

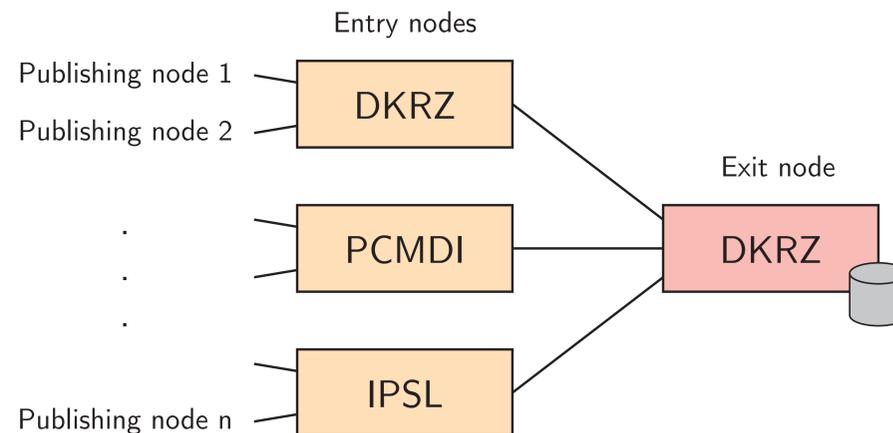
T. Weigel<sup>1</sup>, M. Stockhause<sup>1,2</sup>

<sup>1</sup>Deutsches Klimarechenzentrum, <sup>2</sup>World Data Center for Climate

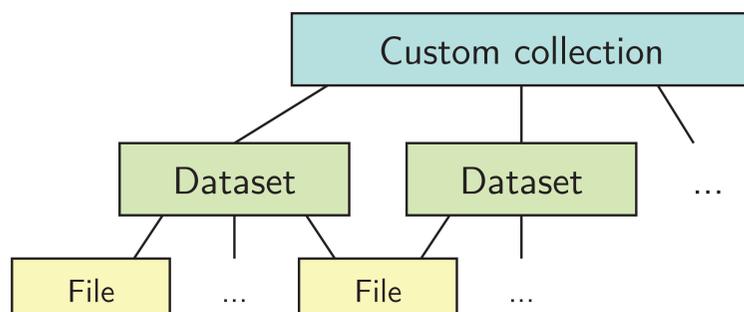
Every data provider for CMIP6 will issue Handle generation requests for CMIP6 files and datasets. These are submitted to a distributed queueing system with currently 3 entry nodes and 1 exit node that also hosts the Handle servers.

The success of this setup depends on adequate operational organization:

1. Proper use of the ESGF publisher to ensure correct versioning. No unmonitored exchange of files.
2. Load/availability monitoring of entry and exit nodes and adjustment as required over the course of CMIP6. SLAs are not required at this point due to redundancies already built in.
3. Operational oversight by CDNOT, distribution of access codes and enforcement of these policies.



Obs4MIPs Information View	
<b>Dataset</b> obs4MIPs.FUB-DWD.SSMI-MERIS.prw.mon	
<b>General Information</b>	
Dataset Id	obs4MIPs.FUB-DWD.SSMI-MERIS.prw.mon
Persistent Identifier	hdl:10876/ESGF/4ee9d37b-6454-44bf-b3ef-e738b2ecedb4
Version	20140616
<b>Data host(s)</b>	
esgf1.dkrz.de	<span>Original</span>
<b>Files belonging to this dataset</b>	
prwErr_SSMI-MERIS_L3_v1-00_200301-200812.nc	10876/ESGF/a9b1dfbc-4b73-4295-8ed6-6b586bf1be02
prwFlag_SSMI-MERIS_L3_v1-00_200301-200812.nc	10876/ESGF/4c4bd2ab-08a8-4d21-be1e-bc93cac9756a
prwStddev_SSMI-MERIS_L3_v1-00_200301-200812.nc	10876/ESGF/21ca953d-7bf3-43fe-84c0-3f36c899b22a
prw_SSMI-MERIS_L3_v1-00_200301-200812.nc	10876/ESGF/a4d6d485-1eae-48d7-9925-1a4ac12fe816

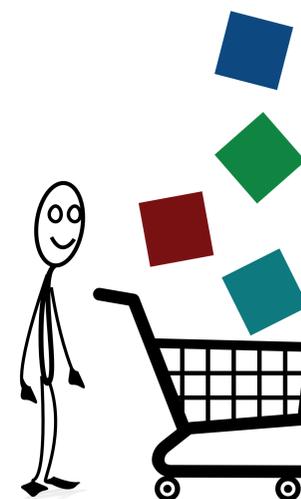


## Preservation policies:

PID preservation and data preservation are separate concerns. Only data in the Reference Data Archive of the IPCC DDC will be preserved and accessible with coarse granular DataCite DOIs.

File Handles: Kept even if objects are gone (retracted, obsoleted, broken). No guarantee given for data objects themselves.

Dataset and custom collection Handles: PIDs and structure remain, no guarantee for data. If there is an archived, latest version, it is discoverable from the PIDs.



## Options for future work:

1. Come to unified PID policies across use cases – the current set is quite make-shift and specific to ESGF CMIP6.
2. Prototype fine-granular provenance tracing, particularly for processing services. Balance complexity & expressiveness.
3. Align custom collections with RDA recommendation on Research Data Collections.

## Greater challenge: How and by whom are data used?

1. Fine-granular data usage tracking at a level lower than citation granularity, complementing Scholix. How should the Handle-based approach be extended (metadata, policies, services) so that not only the toolchain, but persons are covered?
2. Community-wide policy on how to refer to custom data collections in articles. This should be extended to include fragments and other data subsetting techniques. But how to refer to data without preservation policies in scholarly publications?

The screenshot shows the 'My DataCart' interface. At the top, it says 'Number of Items (4)'. Below that, there are several dataset entries with details like 'Description', 'Data Node', 'Version', and 'Total Number of Files'. Each entry has a 'Remove' button. A 'Collection PID for the selected datasets' is displayed, along with a note: '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.'