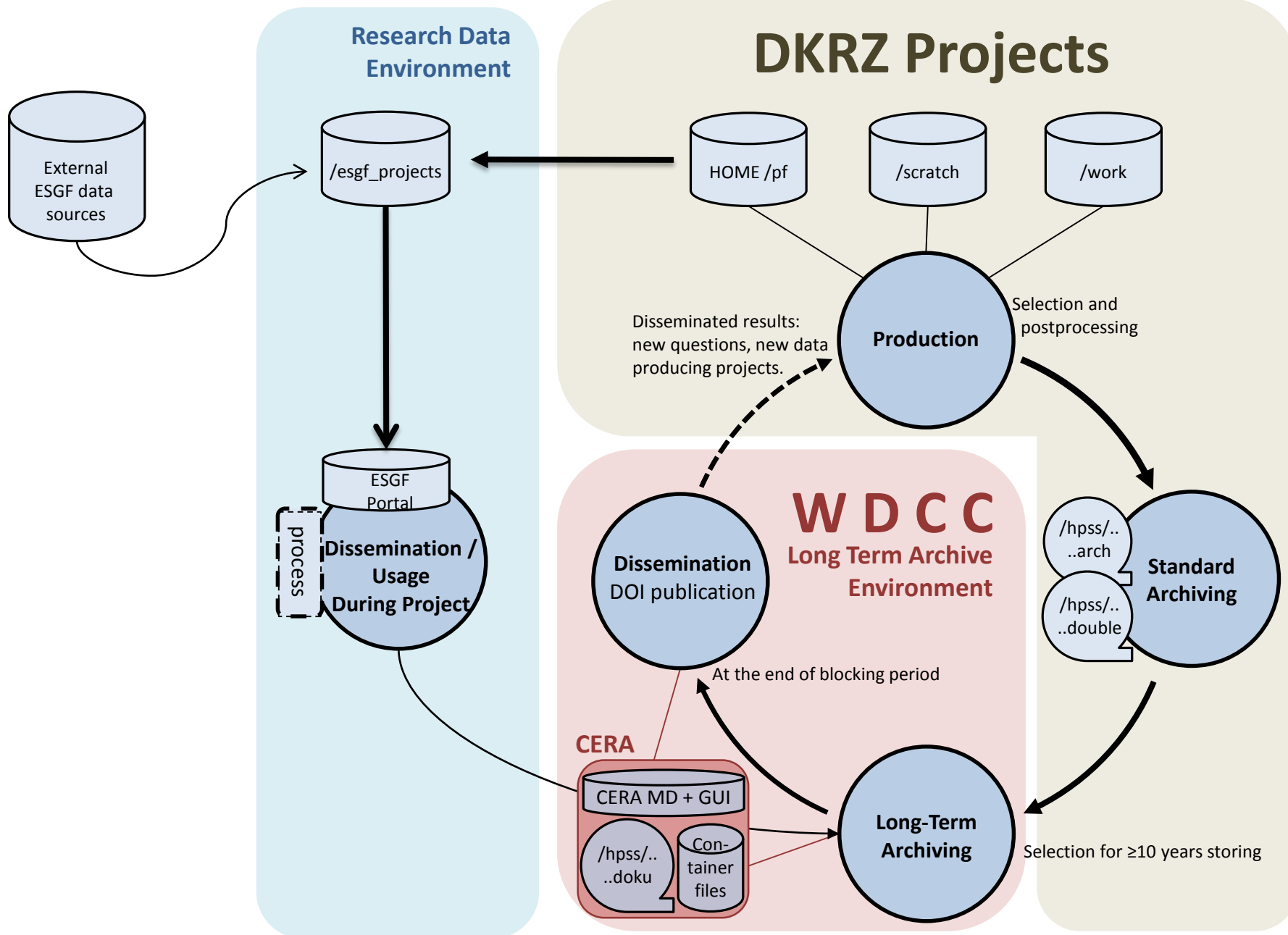


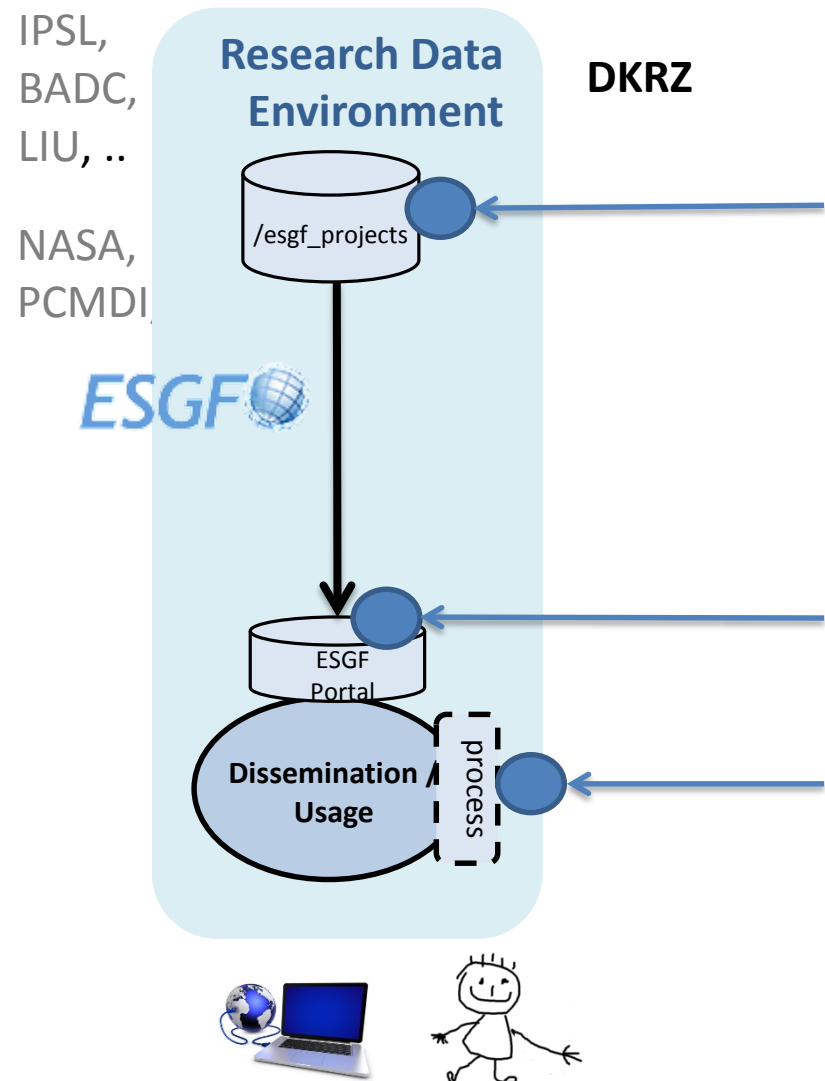
Workshop DKRZ Datendienste

Datenbereitstellung in ESGF¹ und internationale Nutzung

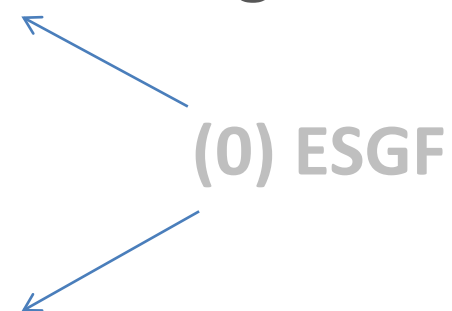
¹) Earth System Grid Federation



Übersicht



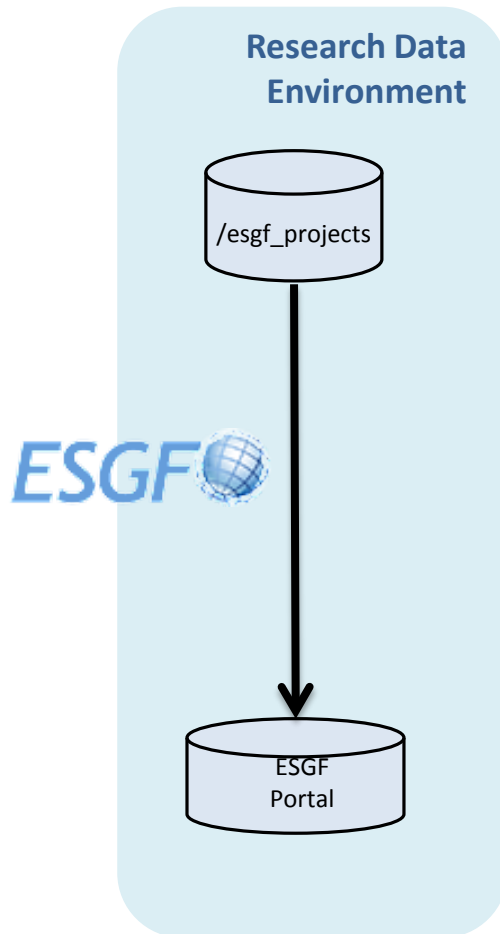
(1) Datenbereitstellung



(2) Datensuche / Zugriff

(3) Datenauswertung

ESGF

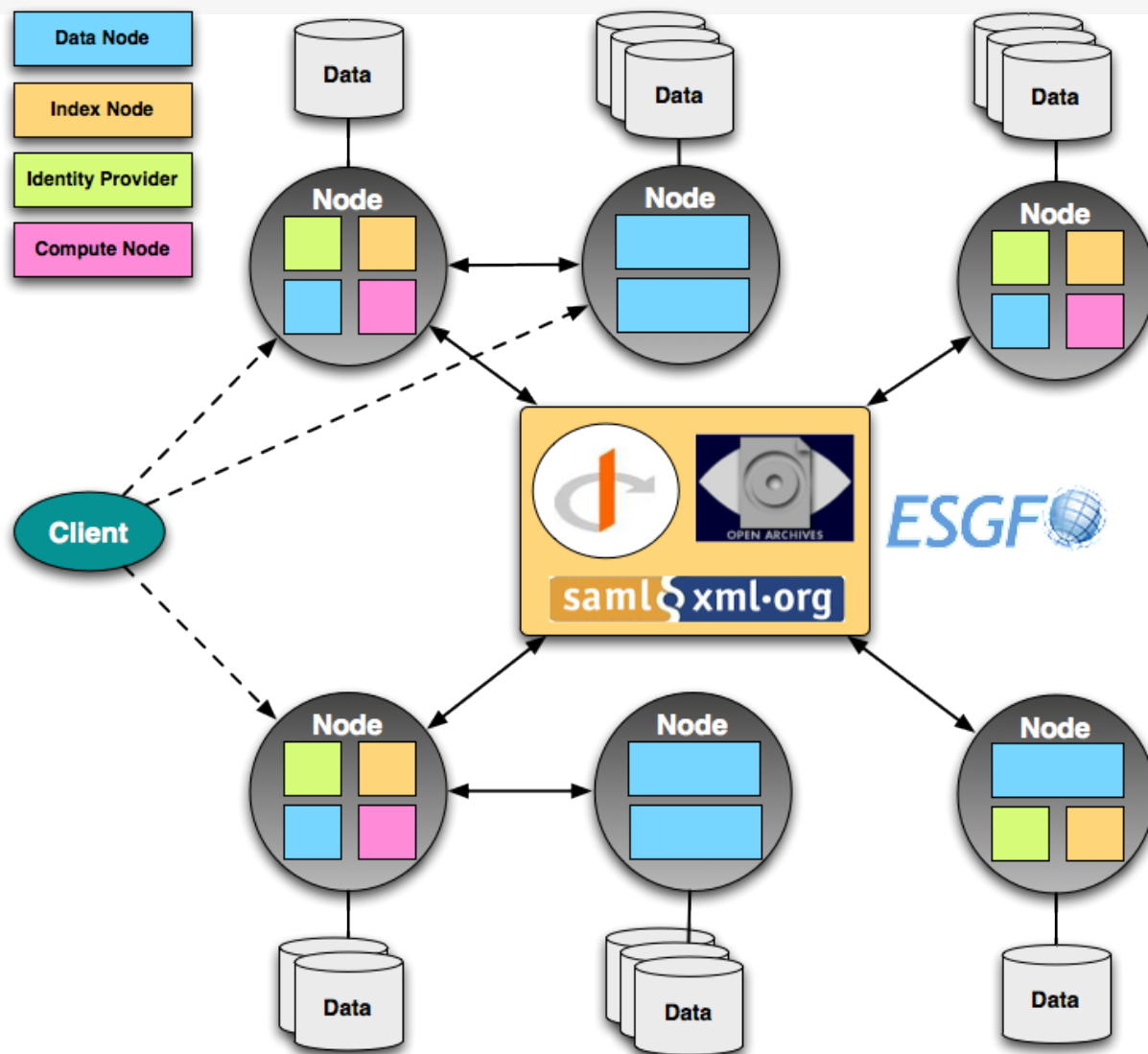


ESGF (Earth System Grid Federation, <http://esgf.org>):

