

Verbundprojekt CMIP6

**Bereitstellung des nationalen Beitrags zur Datenbasis des IPCC/AR6
und Unterstützung der CMIP6⁺-Aktivitäten in Deutschland**

(Juli 2016 - Juni 2020)



Verbund 1: CMIP6-DICAD

Abschlusstreffen

22.06.2020

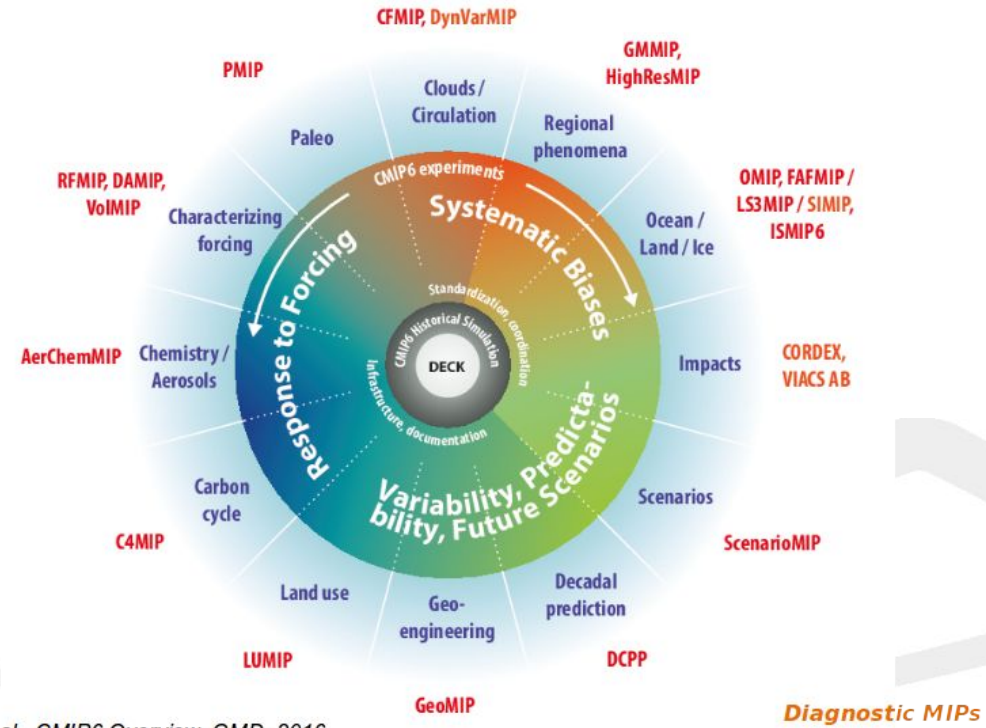
(Kostenneutrale Verlängerung bis Ende 2020 beantragt.)

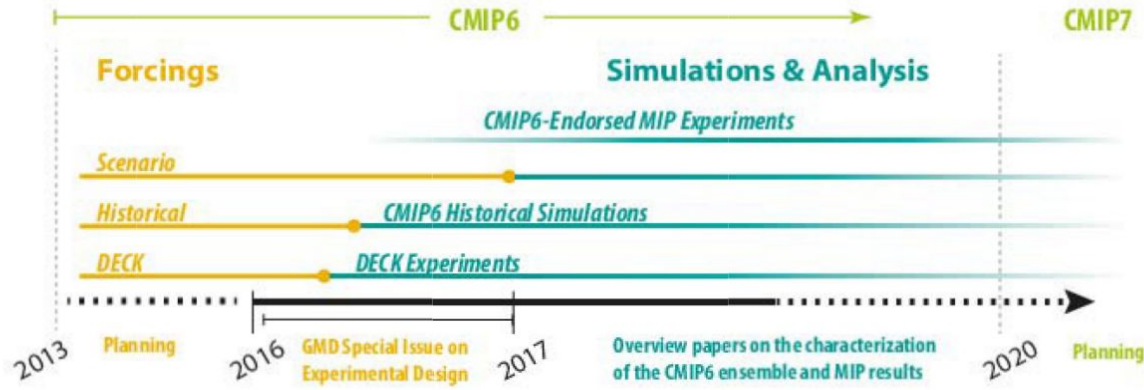
Allgemeines

- Abgesehen vom Vortragenden, Webcams und Mikrofone bitte ausstellen.
- Fragen zu den Vorträgen bitte in dem in der Agenda verlinkten GoogleDoc stellen. Bitte dabei den Namen angeben.
- Fragen werden im Anschluss an den Vortrag soweit es die Zeit erlaubt mündlich beantwortet. Diese und wenn nötig weitere Antworten werden später in das GoogleDoc nachgetragen.

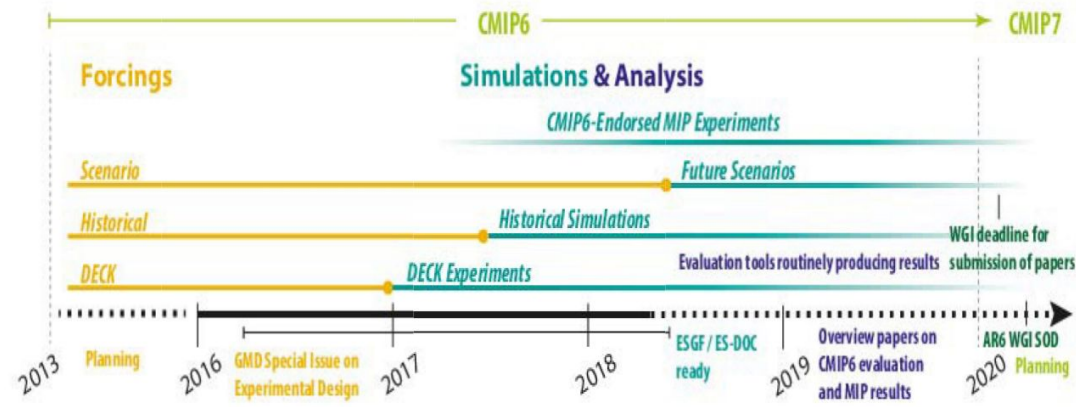
CMIP6 Ziele und Organisation

- CMIP ist ein Projekt der Working Group on Coupled Modeling (WGCM) des World Climate Research Programme (WCRP).
- Das Hauptziel ist den wissenschaftliche Verständnis des Erdsystems zu verbessern.
- CMIP koordiniert seit 1995 internationale Modellvergleichsprojekte.
- So konnte zu einem besseren Verständnis vergangener, gegenwärtiger und zukünftiger Klimaveränderungen und der Variabilität in einem Multi-Modell-Rahmen beitragen
- Die zu den einzelnen CMIP-Phasen erstellten Modellsimulationen wurden auch regelmäßig als Teil der IPCC-Klimasachstandsberichte ausgewertet.





[Eyring et al. 2016](#)



[Eyring 2019](#)

Verbundprojekt CMIP6

Bereitstellung des nationalen Beitrags zur Datenbasis des IPCC/AR6 und Unterstützung der CMIP6⁺-Aktivitäten in Deutschland

(Juli 2016 - Juni 2020)

Verbund-1 DICAD: "Bereitstellung des nationalen Beitrags zur Datenbasis des IPCC/AR6 und Unterstützung der CMIP6⁺-Aktivitäten in Deutschland" (DKRZ, DLR, DWD, FUB)

2.200.000 €

Verbund-2 Chemie: "Unterstützung des aktuellen und zukünftigen nationalen Beitrags zum IPCC/AR mit Fokus Atmosphärenchemie" (DLR, Uni Bonn)

1.200.000 €



Verbundprojekt CMIP6



Bereitstellung des nationalen Beitrags zur Datenbasis des IPCC/AR6 und Unterstützung der CMIP6⁺-Aktivitäten in Deutschland

Verbund 1: CMIP6-DICAD

- Erzeugung von Klimaänderungsdaten im Rahmen von CMIP6
- Kompetenzbildung zur Datenaufbereitung entspr. des CMIP-Standards
- Bereitstellung des deutschen Beitrages zur IPCC/AR6-Datenbasis unter Gewährleistung der Qualität und frühzeitiger Zitierbarkeit der Daten
- Bereitstellung/Replikation wichtiger internationaler Daten
- Vorbereitung der Integration der Daten in das IPCC-Referenzdatenarchiv des WDCC
- Datengrundlage für Forschung zum Klimawandel und für Regionalisierung der Klimaprojektionen und damit für nationale Entwicklungen von Vermeidungs- und Anpassungsstrategien sowie Klimafolgenforschung.
- Aufbau nachhaltiger Infrastruktur zur Unterstützung (inter)nationaler Aktivitäten zur CMIP6-Datenerzeugung und -nutzung (CMIP6 endorsed MIPs, CORDEX-CMIP6, ESMValTool/Freva)

Verbundprojekt CMIP6

Verbund 1: CMIP6-DICAD - Teilprojekte

Teil-Projekt	Titel	Partner
TP-1	Aufbau und Betrieb des Nationalen CMIP6-Datenarchivs und einer Infrastruktur zur Datenqualitätssicherung	
TP-2	Datenauswertung von CMIP6 Simulationen mit dem Earth System Model Evaluation Tool (ESMValTool)	
TP-3	ICON Klimaprojektionen der Atmosphäre mit einer feineren Gitterweite über Europa	
TP-4	Portal zur Bereitstellung der Evaluierungsergebnisse	

Verbundprojekt CMIP6

Bereitstellung des nationalen Beitrags zur Datenbasis des IPCC/AR6 und Unterstützung der CMIP6⁺-Aktivitäten in Deutschland

Verbund 1: CMIP6-DICAD - Arbeitspakete

AP1 CMIP6⁺-Experimente (*DKRZ, DWD*)

AP2 Kompatibilität der CDOs mit CMIP6⁺-Datenstandards (*DKRZ*)

AP3 Infrakstruktur zur Anpassung an CMIP6 Projektdatenstandards (*DKRZ*)

AP4 Datenqualitätsprüfung (*DKRZ*)

AP5 Nationales CMIP6⁺-Datenarchiv (*DKRZ*)

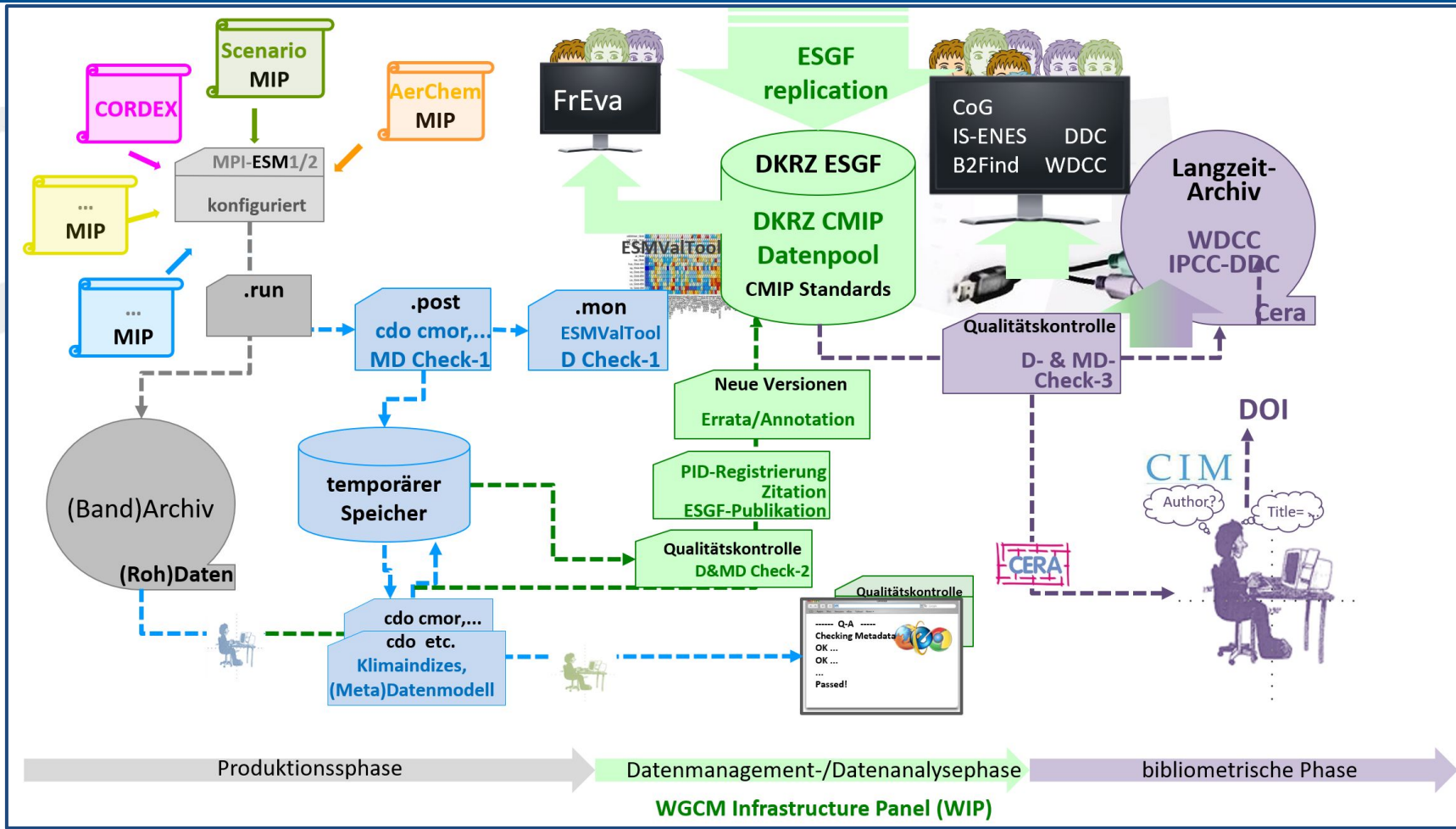
AP6 Standardisierte Diagnostiken und Modellevaluation (*DLR, FUB, DKRZ*)

