaltool.dkrz.de/history/cmip6-results/>)
- Über IS-ENES3 und ESGF ist der DKRZ Datenpool zudem in die international abgestimmte Bereitstellung von CMIP6 Datenreplikaten integriert.

CMIP6 Storage at DKRZ

(replicated & primary)

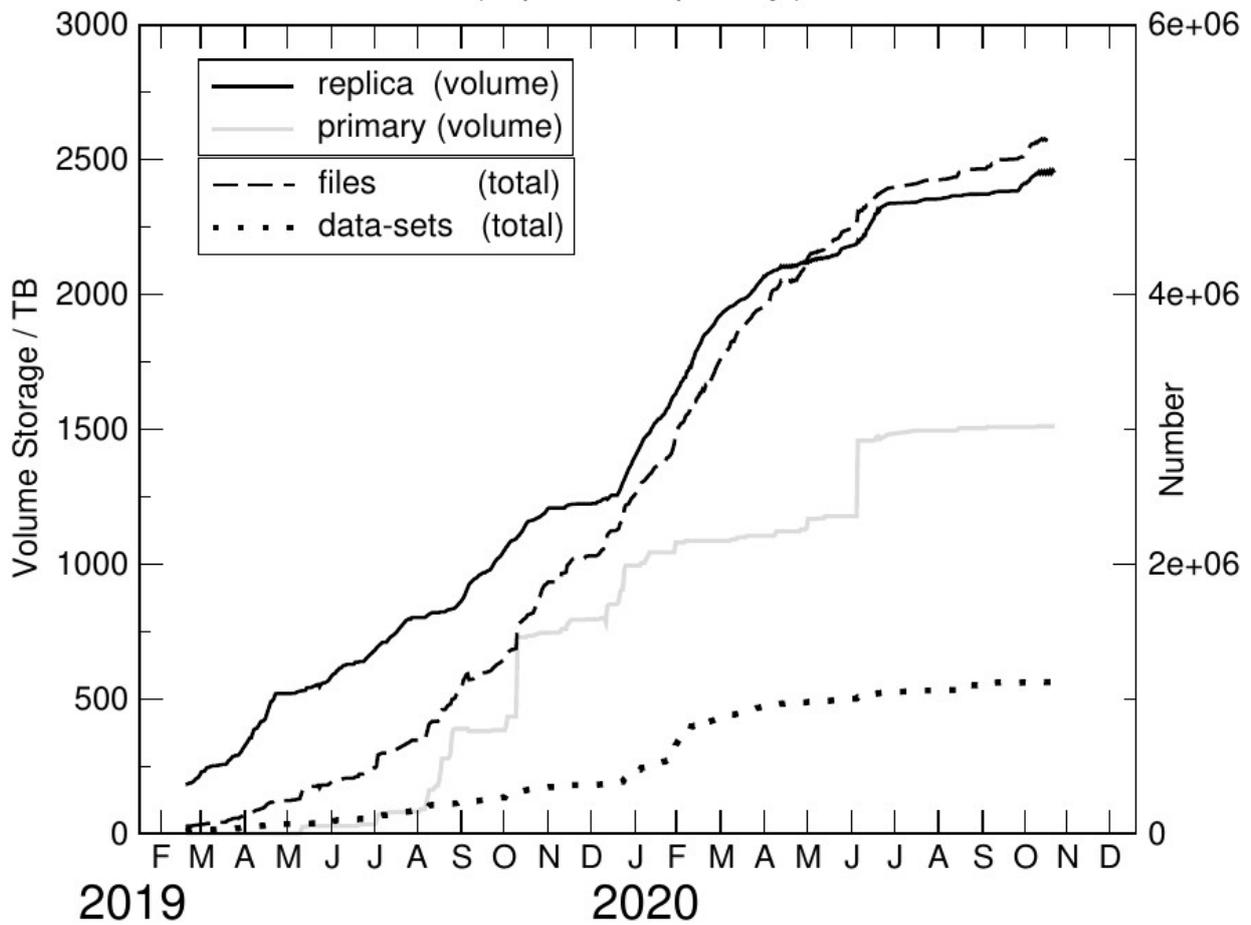


Abbildung 1: Bereitstellung von CMIP6 Datenbeständen als Teil des DKRZ-CMIP6 Datenpools, die bereitgestellten Primärdaten (ca. 1.5 PByte – helle Linie, z.B. CMIP6 Ergebnisse der deutschen Forschungsgemeinschaft) sind getrennt von den bereitgestellten Replikaten (ca 2.5 PByte – dunkle Linie z.B. CMIP6 Ergebnisse internationaler Modellierungsgruppen) dargestellt.