

# Regionalisierung von Klimadaten: Datenmanagement

Stephanie Legutke,

DKRZ/Data Management (DM, Lautenschlager)

DWD, Offenbach, 23.-24. März 2015

# Diskussionspunkte

Was sind die drängendsten **Fragen?**

**Lösungen** und **Werkzeuge?**

Realistische mittelfristige **Ziele** und

benötigte **Unterstützung?**

in Bezug auf

- Metadaten (MD)
- Speicherung
- Zugriff
- Kapazitäten

# Metadaten

**Definition: Beschreibung von Daten**

- formlos, free text
- formatiert, Controlled Vocabulary (VC)

**[[Inter]Nationale]Kooperation\* →**

- Suche
  - Interpretation
  - Herkunft
  - Referenzierbarkeit
  - Diagnostik (leicht anwendbar, automatisch, koordiniert, ...)
- } zur Nachnutzung

Im Folgenden wird meistens angenommen, dass die regionalisierten Daten wg. des großen Volumens in einem gemeinsamen\* verteilten Archiv gespeichert werden.

\* Forschungsgruppe(n), Projekt, weltweit

# Metadaten

## Suche, Interpretation

- CV
  - DRS\*
  - Standards
- } ✓ CORDEX+
- ✓ NetCDF4/CF-1.n
- ✓ CMOR, QC am DKRZ

Modified\*\*  
CMIP5 Standard

## Einbindung vorhandener Archivdaten/Formate

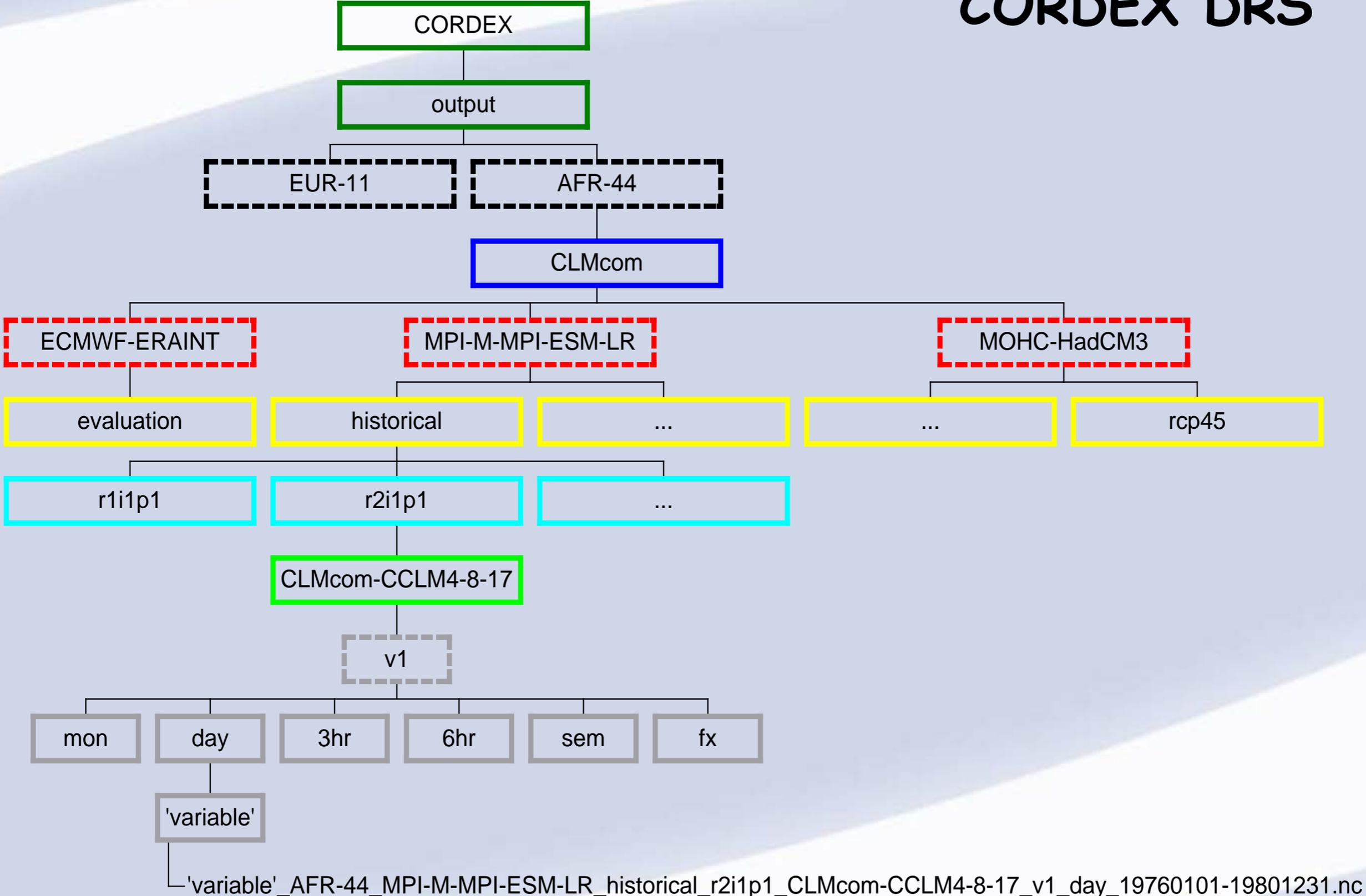
- NetCDF4/CF-1.n ↔ GRIB (ECMWF Aktivität\*\*\*)

