

Data Citation Service for CMIP6

ESGF Conference 2016

M. Stockhause¹, K. Berger¹, G. Levavasseur², M. Lautenschlager¹

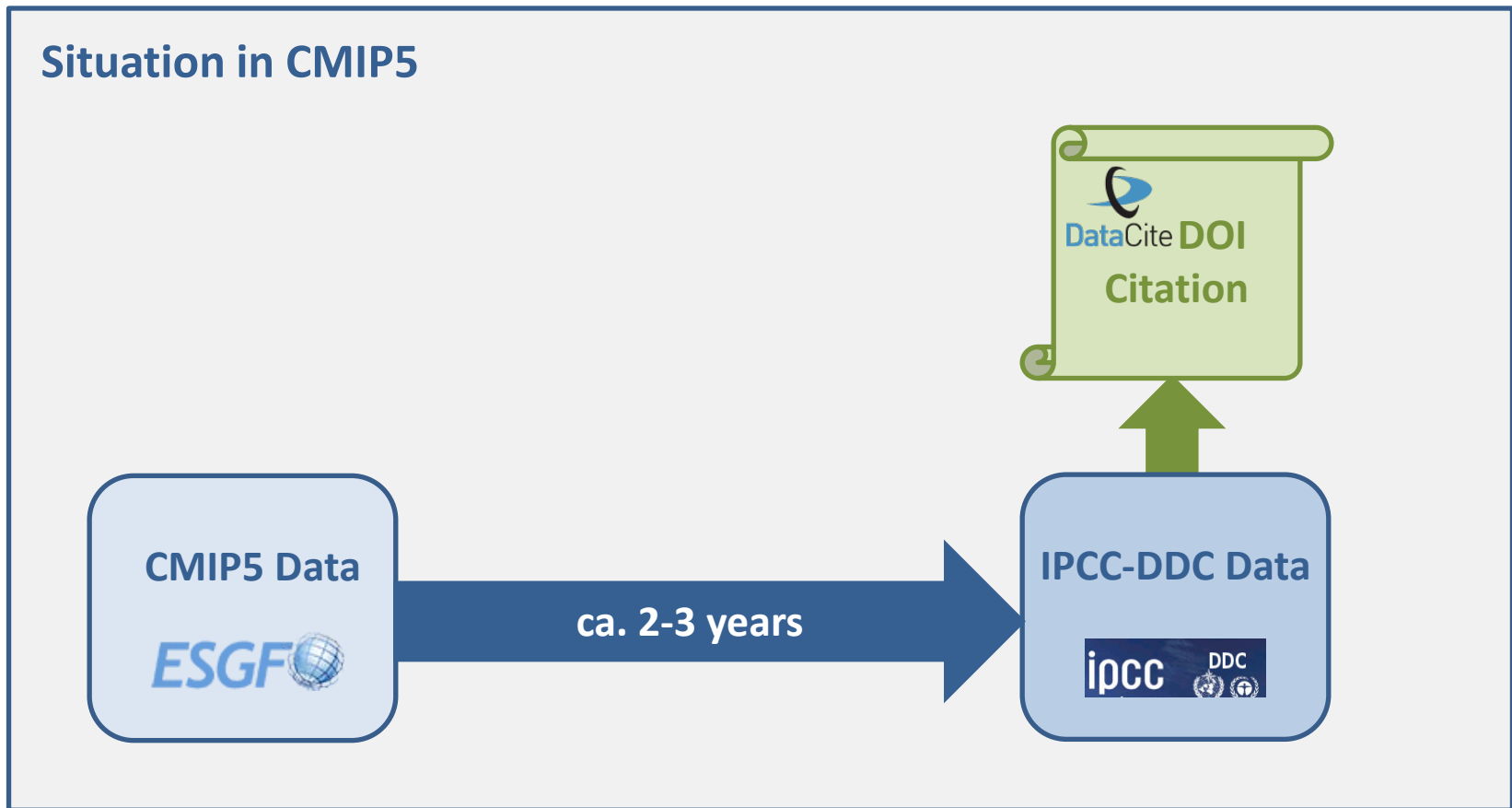
¹Deutsches Klimarechenzentrum (DKRZ), ²Institute Pierre Simon Laplace (IPSL)

