Convergence in France facing Big Data era and Exascale challenges for Climate Sciences

Sébastien Denvil (IPSL)

Objectives

CONVERGENCE :
Project funded by french research agency (2014-2018)

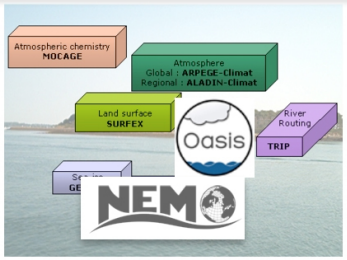
- To develop a platform capable of running large ensembles of simulations with a suite of models
- To handle the complex and voluminous datasets generated
- To facilitate the evaluation and validation of the models and the use of higher resolution models.

T0 : management

T1 : platform

ensemble of tools
different configurations
different resolution
set of simulations
set of diagnostics
assessment

IPSL implementation



GAME-CERFACS implementation

T2 : towards a high-resolution coupled model

- Improving coupled model parallelism in terms of computing and memory
- Managing efficiently input and restart files
- Integrating parallel interpolation mechanisms in XIOS
- Parallel component coupling

T3 : runtime environments

- Process assignment
- Optimization, Load balancing
- Climate Simulations Supervision

T4 : Big Data management and analytics of Climate Simulations

- XIOS implemented within project models
- XIOS a bridge towards standardisation
- Data and metadata services
- Big Data Analytics

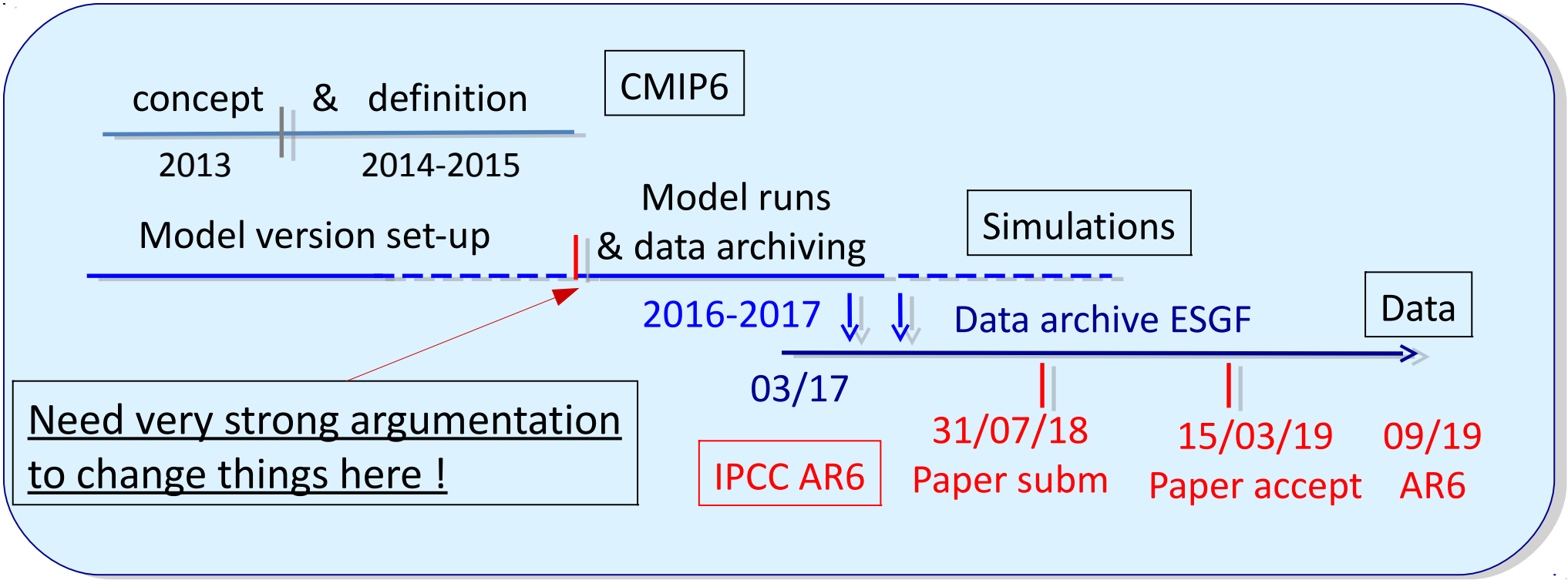
T5 : CliMAF: a framework for climate models evaluation and analysis

- General driver and upstream user interface
- Services layer
- Visualization tools
- Evaluation and monitoring diagnostics

Coupled Model Intercomparison Project – CMIP6

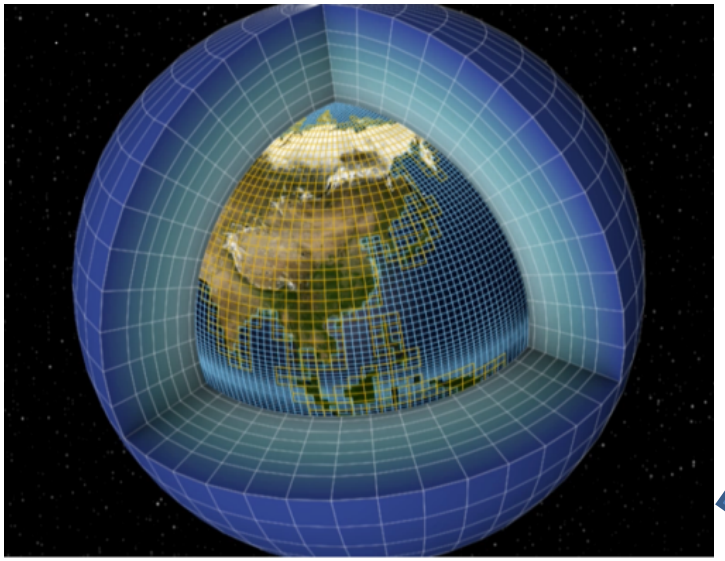
International community under **strong** pressure