\*Dateinamen, Verzeichnisstrukturen

\*\*where necessary

\*\*\*<http://www.ecmwf.int/en/workshop-closing-grib/netcdf-gap>

# CORDEX DRS



## Herkunft

### Metadaten für RCMs und Simulationen:

CommonInformationModell (**CIM**) für RCMs

- EU FP7 Metafor Projekt: CIM für CMIP5
- <http://es-doc.org> (Metafor + ESCurator)
- view** (z.B. ESGF GUI: CIM für CMIP5 ESMs)
- browse**
- compare**
- Questionnaire; pyesdoc**

## Current Selections

(x) [model:MPI-ESM-LR](#)

## Search Categories

Project

Institute

Model

Instrument

Experiment Family

Experiment

1pctCO2 (15)

G1 (19)

G2 (18)

G3 (48)

L1A85 (16)

L1B85 (17)

L2A26 (31)

L2A45 (17)



Search

[Temporal Search](#)  
[Clear search](#)  
[constraints and](#)  
[datacart](#)  
[Search Help](#)  
[Search Controlled](#)  
[Vocabulary](#)

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.  
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

Search All Sites  Show All Replicas  Show All Versions

< 1 2 3 ... 387 388 > displaying 1 to 10 of 3876 search results

Display  datasets per page

[Add All Displayed to Datacart](#)

[Remove All Displayed from Datacart](#)

Results

Data Cart

[project=CMIP5, model=MPI-ESM-LR, Max Planck Institute for Meteorology \(MPI-M\), experiment=1 percent per year CO2, time\\_frequency=day, modeling realm=atmos, ensemble=r1i1p1, version=20120314](#)

Data Node: [bmbf-ipcc-ar5.dkrz.de](#)

**Version: 20120314**

Description: MPI-ESM-LR model output prepared for CMIP5 1 percent per year CO2

Further options: [Add To Cart](#) [Model Documentation](#)

esgf-  
data.dkrz.de

[project=CMIP5, model=MPI-ESM-LR, Max Planck Institute for Meteorology \(MPI-M\), experiment=1 percent per year CO2, time\\_frequency=day, modeling realm=atmos, ensemble=r1i1p1, version=20111006](#)

Data Node: [bmbf-ipcc-ar5.dkrz.de](#)

**Version: 20111006**

Description: MPI-ESM-LR model output prepared for CMIP5 1 percent per year CO2

[Model](#)[Experiment](#)[Simulation](#)[Platform](#)[Grid Spec ▾](#)[Data Object ▾](#)[Dataset QC](#)

# CMIP5 Model : MPI-M - MPI-ESM-LR

[Details](#)[Components](#)

## Overview

<b>Project</b>	cmip5
<b>Institute</b>	MPI-M
<b>Name</b>	MPI-ESM-LR
<b>Long Name</b>	MPI Earth System Model running on low resolution grid
<b>Funder</b>	Bundesministerium fuer Bildung und Forschung
<b>Principal Investigator</b>	<a href="#">Marco Giorgetta</a>
<b>Release Date</b>	2009-01-01
<b>Description</b>	ECHAM6 & JSBACH / MPIOM & HAMOCC coupled via OASIS3. For experiments with the MPI-ESM-LR /MR (compared to MPI-ESM-P): dynamic feedback of vegetation and land use on the climate development is fully included, land cover change data are included from external file, orbital parameters are calculated at every radiation time step.
<b>Doc. Author</b>	Metafor Questionnaire
<b>Doc. Created</b>	2012-07-30