Contributors: H.-H. Winter, M. Kurtz, L. Cinquini

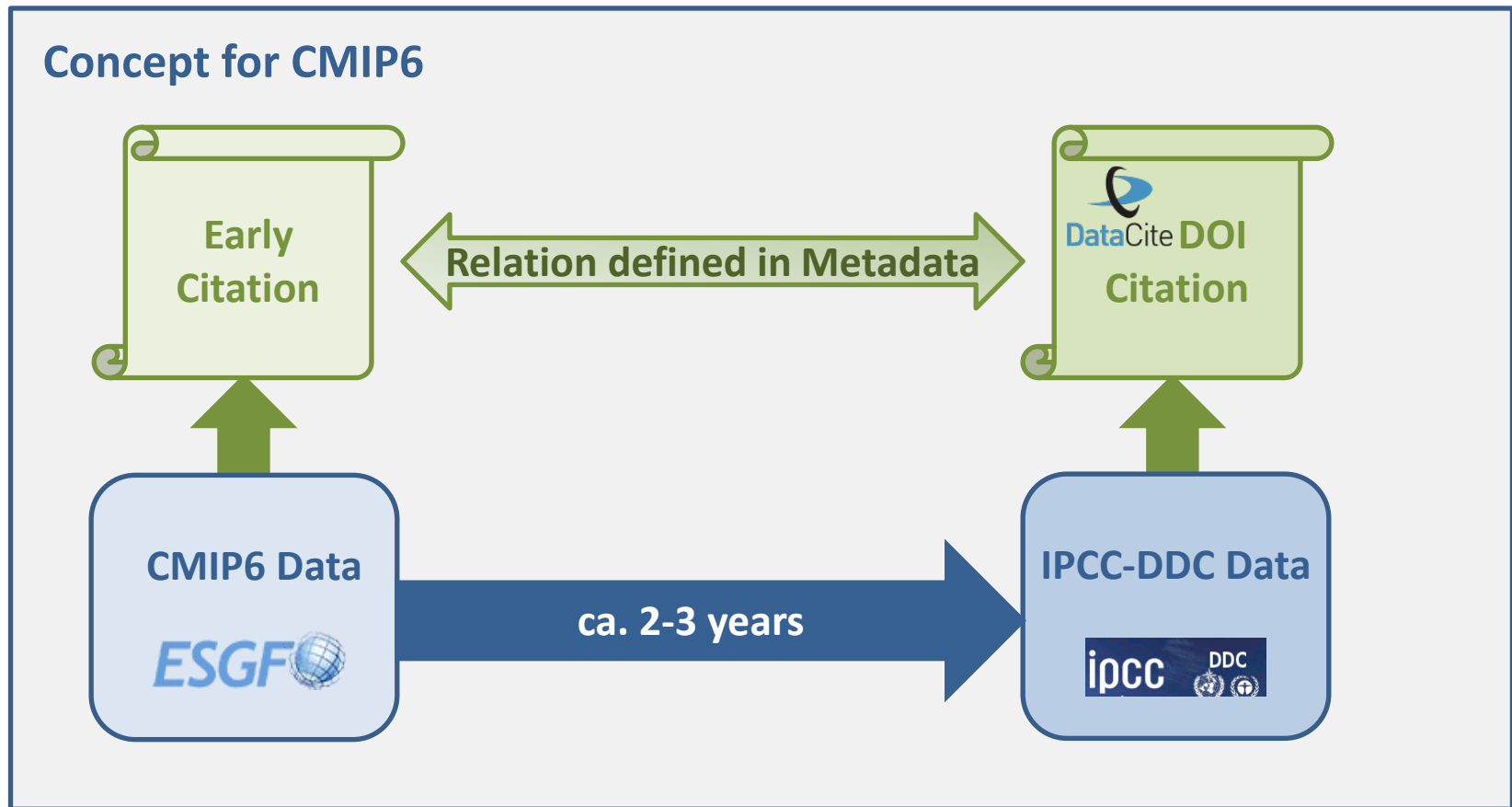
Motivation

Request from WGCM: Make CMIP6 data citable prior to Long-Term Archival in the IPCC-DDC.

Data Citation Concept for CMIP6 (1)

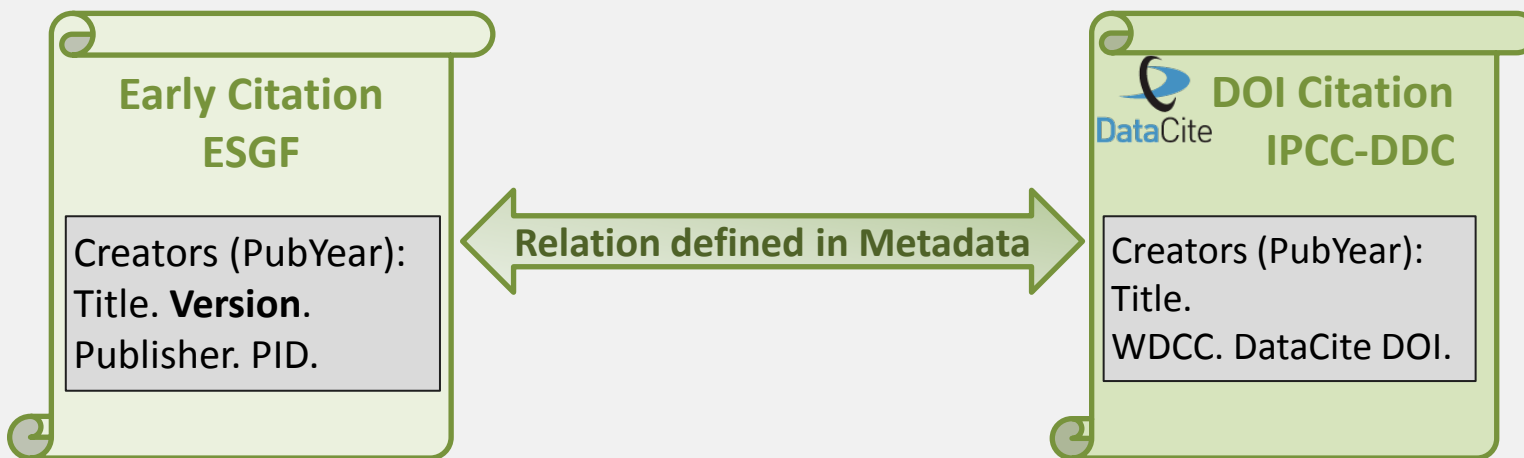


Data Citation Concept for CMIP6 (1)



Data Citation Concept for CMIP6 (2)

