

Project: **21**

Project title: **ESGF - Datenintegration**

Project lead: **Stephanie Legutke**

Allocation period: **1.1.2016 - 31.12.2016**

Project overview

Für die Archivierung der CMIP5 Daten wurde ein Datenformat entwickelt, das Suche und Herunterladen auf den Portalen der ESG-Föderation unterstützt. Dieses Format hat sich zu einem Quasi-Standard entwickelt, der von immer mehr MIPs (Modellvergleichsprojekten) übernommen wurde (z. B. CORDEX, EUCLIPSE, CFMIP, PMIP3, MiKlip, LUCID, GeoMIP, TAMIP, CDMIP). Auch Beobachtungs- und (Re-)Analysedaten werden mittlerweile in diesem Format erzeugt (z. B. obs4MIPs, ana4MIPs). CMIP6 einschließlich der unterstützten MIPs ('endorsedMIPs') wird bis auf einige Modifikationen bei dem Standard bleiben.

Für die CMIP-Datenformatierung wurde am PCMDI (LLNL) eine Software entwickelt (CMOR), deren Anwendung und Erweiterung einerseits die Fehlerquote vermindert, deren Benutzung aber auch einige Expertise verlangt. Wegen der speziellen Archivsoftware und der damit einhergehenden Nutzerunterstützung muss dieser Standard strikt eingehalten werden.

Die benötigte Expertise hat die Abteilung DM am DKRZ während der Aufbereitung der CMIP5 und CORDEX Daten erworben, und wird oft um Unterstützung bei der Datenaufbereitung gebeten wird.

Range of planned work from the scientific view

Die Arbeiten im Projekt 21 haben sich 2014 und 2015, und werden sich auch 2016 weiter mit der Nutzerunterstützung bei der Datenaufbereitung befassen. Um entsprechende Anfragen zeitnah erledigen zu können, wird für das Projekt 21 RZ und Plattenplatz benötigt. Soweit wie möglich wird der /scratch Plattenbereich dafür benutzt, aber wegen der erforderlichen Kommunikation mit den Datenlieferanten oder mit der betreffenden Projektkoordination können längere Pausen in der Bearbeitung entstehen, wodurch die Benutzung von /scratch/... nicht praktikabel ist.

Der Schwerpunkt der Aktivitäten in Projekt 21 wird allerdings in 2016 auf der Vorbereitung von Unterstützungsaktivitäten der deutschen ESM-Forschung im Rahmen des CMIP6-Projekts liegen, für das gegenwärtig Förderung beim BMBF beantragt wird.

Die benötigten Ressourcen sind schwer abzuschätzen, und richten sich deshalb in ihrer Höhe nach den in 2015 auf der Blizzard verbrauchten Ressourcen.

Mathematical or computational aspects; algorithmic /mathematical/numerical methods and solution procedures;

Wir benutzen bei der Arbeit diejenige Software, die von Projektseite vorgegeben ist.

Particular suitability to solve the problem with help of HLRE-3

Die Daten werden meistens auf den Plattformen des DKRZs, also jetzt HLRE3 erzeugt. Es ist teilweise wg. Platzproblemen teilweise wg. des Datentransfers nicht ohne weiteres möglich, die Daten auf anderen Plattformen aufzubereiten.

Performance benefits depending on the number of used CPUs (scalability)

Im Projekt werden nur Modelle ausgeführt, die das Hauptwerkzeug in anderen DKRZ-Projekten sind, und in dem Zusammenhang ihre Skalierbarkeit nachweisen. Darüber hinaus wird bei der Datenaufbereitung OpenMP-Parallelität im Rahmen der Benutzung der CDOs aufgerufen, oder

Parallelität über Background-Prozesse ausgenutzt. Auf der Blizzard haben wir gesehen, dass eine Anzahl von 16 Cores (1/2 Knoten) die Skalierungsgrenze für die Verarbeitung der Daten einer Modellkomponente ist. Eine höhere Anzahl von CPUs war für die Datenverarbeitung nicht nötig, da sie nicht schneller sein muss als die Modellintegration. Auf der Mistral ist die Auslotung der optimalen Anzahl von Cores bei der Datenaufbereitung für CMIP6 Teil der 21-Projektarbeit.

Required computing time and amount of storage space

(Make clear how each part of your scientific problem translates into specific requirements for computing time and storage space.)

Hochgerechnet auf 12 Monate, wurden für die Arbeiten in diesem Bereich **15.000 CPUh an RZ** verbraucht. In diese Zahl sind auch die Verbräuche von ik0555 eingeflossen, da in vielen Fällen die Skripte aus der CMIP5-Projektphase nicht umgestellt wurden. Voraussichtlich wird der Verbrauch unter dem bm0021-Account also diese Größe haben.

Z.Z. sind im /work/bm0021 – Bereich 4 TB Daten gespeichert. Das ist das Minimum der benötigten Daten, die gerade vom temporären Bereich /mnt/... auf /work/ik0021 kopiert wurden. Die maximale Belegung war in 2015 bei 11 TB. Deswegen wird ein **Speicherplatz von 15 TB** als angemessen erachtet. Dabei ist zu bedenken, dass sich die Auflösung der Modelle für die nächste CMIP-Runde deutlich erhöhen wird.

Speicherplatz im HPSS wird benötigt, um Testdaten auch auf längere Sicht im Zugriff zu haben. Hierbei sollten **5 TB** ausreichend sein.

Additional value compared to other projects

Die Aktivitäten dienen dazu, andere Projekte von nicht-wissenschaftlichen Arbeiten zu entlasten, und für einen reibungsfreien Eingang der Daten in den ESGF-Datenknoten am DKRZ zu sorgen. Hierdurch werden die Daten (DKRZ-)externen Nutzern zugänglich.

Es wird RZ von

- 625 Knotenstunden beantragt,

sowie Speicherplatz im Lustre Filesystem für

- 15.000 GB,

und im HPSS für

- 5.000 GB.