AP7 Zitierfähigkeit und Dokumentation im CMIP6⁺-Archiv (*DKRZ*)

AP8 Langzeitarchivierung im WDCC/IPCC-DDC und DataCite Datenpublikation (*DKRZ*)

Agenda

09:00	Begrüßung		
09:20	CDO - cmor-Operator und Klimaindizes	Fabian Wachsmann	DKRZ
09:40	Webbasierte Zusammenstellung des Post-Processing-Workflows	Martin Schupfner	DKRZ
10:00	<u>Datenqualitätskontrolle</u> und -replikation	Heinz-Dieter Hollweg / Fabian Wachsmann	DKRZ
10:20	Datenveröffentlichung im ESGF	Stephan Kindermann / Katharina Berger	DKRZ
10:40	---- Kaffeepause ----		
11:00	Zitationservice und Langzeitarchivierung	Martina Stockhause / Amandine Kaiser	DKRZ
11:20	Erweiterung und Verbesserung des <u>ESMValTools</u>	Björn Brötz	DLR
11:40	Erste Ergebnisse der CMIP6 Evaluation mit dem ESMValTool	Lisa Bock	DLR
12:00	<u>CMIP6</u> -Datenportal	Bianca Wentzel	FUB
12:15	---- Mittagspause ----		
13:15	Modellrechnungen MPI-ESM/ICON	Johann Jungclaus / Karl-Hermann Wieners	MPI-M
13:35	Modellrechnungen AWI-CM	Tido Semmler	AWI-B
14:00	Simulationen mit ICON-A und höher aufgelösten Teilgebieten	Vera Maurer / Christian Steger	DWD
14:20	---- Kaffeepause ----		
14:30	Lessons Learned - Vorstellen der gesammelten <u>Punkte</u>	Fabian Wachsmann / Martin Schupfner	DKRZ
15:00	Allgemeine Diskussion	Fabian Wachsmann / Martin Schupfner	DKRZ
	---- ca. 15:30 Ende der Veranstaltung ----		



AP1 CMIP6⁺-Experimente (DKRZ, DWD)

DKRZ-Teil (s. Vortrag Johann Jungclaus, MPI-M):

- Durchführung und Standardisierung der CMIP6-DECK- und historical-Experimente mit MPI-ESM1-2-HR
- Durchführung und Standardisierung ausgewählter ScenarioMIP-Experimente mit MPI-ESM1-2-HR
- Durchgeführte Simulationen (DICAD Rechenprojekt 988 am DKRZ):

https://redmine.dkrz.de/projects/cmip6-dicad-subproject/wiki/Climate_Model_Simulations_-_Status

DWD-Teil (s. Vortrag Vera Maurer, DWD):

- Erweiterung von ICON-ESM: Simulationen mit höherer räumlicher Auflösung über Europa

Verbundprojekt CMIP6

Verbund 1: CMIP6-DICAD - unterstützte Modelle

AP1 DICAD- Vorhabenbeschreibung

MPI-ESM1-2-HR

(Durchführung der DECK-,
historical- und
ScenarioMIP-Simulationen)

DICAD-Rechenprojekt 988

AWI-CM-1-1-MR
EMAC-2-54-AerChem
ICON-ESM-LR
MPI-ESM1-2-HR

(DECK, historical,
ScenarioMIP)

CMIP6-Beitrag der assoziierten DICAD-Partner

AWI-CM-1-1-HR
AWI-CM-1-1-MR
AWI-CM-1-1-LR
AWI-ESM-1-1-LR
(AWI-ESM-2-1-LR)
(EMAC-2-54-AerChem)
(ICON-ESM-LR)
(ICON-MESSy)
MPI-ESM1-2-XR
MPI-ESM1-2-HR
MPI-ESM1-2-LR

CMIP6-Beitrag sonstiger Modelle

MPI-ESM-1-2-HAM
(EMAC-2-53-Vol)

*in Klammern: Bisher noch keine CMIP6-Daten im
ESGF veröffentlicht (Stand 18.6.2020)*

CMIP6 DICAD - RCM Forcing Data

- CMIP6 data that is part of the CORDEX data request for CMIP6 can be retrieved from ESGF nodes:

eg.: <https://esgf-data.dkrz.de/search/cmip6-dkrz/>

The data from above node can also be directly accessed via the filesystem of DKRZ supercomputer mistral:

“CMIP Data Pool” (up to 5 PB CMIP6 data): `/pool/data/CMIP6/data`

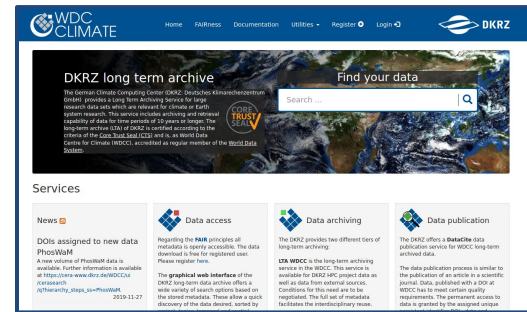
- As requested by the RCM-community additional forcing data has been published in the DKRZ CERA archive:

https://cera-www.dkrz.de/WDCC/ui/ceraresearch/project?acronym=CMIP6_RCM_forcing_MPI-ESM1-2

At the moment, the simulations performed by DKRZ and DWD with MPI-ESM1-2-HR (all scenario simulations and two amip simulations [r2/3i1p1f1]) are available. One more amip (r1i1p1f1) and historical simulation (r2i1p1f1) performed by the MPI-M will follow soon.

Contact: Martin Schupfner, DKRZ - schupfner@dkrz.de

- AWI-CM model output stored in DKRZ hpss tape archive:
`/hpss/arch/bk0988/awicm` and `/hpss/double/bk0988/awicm`
 Contact: Tido Semmler, AWI - Tido.Semmler@awi.de



Extra-Folien

CMIP6 Experiments of MPI-M and DKRZ and general information

- Status of the climate model simulations within DICAD-RZ988 is being tracked under:
https://redmine.dkrz.de/projects/cmip6-dicad-subproject/wiki/Climate_Model_Simulations_-_Status
- Monitoring (MPI-M, DKRZ): <https://modvis.dkrz.de/mh0469/cmip6/>
- The ESGF data pool is mounted on mistral under
`/work/ik1017/CMIP6/data/CMIP6/`
and contains all CMIP6 data published and replicated in the DKRZ ESGF node
DKRZ ESGF node for CMIP6: <https://esgf-data.dkrz.de/search/cmip6-dkrz/>
DKRZ ESGF node for all projects: <https://esgf-data.dkrz.de/projects/esgf-dkrz/>
- Quick looks of some published variables will be available under
<https://cmip-esmvaltool.dkrz.de/> soon
(Access soon possible with the standard DKRZ / LDAP account that is used for the mistral login.)
- Errata information about unpublished files are under
<https://errata.es-doc.org/static/index.html>

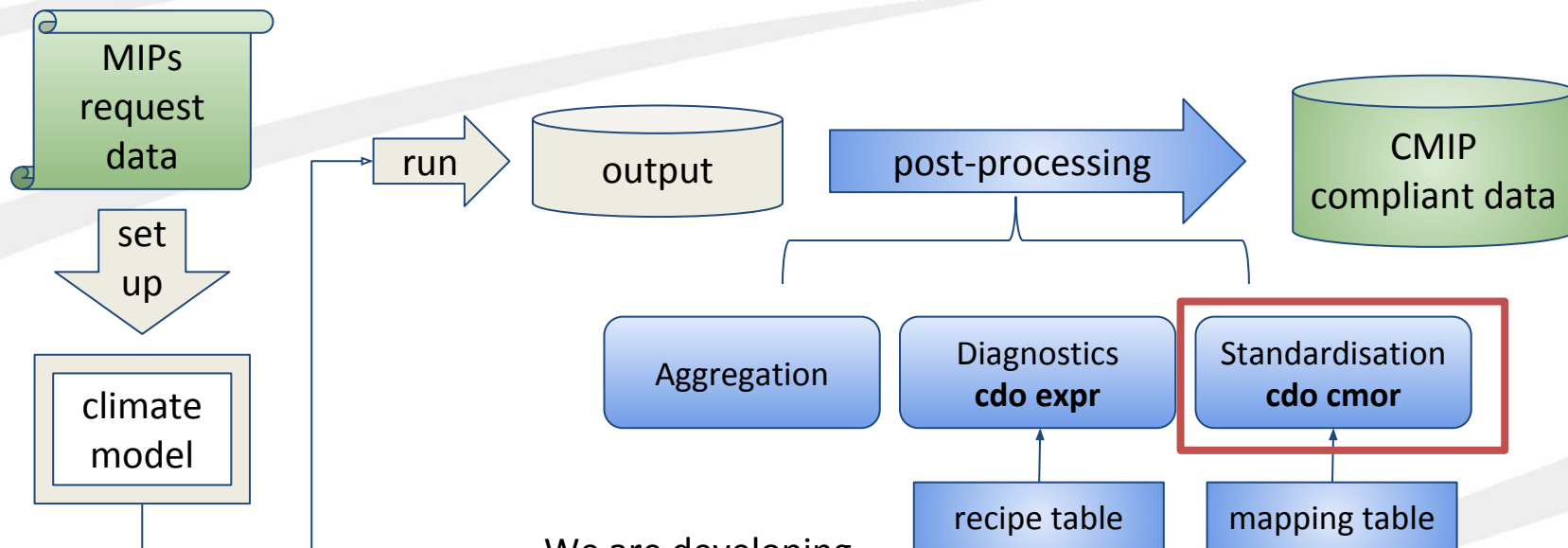
AP2 Kompatibilität der CDOs mit CMIP6⁺-Datenstandards (DKRZ)

1. Develop a CDO-Operator for standardising earth system model raw output by linking CMOR
2. Adapt the CDO-data model and metadata model to den CMIP6 Standard
3. Adapt the CDO-climate extremes indices operators to international standards

AP2 Kompatibilität der CDOs mit CMIP6⁺-Datenstandards (DKRZ)

1. **Develop a CDO-Operator for standardising earth system model raw output by linking CMOR**
2. Adapt the CDO-data model and metadata model to den CMIP6 Standard
3. Adapt the CDO-climate extremes indices operators to international standards

CMIP6 data production workflow



We are developing

- the **cdo cmor** operator
- the **c6dreq-WebGUI** to
 - create recipe/mapping tables
 - create a customized DReq
 - generate script fragments for diagnostic & standardisation

*Further steps are not displayed
(Monitoring, QA, Citation, Publication,
Long-term archiving, ...)*

CMIP6 data standard motivation

CMIP Scope

Both Scientists and their programs must be able to rely on a common data standard for **2000 unique variables** from 100 Experiments of 25 Model Intercomparison Projects (MIPs) and conducted by 100 models.