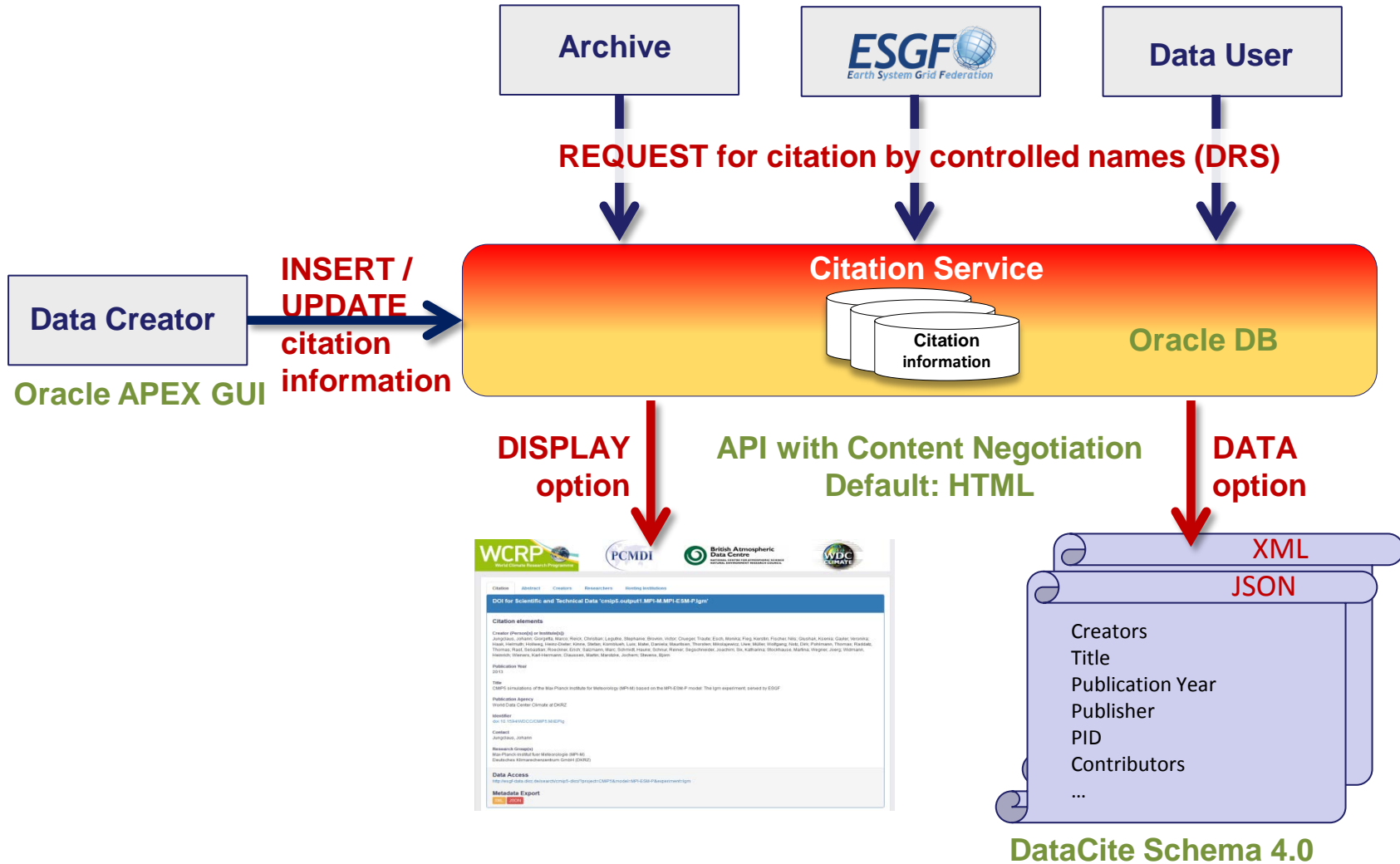
Concept for CMIP6



Citation Granularities: Model/MIP Data and Simulation Data

Early Citations include Versions for the identification of used datasets.

Data Citation Service



Use Cases (1)



Data Creator:

insert/update citation information



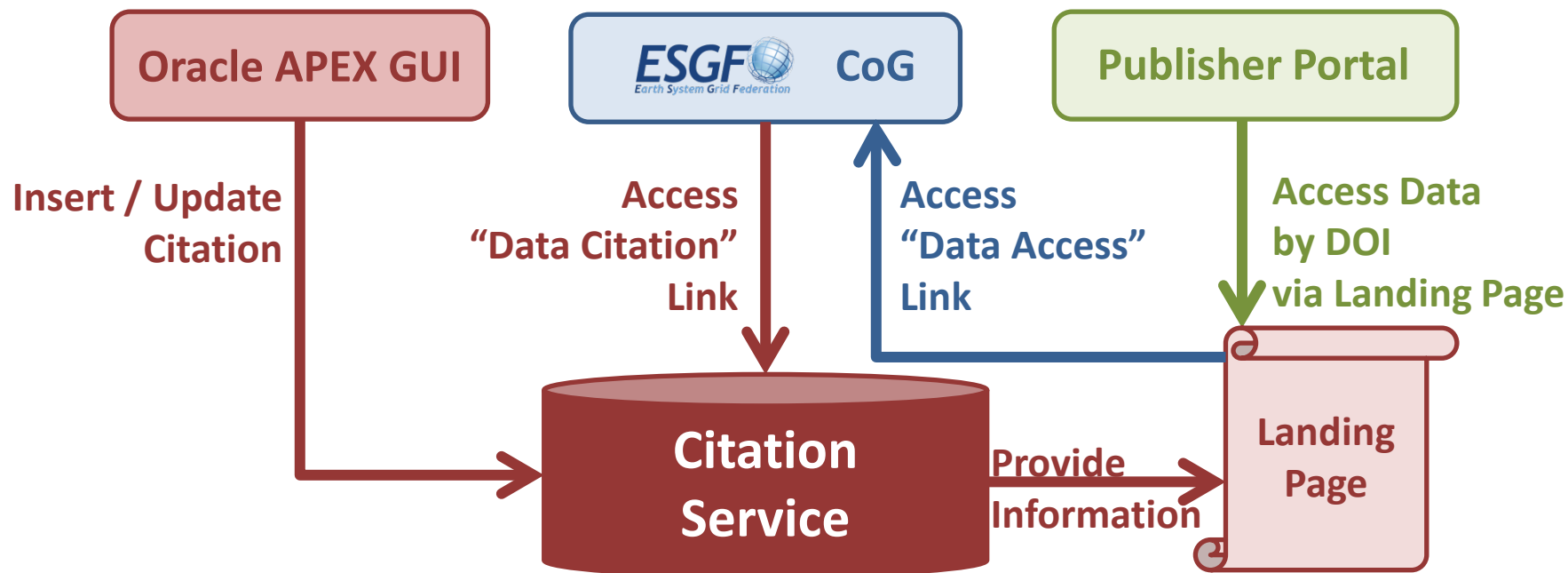
Data User / Article Writer:

access citation information



Data Reviewer / Article Reader:

access cited data



Use Cases – Data Creators (2)