## Citations

### Giorgetta\_etal\_2012

<b>Long Title</b>	M. A. Giorgetta, E. Roeckner, T. Mauritsen, B. Stevens, T. Crueger, M. Esch, S. Rast, L. Kornblueh, H. Schmidt, S. Kinne, B. Möbis, T. Krismer, C. Reick, T. Raddatz and V. Gayler (2012) "The atmospheric general circulation model ECHAM6 - Model description"
<b>Location</b>	<a href="http://www.mpimet.mpg.de/en/science/models/echam.html">http://www.mpimet.mpg.de/en/science/models/echam.html</a>

### HAMMOCC

<b>Long Title</b>	HAMOCC: Technical Documentation
<b>Location</b>	<a href="http://www.mpimet.mpg.de/fileadmin/models/MPIOM/HAMOCC5.1_TECHNICAL_REPORT.pdf">http://www.mpimet.mpg.de/fileadmin/models/MPIOM/HAMOCC5.1_TECHNICAL_REPORT.pdf</a>

## Referenzierbarkeit

- ❑ **Checksummen** (aufwändig RZ)
- ❑ **Archiv Versionierung** (kann verloren gehen)
- ❑ **tracking ID** (CMIP5/CORDEX: NetCDF attribute; **CMOR, UUID**)
- ❑ **PID**
- ❑ **DOI** (aufwändiger WF, zeitverzögert; **DataCite DOI DKRZ**)
  - ❑ **PermanentIdentifizier**
    - **handle server und Infrastruktur**
    - **Lokalisierbarkeit**
    - **versch. Aggregationsebenen**
    - **beim ESGF Publikationsprozess**
    - **vorgeschlagen für CMIP6 (DKRZ)**
    - **Bearbeitung am DKRZ in Eudat**

## Diagnostik

bei verteilter Datenhaltung, Ensembles (z.B. ESGF) →

### lokale zentrale Ausführung

☹️ download großer Datenmengen

(Transferraten, Speicherkapazitäten)