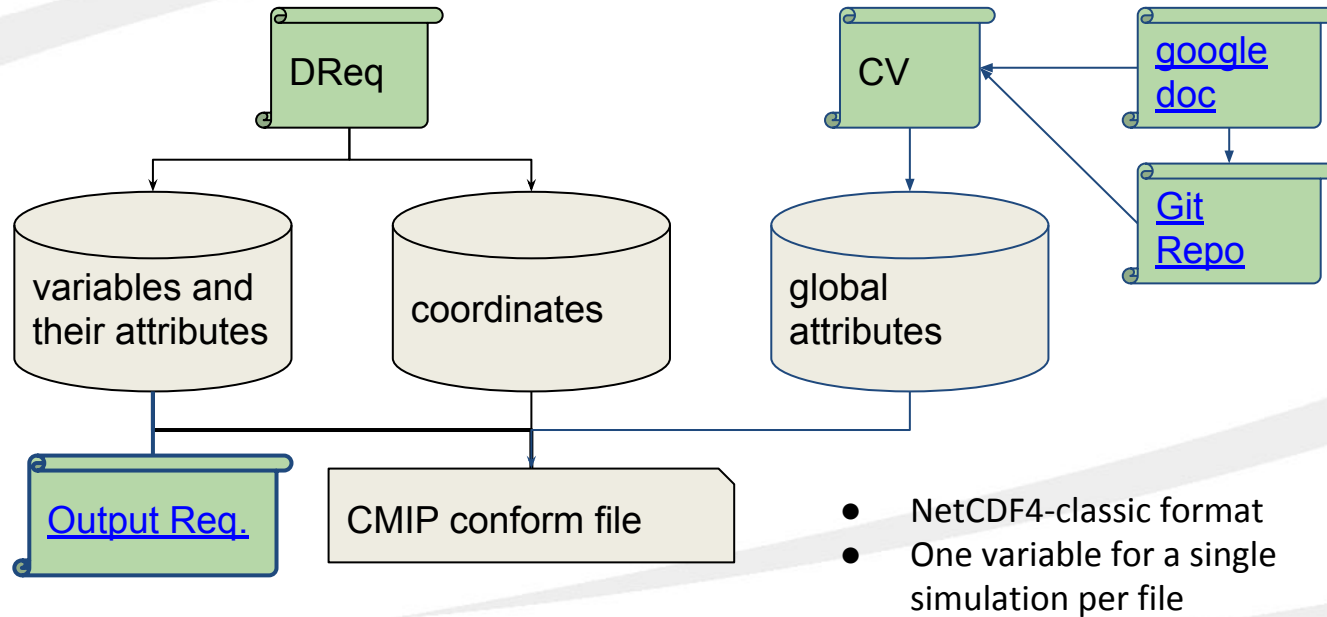
"**Systematic analysis** across models only easy to do if model output is written in

- a common format
- with files structured similarly
- and with sufficient metadata uniformly stored"

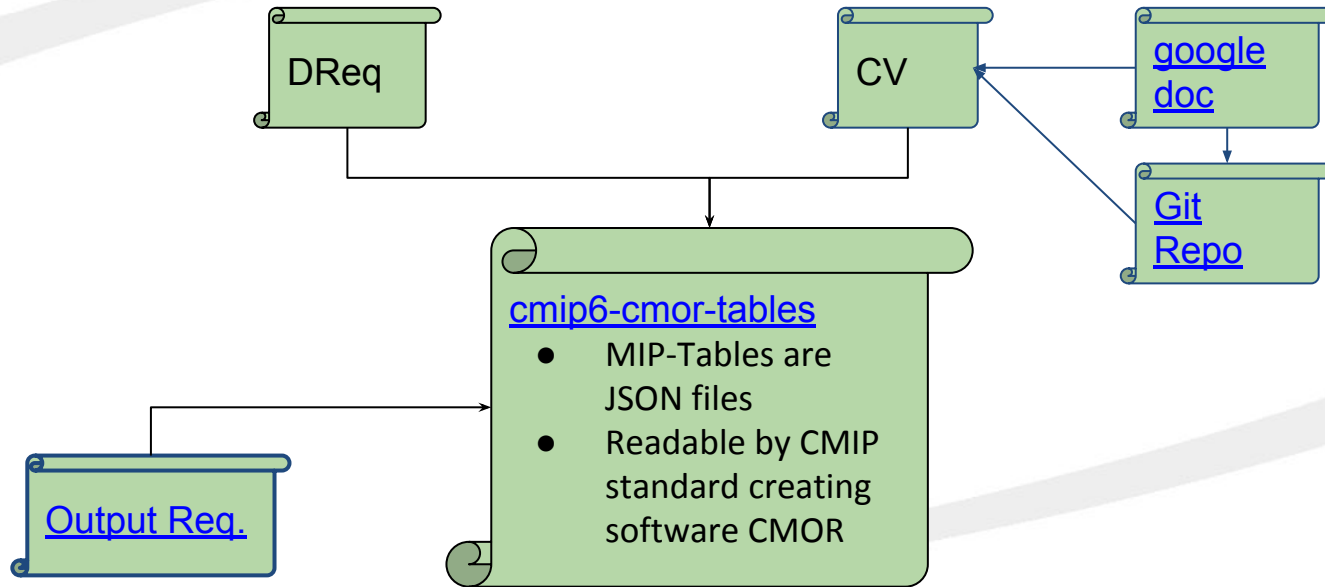
Reusability

Climate model output files should be **self descriptive**. E.g. statistical operations and interpolations over space and time are only processable if all temporal and spatial information including cells and interval bounds and a full description of the vertical axis is available in the input file.

Scheme of the CMIP6 data standard

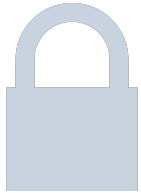


Motivation for cdo cmor



Motivation for cdo cmor

- Why use CMOR to create CMIP standard?
- Why integrate CDO with CMOR?



- CMOR ensures that output is CMIP compliant.
- Different CMIP standards can be produced
- Use synergies, avoid to repeat work

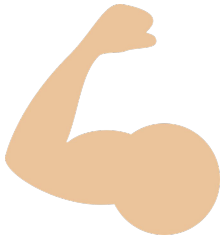
*No user side preparation of
CMIP format description*

Motivation for cdo cmor

- Why use CMOR to create CMIP standard?
- **Why integrate CDO with CMOR?**



Use the power
of CDOs...



- CDO is a well known tool with an active support
- The CDO's interface allows
 - netCDF as well as GRIB input
 - enables access to all infile information

Motivation for `cdo cmor`

... to simplify the
CMOR usage:

```
cmor_setup();  
cmor_dataset_json();  
cmor_load_table();  
cmor_set_table();  
cmor_axis();  
cmor_grid();  
cmor_set_grid_mapping();  
cmor_time_varying_grid_coordinate();  
cmor_zfactor();  
cmor_variable();  
cmor_set_deflate();  
cmor_set_variable_attribute();  
cmor_create_output_path();  
cmor_write();  
cmor_close();
```

are included by one **`cdo cmor`** operator

- Why use CMOR to create CMIP standard?
- **Why integrate CDO with CMOR?**



cdo cmor implementation

CMIP6_Amon.json

contains parts of the data request in a CMOR-readable format

grid_info.nc

contains a grid description including coordinates and bounds

variables:

```
double lat(lat);
double lat_bnds(lat,bnds);
```

```
cdo cmor, CMIP6_Amon.json \
  gi=grid_info.nc, \
  i=config.txt, \
  mt=mapping_table.txt \
```

infile

config.txt

contains the user configuration

```
activity_id="CMIP"
```

can be created with

<https://c6dreq.dkrz.de/cdocmorinfo/index.html>

mapping_table.txt

links model output variables with CMOR variables

```
&parameter pmt=Amon cmor_name=tasmax code=201 /
```

Installation of CDO with CMOR

Instruction\Target format:	CMIP5/CORDEX	CMIP6/CORDEX2
<i>Use DKRZ HPC mistral or install CDO..</i>	locally via script	locally via script or use conda
<i>Install CDO with...</i>	CMOR2	CMOR3

```
#update conda
conda update conda
#name the environment for cdo with cmor
cdoenv=cdocmor
#install develop-cdo which contains CMOR "conda-forge/label/dev::cdo"
#use conda-forge channel with "-c conda-forge"
#set $cdoenv as environment with --name
conda create --name ${cdoenv} conda-forge/label/dev::cdo -c conda-forge
#then activate it
source activate ${cdoenv}
```

Users, use cases and contributions

The operator is used within

- CMIP6 for AWI-CM, AWI-ESM, EMAC, MPI-ESM1-2
- MPI-GE for the large ensemble of MPI-ESM
- CORDEX for REMO at GERICS
- Other institutions: at KIT (Karlsruhe), at UK OX (Oxford), at INPE (Rio, Brazil)

Contributions to conferences and community workshops in addition to self-organized ones:

- CMIP6 Analysis workshop
- EGU, AGU
- CLM-Assembly, ICCARUS

Possible future applications

- Dyamond, AtMoDat, Palmod...

Lessons learned

- Cascade of versioning of the data standard documents causes incompatibility (QA also affected) → For CMIP7
 - Versioning and deployment of cdo cmor improvable
- Set of global attributes are hard to understand. → For CMIP7
- Axes:
 - Discrete axes
Could only be produced with much effort on configuration since they are implemented as a 5. dimension or substitution for a grid. Raw output variables needed to be split up and remerged.
 - Vertical Achsen
Must be submitted completely. This requires the integration of an additional data variable into the file (ECHAM: ps). This needs to be considered during aggregation.
 - Time axes
Variables requested instantaneously needs to be written separately from the others.

AP2 Kompatibilität der CDOs mit CMIP6⁺-Datenstandards (DKRZ)

1. Develop a CDO-Operator for standardising earth system model raw output by linking CMOR
2. **Adapt the CDO-data model and metadata model to den CMIP6 Standard**
3. Adapt the CDO-climate extremes indices operators to international standards

Adaptation of the CDO (meta-)data model

- + Character axes can generally be processed by CDOs
- + Global attributes can be modified with new operators
- + Global attributes are not deleted by any processing

- Complete adaptation of the CDO data model to the CMIP6 data standard not possible because
 - cdo's data model relies on gridded coordinates with spatial directions whereas the CMIP6 data standard
 - combines single axes
 - allows discrete coordinates

- The adaptation todos are based on incomplete spot checks in CMIP5. There is a need to check systematically for CMIP6 compatibility by a test suite. This requires a big pool of test and validation data that becomes available only recently.

AP2 Kompatibilität der CDOs mit CMIP6⁺-Datenstandards (DKRZ)

1. Develop a CDO-Operator for standardising earth system model raw output by linking CMOR
2. Adapt the CDO-data modell und metadata model to den CMIP6 Standard
3. **Adapt the CDO-climate extremes indices operators to international standards**
 - a. Slides on how to use **cdo cei** operators:
<https://slides.com/wachsylon/cdoetccdi>
 - b. A Jupyter-notebook with an example and scripts for processing all 27 indices are provided in this gitlab repo:
https://gitlab.dkrz.de/k204210/cdo_cei/

cdo cei

- The Expert Team on Climate Change Detection and Indices (ETCCDI) defines 27 core Climate Extremes Indices (CEI) which have been implemented in the CDOs and made available in a set of etccdi-operators. The following features features are new in CDO and ensure ETCCDI compliant processing.

Absolute values and threshold exceedances

- Operators allow to set the output frequency (yearly or monthly).

Duration CEIs

- CDO will not stop counting consecutive days when years change.

Percentile based CEIs

- Percentile calculations take the edges of the time series into account.
- The R percentile calculation method “type 8” is applied if sufficient memory is available.
- Bootstrapping is applied for the reference period.

The greatest success aeroptbnd in E3hrPt.json

CMIP6/CMIP/AWI/AWI-ESM-1-1-LR/historical/r1i1p1f1/E3hrPt/aeroptbnd/gn/v20200212/
aeroptbnd_E3hrPt_AWI-ESM-1-1-LR_historical_r1i1p1f1_gn_185001010300-185101010000.nc

<https://furtherinfo.es-doc.org/CMIP6.AWI.AWI-ESM-1-1-LR.historical.none.r1i1p1f1>

The greatest success aeroptbnd in E3hrPt.json

```
"aeroptbnd": {  
  "frequency": "3hrPt",  
  "modeling_realm": "atmos",  
  "standard_name": "atmosphere_absorption_optical_thickness_due_to_ambient_aerosol_particles",  
  "units": "1",  
  "cell_methods": "area: mean time: point",  
  "cell_measures": "area: areacella",  
  "long_name": "Aerosol Level Absorption Optical Thickness for Each Band",  
  "comment": "Optical thickness of atmospheric aerosols in wavelength bands.",  
  "dimensions": "longitude latitude alevel spectband time1",  
  "out_name": "aeroptbnd",  
  "type": "real",  
  "positive": "",  
  "valid_min": "",  
  "valid_max": "",  
  "ok_min_mean_abs": "",  
  "ok_max_mean_abs": ""  
},
```

The greatest success aeroptbnd in E3hrPt.json

```
"aeroptbnd": {  
  "frequency": "3hrPt",  
  "modeling_realm": "atmos",  
  "standard_name": "atmosphere_absorption_optical_thickness_due_to_ambient_aerosol_particles",  
  "units": "1",  
  "cell_methods": "area: mean time: point",  
  "cell_measures": "area: areacella",  
  "long_name": "Aerosol Level Absorption Optical Thickness for Each Band",  
  "comment": "Optical thickness of atmospheric aerosols in wavelength bands.",  
  "dimensions": "longitude latitude alevel spectband time1",  
  "out_name": "aeroptbnd",  
  "type": "real",  
  "positive": "",  
  "valid_min": "",  
  "valid_max": "",  
  "ok_min_mean_abs": "",  
  "ok_max_mean_abs": ""  
},
```

The greatest success aeroptbnd in E3hrPt.json

```
"aeroptbnd": {  
  "frequency": "3hrPt",  
  "modeling_realm": "atmos",  
  "standard_name": "atmosphere_absorption_optical_thickness_due_to_ambient_aerosol_particles",  
  "units": "1",  
  "cell_methods": "area: mean time: point",  
  "cell_measures": "area: areacella",  
  "long_name": "Aerosol Level Absorption Optical Thickness for Each Band",  
  "comment": "Optical thickness of atmospheric aerosols in wavelength bands.",  
  "dimensions": "longitude latitude alevel spectband time1",  
  "out_name": "aeroptbnd",  
  "type": "real",  
  "positive": "",  
  "valid_min": "",  
  "valid_max": "",  
  "ok_min_mean_abs": "",  
  "ok_max_mean_abs": ""  
},
```

The greatest success aeroptbnd in E3hrPt.json

```
"aeroptbnd": {  
  "frequency": "3hrPt",  
  "modeling_realm": "atmos",  
  "standard_name": "atmosphere_absorption_optical_thickness_due_to_ambient_aerosol_particles",  
  "units": "1",  
  "cell_methods": "area: mean time: point",  
  "cell_measures": "area: areacella",  
  "long_name": "Aerosol Level Absorption Optical Thickness for Each Band",  
  "comment": "Optical thickness of atmospheric aerosols in wavelength bands.",  
  "dimensions": "longitude latitude alevel spectband time1",  
  "out_name": "aeroptbnd",  
  "type": "real",  
  "positive": "",  
  "valid_min": "",  
  "valid_max": "",  
  "ok_min_mean_abs": "",  
  "ok_max_mean_abs": ""  
},
```

The greatest success aeroptbnd in E3hrPt.json

```
"aeroptbnd": {
```

```
"dimensions": "longitude latitude alevel spectband time1",
```

```
},
```

The greatest success aeroptbnd in E3hrPt.json

"aeroptbnd": {

Vars with 5 Dimensions are not processable with cdo, "therefore" they do not exist.