Use Case 1: Providing Citation Information



Data Creator

Oracle APEX GUI

Insert / Update
Citation

Citation
Service

Three Steps to provide citations:

1. Enter Title for entry (model MIP/sim.)
2. Prepare author list [and contributor lists]
3. Connect author list to entry

World Data Center for Climate Hamburg

MARTINA_STOCKHAUSE Feedback Logout

Home CMIP6 Citation Institute Person Completed Citations References Upload Area

Create Citation for CMIP6 Data

Three steps are required to create a citation for your CMIP6 data:

1. Create/Edit Citation Title: Enter titles for the data of the created CMIP6 models and experiments.
2. Create Lists of Persons and Institutes: Enter lists of persons and institutes to connect in step 3 with a special role to a data citation.
Note: Required for step 3 are a list of Creators (author) and a list of ContactPersons. Lists are ordered.
3. Build/Edit Data Citation: Connect your Creator (author) list and your ContactPerson (corresponding author) list from step 2 to a citation title from step 1 (mandatory). Optionally, a citation can be linked to other lists of contributors from step 2, e.g. Funders or ResearchGroups, who contributed to the creation of the selected model or simulation data.
Mandatory roles are Creator (Author) and ContactPerson. More information on roles is provided in the DataCite Metadata Schema documentation at schema.datacite.org.

Citation Title	Title	Date/Version	Edit List	List Name	Modification Date	List Description
CMIP6 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model, served by ESGF		2015-04-02	✓	MPI-ESM2_sstasasensinfinal	20-JUN-2016	MPI-ESM2_sstasasensinfinal
CMIP6 ScenarioMIP data for ssp245 experiment from model MPI-ESM2		2016-06-20	✓	KNMI_CORDEX_contact	03-NOV-2016	KNMI contact person
KNMI-CORDEX EUR-11 experiment data		2015-11-03	✓	mp list list	21-JUL-2016	
NICAM model output prepared for CMIP6		2010-07-21	✓	MPI-ESM-P_authors2	08-APR-2016	
CMIP6 simulator Meteorology (MP model): The Igm			✓	MPI-ESM-P_authors	02-APR-2016	
CMIP6 simulator Meteorology (MIP model): The Ipcd			✓	MPI-ESM-P_authors	02-APR-2016	
CMIP6 simulator Meteorology (MIP model): The Ipcd-ESGF			✓	MPI-ESM-P_authors	02-APR-2016	
CMIP6 simulator Meteorology (MP model): The Ipcd			✓	MPI-ESM-P_authors	02-APR-2016	
CMIP6 simulator Meteorology (MP model): The Ipcd			✓	MPI-ESM-P_authors	02-APR-2016	
CMIP6 simulator Meteorology (MP model): The Ipcd			✓	MPI-ESM-P_authors	02-APR-2016	

World Data Center for Climate Hamburg

MARTINA_STOCKHAUSE Feedback Logout

Home CMIP6 Citation Institute Person Completed Citations References Upload Area

Q v Go Actions v Cancel

DRS Name	URL of Citation Landing Page	Citation Landing Page
cmip5.output1.MPI-M.MPI-ESM-P	http://cera-www.dkrz.de/WDCC/meta/CMIP6/cmip5.output1.MPI-M.MPI-ESM-P	View
CMIP6.CMIP.MOHC.HadGEM3.1pctCO2	http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.MOHC.HadGEM3.1pctCO2	View
CMIP6.ScenarioMIP.institute.model	http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.institute.model	View
CMIP6.ScenarioMIP.MPI-M.MPI-ESM2	http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.MPI-M.MPI-ESM2	View
CMIP6.ScenarioMIP.MPI-M.MPI-ESM2.ssp126	http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.MPI-M.MPI-ESM2.ssp126	View
CMIP6.ScenarioMIP.MPI-M.MPI-ESM2.ssp245	http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.MPI-M.MPI-ESM2.ssp245	View
cmip5.output1.MPI-M.MPI-ESM-P.1pctCO2	http://cera-www.dkrz.de/WDCC/meta/CMIP6/cmip5.output1.MPI-M.MPI-ESM-P.1pctCO2	View
cmip5.output1.MPI-M.MPI-ESM-P.abrupt4xCO2	http://cera-www.dkrz.de/WDCC/meta/CMIP6/cmip5.output1.MPI-M.MPI-ESM-P.abrupt4xCO2	View
cmip5.output1.MPI-M.MPI-ESM-P.historical	http://cera-www.dkrz.de/WDCC/meta/CMIP6/cmip5.output1.MPI-M.MPI-ESM-P.historical	View
cmip5.output1.MPI-M.MPI-ESM-P.lgm	http://cera-www.dkrz.de/WDCC/meta/CMIP6/cmip5.output1.MPI-M.MPI-ESM-P.lgm	View

<http://cera-www.dkrz.de/citeXA>

Use Cases – Data Users (3)