- Weltweites, offenes Konsortium von Institutionen, Laboren und Zentren, die Klimaforschung unterstützen.
- ESGF wird aus unterschiedlichen, unabhängigen Quellen gefördert (z.B. NASA, NOAA, NSF in den USA and IS-ENES in Europe)
- Modulare open source Komponenten zur Datenbereitstellung / Datensuche und zum Datenzugriff



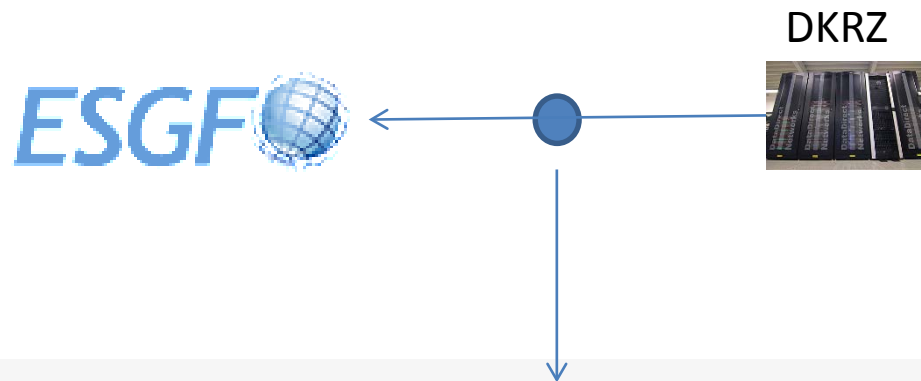
ESGF P2P Architecture: Data Node, Index Node, Compute Node, IdP



ESGF: mehr als CMIP5 ...

Über ESGF angebotene Projekt-Daten:

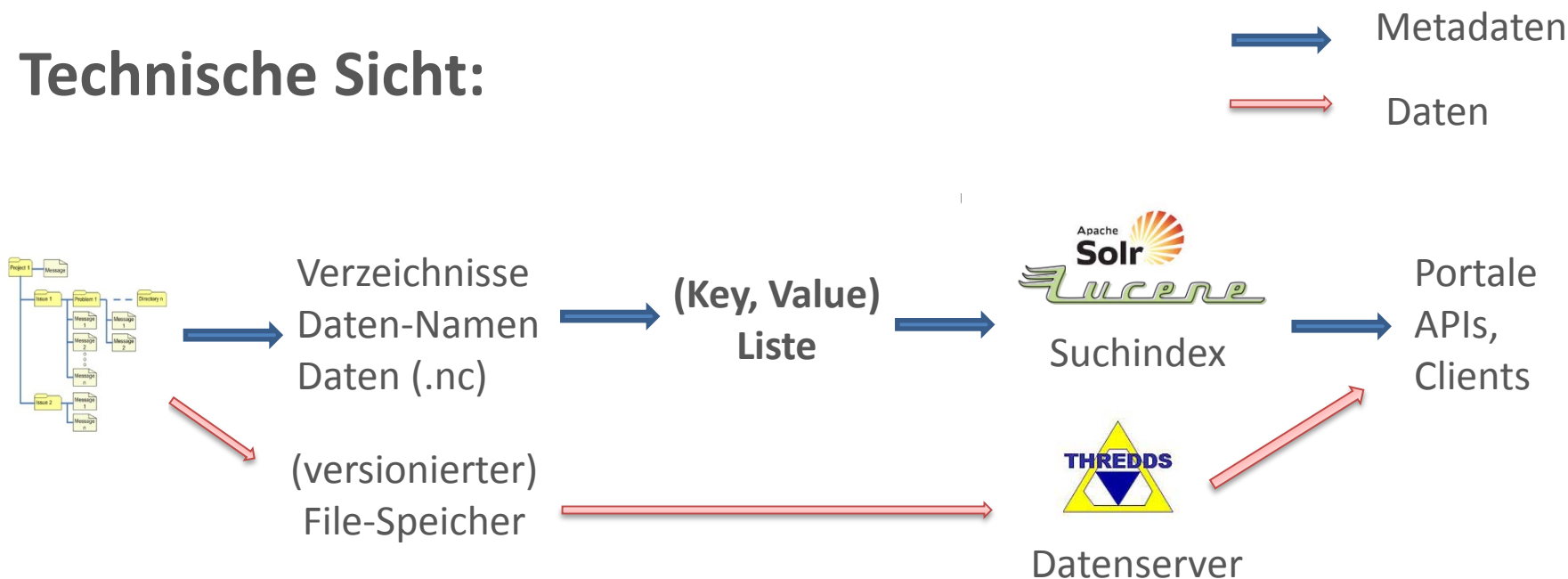
- **Phases 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5)**
- Coordinated Regional climate Downscaling Experiment (CORDEX)
- Climate Science for a Sustainable Energy Future (CSSEF)
- European Union Cloud Intercomparison, Process Study & Evaluation Project (EUCLIPSE)
- Geo-engineering Model Intercomparison Project (GeoMIP)
- Land-Use and Climate, Identification of robust impacts (LUCID)
- Paleoclimate Modeling Intercomparison Project (PMIP)
- Transpose-Atmospheric Model Intercomparison Project (TAMIP)
- Clouds and Cryosphere (cloud-cryo)
- Observational products more accessible for coupled model intercomparison (obs4MIPs)
- Reanalysis for the coupled model intercomparison (ana4MIPs)
- Dynamical Core Model Intercomparison Project (DCMIP)
- Community Climate System Model (CCSM)
- Parallel Ocean Program (POP)
- North American Regional Climate Change Assessment Program (NARCCAP)
- Carbon Land Model Intercomparison Project (C-LAMP)
- Atmospheric Infrared Sounder (AIS)
- Microwave Limb Sounder (MLS)



Datenbereitstellung in der ESGF

Datenbereitstellung in ESGF

Technische Sicht:

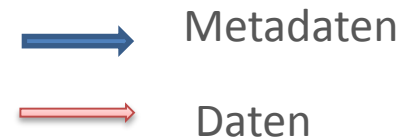


→ Basis-System sehr flexibel

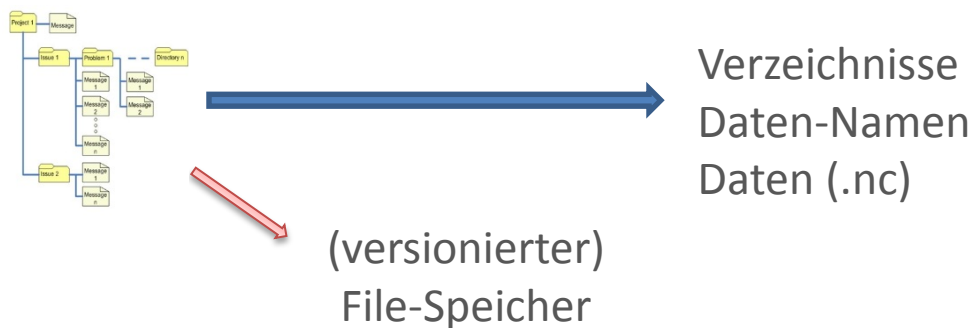
→ Stark kontrollierte Datenbereitstellung notwendig !

Datenbereitstellung in ESGF

Datenprojekt Sicht: Festlegungen notwendig !!



- **Festlegung** einer Schablone für die Directory Struktur zur Ablage von files
z.B. `../myproj/institute/model/m_version/run/variable/filename`
- **Festlegung** einer Daten-Namens Schablone (DRS mit Vokabularen)
z.B. `pr_day_MPI-ESM-LR_1pctCO2_r1i1p1_18500101-18591231.nc`
- **Festlegung** der File-Metadaten (z.B. netcdf header info)
z.B. netcdf CF + Festlegung von Vokabularen für bestimmte Attribute



Beispiel: CORDEX Festlegungen

- Mehrere Dokumente: “Archive design”, “quality control”, “node admin” ...

- File naming conventions:

VariableName_Domain_GCMModelName_CMI P5ExperimentName_CMI P5EnsembleMember_RCMModelName_RCMVersionID_Frequency_StartTime-EndTime.nc

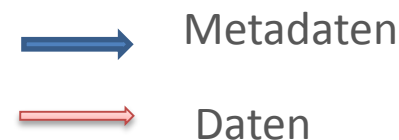
- Required attributes:, NetCDF and CF conventions:

We require the global attributes `institute_id`, `model_id` (identifying the RCM, the same as `RCMModelName` in the file name), `experiment_id`, `experiment`, `contact`, `product`, `Conventions`, `creation_date`, `frequency` as defined in CMI P5

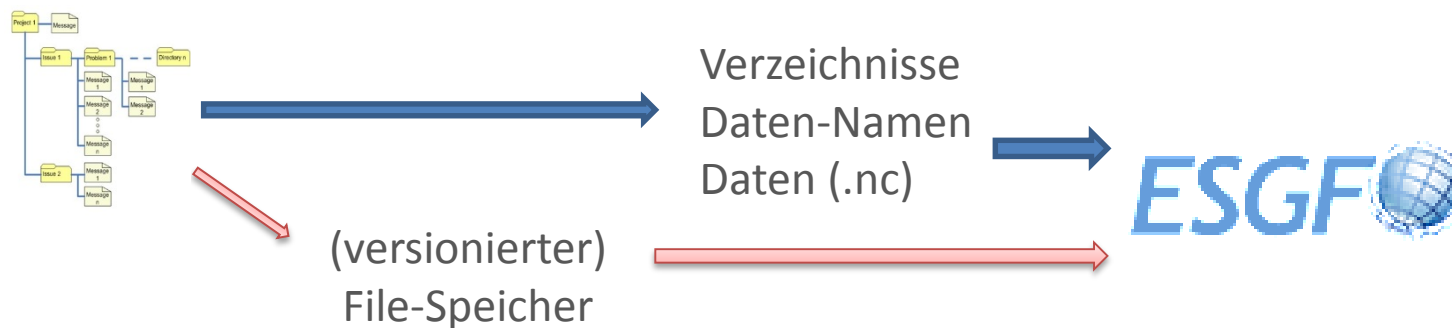
```
// global attributes:
:Conventions = "CF-1.4" ;
:contact = "rossby.data@smhi.se" ;
:creation_date = "2012-05-31-T16:01:08Z" ;
:experiment = "Evaluation run with reanalysis forcing" ;
:experiment_id = "evaluation" ;
:driving_experiment = "ECMWF-ERAINT, evaluation, r1i1p1" ;
:driving_model_id = "ECMWF-ERAINT"
;
:driving_model_ensemble_member = "r1i1p1" ;
:driving_experiment_name = "evaluation" ;
:frequency = "day" ;
:institution = "Swedish Meteorologi
cal and Hydrological Institute, Rossby Centre" ;
:institute_id = "SMHI" ;
:model_id = "SMHI-RCA4" ;
:rcm_version_id = "v1" ;
:project_id = "CORDEX" ;
:CORDEX_domain = "AFR-44" ;
:product = "output" ;
:references = "http://www.smhi.se/en/Research/Research-departments/climate-
research-rossby-centre" ;
:tracking_id = "d3afe777-5792-4e07-8781-ca469c0078ec" ;
:rossby_comment = "201106: CORDEX AFRICA 0.44 deg | RCA4 v1 | ECMWF ERA-
INTERIM | L40" ;
:rossby_run_id = "201106" ;
:rossby_grib_path = "/nobackup/ross
by15/rossby/joint_exp/cordex/201106/raw/" ;
```