Warning (cdfCheckVars): 5 dimensional variables are not supported, skipped variable aeroptbnd!

However, we can create those.

alevel:

- hybrid_sigma_pressure in echam → requires **ps** in file

"dimensions": "longitude latitude alevel spectband time1",

spectband:

- spectral bands are saved as individual variables → need to be merged
- Bounds required

time1:

- Instantaneous variables require CMIP time stamp → move (about minutes)

},


```

netcdf aeroptbnd_E3hrPt_MPI-ESM1-2-HR-dcppC-hindcast-noEiChichon_s1981-r2i1p1f1_gn_201501010300-201501010900 (
dimensions:
  time = UNLIMITED ; // (3 currently)
  spectband = 2 ;
  lev = 95 ;
  lat = 192 ;
  lon = 384 ;
  bnds = 2 ;
variables:
  double time(time) ;
      time:units = "days since 1850-01-01 00:00:00" ;
      time:calendar = "proleptic_gregorian" ;
      time:axis = "T" ;
      time:long_name = "time" ;
      time:standard_name = "time" ;
  double spectband(spectband) ;
      spectband:bounds = "spectband_bnds" ;
      spectband:units = "m-1" ;
      spectband:long_name = "Spectral Frequency Band" ;
      spectband:standard_name = "sensor_band_central_radiation_wavenumber" ;
  double spectband_bnds(spectband, bnds) ;
  double lev(lev) ;
      lev:bounds = "lev_bnds" ;
      lev:units = "1" ;
      lev:axis = "Z" ;
      lev:positive = "down" ;
      lev:long_name = "hybrid sigma pressure coordinate" ;
      lev:standard_name = "atmosphere_hybrid_sigma_pressure_coordinate" ;
      lev:formula = "p = ap + b*ps" ;
      lev:formula_terms = "ap: ap b: b ps: ps" ;
  double lev_bnds(lev, bnds) ;
      lev_bnds:formula = "p = ap + b*ps" ;
      lev_bnds:standard_name = "atmosphere_hybrid_sigma_pressure_coordinate" ;
      lev_bnds:units = "1" ;
      lev_bnds:formula_terms = "ap: ap_bnds b: b_bnds ps: ps" ;
  double ap(lev) ;
      ap:long_name = "vertical coordinate formula term: ap(k)" ;
      ap:units = "Pa" ;
  double b(lev) ;
      b:long_name = "vertical coordinate formula term: b(k)" ;
  float ps(time, lat, lon) ;
      ps:long_name = "vertical coordinate formula term: ps" ;
      ps:units = "Pa" ;
  double ap_bnds(lev, bnds) ;
      ap_bnds:long_name = "vertical coordinate formula term: ap(k+1/2)" ;
      ap_bnds:units = "Pa" ;
  double b_bnds(lev, bnds) ;
      b_bnds:long_name = "vertical coordinate formula term: b(k+1/2)" ;
  double lat(lat) ;
      lat:bounds = "lat_bnds" ;
      lat:units = "degrees_north" ;
      lat:axis = "Y" ;
      lat:long_name = "Latitude" ;
      lat:standard_name = "latitude" ;
  double lat_bnds(lat, bnds) ;
  double lon(lon) ;
      lon:bounds = "lon_bnds" ;
      lon:units = "degrees_east" ;
      lon:axis = "X" ;
      lon:long_name = "Longitude" ;
      lon:standard_name = "longitude" ;
  double lon_bnds(lon, bnds) ;
  float aeroptbnd(time, spectband, lev, lat, lon) ;
      aeroptbnd:standard_name = "atmosphere_absorption_optical_thickness_due_to_ambient_aerosol_particles" ;
      aeroptbnd:long_name = "Aerosol Level Absorption Optical Thickness for Each Band" ;
      aeroptbnd:comment = "Optical thickness of atmospheric aerosols in wavelength bands." ;
      aeroptbnd:units = "1" ;
      aeroptbnd:cell_methods = "area: mean time: point" ;
      aeroptbnd:cell_measures = "area: areacella" ;
      aeroptbnd:history = "2019-08-13T09:57:13Z altered by CMOR: Reordered dimensions, original order: time lat lon lev spectband. 2019-08-13T09:57:13Z altered by CMOR: replaced missing value flag (-5.40299e-35) and corresponding data with standard missing value (1e+20). 2019-08-13T09:57:13Z altered by CMOR: Inverted axis: lev. 2019-08-13T09:57:13Z altered by CMOR: Inverted axis: lat." ;
      aeroptbnd:missing_value = 1.e+20f ;
      aeroptbnd:_FillValue = 1.e+20f ;

```

The greatest success aeroptbnd in E3hrPt.json

Schriftgröße 4
ohne globale
Attribute

```
netcdf aeroportbnd_E3hrPt_MPI-ESM1-2-HR_dcppC-hindcast-noElChichon_s1981-r2i1p1f1_gn_201501010300-201501010900 {  
dimensions:
```

The greatest success aeroptbnd in E3hrPt.json

Schriftgröße 12
ohne Attribute

```
    time = UNLIMITED ; // (3 currently)  
    spectband = 2 ;  
    lev = 95 ;  
    lat = 192 ;  
    lon = 384 ;  
    bnds = 2 ;  
variables:  
    double time(time) ;  
    double spectband(spectband) ;  
    double spectband_bnds(spectband, bnds) ;  
    double lev(lev) ;  
    double ap(lev) ;  
    double b(lev) ;  
    float ps(time, lat, lon) ;  
    double ap_bnds(lev, bnds) ;  
    double b_bnds(lev, bnds) ;  
    double lat(lat) ;  
    double lat_bnds(lat, bnds) ;  
    double lon(lon) ;  
    double lon_bnds(lon, bnds) ;  
    float aeroportbnd(time, spectband, lev, lat, lon) ;
```



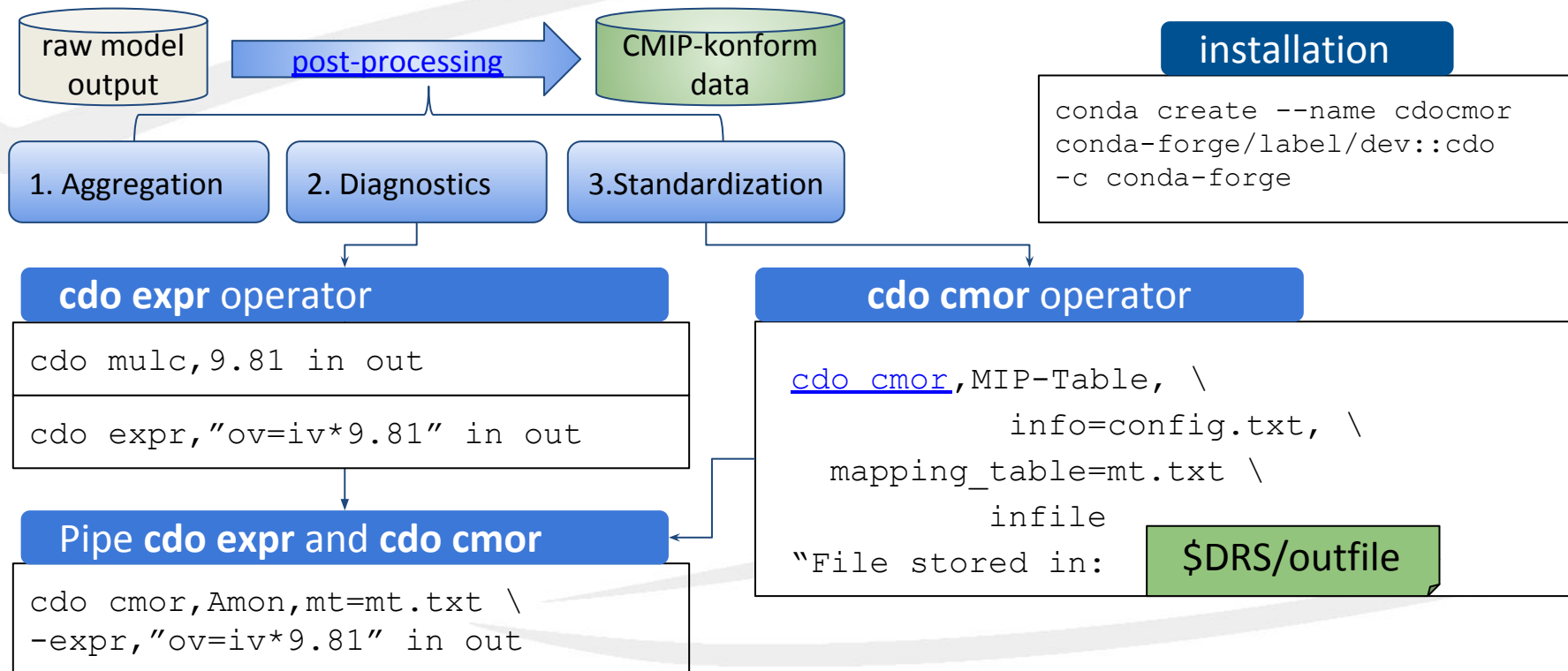
Danke für die Aufmerksamkeit!



Extra-Folien

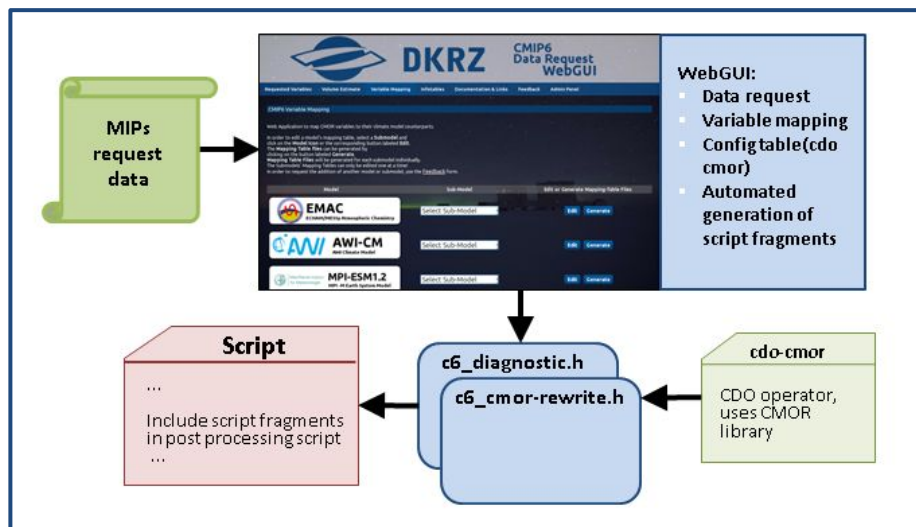
Milestones	Erledigt [%]	Erledigte Aufgaben
2.1. cdo cmor ist für CMIP5 nutzbar	100	Eingesetzt für verschiedene Projekte, Modelle, Experimente und auf verschiedenen Plattformen.
2.2. cdo cmor ist für CMIP6 nutzbar	100	
2.3. Anpassung des CDO-Metadatenmodells an das CMIP-Metadatenmodell	100	Attribute werden nicht mehr gelöscht und können beliebig geändert werden. Fehlender Achsentyp wurde ergänzt. Anleitung zur Wiederherstellung des CMIP Standards begonnen.
2.4. Anpassung des CDO-Datenmodells an das CMIP-Datenmodell	60	
2.5. Anpassung der Klimaindizes-Operatoren an internationale Standards	100	ETCCDI konforme Indices wurden eingebaut bzw. ergänzt.

CDO can be linked with **CMOR** which creates CMIP compliant output.
 Users can combine other operators with `cdo cmor`.