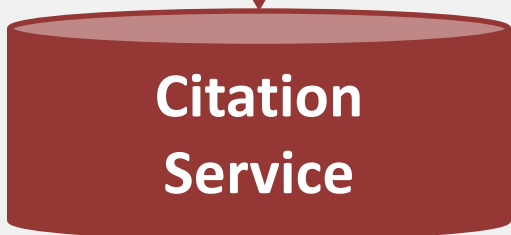
Use Case 2: Preparing an Article



Data User / Citer



Access
"Show Citation"
Link



Citation
Service

Citation Information in JSON

CMIP5

ESGF CoG Portal

You are at the ESGF-DEV.DKRZ.DE node

Home About Us Resources Contact Us Technical Support

Enter Text: Search Reset Display 10 results per page

Search Constraints: CMIP5 | MPI-ESM-P | Igm

Total Number of Results: 21
-1- 2 3 Next >>
Please login to add search results to your Data Cart
Expert Users: you may display the search URL and return results as XML or return results as JSON

- project=CMIP5, model=MPI-ESM-LR, Max Planck Institute for Meteorology (MPI-M), experiment=last glacial maximum, time_frequency=6hr, modeling_realm=atmos, ensemble=r11p1, version=20111028
Description: MPI-ESM-P model output prepared for CMIP5 last glacial maximum
Data Node: esgf-dev.dkrz.de
Version: 20111028
Total Number of Files (for all variables): 120
[Show Metadata] [Show Files] [THREDDS Catalog] [WGET Script] [LAS Visualization] [Hide Citation]

Data Citation
Identifier DOI: <http://doi.org/10.1594/WDC/CMIP5.MXEPig>
Creators: JungCLAUS, Johann, GiorGETTA, Marco, REICK, Christian et al.
Title: CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model. The Igm experiment, served by ESGF
Publisher: World Data Center Climate at DKRZ
Publication Year: 2013

Citation Landing Page

World Climate Research Programme British Atmospheric Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL WDC CLIMATE

Citation Abstract Creators Researchers Hosting Institutions

DOI for Scientific and Technical Data 'cmip5.output1.MPI-M.MPI-ESM-P.Igm'

Citation elements

Creator (Person[s] or Institute[s])
JungCLAUS, Johann, GiorGETTA, Marco, REICK, Christian, LEQUETE, Stephane, BROVINK, Victor, CRUAGER, Traute, EACH, Monika, FIEG, Kerstin, FISCHER, Hils, GLUSHAK, Kraina, GAYLER, Veronika, HASAK, Helmut, HOLLING, HERRIG, Dieter, KINKE, Stefan, KOMBIAKH, Luis, MAJE, Daniela, MAUNTTIN, Thorsten, MIKOSZEWICZ, Uwe, MÜLLER, Wolfgang, HOLT, Detl, POHLMANN, Thomas, RASCHKE, Thomas, RAST, Sebastian, ROECKNER, Erika, SALZMANN, Marc, SCHMIDT, Hauke, SCHNUR, Reinier, SEGETHNEIDER, Joachim, SIX, Katharina, STOCKHAUSE, Martina, WEGNER, Joerg, WIDMANN, Heinrich, WIENERS, Karl-Hermann, CLAUSSEN, Martin, MAROTZKE, Jochem, STEVENS, Bjorn

Publication Year
2013

Title
CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model. The Igm experiment, served by ESGF

Publication Agency
World Data Center Climate at DKRZ

Identifier
doi: 10.1594/WDC/CMIP5.MXEPig

Contact
JungCLAUS, Johann

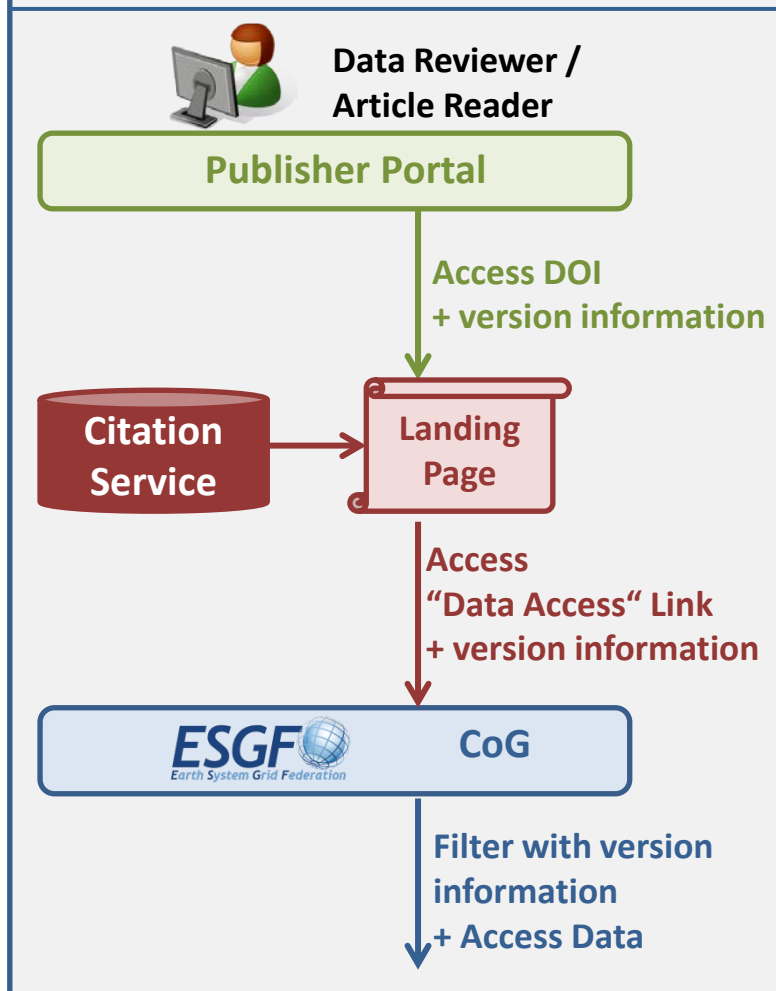
Research Group(s)
Max-Planck-Institut fuer Meteorologie (MPI-M)
Deutsches Klimarechenzentrum GmbH (DKRZ)

Data Access
<http://esgf-data.dkrz.de/search/cmip5-dv27/project=CMIP5,model=MPI-ESM-P&experiment=Igm>

Metadata Export
[XML](#) [JSON](#)

Use Cases - Article Readers (3)

Use Case 3: Access Data underlying an Article



Find Data Citation in reference list of an article, e.g.:

Jungclaus et al. (2012): The MPI..... v20130807. WDCC.
[doi:10.1594/WDC/CMIP5.MXEPc1](https://doi.org/10.1594/WDC/CMIP5.MXEPc1).