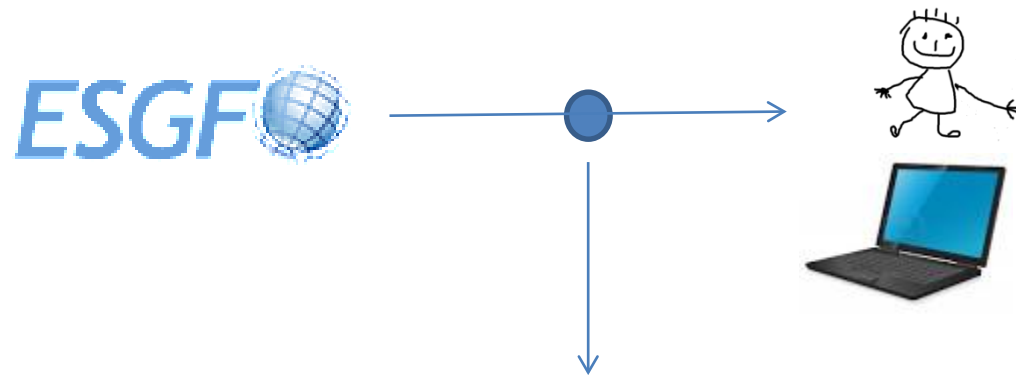
Datenbereitstellung (in ESGF)

DKRZ Dienstleistungen:



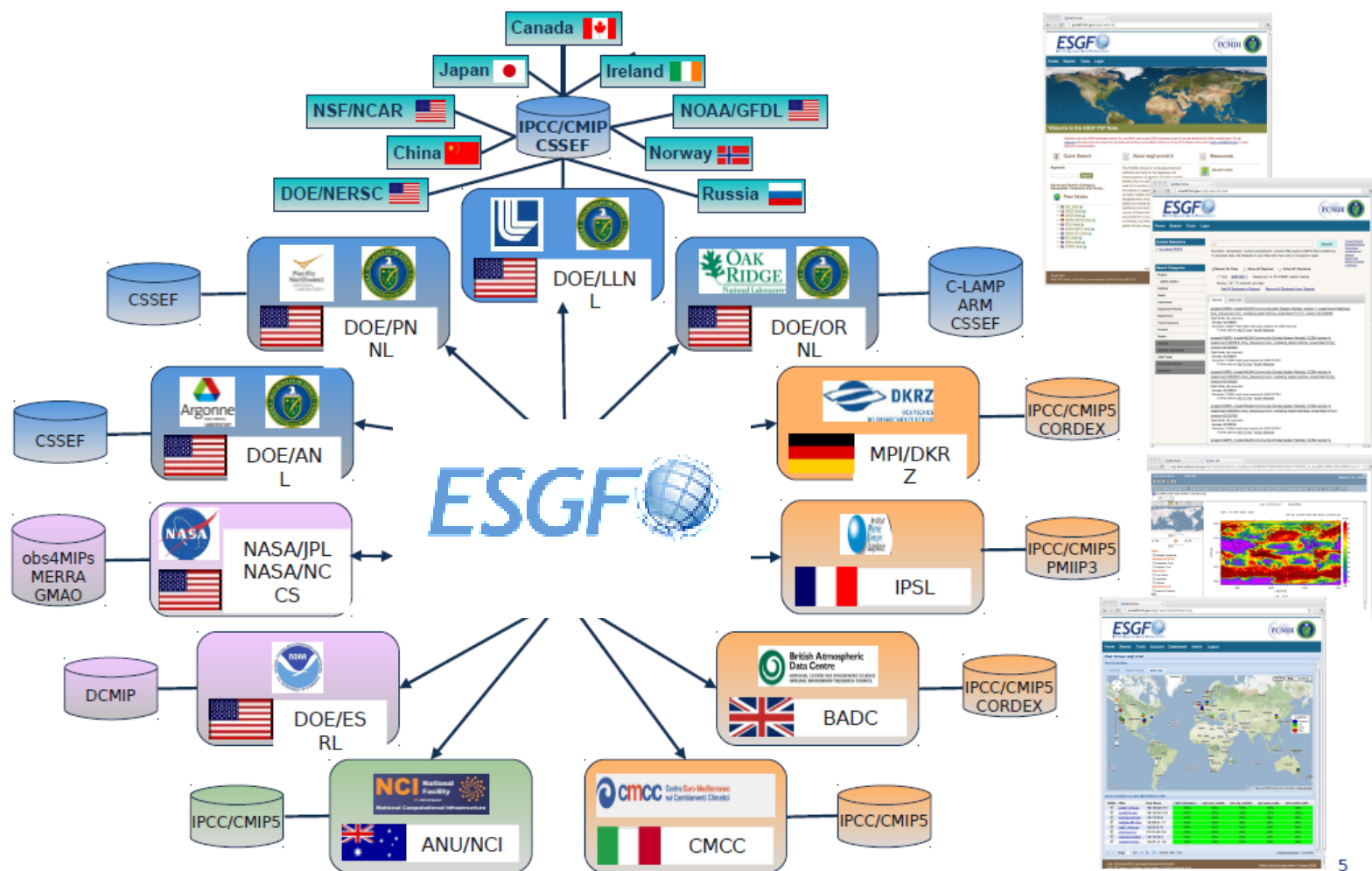
- **Unterstützung** bei Projekt-Festlegungen von Schablonen
- **Entwicklung** von automatisierten Daten-Prüf Routinen der Projektfestlegungen (bzw. Entwicklungsunterstützung)
- **Veröffentlichung** der Daten auf DKRZ ESGF Datenknoten