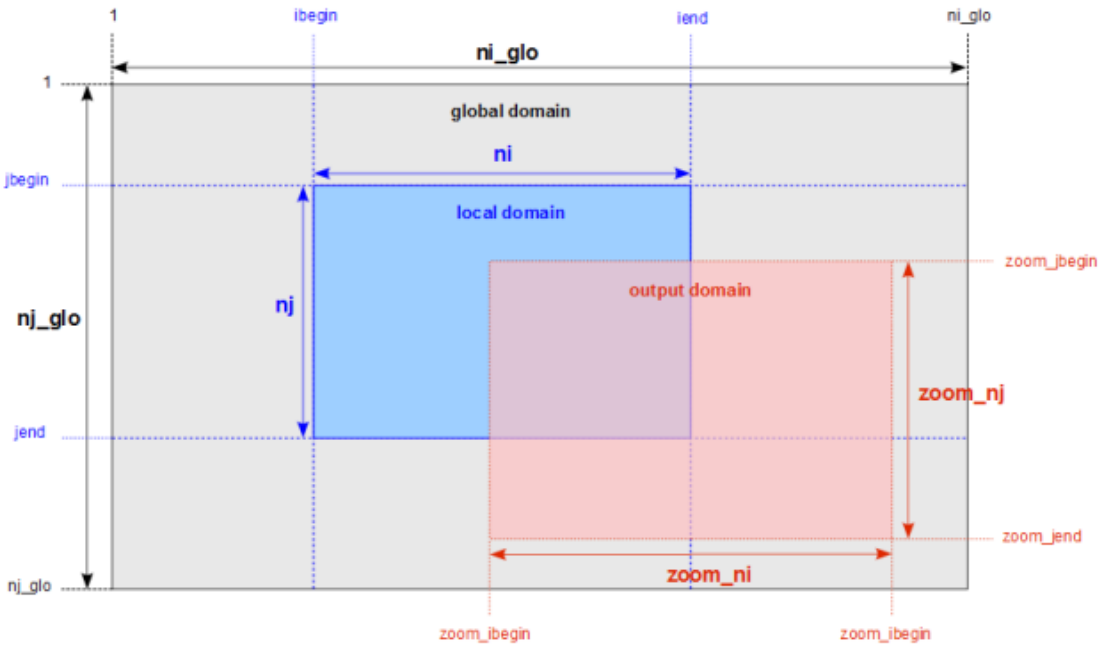
CMIP6/AR6 cycle

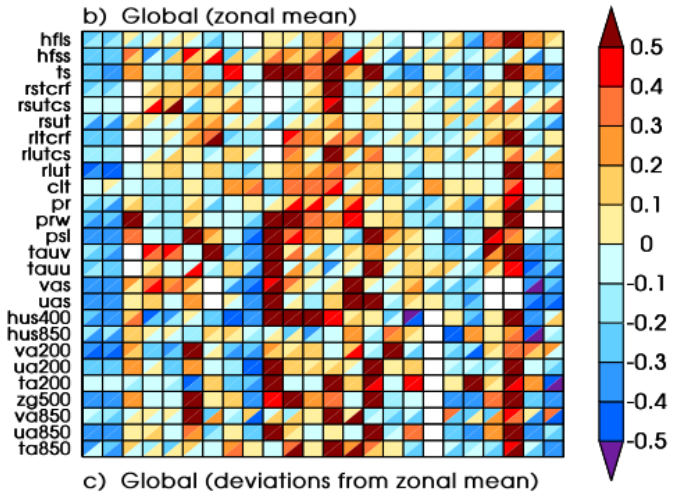


Sciences de l'environnement
Institut Pierre Simon Laplace

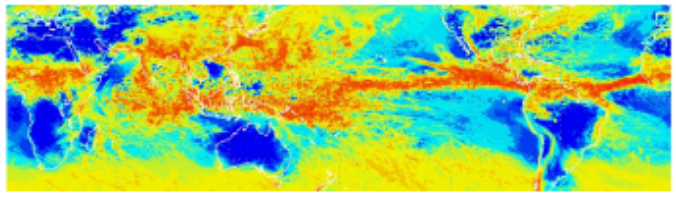
WCRP World Climate Research Programme

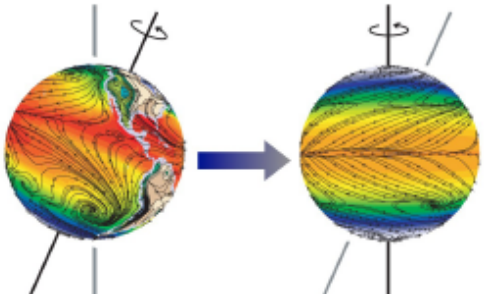


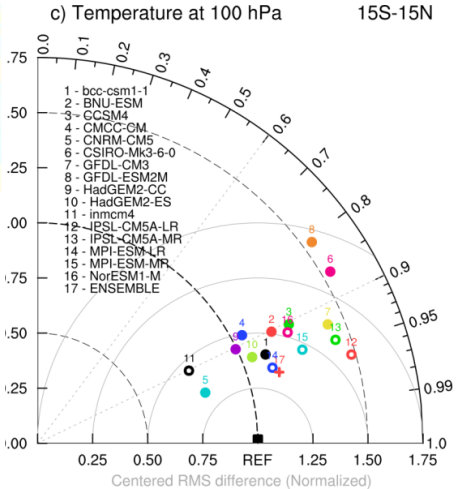