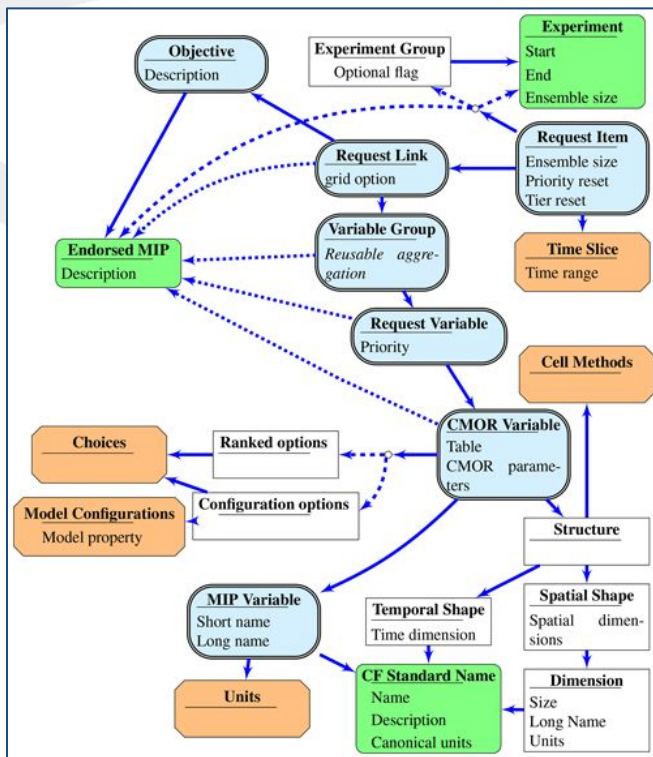


AP3 Infrastruktur zur Anpassung an CMIP6-Projektstandards (DKRZ)

- Entwicklung von Skripten zur diagnostischen Nachbereitung und Re-Formatierung der Modelldaten entsprechend dem Projektstandard
- Modulare Durchführung der einzelnen Workflow-Schritte
- Unterstützung der CMIP6-endorsed MIPs
- Anwendung der Infrastruktur außerhalb CMIP6 ermöglichen



CMIP6 Data Request



Variable Lists:

- Variable Definition > 1200 variables
- Output specification (-> CMOR variable) > 2000 CMOR variables

Output Requirements:

- Experiments and time slices > 10.000 different requests
- Objectives supported > 20 MIPs
- Priority of variable

Experiment Specifications:

- Duration of simulation > 300 experiments
- Tier
- Number of ensemble members

[Juckes, M. 2019](#)

Das WebGUI führt durch den Arbeitsablauf projektkonforme Daten zu erstellen

Datenrequest (Dreq)

1 – Welche Variablen werden für die geplanten Experimente angefordert?

- Das WebGUI nutzt Martin Juckes' für CMIP6 entwickelte [DreqPy API](#).
- Datenrequest lässt sich als CSV oder EXCEL-Tabelle den Nutzerbedürfnissen entsprechend erstellen und herunterladen.
- Eine Volumenabschätzung für die zu erwartende Datenmenge kann berechnet werden.

Variablen-Mapping

2 – Welche Modellvariable entspricht der jew. Variable des Datenrequests?

- Zuordnung der CMOR-Variable zum passenden Gegenstück in der Modellausgabe.
- Spezifizieren von Diagnostiken. Unterstützt wird der Climate Data Operators (CDO) [expr Operator](#).
- Die Mapping-Informationen werden in einer Datenbank gespeichert. Änderungen sind nachverfolgbar und revertierbar.
- Gleichzeitige Bearbeitung der Mapping-Informationen durch verschiedene Nutzer möglich.
- Genauere Volumenabschätzung möglich durch Ausschluss nicht zugeordneter Variablen.

DRS / CV

3 – Zusammenstellung aller Metadaten-Informationen für die Anwendung von CMOR / [CDO cmor](#).

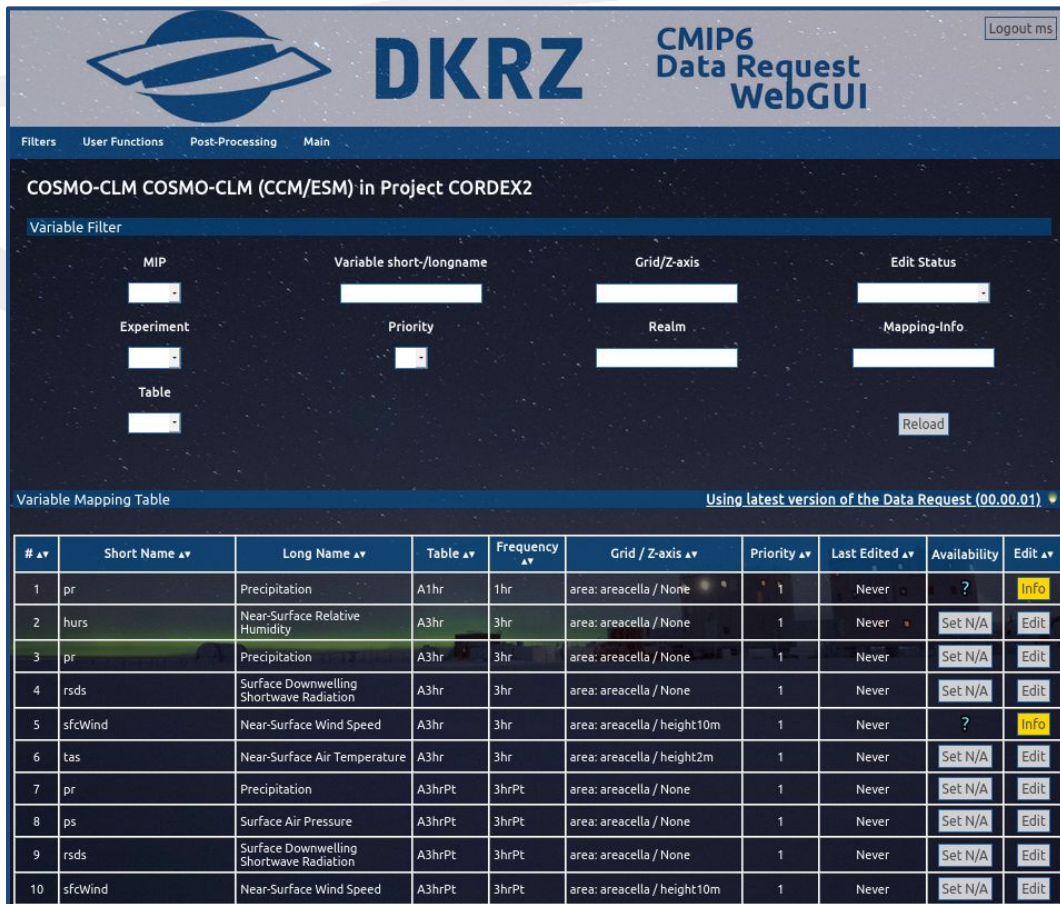
- Auswahl der geeigneten Metadateneinträge aus der CV (Controlled Vocabulary) und den DRS-Vorgaben (Data Reference Syntax) des Projektes.
- Auswahl aus vordefinierten Einstellungen möglich.
- Die Metadaten-Informationen können in für CDO cmor interpretierbarem Format heruntergeladen werden.

Diagnostik & CMOR-Rewrite

4 – Automatische Erstellung und Herunterladen von Diagnostik- und CMORisierungsskripten

- Automatische Erzeugung aus den eingegebenen Variablen-Mapping-Informationen.
- Verwendung der CDO-Operatoren expr und cmor.
- Die Skripte werden gespeichert und können jederzeit heruntergeladen / geteilt werden.
- Einzelne Variablen können abh. von Datenrequest oder Nutzerspezifikationen von der Prozessierung ausgeschlossen werden.
- Die Skripte können in das Modell-eigene Scripting integriert werden.

Post-Processing-Infrastruktur für andere Projekte verwendbar



The screenshot shows the 'CMIP6 Data Request WebGUI' interface. At the top, there is a navigation bar with 'Filters', 'User Functions', 'Post-Processing', and 'Main'. Below this, the main content area is titled 'COSMO-CLM COSMO-CLM (CCM/ESM) in Project CORDEX2'. A 'Variable Filter' section contains several input fields: 'MIP', 'Variable short-/longname', 'Grid/Z-axis', 'Edit Status', 'Experiment', 'Priority', 'Realm', and 'Mapping-Info'. A 'Table' dropdown and a 'Reload' button are also present. Below the filter section is a 'Variable Mapping Table' with a dropdown for 'Using latest version of the Data Request (00.00.01)'. The table lists 10 variables with columns for #, Short Name, Long Name, Table, Frequency, Grid / Z-axis, Priority, Last Edited, Availability, and Edit actions.

#	Short Name	Long Name	Table	Frequency	Grid / Z-axis	Priority	Last Edited	Availability	Edit
1	pr	Precipitation	A1hr	1hr	area: areacella / None	1	Never	?	Info
2	hurs	Near-Surface Relative Humidity	A3hr	3hr	area: areacella / None	1	Never	Set N/A	Edit
3	pr	Precipitation	A3hr	3hr	area: areacella / None	1	Never	Set N/A	Edit
4	rsds	Surface Downwelling Shortwave Radiation	A3hr	3hr	area: areacella / None	1	Never	Set N/A	Edit
5	sfcWind	Near-Surface Wind Speed	A3hr	3hr	area: areacella / height10m	1	Never	?	Info
6	tas	Near-Surface Air Temperature	A3hr	3hr	area: areacella / height2m	1	Never	Set N/A	Edit
7	pr	Precipitation	A3hrPt	3hrPt	area: areacella / None	1	Never	Set N/A	Edit
8	ps	Surface Air Pressure	A3hrPt	3hrPt	area: areacella / None	1	Never	Set N/A	Edit
9	rsds	Surface Downwelling Shortwave Radiation	A3hrPt	3hrPt	area: areacella / None	1	Never	Set N/A	Edit
10	sfcWind	Near-Surface Wind Speed	A3hrPt	3hrPt	area: areacella / height10m	1	Never	Set N/A	Edit

- Variablen-Mapping und Skriptterzeugung kann für Projekte erweitert werden, welche einen CMOR-Standard unterstützen.
- Neben CMIP6 ist bisher die Unterstützung folgender Projekte eingerichtet bzw. in Arbeit:
 - [CORDEX2 \(CORDEX-CMIP6\)](#)
Vorläufige Variablendatenbank eingerichtet.
 - [PRIMAVERA](#)
Zur Unterstützung des AWI-Beitrages zu PRIMAVERA/HighResMIP eingerichtet
 - [DYAMOND](#)
Variablendatenbank eingerichtet. Weitere Erfordernisse bzgl. des Scriptings werden erwartet.

DICAD - Ein paar Zahlen

WebGUI

Nutzer-Registrierungen	43
Variablen (EMAC-2)	500
Submits (EMAC-2)	1500
Variablen (MPI-ESM1-2)	900
Submits (MPI-ESM1-2)	2600

ESGF

Variablen (MPI-ESM historical) - CMIP5	250
Variablen (MPI-ESM historical) - CMIP6	664

- Mapping-Informationen je Variable werden oft nicht beim ersten Bearbeiten vervollständigt / finalisiert:
 - Erfassen der CMOR-Variableninformationen
 - Recherche nach passender Diagnostik, Modellvariablen-Informationen (units, comment)
 - Änderungen im Datenrequest / Modell-Namelist machen spätere Anpassungen nötig
- AWI-CM: Mappinginformationen hauptsächlich von MPI-ESM (echam6, jsbach) übernommen und angepasst.
- Vergleich zur Erzeugung des Modelldatenstandards inkl. CMIP6-Mapping mit dem WebGUI: GitHub-Projekt [CMIP6-CV](#) ~1500 Commits

DICAD - Ein paar Zahlen

	ESGF-Volumenabschätzung 2017 mittels DreqPy/WebGUI [TB]	ESGF - Bisher veröffentlichtes Datenvolumen (neueste Version) [TB]
AWI-CM/AWI-ESM	200	125
MPI-ESM	720	1000
EMAC-2	3	-

- Abschätzung mittels früher Version des Datenrequests und DreqPy-Software (mittlerweile komplett überarbeitet).
- "Abgabepolicy" entscheidend (alle möglichen Variablen, alle angeforderten Variablen, über alle Experimente vereinigter Datenrequest, ...)
- EMAC-2: Chemiemodul ist bezogen auf die Rechenzeit sehr teuer - Modellrechnungen auf T42 durchgeführt (gegenüber T63 und T127 für die meisten AWI-CM/MPI-ESM-Rechnungen)

DICAD - Ein paar Zahlen

<i>Jeweils pro Jahr</i>	MPI-ESM1-2-LR Datenvolumen [GB]		MPI-ESM1-2-HR Datenvolumen [GB]	
	Rohdaten	75		500
Aggregation	25		165	
ESGF	<30		170	
	Dauer [min]	Rechenzeit [Nh]	Dauer [min]	Rechenzeit [Nh]
Modelllauf	25	12	105	150
Aggregation	8	0.13	45	0.75
Diagnostik +CMOR-Rewrite	5	0.08	60	0.75

- Rechenzeit des Post-Processings gegenüber dem Modelllauf vernachlässigbar.
- Post-Processing kann bei Bedarf über die Jahre parallelisiert werden.