### verteilte Ausführung

☹️ Zugang, Portierung der Werkzeuge

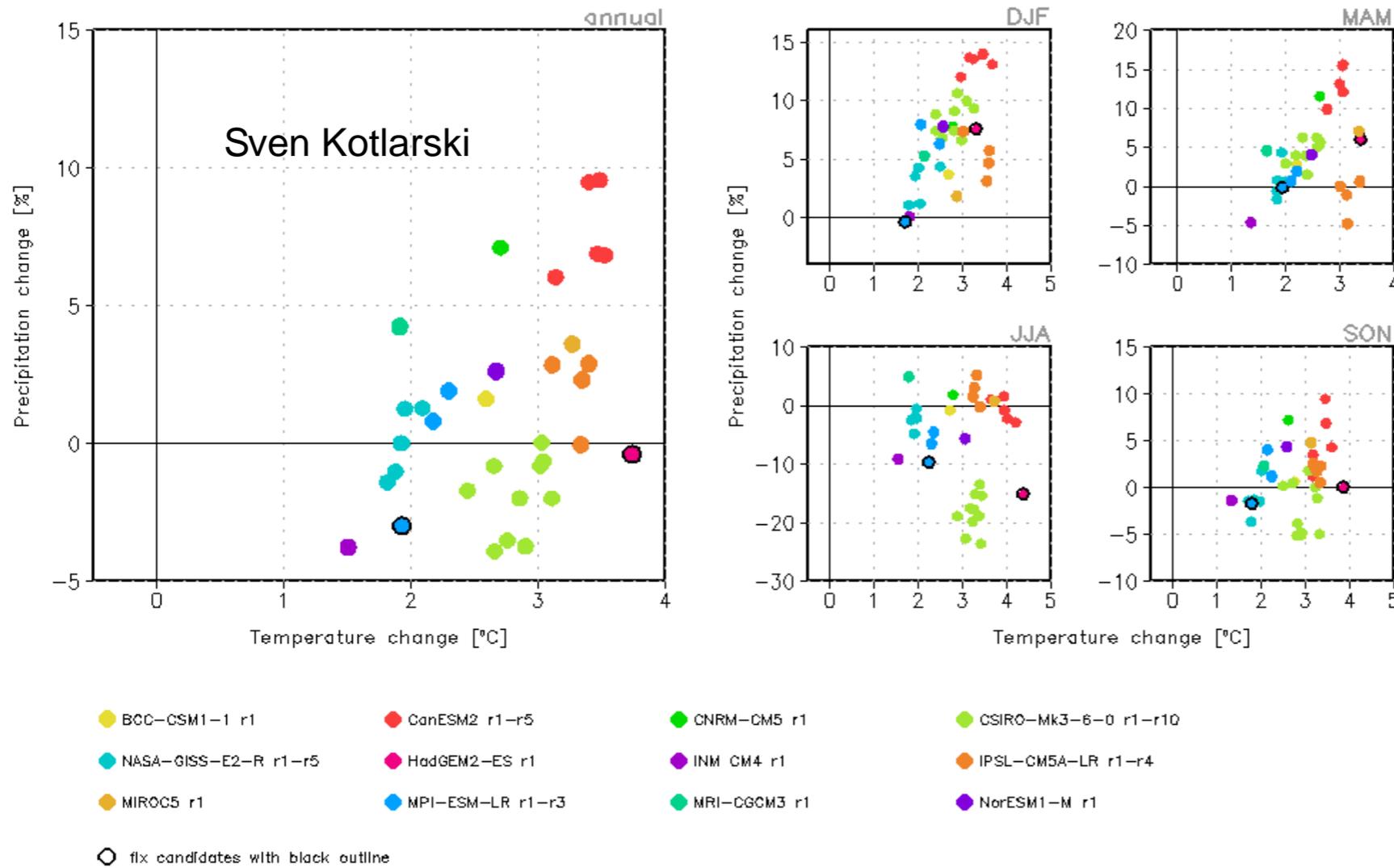
☹️ Koordinierung von Methoden

### koordinierte Ausführung von Standarddiagnostiken

😊 **ESMValTool** (DLR, Veronika Eyring)

○ auf ESGF Datenknoten

○ interfaced with ESGF (DKRZ)



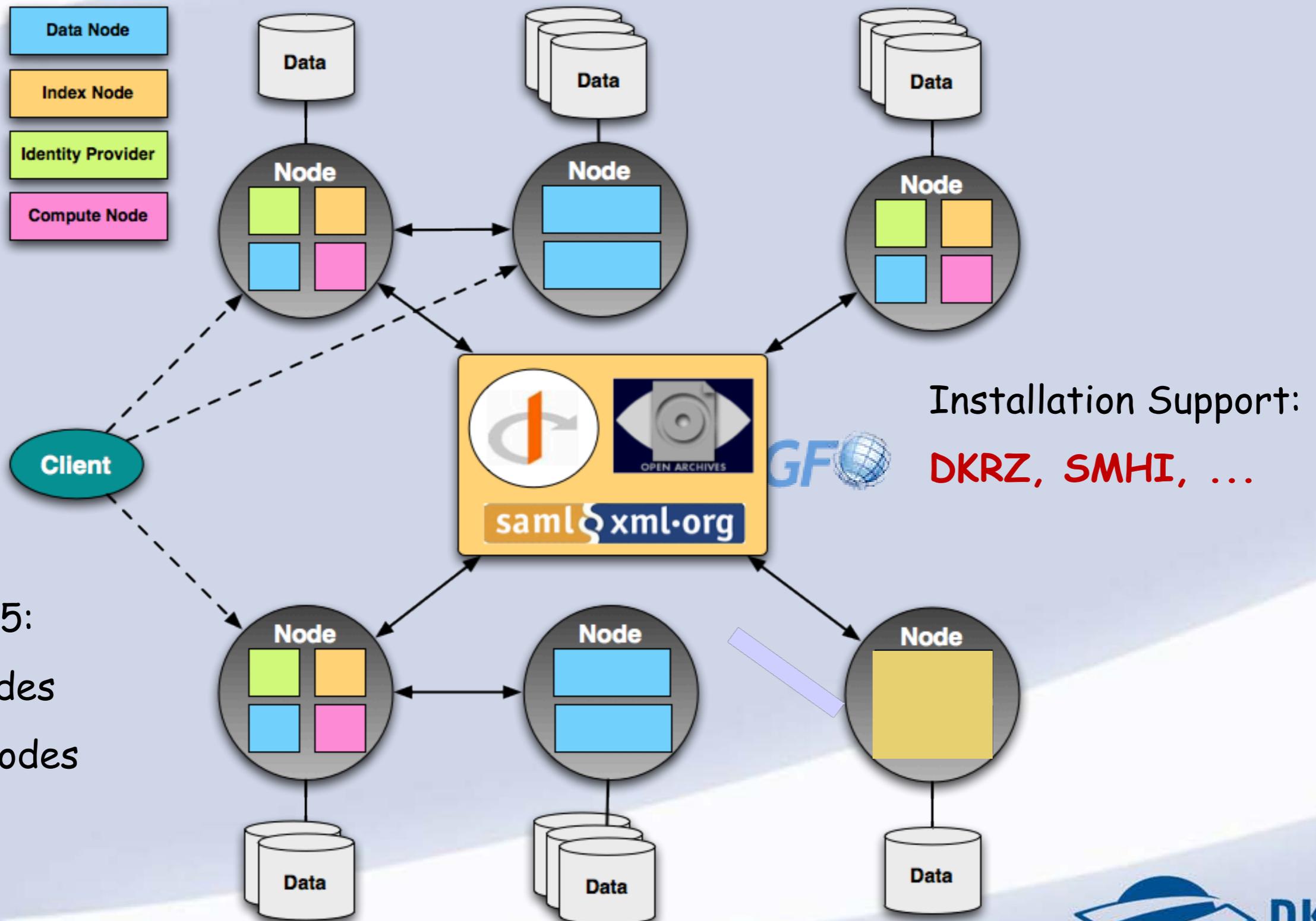
## possible ESMValTool Diagnostic:

Support ESM selection for downscaling

- help choosing ESM forcing (cold/warm, wet/dry model, ...) per domain
- maximize ESM result spread



## Climate Model Data Federation (ESGF)



March 2015:  
41 data nodes  
22 index nodes

# Speicherung & Zugriff

- ESGF Datenknoten
  - saml Zugriffskontrolle:
    - ,unrestricted' / ,non-commercial only'
  - CMIP6 wünscht ,unrestricted' access
  - MoU in CORDEX : ,use CMIP6 ToU' (Problem für CLMcom?)
  
- WDCC (WorldDataCenter of Climate; Cera)
  - Langzeitarchivierung (LZA): > 10 Jahre
  - DataCite DOI Daten
  - Bestand (4 PB); bei größeren Datenvolumen (~20TB) ist ein DataManagementPlan gefordert
  - Zugriff nur mit account; frei oder Datenfreigabe mit Zustimmung

# Speicherung & Zugriff

- HPSS Bandarchiv am HLRE-2/3:
  - /arch/<proj> 90%; Speicherung bis 1 Jahr nach Projektende garantiert
  - 10% in /doku/<proj> für  $\geq 10$  Jahre
  - owner, group, others permissions separately
  
- Cloud Speichersystem Swift-based on OpenStack:
  - für unstrukturierte Daten; ohne DKRZ account  
(z.B. den CORDEX Datenpublikationsprozess)
  - 300 TB auf 2x3 Datenknoten redundant
  - nach 2 Wochen nur noch 1 Kopie  
(kein Zugriff wenn server down)
  - 5 GB per account; project space proportional to HLRE-3
  - <https://swiftbrowser.dkrz.de>

# Kapazität

	HLRE-2 ‚blizzard‘	HLRE-3 ‚mistral‘	Phase I Jun 15	Phase II Jun 16
<b>Vendor</b>	IBM	Bull		
<b>Architecture</b>	AIX	Linux cluster		
<b>File System</b>	GPFS	Lustre		
<b>Processors per node</b>	16 dual core	24 cores		
<b>Performance</b>	0.158 PF/s	3.000 PF/s	8 x	19 x
<b>Cores</b>	8,400	60,000	2.4 x	7.1 x
<b>System memory</b>	20 TB	170 TB	3.7 x	8.5 x
<b>Disk storage capacity</b>	6 PB	50 PB	2.9 x	8.3 x
<b>Disk storage throughput</b>	30 GB/s	400 GB/s		13.3 x
<b>Disk drives</b>	7,200	8,500		1.2 x
<b>Archive capacity</b>	53 PB	335 PB	3.6 x	6.3 x
<b>Archive throughput</b>	9.6 GB/s	21 GB/s		2.2 x
<b>Power consumption</b>	1.6 MW	1.3 MW		0.8 x
<b>Investment</b>	30 M €	41 M €		1.4 x
<b>HPSS</b>	8 PB/yr	75 PB/yr		9.4 x
<b>WDCC</b>	1 PB/yr	8 PB/yr		8.8 x

# Fund rising initiative

**Partners:** BTU/CLMcom, CS2.0, DKRZ, DLR, DWD, MPI-M, ..?

**Funding Agency:** BMBF?

**Working Title:** Services around CMIP6 in Hamburg/Germany

Activities:

- configure, install, and use of CMOR (all MIPs)
- compatibility of CDOs with CMIP6+ standards
- CMOR-lite (easy to use toolkit to ,CMORize')
- CMIP6+ simulations (ScenarioMIP, DAMIP, CORDEX, ...)
- Processing/formatting data for (ESGF) CMIP6 archive
  - simulations
  - observations
- ESGF services
- additional data products (ECV, climate indices, ..?)
- ESMValTools

**Ende**