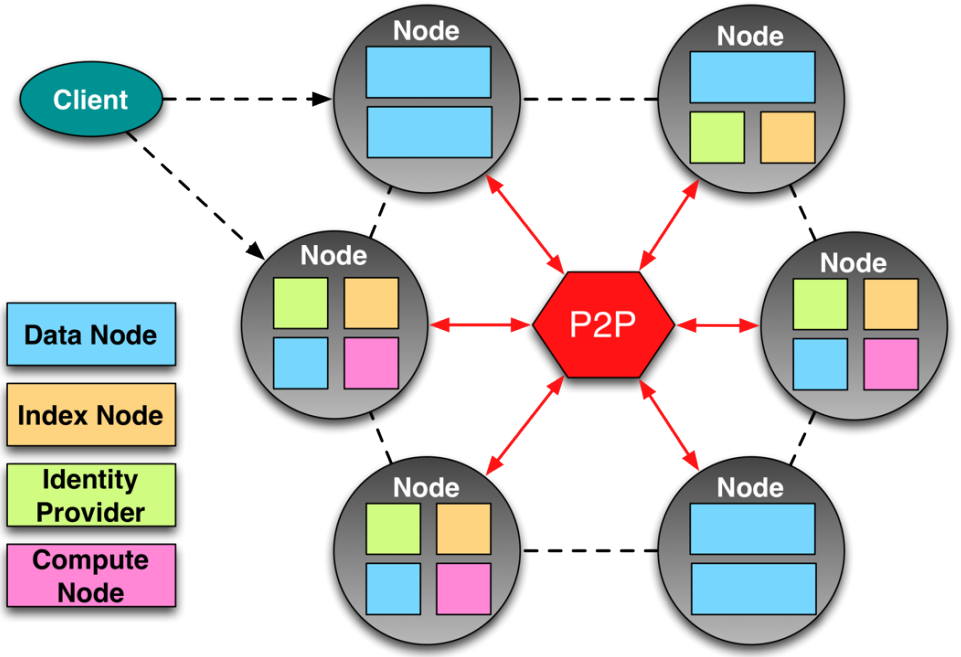


- Data producer
- ~~Distribution and archiving~~
- Analysis







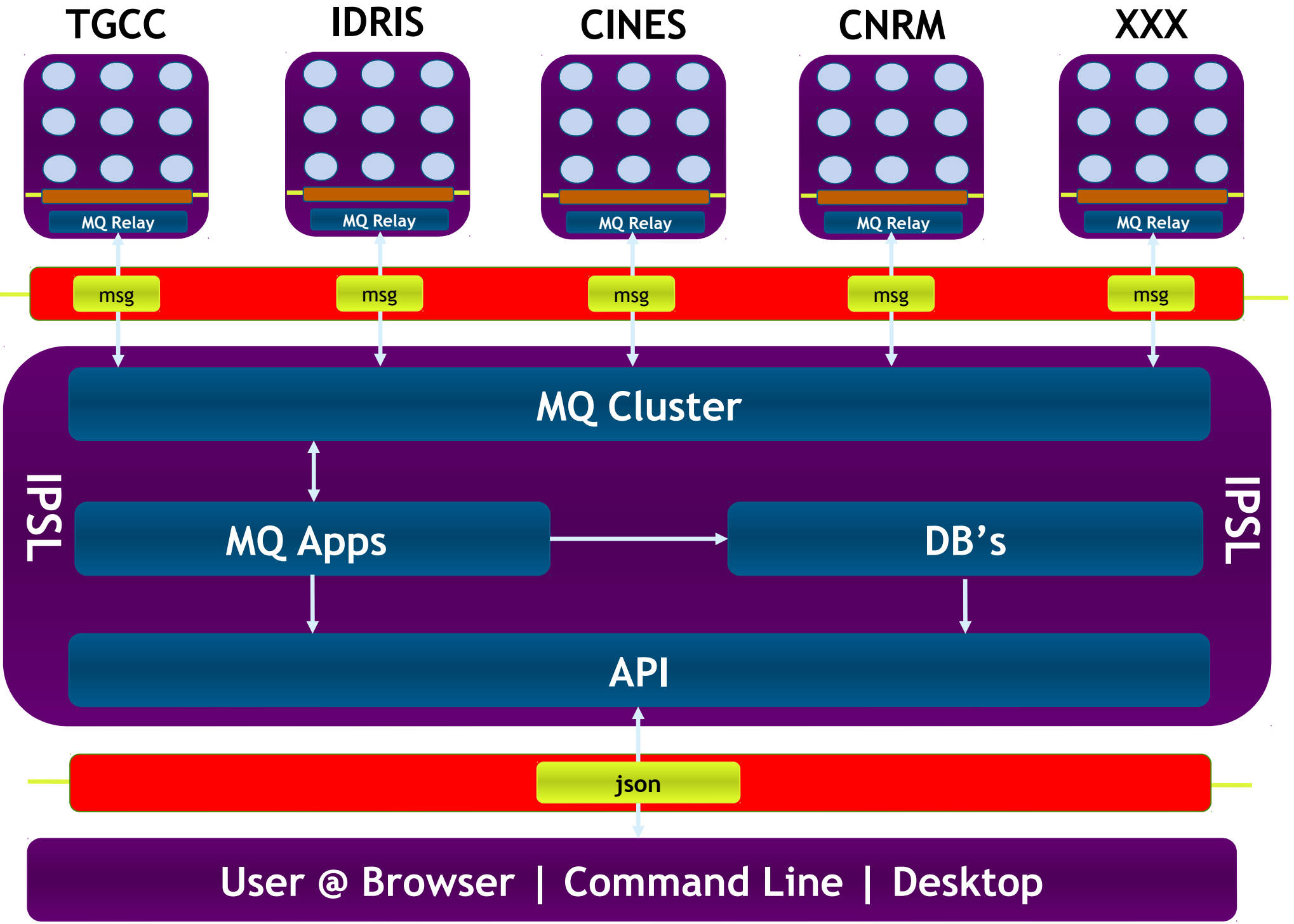


To keep in mind

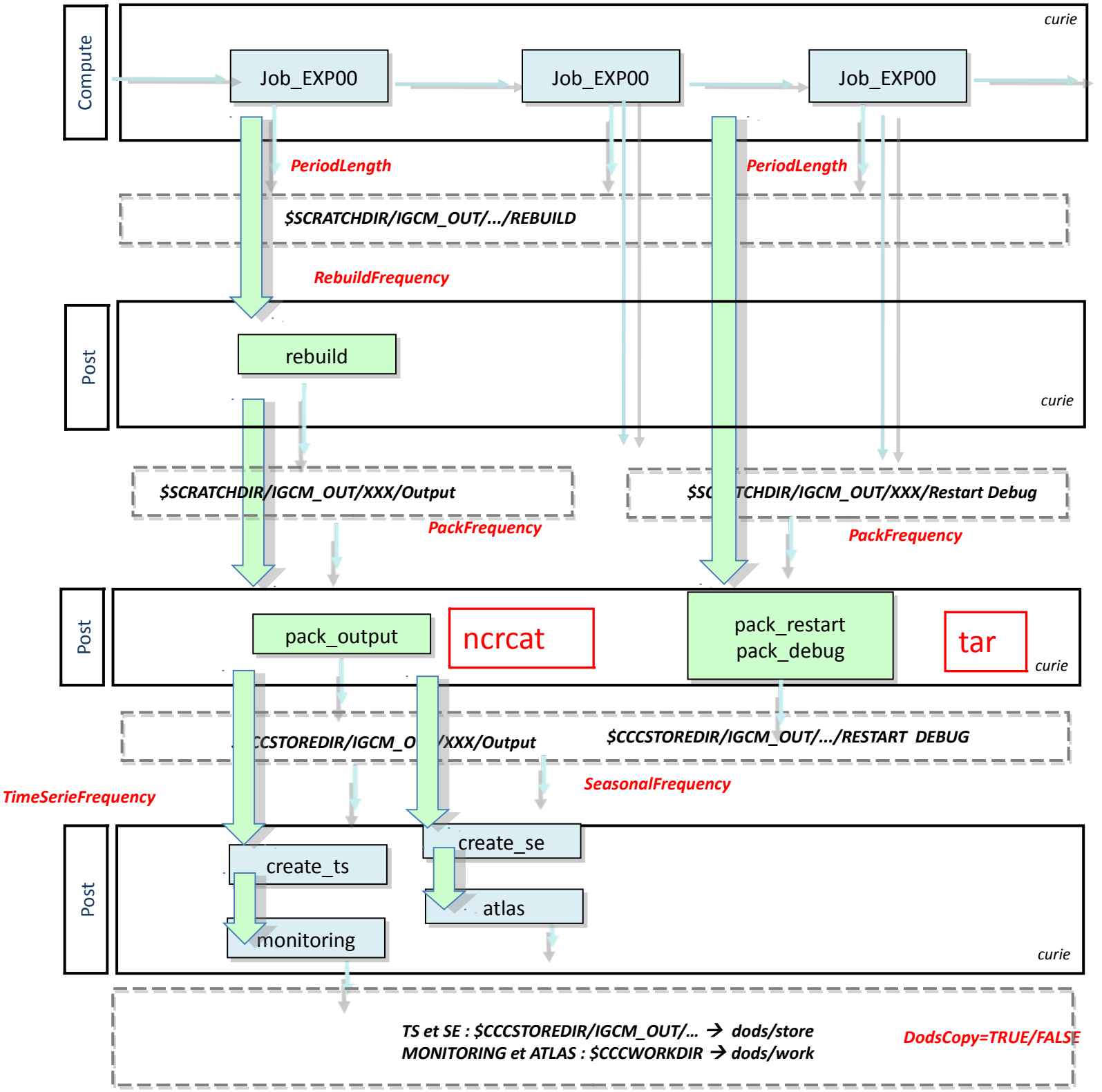
Computation useless if results cannot be stored and distributed and analyzed !