Datensuche und Zugriff

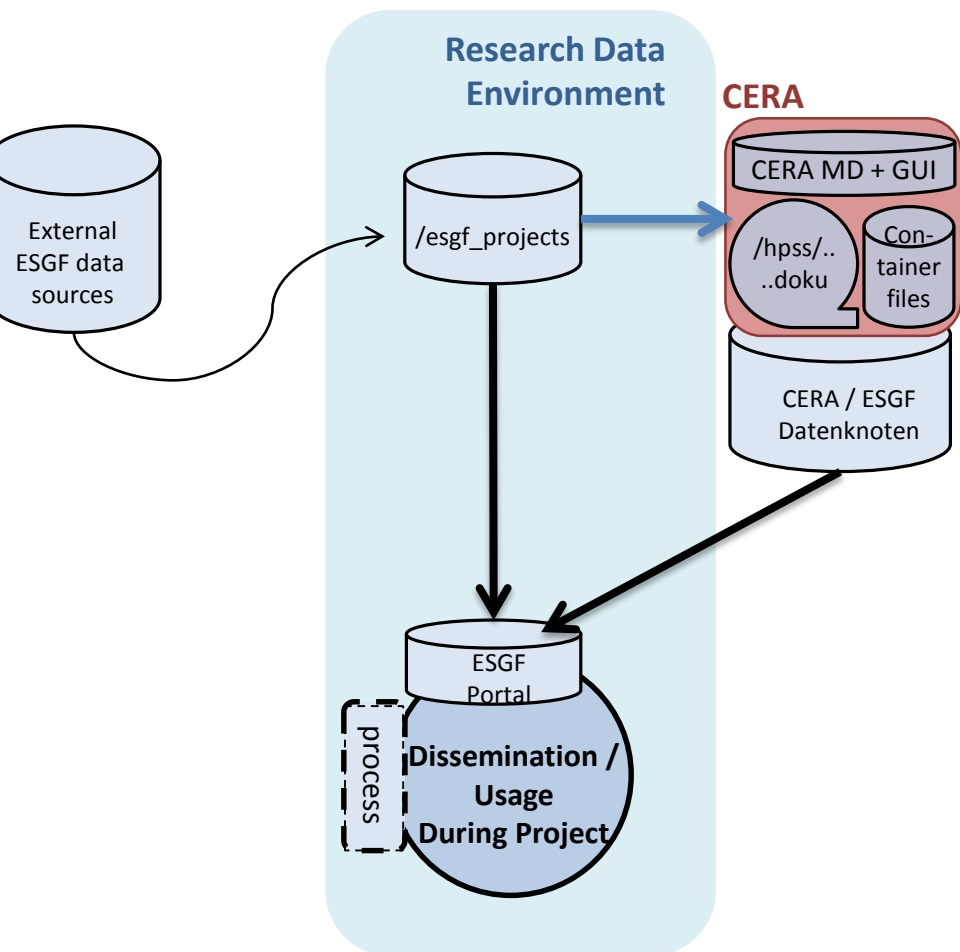
ESGF



5

DEMO

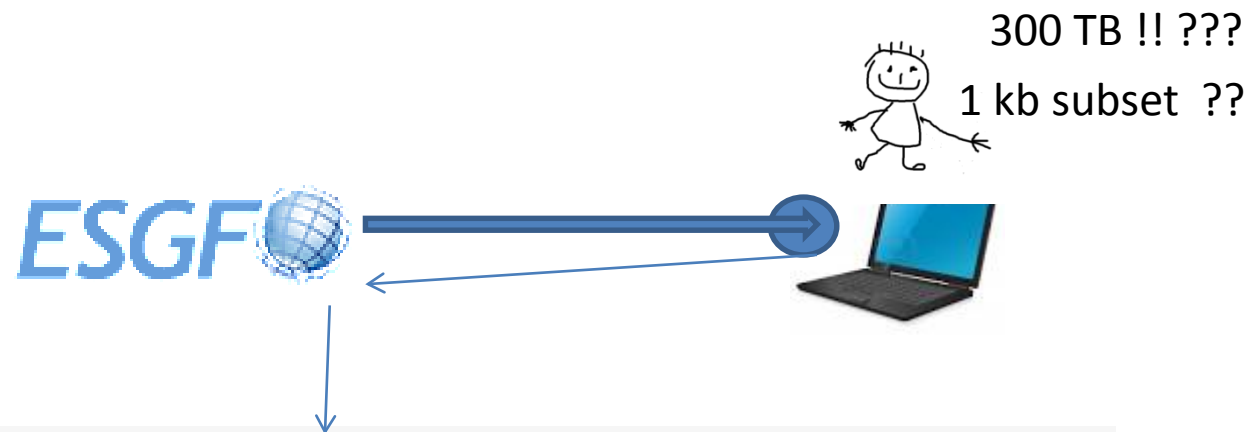
ESGF / CERA – WDCC Integration



WDCC / CERA kann als ESGF Datenknoten fungieren (für bestimmte Projekt-Daten)

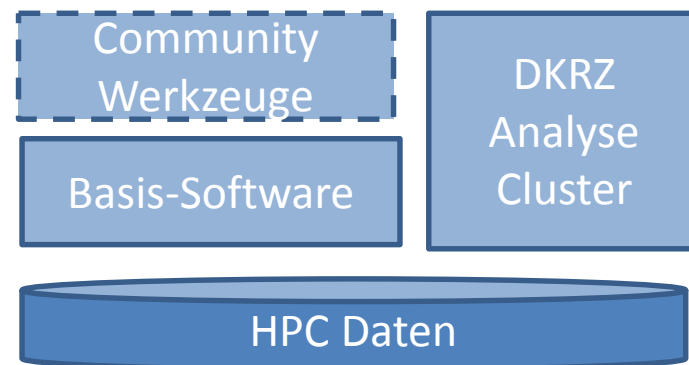
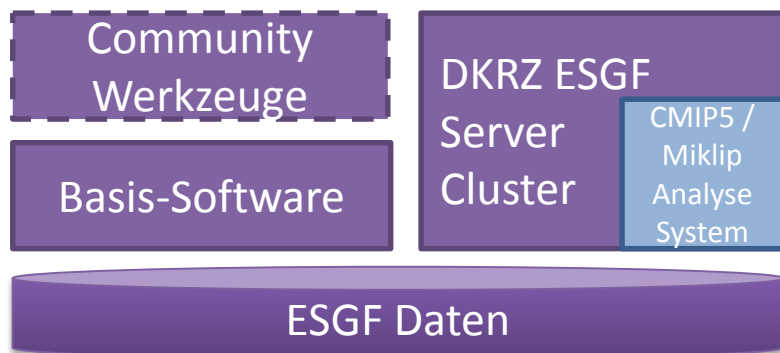
ESGF / CERA – WDCC Integration

- Auch langzeitarchivierte Daten können über ESGF such- und zugreifbar gemacht werden
- Für CMIP5 Daten umgesetzt
- Bei Bedarf auch für andere ESGF Daten möglich
(benötigt Anpassungs-arbeiten des DKRZ internen workflows)



Datenauswertung und Nutzung

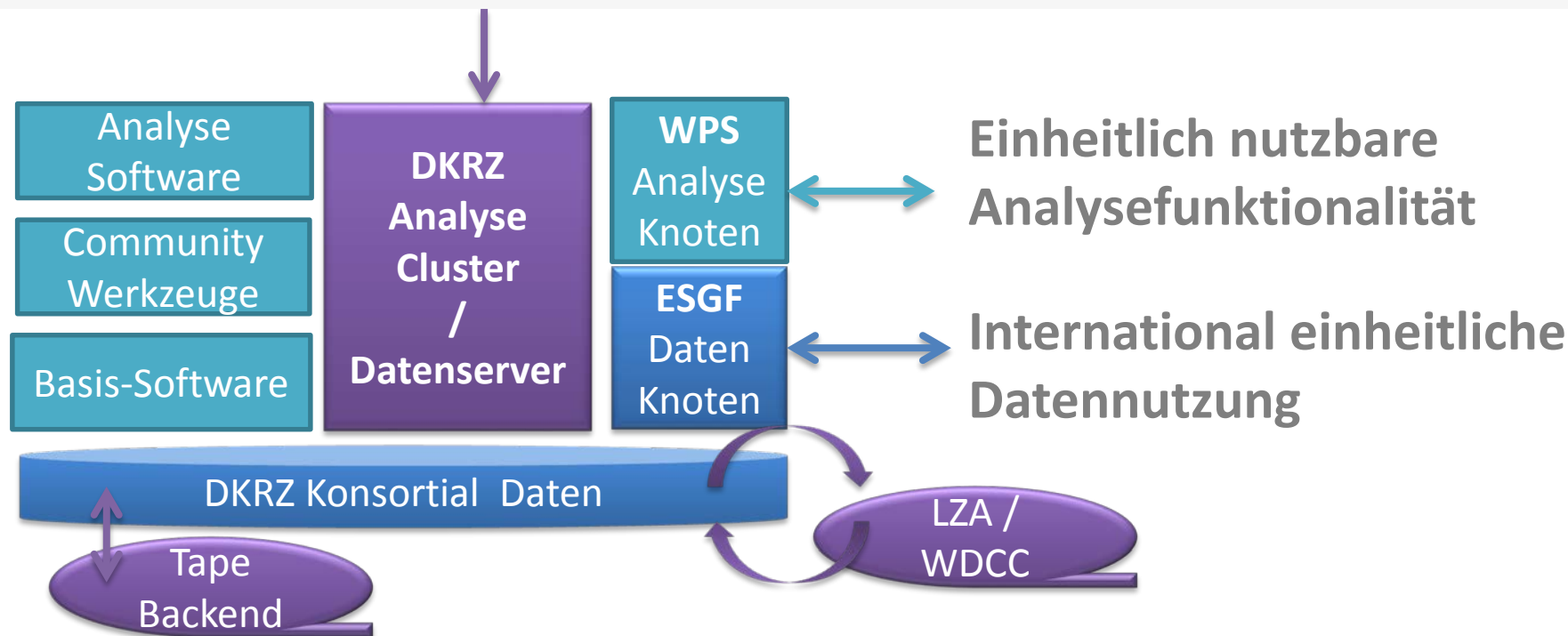
Datenauswertung @ DKRZ: Status



“Basis-Software” ist z.B.:

Name	Bereitsteller	Beschreibung	ESGF Verwendung
Netcdf4 (+opendap)	Unidata	Basis-Bibliothek	Arbeiten auf heruntergeladenen files (remote netcdf file Zugriff)
Netcdf4 linked clients z.B. netcdf4-python, matlab, R, ..	Verschiedene	Community Werkzeuge	Direkter ESGF Zugriff über opendap der netcdf4 Bibliothek
cdo nco	MPI-M NCAR	Auswerte ToolBox	Arbeiten auf ESGF files
UV-CDAT	PCMDI	Auswerte und Visualisierungs-Toolbox	Arbeiten auf und Visualisieren von ESGF files

Datenauswertung @ DKRZ: nächste Schritte



Einheitlich nutzbare
Analysefunktionalität

International einheitliche
Datennutzung

- Daten: Projekt-einheitliche Ablage → ESGF Veröffentlichung
 dynamische Datenprodukte → WPS Veröffentlichung
- Analysefunktionalität: Projekt-einheitliche Ablage → WPS Veröffentlichung

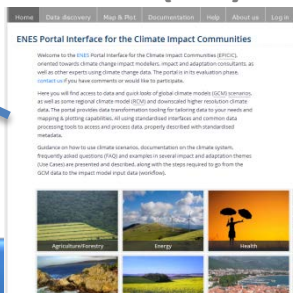
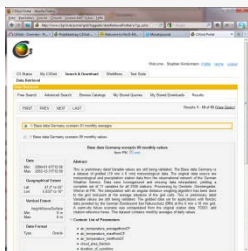
Datenauswertung: nicht-lokale Sicht

Portale

Downscaling Portal (UC)

Climate4impact (IS-ENES)

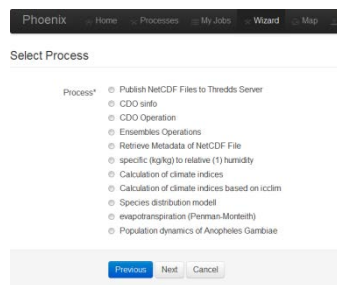
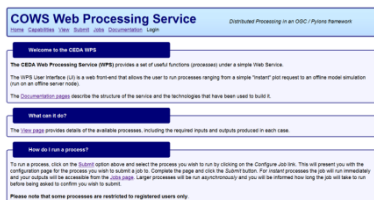
C3Grid



WPS



WPS Dienste



COWS framework (BADG)

ClimDaps Framework (DKRZ)

WPS Frameworks

Datenauswertung und Nutzung

Mögliche DKRZ Dienstleistungen:

- **Unterstützung** beim Einbinden von Auswerte-Funktionalität in WPS
- **Unterstützung** beim Betrieb von am DKRZ gehosteten WPS Diensten (VM basiert oder auf physikalischen Ressourcen)

(DKRZ: Basisbetrieb von WPS Diensten,
Projekt/Endnutzer: Pflege der Funktionalität)

Datenauswertung und Nutzung

Mögliche DKRZ Dienstleistungen:

- **Unterstützung** beim Einbinden von Auswerte-Funktionalität in WPS
- **Unterstützung** beim Betrieb von am DKRZ gehosteten WPS Diensten (VM basiert oder auf physikalischen Ressourcen)

(DKRZ: Basisbetrieb von WPS Diensten, Projekt/Endnutzer: Pflege der Funktionalität)

Zusammenfassung:

Das DKRZ bietet rund um ESGF einheitliche Dienste zur Bereitstellung von Projektdaten und Analysefunktionalitäten

- **Datenbereitstellung:** ESGF Vorbereitung / Qualitätsprüfung / ESGF Datenpublikation
- **Datenarchivierung:** Archivierung / ESGF Publikation archivierter Daten
- **Bereitstellung von Auswerte-Funktionalität:** “Patenschaft” für die Funktionalität durch Projekt/Endnutzer – DKRZ bietet Basis-Plattform

Zusammenfassung:

- “Konsortialdaten” Projekt Beantragung notwendig
- Abstimmung der Datenbereitstellung:

data@dkrz.de oder
esgf@dkrz.de
- Bereitstellung von Analysefunktionalität über WPS

bis Mitte 2015 “prä-operativ”
– ab Mitte 2015 Produktionsbetrieb möglich

Danke

Fragen ?

Diskussion !

Anhang

ESGF nodes deployment (map)



ESGF (P2P) Portale

Home Search Tools Login

Welcome to the ENES archive at

Welcome to the new CMIP5 distributed archive. [gateways](#) will remain active and output from all report bugs and provide feedback.

Quick Search

Keyword:

Advanced Search (Category, Geospatial, Temporal, and more)...

Peer Nodes

- ANL Node
- BADC Node
- CMCC Node
- DKRZ Node
- DKRZ CMIP5 Node
- IPSL Node
- NASA-GSFC Node
- NASA-JPL Node
- NCI Node
- ORNL Node
- PCMDI Node

Home Search Tools Login

selected search facets

search across all sites or locally

search results

search facet selection

select data for later retrieval

Current Selections

- [remove all](#)
- [\(x\) project:CMIP5](#)
- [\(x\) institute:CNRM-CERFACS](#)
- [\(x\) time frequency:mon](#)
- [\(x\) realm:atmos](#)
- [\(x\) ensemble:r1i1p1](#)
- [\(x\) variable:tas](#)

Search Categories

Project
Institute
Model
Instrument
Experiment Family
Experiment
Time Frequency
Product
Realm
Variable
Variable Long Name
CMIP Table

Examples: *temperature*
To download data: add *CMIP5 AND variable:hus* to search criteria. Click on *Expand* or *wget*.

Search All Sites Show All Replicas Show All Versions

< 1 2 3 4 > displaying 1 to 10 of 39 search results

Display datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Results Data Cart

Replica [project=CMIP5_model=CNRM-CM5_Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique \(CNRM/CERFACS\)_experiment=1_percent_per_year_CO2_time_frequency=mon_modeling_realm=atmos_ensemble=r1i1p1_version=20110701](#)
Data Node: [cmip3.dkrz.de](#)
Version: **20110701**
Description: CNRM-CM5 model output prepared for CMIP5 1 percent per year CO2
Further options: [Remove From Cart](#) [Model Metadata](#)

[project=CMIP5_model=CNRM-CM5_Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique \(CNRM/CERFACS\)_experiment=abrupt_4XCO2_time_frequency=mon_modeling_realm=atmos_ensemble=r1i1p1_version=20110701](#)
Data Node: [cmip3.dkrz.de](#)
Version: **20110701**
Description: CNRM-CM5 model output prepared for CMIP5 abrupt 4xCO2
Further options: [Add To Cart](#) [Model Metadata](#)

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

Current Selections

(x) project:CORDEX

Examples: *temperature*, "*surface*"
To download data: add datasets to your Data Cart

Search Categories

- Project < 1 2 3 ... 1252 1253 >
- Institute Display 10 datasets per page
- Model [Add All Displayed to Data Cart](#)
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name
- CMIP Table
- CF Standard Name
- Ensemble
- Domain
- Driving Model
- Data Node

Search All Sites Show All Replicas

Results Data Cart

- [cordex.EUR-44.SMHI.CCCma-CanESM2.historical.r0i0p0.RCA4-v1.fx.orog.v20130927](#)
Data Node: esg-dn1.nsc.liu.se
Version: 20130927
No description available.
Further options: [Add To Cart](#)
- [cordex.EUR-44.SMHI.CCCma-CanESM2.historical.r0i0p0.RCA4-v1.fx.orog.v20130927](#)
Data Node: esg-dn1.nsc.liu.se
Version: 20130927
No description available.
Further options: [Add To Cart](#)
- [cordex.EUR-44.SMHI.CCCma-CanESM2.historical.r1i1p1.RCA4-v1.day.cliv](#)
Data Node: esg-dn1.nsc.liu.se
Version: 20130927
No description available.