Erfahrung und Feedback

- Zeitmangel durch Verzögerung bei den Klimamodellsimulationen und gleichzeitiger Druck durch IPCC-Deadlines lassen den Nutzern nur wenig Zeit und Nerven sich mit der Post-Processing-Infrastruktur zu beschäftigen
 - Erst spät Testdaten der Modelle / Modell-Namelisten verfügbar
 - Erst spät Feedback:
 - Ende 2018 allgemeineres Feedback zum Nutzerinterface
 - Ende 2019 spezielleres Feedback zu den generierten Skripten
 - Kompetenzbildende Maßnahmen haben nicht den gewünschten Effekt:
 - Andere als die anwesenden Personen führen am Ende die Standardisierung durch und müssen eingewiesen werden
 - Personen nehmen nicht teil (Zeitmangel, Unterschätzen des Aufwandes der Standardisierung)
 - Nutzerunterstützung sehr zeitaufwändig, Umsetzen des Feedbacks verzögert sich

Erfahrung und Feedback

- Zielvorstellungen des Projektes und Wunschvorstellungen der Nutzer weichen (nur scheinbar) voneinander ab:
 - Projekt: Flexibles und allgemeines Tool, maximal modular aufgebaut. Integrierung in das Skripting der Modellierungs-Workflows übernehmen die Modellierer mit unserer Unterstützung.
 - Nutzer/Modellierer: Modellspezifisches Tool, weniger modular konfigurierbar/ausführbar (wenn dann versteckt). Integriert in das Skripting des Modells.
- Um Nutzer- und Projektvorstellungen in Einklang zu bringen, ist von Anfang an ein Ansprechpartner auf Modellierseite nötig, der die Integrierung in das modelleigene Skripting übernimmt, wie im Projekt vorgesehen. Teilweise war das nicht gegeben.
- Umgesetzt für MPI-ESM (sowohl integriert ins Skripting des MPI-M [MakeExperiments - mkexp] und des DKRZ [IMDI] als auch flexible Standalone-Anwendung mit dementsprechend mehr Konfigurationsaufwand)
 - Standalone-Anwendung umgesetzt für EMAC (allerdings nur 1-2 Anwender)
 - Standalone-Anwendung umgesetzt für AWI-CM, allerdings mehrere Anwender, die teilweise lieber eine Integration in das modelleigene Skripting hätten (esmtools).
 - Bisher keine Unterstützung von ICON-ESM, evtl. auch nicht notwendig (bereits standardisierte Modellausgabe geplant).

Ausblick

Anstehende Arbeiten

- Vereinfachung für AWI-CM/MPI-ESM: Bestehendes Mapping für abgeänderte Modelloutput-Namelist übernehmen, inkl. Aggregation
- Umzug SVN→Git, inkl. möglicher Automatisierungen/Tests
- Support für Post-Processing-Infrastruktur
- Support / Korrektur bei errata-Fällen
- Unterstützung PostProcessing-Infrastruktur für DYAMOND, CORDEX wenn nötig
- Veröffentlichung / Nachprozessierung bisher noch nicht veröffentlichter RCM-Forcingdaten und CMIP6-Daten für MPI-ESM-Simulationen aus DICAD/RZ988

Vermutlich nicht mehr umzusetzen (→CMIP7?)

- “Flächenmäßige” Änderung einer Mapping-Informationen für versch. Variablen (z.B. Dateiname, Einheit)
- Vereinfachung der Einbindung von benutzerdefinierten Diagnostiken (Diagnostiken, die nicht mit cdo expr abbildbar sind)
- Aggregationskripte auf Wunsch automatisch erstellen (hier wären allerdings deutlich mehr Nutzerangaben nötig)
- Unterstützung von cylc zur Ablaufsteuerung des Workflows
- Allgemeines Namelist-Mapping (modellunabhängig, deutlich mehr Nutzerangaben wären nötig).
- Definition eigener Variablen oder Datenanforderungen über das WebGUI (bisher nur für mich von außen möglich)



Danke für die Aufmerksamkeit!



Extra-Folien

CMIP6 Data Request

“Request Harvesting”:

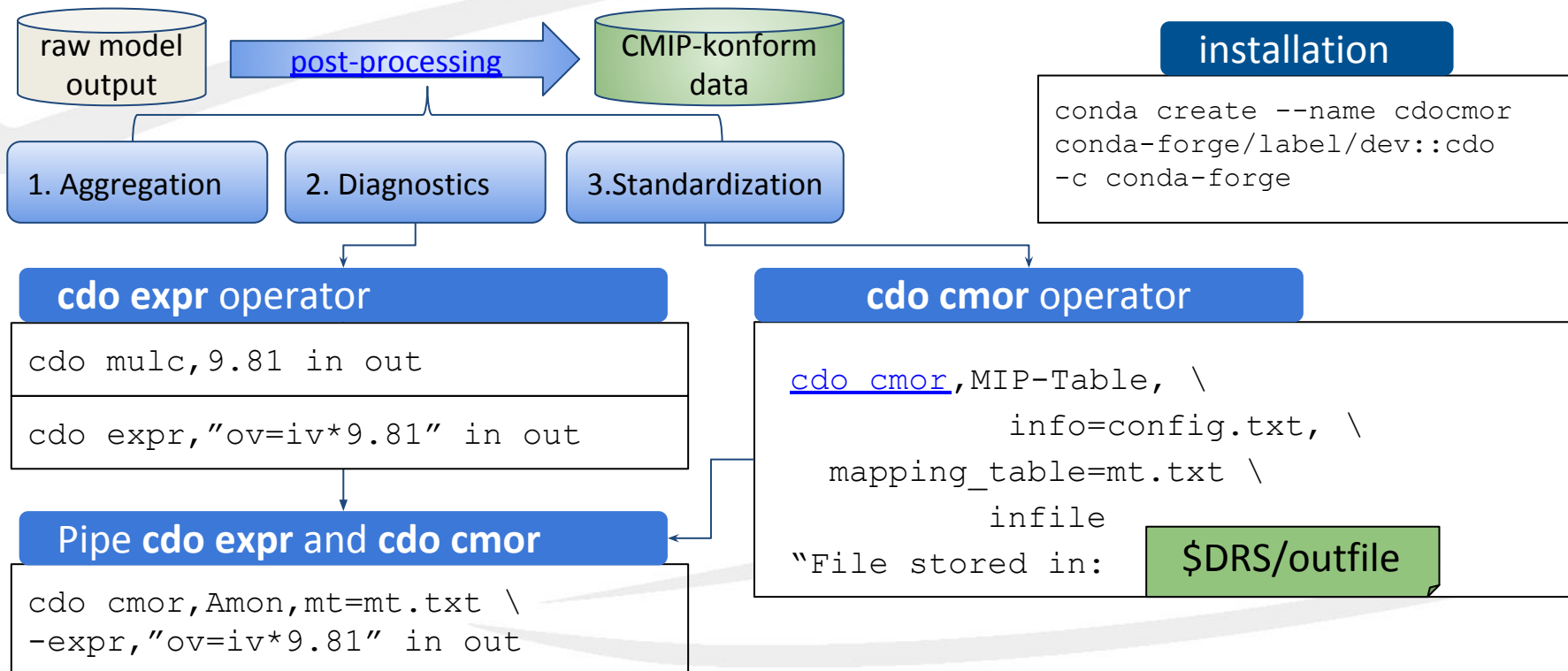
MIP officials define variable lists, output requirements & experiment specifications. Complete list collected centrally. Original date for final release of the data request: 15th October 2015. [see also: Veronika Eyring, 2014](#)

Problems:

- complex structure of different request objects
 - very large number of requests and MIPs
 - No official deadline: waiting for the final definitions of “slower” MIPs
 - “definition” rather than “selection” of variables lead to multiple mentions of the same variable
 - continuous changes of variable definitions and requests that were hard to track
- > “unstable” data request that slows down the entire project

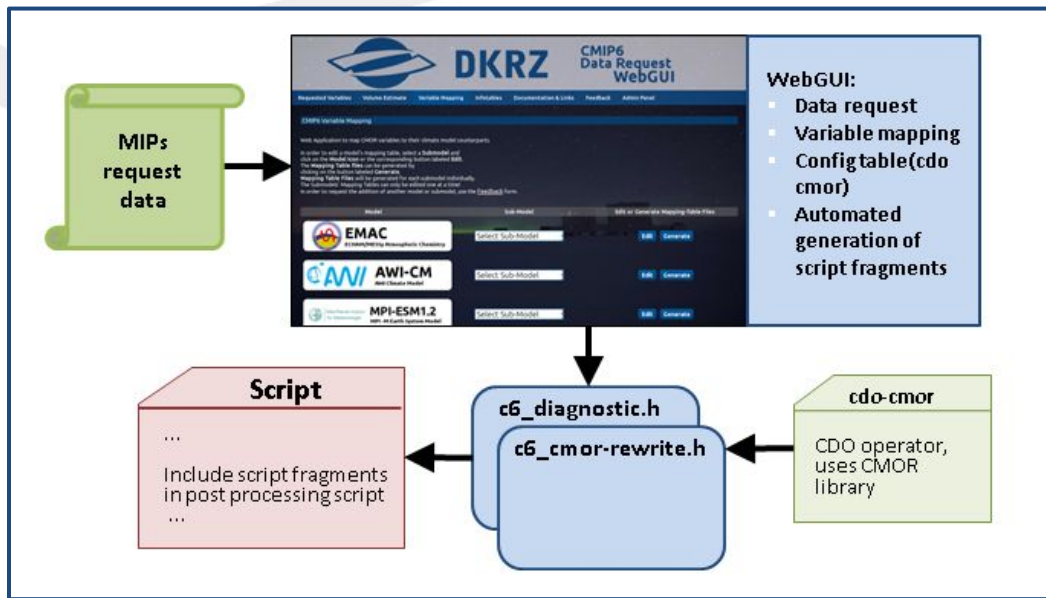
But now the data request is quite stable and the definition process has been completed.

CDO can be linked with **CMOR** which creates CMIP compliant output.
Users can combine other operators with `cdo cmor`.



Unterstützungsinfrastruktur: Status

Infrastruktur zur Anpassung an CMIP6 Projektdatenstandards



WebGUI-unterstützter Workflow:

- Manuell erzeugter Aggregationsteil
- Generierung eines **Datenrequests** u. einer **Volumenabschätzung**
- **“Variablenmapping”** (Zuordnung CMIP zu Modellvariable)
- Automatische Erstellung von **Diagnostik- und CMORisierungs-Skriptfragmenten** (Verwendung von cdo expr, cdo cmor) aus dem jew. Variablenmapping
- Bereitstellung von **Wrapper-Skript** und **Kornshell-Funktionen** zum **modularen Ablauf** des Workflows:
 - Workflowschritt (Agg., Diag., CMOR)
 - Submodell - Realisierung - Variable und zum Abgleich der zu standardisierenden Variable mit dem Datenrequest bzw. Nutzervorgaben
- Möglichkeit zur Benutzung für andere Projekte (PRIMAVERA, CORDEX, DYAMOND, ...)

Unterstützungsinfrastruktur: Status

WP3 - Infrastruktur zur Anpassung an CMIP6 Projektdatenstandards

Milestones	Erledigt [%]
M1 Skript für Diagnostik mit MPI-ESM1-2	100
M2 Werkzeug zum Überprüfen des pot. Variablenumfangs	100
M3 Skript zur Standardisierung des Modelloutputs	100
M4 Software für modulare Steuerung des Workflows	100
M5 Nutzbarkeit der Infrastruktur außerhalb CMIP	90
M6 Nutzerworkshop	100
M7 Unterstützungsinfrastruktur	100
M8 Verbesserte nachhaltige Infrastruktur	50

Script fragments

- Automatic creation of diagnostic and cmor rewrite script fragment out of variable mapping information
- Automatic creation of data request configuration out of CMIP6 data request, further customizable by user

Diagnostic

c6_diagnostic.h

- **cdo merge** (of multiple inputfiles)
- **cdo expr**

```
#!/bin/bash
#-- Diagnostic for echam6 (ESM: AWI-CM-1-0-HR) variable rlds / table 3hr
{ (if requested $member $atmmod 3hr rlds $chunk && {
  find_file -e -p $period "$sdir" "TST_????01.01_echam.grb" ifile1
  find_file -e "$sdir" "rlds_3hr_${period}*" ifile2
  $cdo -f nc -O \
    expr,'rlds=var177-var205;' \
    -merge -selcode,177 $ifile1 -selcode,205 $ifile2 \
    in_cmor/3hr_rlds_$period.nc || echo ERROR
}; )&& }>$err.rlds.3hr 2>&1
```

CMOR rewrite

c6_cmor-rewrite.h

- **cdo cmor** call

```
#!/bin/bash
#-- CMOR-rewrite for echam6 (ESM: AWI-CM-1-0-HR) 3hr
cn='rlds hfls'
for var in $cn; do
  { (if requested $member $atmmod 3hr $var $chunk || continue
    ifile=in_cmor/3hr_${var}_${period}.nc
    $cdo cmor,3hr,mt=$mt,dr=$dr,cn=$var $ifile || echo ERROR
  )&& }>>$err.$var.3hr 2>&1
done
```

The script fragment does ...