Sciences de l'environnement

Institut Pierre Simon Laplace


TGCC
IDRIS
CINES
CNRM
XXX
MQ Relay
MQ Relay
MQ Relay
MQ Relay
MQ Relay
msg
msg
msg
msg
msg
MQ Cluster
IPSL
MQ Apps
DB's
IPSL
API
json
User @ Browser | Command Line | Desktop

TGCC


Compute
curie
Job_EXP00
Job_EXP00
Job_EXP00
PeriodLength
PeriodLength
$SCRATCHDIR/IGCM_OUT/.../REBUILD
RebuildFrequency
Post
rebuild
curie
$SCRATCHDIR/IGCM_OUT/XXX/Output
$SC TCHDIR/IGCM_OUT/XXX/Restart Debug
PackFrequency
PackFrequency
Post
pack_output
ncrcat
pack_restart
pack_debug
tar
curie
CCSTOREDIR/IGCM_O /XXX/Output
$CCCSTOREDIR/IGCM_OUT/.../RESTART DEBUG
SeasonalFrequency
TimeSerieFrequency
Post
create_se
create_ts
atlas
monitoring
curie
TS et SE : $CCCSTOREDIR/IGCM_OUT/... → dods/store
MONITORING et ATLAS : $CCCWORKDIR → dods/work
DodsCopy=TRUE/FALSE

Why is it good to log « all around » ?

THEN...

Log files

```
[Sun Dec 21 09:17:09 2008] [error]
[Sun Dec 21 10:04:53 2008] [error]
[Sun Dec 21 10:45:50 2008] [error]
[Sun Dec 21 11:14:09 2008] [error]
[Sun Dec 21 12:26:04 2008] [error]
[Sun Dec 21 13:41:05 2008] [error]
[Sun Dec 21 14:37:16 2008] [error]
[Sun Dec 21 15:19:39 2008] [error]
[Sun Dec 21 15:26:05 2008] [error]
[Sun Dec 21 15:51:07 2008] [error]
```

NOW...

Operational intelligence

Sciences de l'environnement

Institut Pierre Simon Laplace

Trusting web service

Trusting

Check reliability of configurations on the SX9 from the CCRT.

Release: 0.60

Ask for support
Discover other services

Configuration IPSLCM5_v2

2010-06-07T06:10

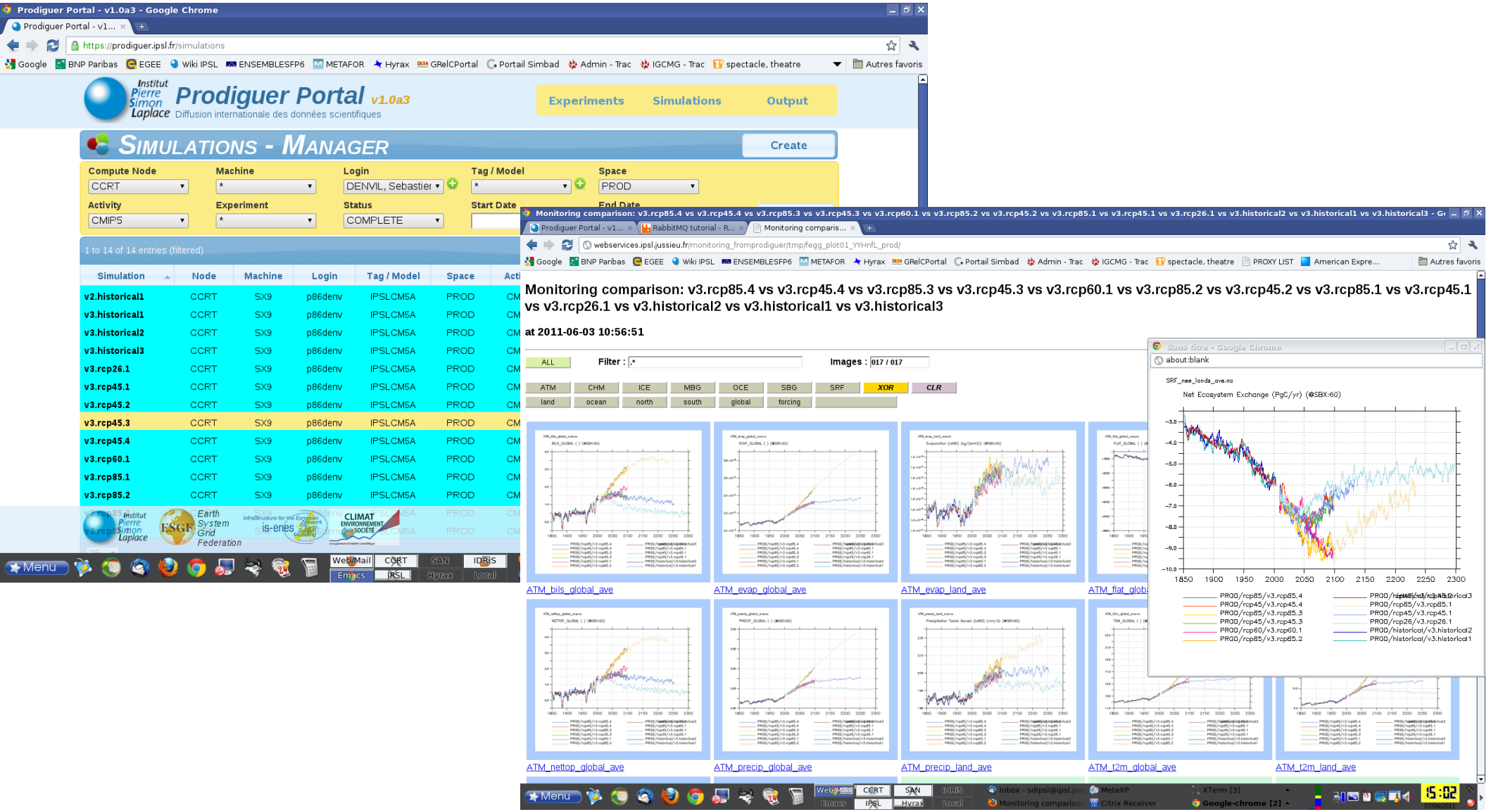
IOIPSL/src	svn checkout -r HEAD http://forge.ipsl.jussieu.fr/igcmg/svn/IOIPSL/tags/v2_1_9/src IOIPSL/src
ORCHIDEE	cvs -d :pserver:sechiba@cvs.ipsl.jussieu.fr:/home/ssipsl/CVSREP checkout -r orchidee_1_9_4_2 ORCHIDEE
OASIS3	cvs -d :pserver:anonymous@cvs.ipsl.jussieu.fr:/home/ioipsl/CVSROOT checkout -r ipslcm5a -d prism OASIS3
LMDZ4	svn checkout -r 1329 http://svn.lmd.jussieu.fr/LMDZ/LMDZ4/trunk LMDZ4
IPSLCM5	svn checkout -r HEAD http://forge.ipsl.jussieu.fr/igcmg/svn/CONFIG/IPSLCM/IPSLCM5/branches/IPSLCM5_v2 IPSLCM5
libIGCM	svn checkout -r 246 http://forge.ipsl.jussieu.fr/libigcm/svn/trunk/libIGCM libIGCM
NEMO	svn checkout -r HEAD http://forge.ipsl.jussieu.fr/nemo/svn/tags/nemo_v3_2/NEMO
UTIL	svn checkout -r HEAD http://forge.ipsl.jussieu.fr/nemo/svn/tags/nemo_v3_2/UTIL
XMLF90	svn checkout -r 54 http://forge.ipsl.jussieu.fr/ioserver/svn/XMLF90
XMLIO_SERVER	svn checkout -r 54 http://forge.ipsl.jussieu.fr/ioserver/svn/XMLIO_SERVER/trunk XMLIO_SERVER