Current Selections

(x) project:CORDEX

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*.

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

Search Categories

- Project < 1 2 3 ... 1252 1253 >
- Institute
- Model
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name
- CMIP Table
- CF Standard Name
- Ensemble
- Domain
- Driving Model
- Data Node

Search All Sites Show All Replicas Show All Versions

< 1 2 3 ... 1252 1253 > displaying 1 to 10 of 12523 search results

Display 10 datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Results Data Cart

Show all Filter over text

[Globus Online All Selected](#) [WGET All Selected](#) [Remove All](#)

[Hide Files](#) [WGET](#) [remove](#)

[cordex.EUR-44.SMHI.CCCma-CanESM2.historical.r0i0p0.RCA4-v1.fx.orog.v20130927|esg-dn1.nsc.liu.se](#)
(Total Number of Files for All Variables: 1)

[cordex.EUR-44.SMHI.CCCma-CanESM2.historical.r0i0p0.RCA4-v1.fx.orog.v20130927.orog_EUR-44_CCCma-CanESM2_historical_r0i0p0_SMHI-RCA4_v1_fx.nc|esg-dn1.nsc.liu.se](#)
tracking_id: N/A
checksum: N/A (N/A)

[HTTP](#)

ESGF / CERA – WDCC Integration






Home Search Tools Login Help

[Follow this link to skip to the main content](#)







Home Search Tools Login Help

Current Selections

(x) [data_node.wdcc-esgf.dkrz.de](#)

 Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hu*
 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 ... 220 221 > displaying 1 to 10 of 2209 search results

Display datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Search Categories

Project

CMIP5 (2209)

Institute

Model

Instrument

Experiment Family

Experiment

Time Frequency

Product

Realm

Variable

Variable Long Name

CMIP Table

CF Standard Name

Ensemble

Domain

Driving Model

Downscaling realisation

Results Data Cart

Replica [project=CMIP5_model=Institute for Numerical Mathematics_experiment=ESM RCP8.5_time_frequency=mon_modeling_realm=land_ensemble=r1i1p1_version=20110323](#)

Data Node: [wdcc-esgf.dkrz.de](#)

Version: 20110323

Description: inmcm4 model output prepared for CMIP5 ESM RCP8.5

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

Replica [project=CMIP5_model=Institute for Numerical Mathematics_experiment=ESM RCP8.5_time_frequency=mon_modeling_realm=atmos_ensemble=r1i1p1_version=20130207](#)

Data Node: [wdcc-esgf.dkrz.de](#)

Version: 20130207

Description: inmcm4 model output prepared for CMIP5 ESM RCP8.5

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

Replica [project=CMIP5_model=Institute for Numerical Mathematics_experiment=ESM RCP8.5_time_frequency=mon_modeling_realm=ocean_ensemble=r1i1p1_version=20110323](#)

Data Node: [wdcc-esgf.dkrz.de](#)

Version: 20110323

Description: inmcm4 model output prepared for CMIP5 ESM RCP8.5

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

Replica [project=CMIP5_model=Institute for Numerical Mathematics_experiment=ESM RCP8.5](#)

**Dataset: project=CMIP5,
 model=Institute for Numerical
 Mathematics, experiment=ESM
 RCP8.5, time_frequency=mon,
 modeling_realm=atmos,
 ensemble=r1i1p1, version=20130207**

Metadata

Show/Hide Properties cf_standard_name

	Property	Value
expand	cf_standard_name	toa_incoming_shortwave_flux ; surface_downwelling_longwave_flux_in_air_assuming_clear_sky ; surface_do...
	cmor_table	Amon
	data_node	wdcc-esgf.dkrz.de
	datetime_start	2006-01-16T12:00:00Z
	datetime_stop	2100-12-16T12:00:00Z
collapse	description	inmcm4 model output prepared for CMIP5 ESM RCP8.5
	doi	doi:10.1594/WDCC/CMIP5.INC4e8
	doi_author	Volodin, Evgeny, Diansky, Nikolay
	doi_publication_year	2013
	doi_publisher	WDCC
collapse	doi_title	inmcm4 model output prepared for CMIP5 ESM RCP8.5, served by ESGF
expand	drs_id	cmip5.output1.INM.inmcm4.esmrcp85.mon.atmos.Amon.r1i1p1...

Providing Data

ESGF index node

- search facets

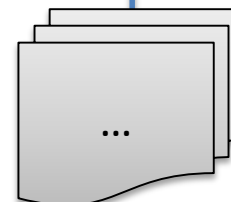
ESGF data publication

- File conventions
- Attribute conventions
- Directory conventions

QC check



ESGF Data node



Providing Data

Directory structure:

The CMI P5 directory structure has been defined as

```
<activity>/<product>/<institute>/<model>/<experiment>/<frequency>/<modelingrealm>/<MIPtable>/<ensemblemember>/<versionnumber>/<variablename>/.
```

In CORDEX this is modified as follows:

```
<Domain>/<Institute>/<GCMModelName>/<CMI P5ExperimentName>/<CMI P5EnsembleMember>/<RCMModelName>/<RCMVersionID>/<frequency>/<variablename>/
```

This “cmor-ization” process is oftenly time consuming and boring for scientists

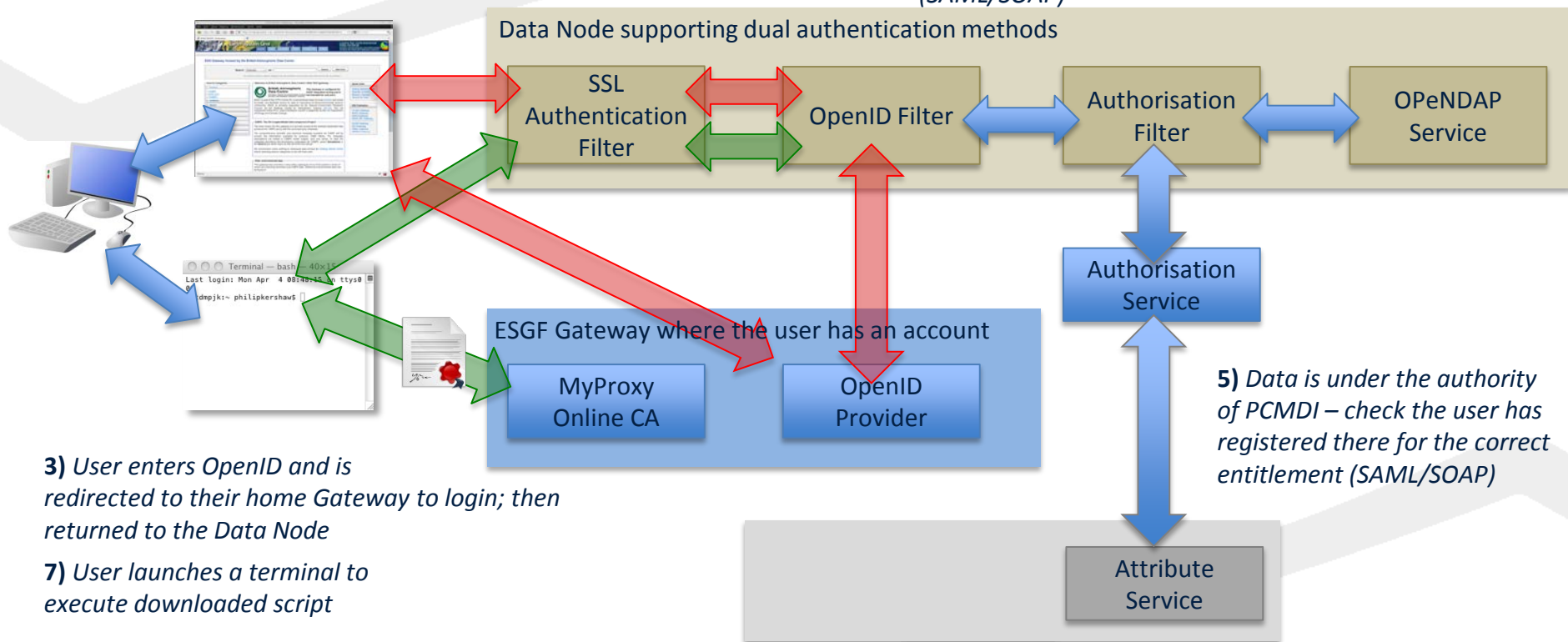
A short look behind the curtain

1) Access via Browser

2) SSL Filter intercepts request (HTTPS) - no certificate passed – pass on to OpenID Filter