- ... test if variable is requested (data request, timeslice, user specifications)
- ... find inputfile
- ... call cdo (cmor)

Data Request Config - *CMIP6_historical_requested_vars.conf*

```
#####  
EXP=historical  
#####  
  
DREQSETTINGS  
SIday      : SIday      = slice: piControl030,piControl050,piControl100,piControl140  
sispeed    : SIday      = slice: piControl100  
Emon       : Emon       = slice: TOTAL  
Emon       : hus        = slice: piControl100  
thetaot300 : Emon       = False  
EmonZ      : EmonZ      = False  
Amon       : no2        = False  
  
USERSETTINGS  
# ---> Specify your settings for Experiment historical here  
sispeed    : SIday      = slice: piControl100,1900010100-1914123124  
day        : day = False  
Lmon       : echam6     = False  
Elhr       : r2ilp1f1   = False  
# <---- Specify your settings for Experiment historical here
```

Post processing script - *.runpp

- Define certain variables (experiment, path, ...)

```
#Models
atmmod=echam6
bgcmmod=hamocc
ocemod=mpiom
srfmod=jsbach
esmod=MPI-ESM1-2

#Experiment etc
experiment=abrupt-4xCO2
member=r1ilp1f1
mip=CMIP6
```

- Initialize if_requested function and read data request

```
. $path/function_if_requested
. $path/function_Read_request_config
. $path/function_find_file

#Load predefined timeslices (depends on $iniyear/$finyear)
. $cpath/TimeSlices.h

#Initialize DataRequest/User Configuration for the if_requested function
# This will read all SettingsContainers, TimeSlices and the configuration file
Read_request_config -s $experiment ./conf/${mip}_requested_vars_${experiment}.conf || exit 1
```

- Run diagnostic and cmor-rewrite script fragments in a time loop

```
#Loop over file output periods and load CMOR-Rewrite script fragment
for period in `seq ${iniyear} ${finyear}`; do
  #Define the chunk that is tested by if_requested (YYYYMMDDHH-YYYYMMDDHH)
  chunk=${period}010100-${period}123124

  for submodel in $atmmod $ocemod $srfmod $bgcmmod; do

    #Location of Mapping table:
    mt="./tables/${esmod}_${submodel}_${mip}_mapping.txt"

    #Load and run the CMOR-Rewrite ScriptFragment
    echo "Performing cdo CMOR operations for $submodel $chunk..."
    . ./c6_cmor-rewrite_${esmod}_${submodel}_auto.h
    wait
  done |
done
```

Benchmarks from *RZ request* and *sbatch-log* files

Model	MPI-ESM1-2-HR
<i>Runtime experiment [Nhour/a]</i>	197
Experiment	1pctCO2 (165 sim. years)
Partition for post-processing	PrePost, Haswell
Runtime total post-processing [Nhour/a]	2,16
Of PP: Aggregation (mainly ECHAM) [Nhour/a]	1,2
Of PP: Diag/CMOR [Nhour/a]	0,96
ECHAM of Diag/CMOR [Nhour/a]	0,9



AP4 Datenqualitätsprüfung (DKRZ)

WP 4.1: Das CV-Prüfprogramm **PrePARE** (PCMDI) wurde in das Prüfprogramm QA-DKRZ eingebunden.

- Leistungsumfang für ESGF: Files müssen DRS kompatibel sein und PrePARE passieren.

WP 4.2: Institute, soweit sie QA-DKRZ selber einsetzen wollten, wurden bei der Installation unterstützt.

- Ansonsten wurde die QA am DKRZ vorgenommen und Rücksprache mit den Instituten gehalten.
- Dokumentation auf <https://readthedocs.org/projects/ga-dkrz>.

WP 4.3: **Spot-Check-Dienst** existiert, ist aber nicht praktikabel wegen der großen Files.

- Einzelne Files sind nicht repräsentativ für die Variabilität des CV insgesamt.

ToDo:

- Die conda Installation funktioniert nicht auf allen Plattformen insbesondere wegen library Probleme.
- Benutzerunabhängige QA-Installation ist in Arbeit.

AP4 QA: Aufgetretene Fehler

Tolerierte Abweichungen vom Standard

A) Fehler in PrePARE / CMOR:

- Zeitintervall in ODec Experimenten.
- Vertauschung von CMOR-Name und Variable-Name.
- der Wert von `_FillValue` wurde nicht getestet (temporär).
- PrePARE verlangt `cell_methods=area: time: mean'`;
nur fuer 3hr, ansonsten ohne 'area:' in Ordnung.
- `<parent_time_units> set to <days since 1-1-1> is invalid.`

B) Ignoranz von CF Conventions Empfehlungen:

- specification of attribute `<lev>:computed_standard_name` is strongly recommended.
- `cell_methods`: falsch formatierte Zusatzinformation.

C) Sonstiges:

- abweichender `_FillValue` Wert geduldet aus Konsistenzgründen, (siehe A).
- `units = <psu>` is not CF compatible with `standard_name`.
- versteckte Files.
- Typ von Variable und `missing_value` ist unterschiedlich.

AP4 QA: Aufgetretene Fehler

Von Veröffentlichung ausgeschlossen (return to sender)

- alle Einreichungen, für die PrePARE 'Error' ausgibt.
- zeitliche Lücke zwischen sub-temporären Files eines Datensatzes.
- sich überlappende Files (zumeist alte und neue Files im gleichen Ordner).
- Files aus Prozesskette sowie invalide NetCDF, die nicht versteckt sind.
- Inkonsistenz zwischen DRS Pfad/Filename und globalen Attributen.
- Einheiten ode Dimensionen änderten sich in einem Datensatz.
- Identische Variablennamen aus zwei Teilmodellen führte zu Inkonsistenzen.
- Fehlende time_bnds.

AP5 Nationales CMIP6+-Datenarchiv (DKRZ)

WP 5.1: Ein CMIP6 Repository im Rahmen des ESGF wurde eingerichtet.

WP 5.2: Primäre Daten wurden über den Leistungsumfang hinaus (siehe WP 4.1) geprüft und publiziert.

- Die **Replikation** aus dem ESGF wurde automatisiert.

WP 5.3: Der **CMIP6 Datenpool** umfasst gegenwärtig* **3.8 PB**, davon

primäre Daten (1.5 PB),

replizierte Daten (2.3 PB) mit einer Transferrate von rund 10 TB/Tag.

- Daten aus den Aktivitäten CMIP und ScenarioMIP wurden in ihrer Gesamtheit berücksichtigt.
- Die anderen Aktivitäten wurden entsprechend einer ESGF Absprache teilweise repliziert.
- Anforderungen von lokalen Anwendern, insbesondere ESMValTool, wurden Priorität eingeräumt.

ToDo:

- Feststellung komplett replizierter Datensätze zwecks Publikation.
- Fortlaufende Überwachung des Replikationsvorgangs.
- Fortlaufende Bereitstellung primärer Datensätze.

**Ende Juni 2020*

ESGF Replikation - Download

- Ansatz: Sogenannte Selektionfiles bestimmen was repliziert wird (*Format: Aktivität_Institut_Modell_Experiment*).
Tägliche ESGF-solr-Abfragen erfassen neue in ESGF publizierte Daten und erzeugen neue Selektionsfiles.
Feiner granulare Spezifikationen, wie Table_id, Frequenz und Variable sowie Prioritäten sind optional.
- Die Prozessketten (auf 2 Blades) sind durch cronjobs gegen Rechner-Downtimes geschützt. Alle Mitglieder der ESGF Gruppe haben mittels Textfiles Zugriff auf aktuellen Lauf.
- Work-flow:
 - Bestimmung der ausgewählten Selektionfiles
 - Suche aller Datensätze zu jeweiligem Selektionfile.
 - Start des Downloads neuer Datensätze mittels synda.
 - Für jeden Selektionfile wird eine eigene DB und Logfile angelegt.
 - Optional auch zyklisches Prozessieren.

Programm Optionen (Auswahl)

Usage: run_synda_cmip6 [--opts] sel-file(s)

Purpose: Run synda for automatic processing of selection files. This program secures against multiple instances on the given machine and locks selection files currently in use.

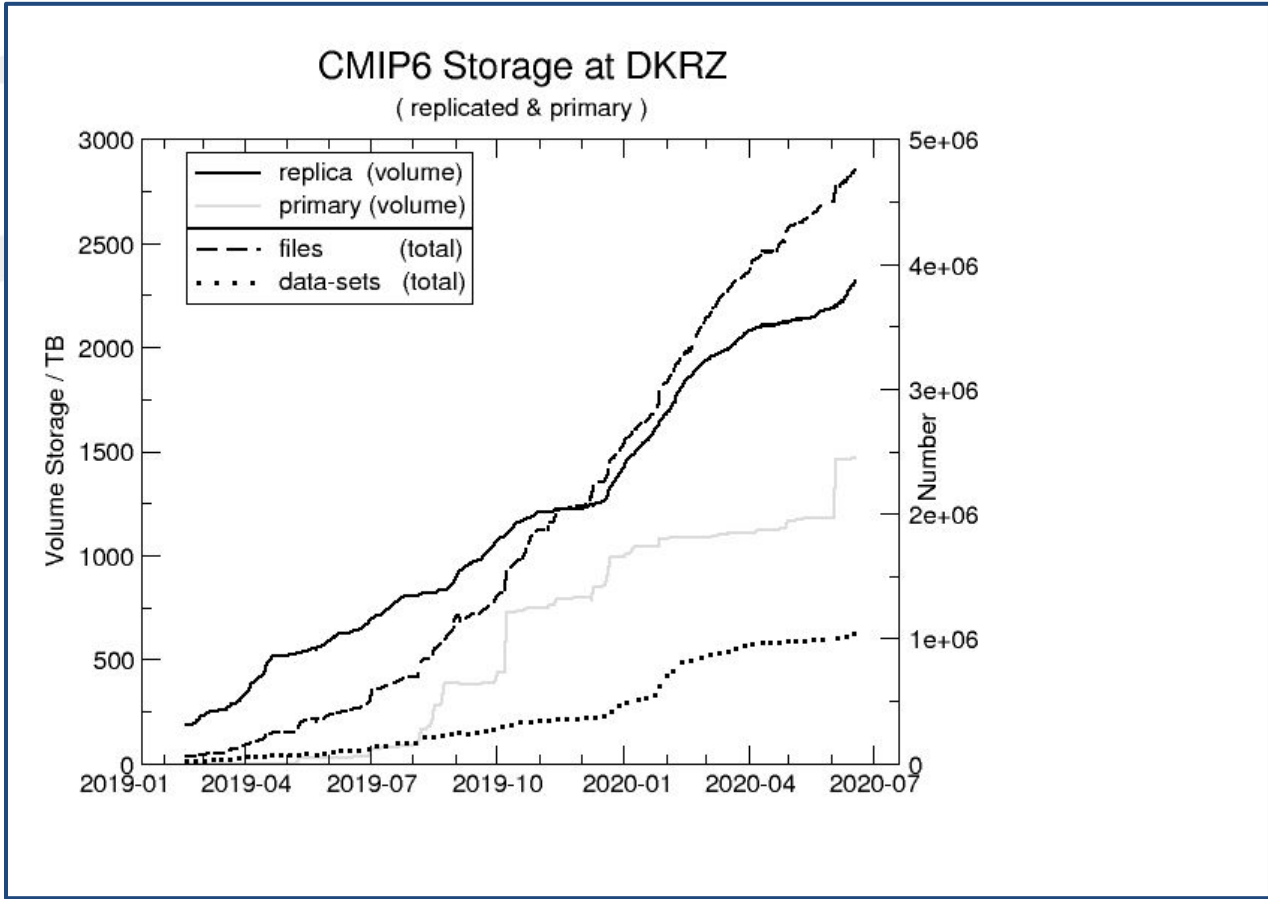
- continue Continue from the last used sel-file.
- cycle After processing of a all chosen selection-files start all-over again.
- lock-completed Keep locking of selection files when download completed.
Locks are released when new versions for a given SF where found.
- r[equest-]f[ile]=csv Special request, e.g. ESMVal.
- sel-file(s) Only run command-line provided selection files.

Signals:

TERM Complete running downloads and exit.