Trusting log

Date	Status	Step	Comments	C++	F90	MPI	CROSSKIT	NETCDF	IOIPSL/src	ORCHIDEE	OASIS3	LMDZ4	IPSLCM5	libIGCM	NEMO	UTIL
2010-06-07T06:10	OK	Code is reliable	-	087	400	8.0.4	18.1/3	3.6.1	HEAD 740	orchidee_1_9_4_2 2010-05-27T10:30:35	ipslcm5a 2010-03-18T15:16:18	1329	HEAD 951	246	HEAD 1779	HEAD 1773
2010-06-06T06:10	OK	Code is reliable	-	087	400	8.0.4	18.1/3	3.6.1	HEAD 740	orchidee_1_9_4_2 2010-05-27T10:30:35	ipslcm5a 2010-03-18T15:16:18	1329	HEAD 951	246	HEAD 1779	HEAD 1773
2010-06-05T06:10	OK	Code is reliable	-	087	400	8.0.4	18.1/3	3.6.1	HEAD 740	orchidee_1_9_4_2 2010-05-27T10:30:35	ipslcm5a 2010-03-18T15:16:18	1329	HEAD 951	246	HEAD 1779	HEAD 1773
2010-06-04T06:10	OK	Code is reliable	-	087	400	8.0.4	18.1/3	3.6.1	HEAD 740	orchidee_1_9_4_2 2010-05-27T10:30:35	ipslcm5a 2010-03-18T15:16:18	1329	HEAD 951	246	HEAD 1779	HEAD 1773

Sciences de l'environnement
Institut Pierre Simon Laplace

Dashboard and data access



Metrics Garden

Metrics Garden User Web Interface

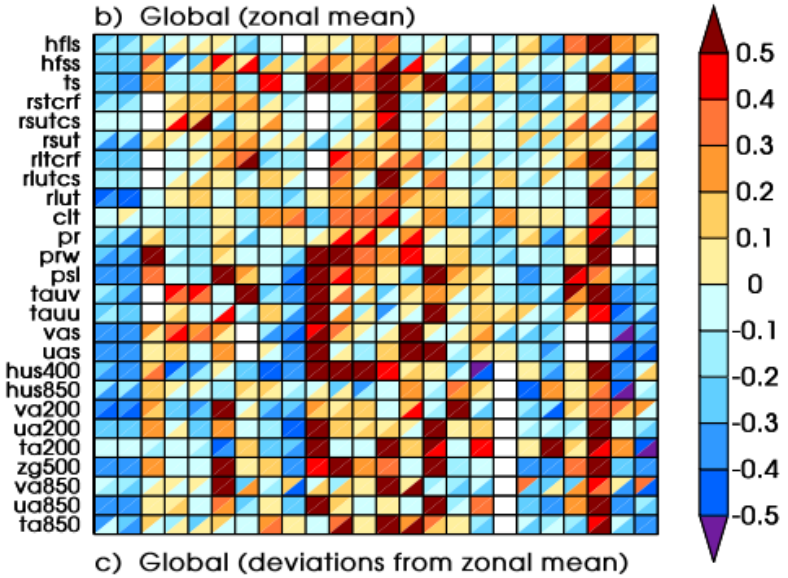
Test Glecker like metrics on CMIP5 version of IPSL models

