

Project: **21**

Project title: **ESGF - Datenintegration**

Project lead: **Stephanie Legutke**

Allocation period: **2018-01-01 to 2018-12-31**

## 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. input4MIPs, 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 von CMIP3, CMIP5 und CORDEX Daten erworben, und wird deshalb oft um Unterstützung bei der Datenaufbereitung gebeten. Rechnungen dazu werden mit dem Account dieses Projekts durchgeführt.

Das BMBF fördert seit Juli 2016 die Entwicklung einer Infrastruktur zur Erzeugung und Formatierung von Modellergebnissen entsprechend dem CMIP6-Standard. Für die im geförderten Vorhaben geplanten Modellexperimente werden im RZ-Antrag 988 Rechnerressourcen beantragt. Die Vorbereitungen der Rechnungen, Tests der zu entwickelnden Infrastruktur werden über den Account bm0021 dieses DKRZ-Projekts abgerechnet. Damit kann und soll auch vermieden werden, dass alle Mitarbeiter, die in die Entwicklung oder die Tests involviert sind, in der Gruppe bk0988 sein müssen.

## Range of planned work from the scientific view

Die Arbeiten im Projekt 21 werden sich auch in 2018, wie in den Vorjahren, weiter mit der Nutzerunterstützung bei der Datenaufbereitung befassen. Der Umfang der geplanten Arbeiten, insbesondere vom wissenschaftlichen Standpunkt her, entspricht also dem der Aktivitäten der unterstützten Nutzer.

## 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. Für die Datenaufbereitung gibt es eine Skalierungsgrenze, da sie wg. des parallelen Workflows von Modellintegration und Datenaufbereitung nicht schneller sein muss als die Integration des Modells. 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

Hochgerechnet auf 12 Monate wurden 2016 für die Arbeiten in diesem Bereich **85 Nh an RZ** auf der Mistral verbraucht. Die im letzten Antrag für 2017 prognostizierte Zunahme der CMIP6-Projektaktivitäten hat ja aus diversen Gründen nicht stattgefunden. Sie wird also vermutlich in 2018 stattfinden, d.h. der Verbrauch an RZ in 2017 ist nicht charakteristisch. Wir beantragen deshalb dieselben Ressourcen wie für 2017, auch wenn sie nicht abgerufen wurden. 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/... risikobehaftet, also nicht praktikabel ist.

Der Schwerpunkt der Aktivitäten in Projekt 21 wird in 2018 auf der Entwicklung von Unterstützungsaktivitäten der deutschen ESM-Forschung im Rahmen der BMBF-geförderten CMIP6-Aktivitäten liegen. Wie für den Speicherplatz gilt auch für die RZ, dass die beantragten Ressourcen in 2017 nicht abgerufen wurden, da bis Oktober praktische keine CMIP6-Experimentrechnungen durchgeführt wurden. Da diese jetzt Ende 2017 endgültig starten werden, werden für 2018 dieselben RZ-Ressourcen wie für 2017 beantragt.

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 für 2018 wie für 2017 RZ von

- 300 Knotenstunden beantragt,

sowie Speicherplatz im Lustre Filesystem für

- 18.000 GB,

und im HPSS für

- 5.000 GB.