Citation Landing Page

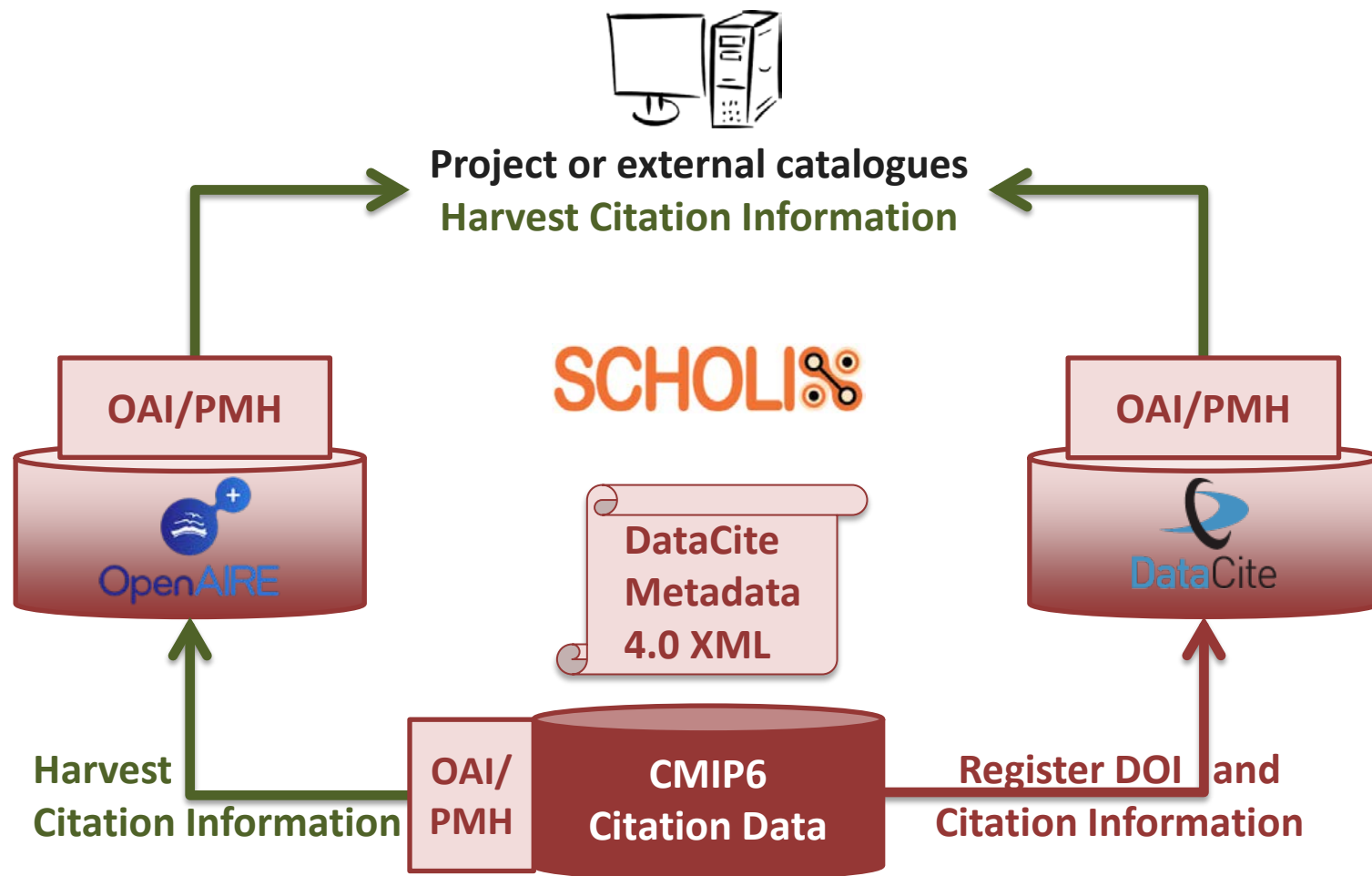
Resolve DOI

The screenshot shows the **WCRP** (World Climate Research Programme) website. It features logos for **PCMDI**, **British Atmospheric Data Centre**, and **WDC CLIMATE**. The page displays the **DOI for Scientific and Technical Data 'cmip5.output1.MPI-M.MPI-ESM-P.Igm'**. Under **Citation elements**, it lists the creator (Jungclaus, Johann; Giorgetta, Marco; Reich, Christian; Legutke, Stephanie; Brovén, Victor; Crueger, Traute; Esch, Monika; Fieg, Kerstin; Fischer, Nils; Glushak, Ksenia; Gayler, Veronika; Haak, Helmut; Hollweg, Heinz-Dieter; Kinne, Stefan; Kombluh, Luis; Mabe, Daniela; Mauritsen, Thorsten; Mikolajewicz, Uwe; Müller, Wolfgang; Nötz, Dirk; Pohlmann, Thomas; Raddatz, Thomas; Rast, Sebastian; Roeckner, Erich; Saltzman, Marc; Schmidt, Hauke; Schnur, Reiner; Segsneider, Joachim; Six, Katharina; Stockhause, Martina; Wegner, Joerg; Wilmann, Heinrich; Wieners, Karl-Hermann; Claussen, Martin; Morcrette, Jochem; Stevens, Eljon). It also shows the **Publication Year** (2013), **Title** (CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model: The Igm experiment, served by ESGF), **Publication Agency** (World Data Center Climate at DKRZ), **Identifier** (doi:10.1594/WDC/CMIP5.MXEPc1), **Contact** (Jungclaus, Johann), and **Research Group(s)** (Max-Planck-Institut fuer Meteorologie (MPI-M), Deutsches Klimarechenzentrum GmbH (DKRZ)). The **Data Access** section provides a URL: <http://esgf-data.dkrz.de/search/cmip5-dkrz?project=CMIP5&model=MPI-ESM-P&experiment=igm>. The **Metadata Export** section offers options for **XML** and **JSON**.