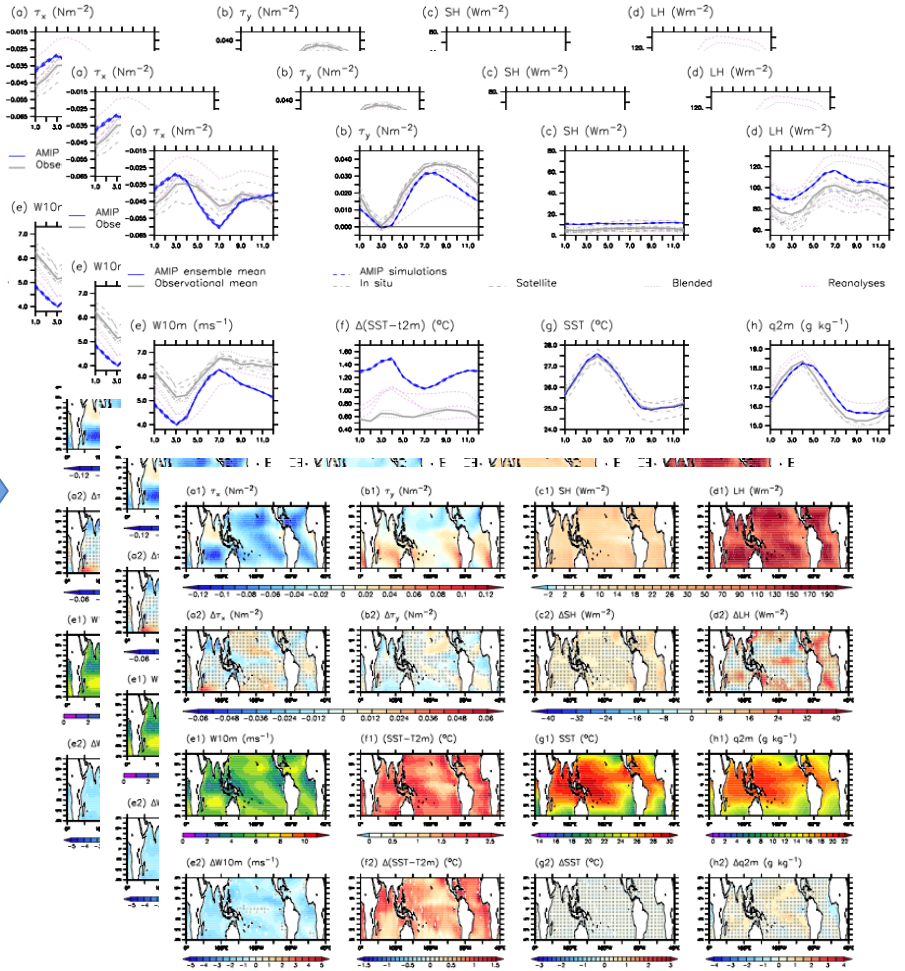
			Result															
		Variable	clt		hfls		hfss		pr		psl		rlds		rldscs		rlus	
		RegionName	Globe	NHEX	Globe	NHEX	Globe	NHEX	Globe	NHEX	Globe	NHEX	Globe	NHEX	Globe	NHEX	Globe	NH
Model	Experiment	SimName																
IPSL-CM5A-LR	historical	r1i1p1	-0.092	-0.118	+0.071	-0.006	-0.044	-0.024	+0.230	+0.241	+0.040	-0.020	+0.092	+0.072	+0.192	+0.092	-0.101	-0.
		r2i1p1	-0.098	-0.135	+0.055	-0.020	-0.062	-0.033	+0.213	+0.083	+0.039	+0.095	+0.026	-0.001	+0.124	+0.029	-0.113	-0.
		r3i1p1	-0.094	-0.126	+0.109	+0.010	-0.047	-0.035	+0.316	+0.238	+0.035	-0.033	+0.198	+0.155	+0.363	+0.230	-0.058	+0.
		r4i1p1	-0.090	-0.117	+0.036	-0.007	-0.071	-0.021	+0.252	+0.242	+0.001	-0.013	+0.136	+0.121	+0.217	+0.122	-0.089	+0.
		r5i1p1	-0.094	-0.115	+0.045	-0.038	-0.105	-0.072	+0.283	+0.197	+0.031	-0.019	+0.113	+0.034	+0.224	+0.056	-0.181	-0.
		r6i1p1	-0.097	-0.115	+0.053	+0.001	-0.087	-0.051	+0.370	+0.251	-0.041	-0.002	+0.116	+0.059	+0.221	+0.091	-0.057	-0.
IPSL-CM5A-MR	historical	r1i1p1	-0.122	-0.070	-0.331	-0.235	-0.167	-0.197	-0.537	-0.519	-0.214	-0.250	-0.409	-0.354	-0.510	-0.413	-0.142	-0.
		r2i1p1	-0.095	-0.030	-0.299	-0.220	-0.125	-0.170	-0.630	-0.593	-0.285	-0.245	-0.406	-0.343	-0.576	-0.383	-0.028	-0.
		r3i1p1	-0.097	-0.040	-0.370	-0.242	-0.125	-0.199	-0.597	-0.547	-0.303	-0.257	-0.433	-0.371	-0.628	-0.411	-0.052	-0.
IPSL-CM5B-LR	historical	r1i1p1	+0.878	+0.865	+0.630	+0.758	+0.833	+0.801	+0.098	+0.407	+0.697	+0.743	+0.567	+0.629	+0.372	+0.587	+0.819	+0.