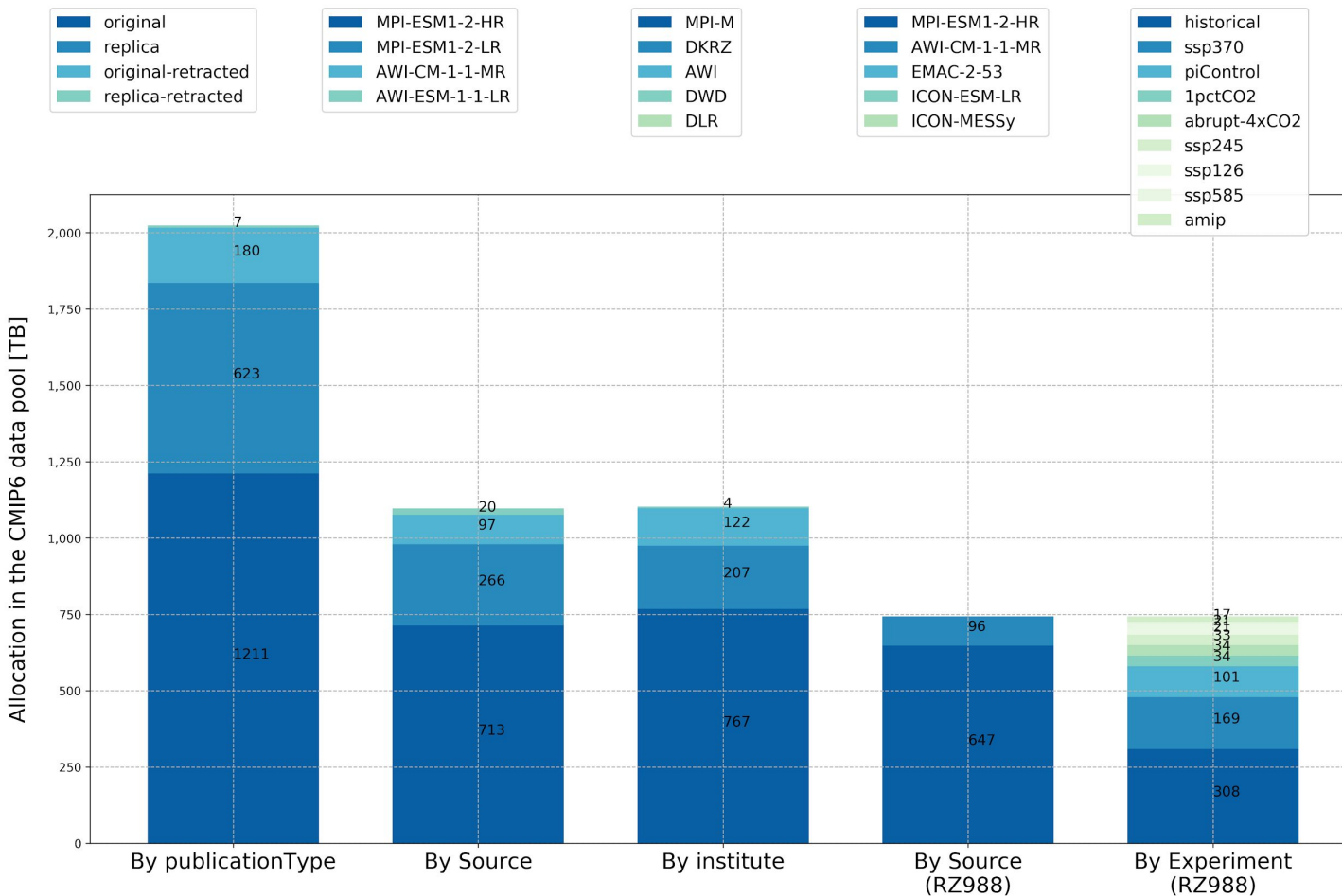
USR1 Skip processing of the current selection file.

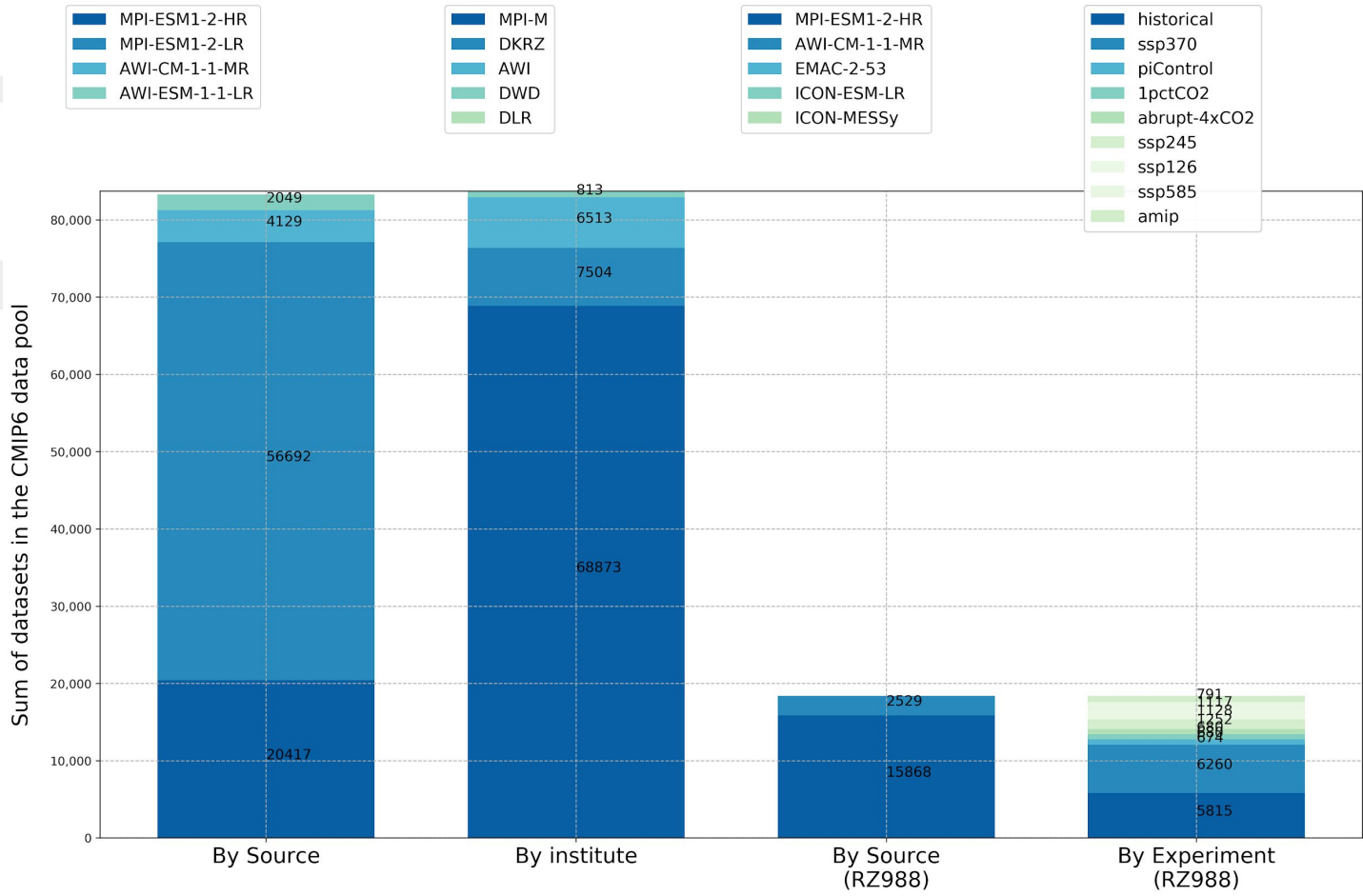
USR2 Exit after the current sel-file.



intake-esm

- Das Python-Paket intake-esm bietet eine Möglichkeit, durch den kompletten CMIP6 Datenpool
 - zu browsen (ähnlich Freva)
 - tabellarisch basierend auf pandas darzustellen
 - im Verlauf mit Xarray zu öffnen.
- Der dazu notwendige Katalog wird täglich neu erstellt, mit einfachen Tests geprüft, zur Verfügung gestellt auf /work/ik1017/Catalogs und alte Versionen archiviert. (Einmal im Monat für CORDEX, für ERA5 in Planung)
- Anwendungsbeispiele und Skripte gibt es hier:
 - <https://gitlab.dkrz.de/mipdata/intake-esm>, [Inklusive Catalogbuilder](#)
 - [Lunchbyte](#)
- Die folgenden Plots basieren u.a. auf diesem [Notebook](#)





Allgemein

- Installation eines eigenen Datenknotens für CMIP6 Daten (esgf3.dkrz.de)
- Support für externe Datenknoten
 - Hilfestellung bei Installation und Betrieb
 - Bereitstellung von Solr Index, Portal und IDP

Publikation

- Publikation der Originale
 - keine größeren Probleme
 - automatische Generierung der CIM Dokumente schlägt häufig fehl: z.B. “Mismatched axis size (1306775 != 1)”
→ wird einfach ignoriert
- Publikation der Replikate
 - Vollständigkeit der Datensätze und Prüfsummen der Files müssen vor Publikation geprüft werden → langwieriger Prozess, wurde weitestgehend automatisiert

CMIP6 Datenpool - Status

CMIP-Datenpool des DKRZ	Datenvolumen [TB]
Gesamtes Datenvolumen	3600¹ 3219²
Primär publizierte Daten	1530¹ 1330²
Replizierte Daten (auf Platte)	~2100¹ ~2000²
Replizierte Daten (ESGF-publiziert)	694¹ 687²
An DICAD beteiligte Institute	1400¹ 1190²

1: *inklusive alter Versionen* | Stand 03.06.2020

2: *exklusive alter Versionen* | Stand 03.06.2020

CMIP6 Datenpublikation

Institute	Anzahl Datensätze	Anzahl zurückgezogene Datensätze	Anzahl Publizierte Replikate
MPI-M	99001	831	73228
INM	8255	0	7880
DKRZ	7504	71	8138
AWI	6718	3	4222
HAMMOZ-Consortium	3920	0	3982
DWD	1080 (davon 10 auf DKRZ Datenknoten)	2	1092
MOHC	24	0	-
AER	12	0	-
UHH	8	0	-
RTE-RRTMGP-Consortium	4	0	-

CMIP6 ESGF Datenpublikation

#	Institute	Number of Datasets [1]	Datenvolumen (in TB) [1]
1	IPSL	1213365	2448.52
2	EC-Earth-Consortium	110741	1806.76
3	MPI-M	174683	1423.05
4	NCAR	1189942	1334.21
...
10	DKRZ	15795	402.26
...
13	AWI	10946	238.53
...
23	DWD	2176	46.06
...
38	UHH	16	0

[1] Originale und Replizierte Daten

Download Statistics

Anzahl der Downloads - Top 10 Source

#	Source	Value
1	AWI-CM-1-1-MR	4518250
2	MPI-ESM1-2-HR	4173330
3	MPI-ESM1-2-LR	2863368
4	INM-CM4-8	1230206
5	INM-CM5-0	811204
6	AWI-ESM-1-1-LR	281914
7	MPI-ESM-1-2-HAM	228650
8	MRI-ESM2-0	179832
9	INM-CM5-H	110602
10	MIROC6	32460

<http://esgf-ui.cmcc.it/esgf-dashboard-ui/cmip6.html>

Download Statistics

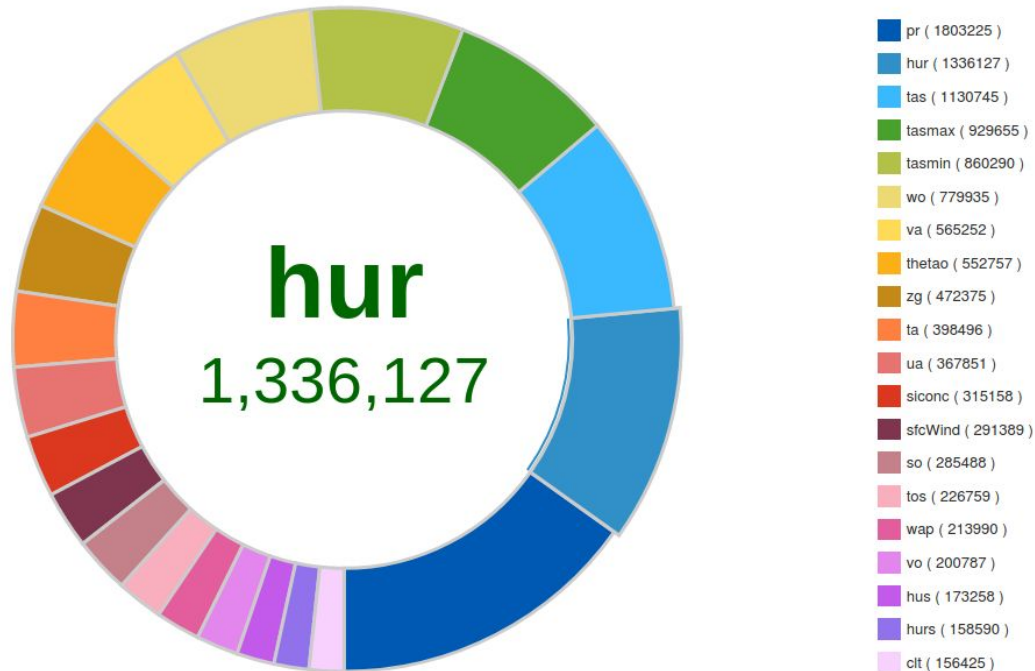
Anzahl der Downloads - Top 10 Experiment

#	Experiment	Value
1	historical	5204304
2	ssp585	2164897
3	ssp126	2056040
4	ssp245	1799325
5	ssp370	1353774
6	piControl	1183697
7	1pctCO2	191316
8	abrupt-4xCO2	165021
9	amip	111714
10	hist-1950	89495

<http://esgf-ui.cmcc.it/esgf-dashboard-ui/cmip6.html>

Download Statistics

Anzahl der Downloads - Top 20 Variable



<http://esgf-ui.cmcc.it/esgf-dashboard-ui/cmip6.html>