ESGF CoG Portal

Version Dependency !

Data Citation Integration



Requirements vs. Status (1)

Identification of early citations for dynamic data in ESGF CoG portal needs to support



- Filtering for **Version ≤ 20150101**, i.e. state of ESGF search request at given date,



- Filtering of information on unpublished datasets, and



- Use of portal functionalities to download selected datasets.

Other Requirement:



- Controlled Vocabulary (CV) on DRS components



- “show citation” link in CoG portal / ancillary MD registration



- Data citation information on furtherInfoURL page



- PID registration and citation metadata publication for early citations

Requirements vs. Status (2)

Data Citation Information Dissemination:

- ✓
- ✓
- Metadata set for early citations on OAI server
- Data Citation API provides DataCite Metadata 4.0 as XML, JSON, and HTML using content negotiation (default: HTML)

Data Citation Insert - GUI

- ✓
- Data Citation GUI – Basic Functionality (End of 12/2016)
Test user account with examples: “cite_demo”
- ✗
- Get list of citation managers to provide citation account
- ⚠
- Automated creation of entries out of CV (including defaults for title and author=modeling center)
- ⚠
- Data Citation GUI – Advanced Functionality

→ see also living document on specifications at: <http://bit.ly/1XsVOoz>

CMIP6 Citation vs. RDA Recommendations (1)

Preparing Data & Query Store

- R1 – Data Versioning
- R2 – Timestamping
- R3 – Query Store

When Resolving a PID

- R11 – Landing Page
- R12 – Machine Actionability

When Data should be persisted

- R4 – Query Uniqueness
- R5 – Stable Sorting
- R6 – Result Set Verification
- R7 – Query Timestamping
- R8 – Query PID (*PID on cited subset*)
- R9 – Store Query (*store MD for cited subset and relation to full data*)
- R10 – Citation Text (*store MD for cited subset and relation to full data*)

Upon Modifications to the Data Infrastructure

- R13 – Technology Migration
- R14 – Migration Verification

research data sharing without barriers
rd-alliance.org

(R15 – PID for whole dataset)

(): suggested by DataCite at RDA 8th Plenary, 09/2016.



Rauber et al. (2016):
http://www.ieee-tcdl.org/Bulletin/v12n1/papers/IEEE-TCDL-DC-2016_paper_1.pdf

CMIP6 Citation vs. RDA Recommendations (2)

CMIP6 Citation Service:

Citation of evolving full Data Superset

- Citations with PIDs and metadata generated at data creation
- **No. of cited datasets > No. of used datasets**

+ Data Cart Idea:

Citation of Used Data Subset

- Citations with PIDs and metadata generated at moment of citation
- **No. of cited datasets = No. of used datasets**
- Include a relations for the data superset in the metadata

= Full implementation of RDA recommendations

CMIP6 Citation vs. RDA Recommendations (2)

CMIP6 Citation Service:

Citation of Full Dataset or Data Superset

- Citations with PIDs and metadata generated at data creation
- **No. of cited datasets > No. of used datasets**

+ Data Cart Idea:

Citation of Used Data Subset

- Citations with PIDs and metadata generated at moment of citation
- **No. of cited datasets = No. of used datasets**
- Include a relations to the Credit PIDs in the metadata

CMIP7

= Half-way implementation of RDA
recommendations for CMIP6

CMIP6 Data References for CMIP6 (1)

CMIP6 Citation for evolving Full Dataset are:

- integrated in reference lists of scholarly literature and thus
 - included in data-literature interlinking or cross-referencing
- data usage is traceable and credit to data creators is given

+ Data Cart Handle IDs for specific Data Subset Identification are:

- no citations because of missing citation information and non-persistent target data subsets,
- not integrated in reference list (acknowledgement?), and
- restricted to CMIP6 dataset collections bearing Handle IDs

= Half-way implementation of RDA-endorsed recommendations for CMIP6

CMIP6 Data References for CMIP6 (2)

You are at the [ESGF-FEDTEST.DKRZ.DE](#) node

Technical Support

Last Search | My Data Cart (3)

My DataCart

Number of Items (3) | [Return to Last Search](#)

Collective Services for All Selected Datasets: [[WGET Script](#)] [[LAS Visualization](#)] [[Globus Download](#)] [[Hide Collection PID](#)]

When 'Show Files' is clicked, or when using WGET or Globus, you may use an optional string to sub-select the filenames:

Collection PID for the selected datasets:
[hdl:21.14100/1f9d6b80-6f2e-395b-aeb4-b00157263733](#)

You can use this [persistent identifier](#) to refer back to your individual data cart. Note that this is no guarantee that the data in the cart will remain available or stable, but only that there will always be a reference back to them. For more information, see [here](#).