Sciences de l'environnement

Institut Pierre Simon Laplace

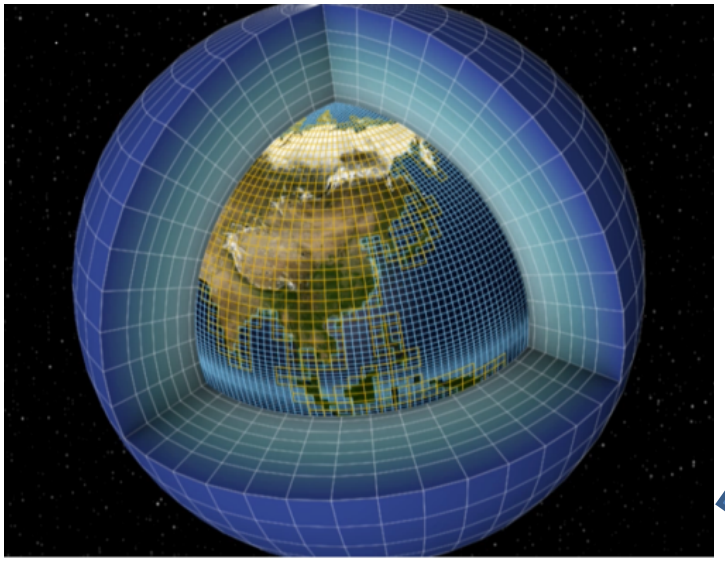
From syntetic metrics to great details

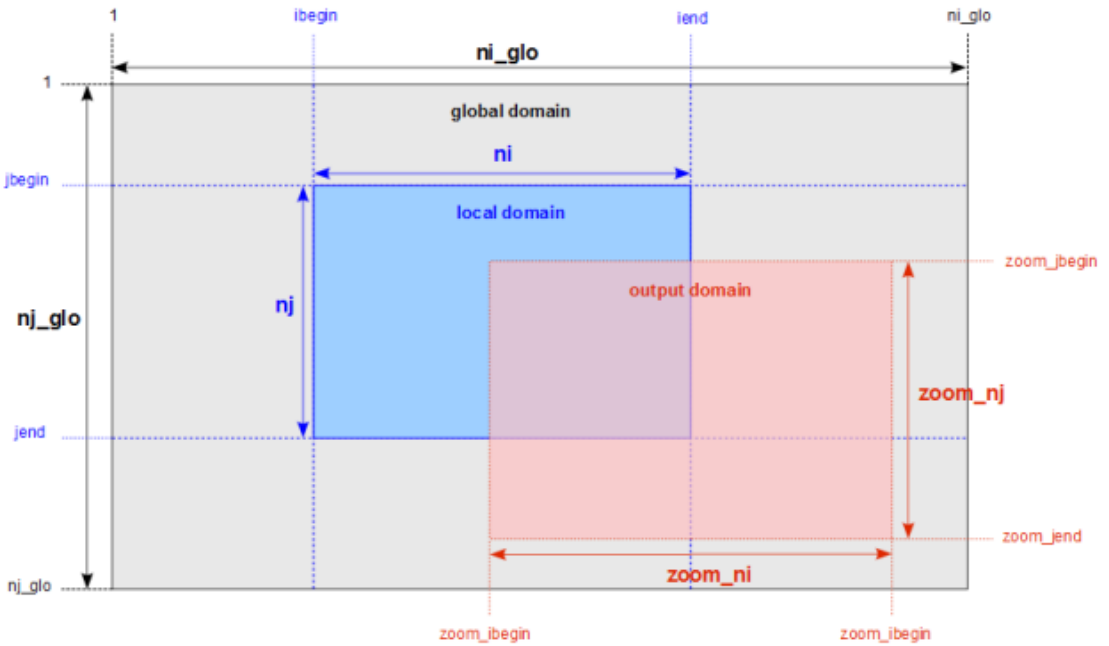


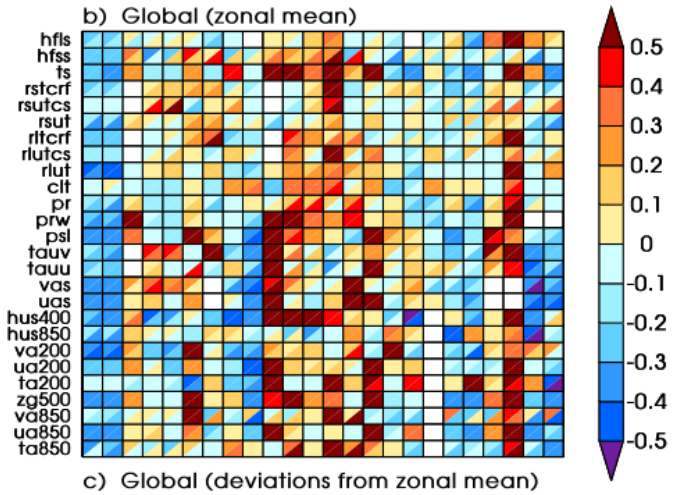


Sciences de l'environnement

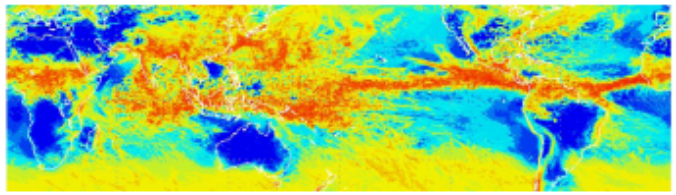
Institut Pierre Simon Laplace

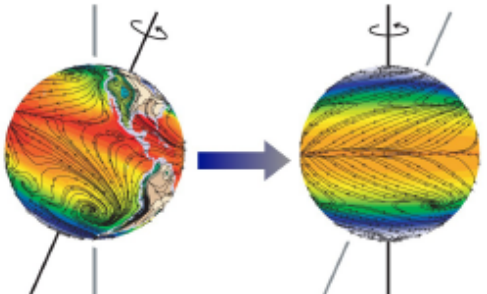


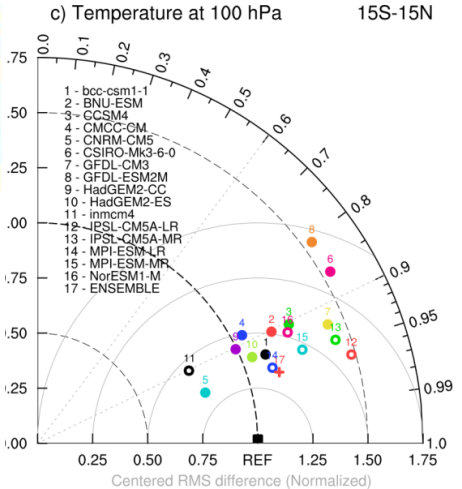


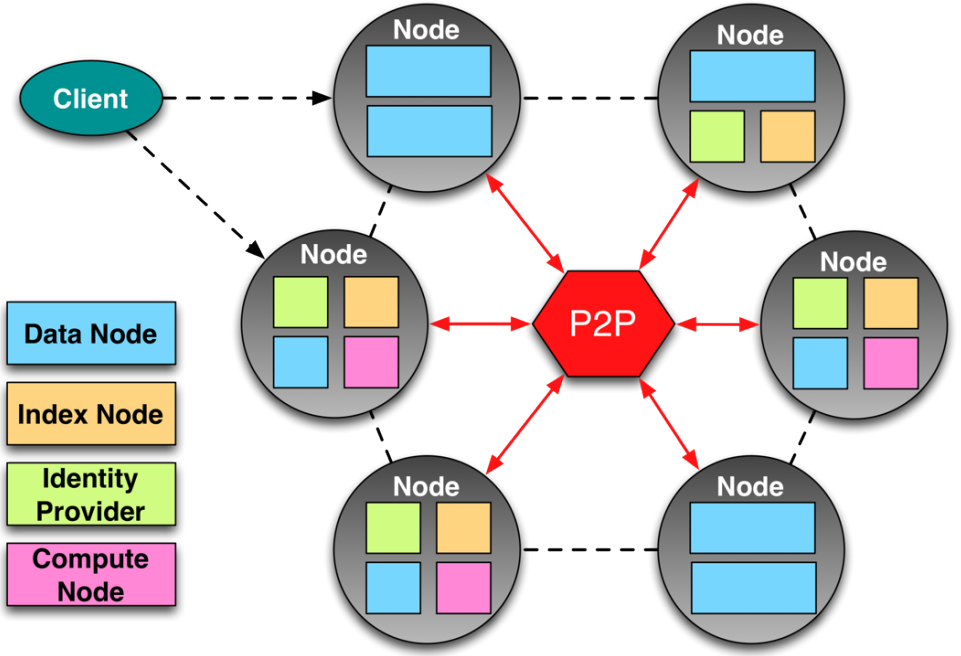


- Data producer
- ~~Distribution and archiving~~
- Analysis









SYNDA : Synchro-data

- Search for data in ESGF
  - File selection using DRS facets (realm, freq, experiment, ensemble, variable). So called template.
  - Incremental search (keep track of what have already been downloaded)
- Transfer files from ESGF to local filesystem
  - Transparent handling of x509 Certificate based security
  - HTTP Parallel download
- Schedule task execution
  - Capacity to define series of task over variables
  - Tracability and ability to pint point consequence of changes on final product