4) The user is authenticated – now check their authorisation (SAML/SOAP)

6) Access granted – pass control to app to serve data



3) User enters OpenID and is redirected to their home Gateway to login; then returned to the Data Node

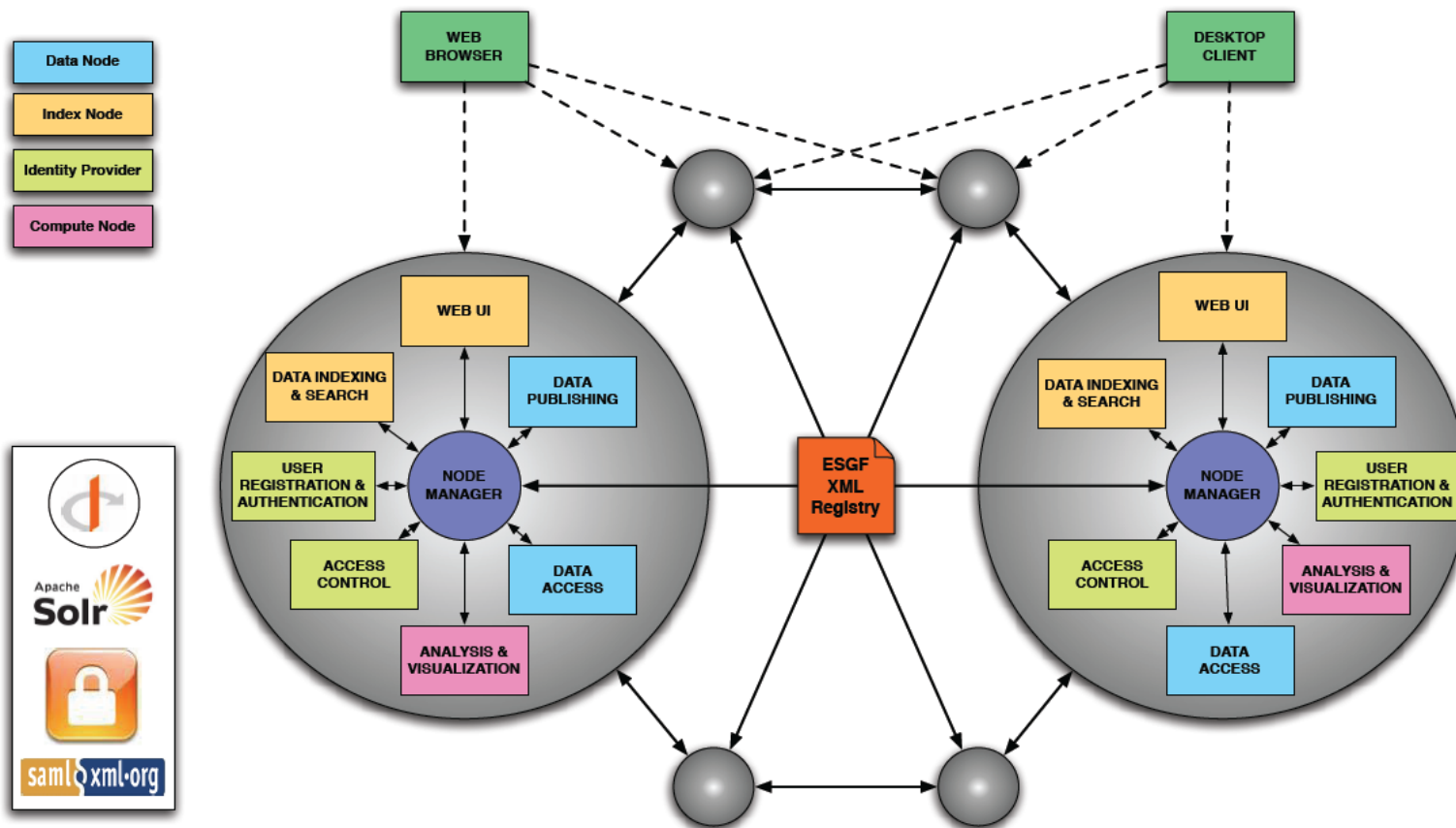
7) User launches a terminal to execute downloaded script

8) Skript Calls MyProxy logon to get a certificate

9) Skript downloads data via wget (using the certificate)

10) Authorisation and access proceeds as before

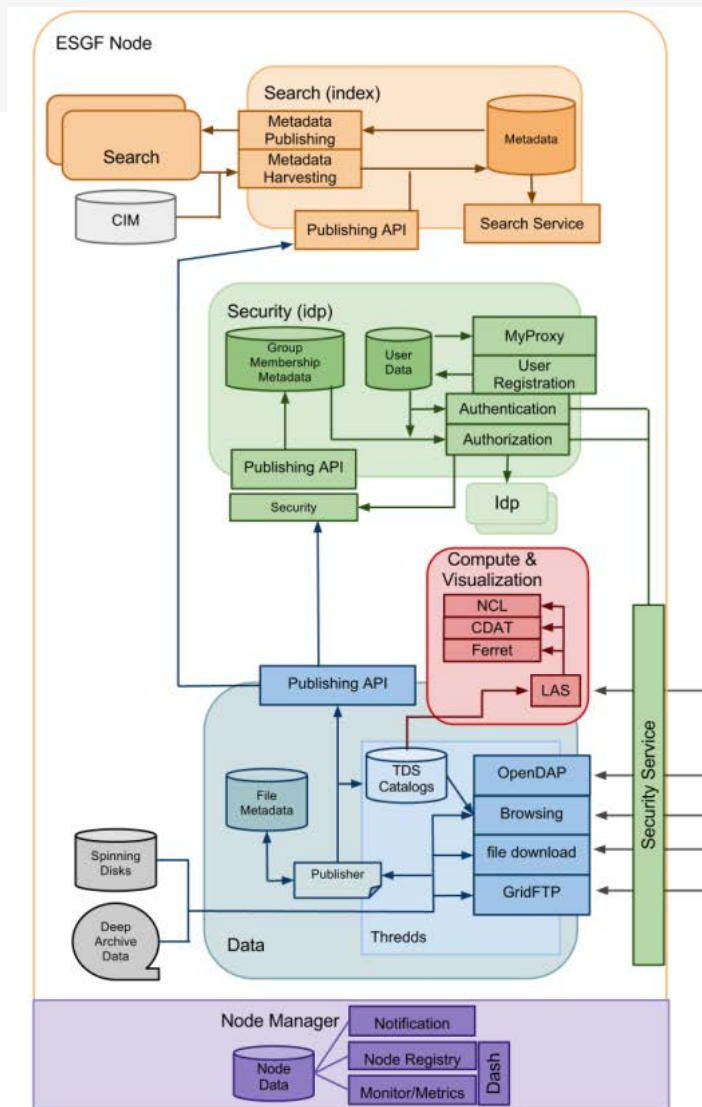
5) Data is under the authority of PCMDI – check the user has registered there for the correct entitlement (SAML/SOAP)



<http://esgf-data.dkrz.de>
<http://cordexesg.dmi.dk>
<http://esg-dn1.nsc.liu.se>

Der ESGF software stack

- Index
- Idp
- Data
- Compute



..

- DKRZ ESGF Team vorstellen .. ? Am Anfang oder am Ende ?