Select All Datasets

Remove All

cmip6.CMIP.CSIRO-BOM.ACCESS-1-0.piControl.r1i1p1f1.6hrLev.ua.gn

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

Version: 20160714

Total Number of Files (for all variables): 1

[[Show Metadata](#)] [[Show Files](#)] [[THREDDS Catalog](#)] [[WGET Script](#)] [[LAS Visualization](#)] [[PID](#)] [[Hide Citation](#)]

Data Citation (Landing Page)
Identifier URL: <http://cera-www.dkrz.de/WDCC/meta/CMIP6/cmip6.CMIP.CSIRO-BOM.ACCESS-1-0>
Creators: Collier, Mark; Oberhuber, Josef Maximilian; Ose, Tomoaki
Title: CSIRO-BOM ACCESS1.0 model output prepared for CMIP6 CMIP
Publisher: Earth System Grid Federation
Publication Year: 2016

Remove

Currently investigated

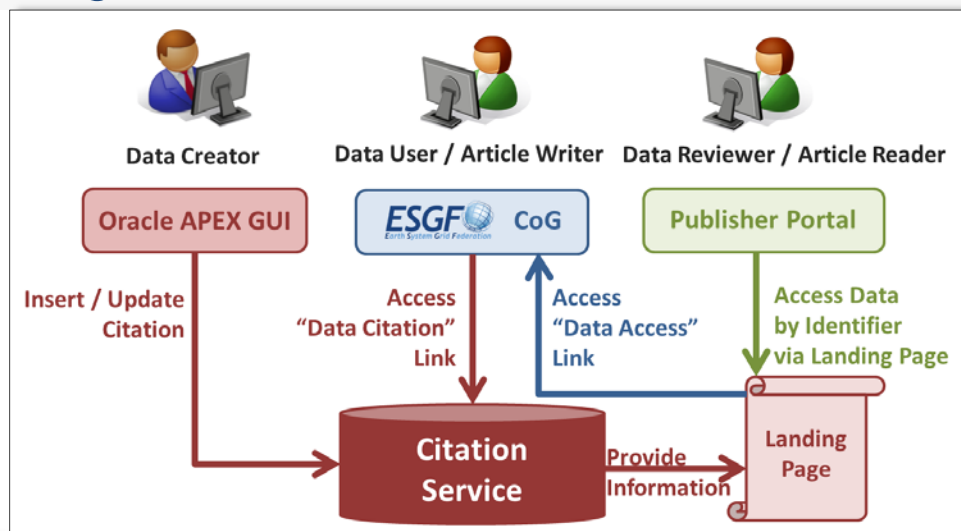
- Evolving data version support by use of **Fragment Identifiers**
→ Point users resolving the PID to the version subset data
- Integration of early citations in **Scholix**, e.g. via **OpenAire's DLI Service** (Data Literature Interlinking) or possibility to **use of DataCite DOIs for early citations** of full evolving datasets

cmip6cite.wdc-climate.de

Summary: Workflow and Status

Workflow for Modeling Centers:

0. Contact Person for citation required
1. Provide at least basic data citations for model data: author lists + titles (defaults: author=institute; title=CMOR title)
2. First data delivered to ESGF
3. Update and extend model data citations and add citations for simulation data
4. At time of long-term archival, citation information will be fixed.



CMIP5 You are at the ESGF-DEV:DKRZ.DE node

Home About Us Resources Contact Us Technical Support

Project: CMIP5 (21)

Variable: +

Institute: +

Model: -

Experiment Family: +

Experiment: -

Time Frequency: +

Realm: +

CMIP Table: +

Ensemble: +

Enter Text: [] Search Reset Display 10 results per page

Search Constraints: CMIP5 | MPI-ESM-P | lgm

Total Number of Results: 21

Please login to add search results to your Data Cart

Expert Users: you may display the search URL, and return results as XML, or return results as JSON

1. project=CMIP5, model=MPI-ESM-LR, Max Planck Institute for Meteorology (MPI-M), experiment=last glacial maximum, time_frequency=6hr, modeling_realm=atmos, ensemble=r11p2, version=2011028
Description: MPI-ESM-P model output prepared for CMIP5 last glacial maximum
Data Node: esgf-dev.dkrz.de
Version: 20111028
Total Number of Files (for all variables): 120
[Show Metadata] [Show Files] [THREDDS Catalog] [WGET Script] [LAS Visualization] [Hide Citation]
Data Citation
Identifier DOI: <http://dx.doi.org/10.1594/WDC/CMIP5.MXEP1g>
Creators: Jungclauss, Johann, Giorgetta, Marco, Reick, Christian et al
Title: CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model: The lgm experiment, served by ESGF
Publisher: World Data Center Climate at DKRZ
Publication Year: 2013
2. project=CMIP5, model=MPI-ESM-LR, Max Planck Institute for Meteorology (MPI-M), experiment=last glacial maximum, time_frequency=6hr, modeling_realm=atmos, ensemble=r11p2, version=20120713
Description: MPI-ESM-P model output prepared for CMIP5 last glacial maximum
Data Node: esgf-dev.dkrz.de
Version: 20120713
Total Number of Files (for all variables): 120
[Show Metadata] [Show Files] [THREDDS Catalog] [WGET Script] [LAS Visualization] [Show Citation]
3. project=CMIP5, model=MPI-ESM-LR, Max Planck Institute for Meteorology (MPI-M), experiment=last glacial maximum, time_frequency=day, modeling_realm=atmos, ensemble=r11p2, version=20111028
Description: MPI-ESM-P model output prepared for CMIP5 last glacial maximum
Data Node: esgf-dev.dkrz.de
Version: 20111028
Total Number of Files (for all variables): 60
[Show Metadata] [Show Files] [THREDDS Catalog] [WGET Script] [LAS Visualization] [Show Citation]

Information and Status:

(responsible party for citation service: DKRZ)

- General information: <http://cmip6cite.wdc-climate.de>
- Graphical User Interface (release: 12/2016): <http://cera-www.dkrz.de/citeXA>
- Support for pre-population of citation GUI
- Data Citation accessible in ESGF CoG portal
- Citations will be integrated in furtherInfo URL landing page → ES-DOC
- Terms of use under discussion in WIP