SYNDA : Synchro-data

- INITIALISATION

  synda cache init : synda certificate renew

- BASIC USAGE

  search datasets : synda search FACET..

  install a dataset : synda install DATASET..

  start the download : synda daemon start

  check download completion : synda queue status

Sciences de l'environnement
Institut Pierre Simon Laplace

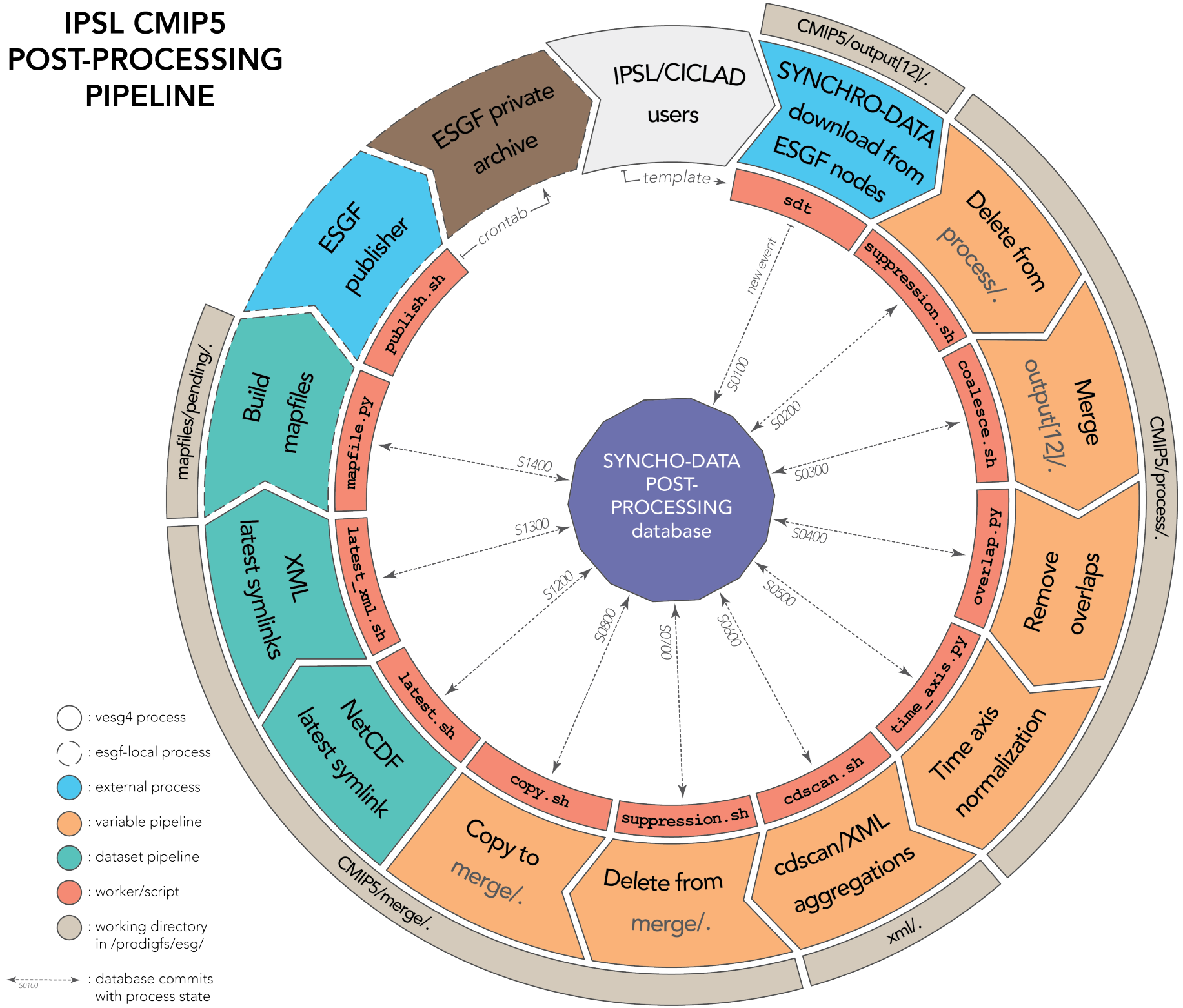
SYNDA : Synchro-data

project="obs4MIPs"

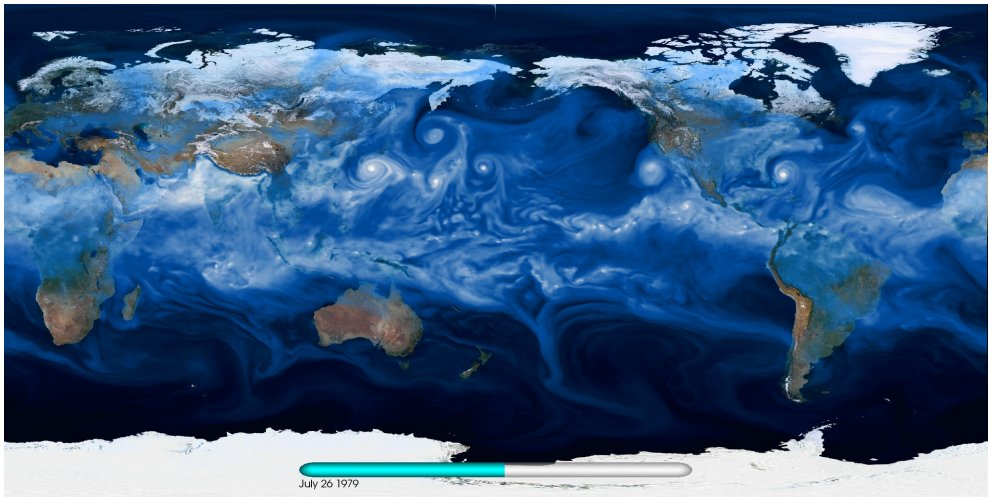
project="ISI-MIP%20Fasttrack"
GCM="GFDL-ESM2M HadGEM2-ES MIROC-ESM-CHEM"
searchapi_host="esg.pik-potsdam.de"

project="CORDEX"
model="HIRHAM5 RACMO22E REMO2009 WRF331F RCA4"
experiment="historical evaluation rcp26 rcp45 rcp85"
ensemble="all"
variable[*][day]="sfcWindmax tasmin tasmax prhmax pr"
variable[*][mon]="tas tasmin tasmax pr sfcWind"
domain="EUR-11"

IPSL CMIP5 POST-PROCESSING PIPELINE



Thank you for your attention

Also, if you have **skills** in High Performace Computing, Information Technology, Grids or Clouds, and if you are **looking for a job** please **ping me** by email. And I will be around ...

sebastien.denvil@ipsl.jussieu.fr