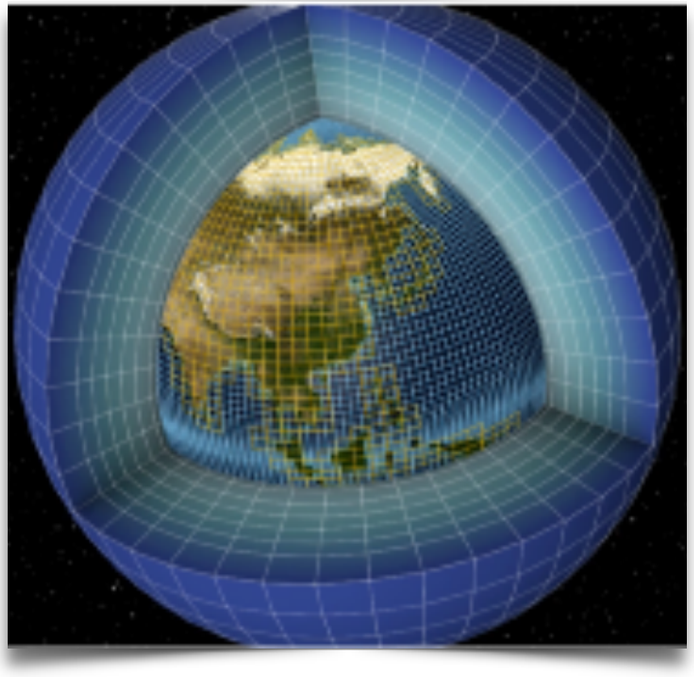
Institut Pierre Simon Laplace
Sciences de l'environnement

ESPRI

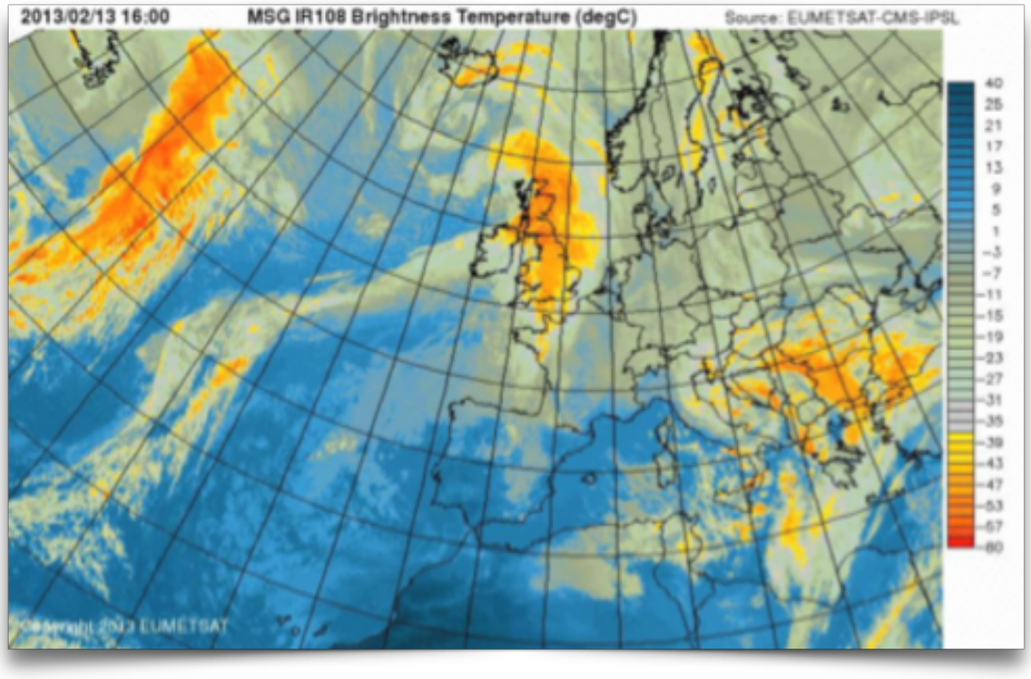
MÉSOCENTRE IPSL

MODULE 2

Data you can find or access



Model data



Observational data

INTRODUCTION

Objectives of data management on the mesocentre:

- The data close to the computing facility used by IPSL Labs
- To ease the use of observation and model data ⇨ IPSL: a unique place
- To allow a mix use of different data types (model data, in-situ, satellites, reanalyses):

- Which data?
  - Data produced by the IPSL scientific community
  - Data from other data centres or communities

The organised archive of data ≈ 1.6Po

INTRODUCTION

Different services proposed by the ESPRI team:

- Data management (acquisition, production, archival)

A centralised management of datasets:
To avoid the duplication of dataset on individual user account.

Open data/protected data:
Some data are non public.

- Tools for distribution and computation on data, access to the data
- Documentation on data (cf web sites)
- User and project support, close to the scientific teams
- Dedicated workspaces for projects (specific storage for data exchange)
- DOI: IPSL is allowed to generate DOI on data (INIST convention)

Observation Data

What we do?

Host, manage, and distribute data...
Deliver by-products, algorithms, software, tools
Offer services to the scientific community and other end users

Observation Data

What we do?

The services around the observation data:

- Different types of observation data: in-situ, satellite, reanalyse, FCDR, etc.
- Data management, a collaborative task:
  - Centralised data of interest on demand
  - If you have dataset, you can propose to share it on the mesocentre
- Data from IPSL team or from other data centres (mirroring)
- Formatting in NetCDF in some cases
- Services for the valorisation of YOUR dataset (website, dedicated access and quick look)

Where are the observation data on the mesocentre?

- `/bdd` — Long term datasets on water cycle, clouds, etc.
- `/etherfs` — Atmospheric chemistry dataset

How to find the information?

- On websites (today): climserv.ipsl.polytechnique.fr or cds-espri.ipsl.fr
- Mail to meso-support@ipsl.fr (Cathy Boonne, Sophie Cloché, Karim Ramage)

OBSERVATION DATA

Examples of observation data on /bdd

In-situ data:

- Campaigns measurement :CAL/VAL MT, balloons, etc.
- Systematic measurements: SIRTA
- Post-processing of in-situ Radiosounding: ARSA, TIGR
- Others.

Satellite L1/L2/L3 products :

- AMSU L1C, TRMM, SSMI, Wentz products (SSMI, TMI), GSSTF,
- CALIOP L1/L2, IIR L1/L2
- GEO (HDF) pour MT, METEOSAT, MSG, ISCCP, CLAUS, SAF-LAND
- ScaRaB 1/2, CERES, ERBE, GERB, NOAA-OLR
- POLDER1/2, PARASOL, MODIS, AVHRR,
- NVAP, GPCP, CPC, HOAPS
- AVISO, OISST (MW), OISST-NOAA, OAFLUX, SAFOSI, QUIKSCAT
- CFMIP-obs, GEWEX/CA, clear sky WV radiance, Emissivité surface MW
- Others

Satellite/model outputs for INDOEX, AMMA, HyMex (+ radar), ChArMEx

Try to centralise WRF forcing data.

OBSERVATION DATA

Examples of observation data on /bdd

Focus on reanalysis dataset: climate and satellite reanalysis

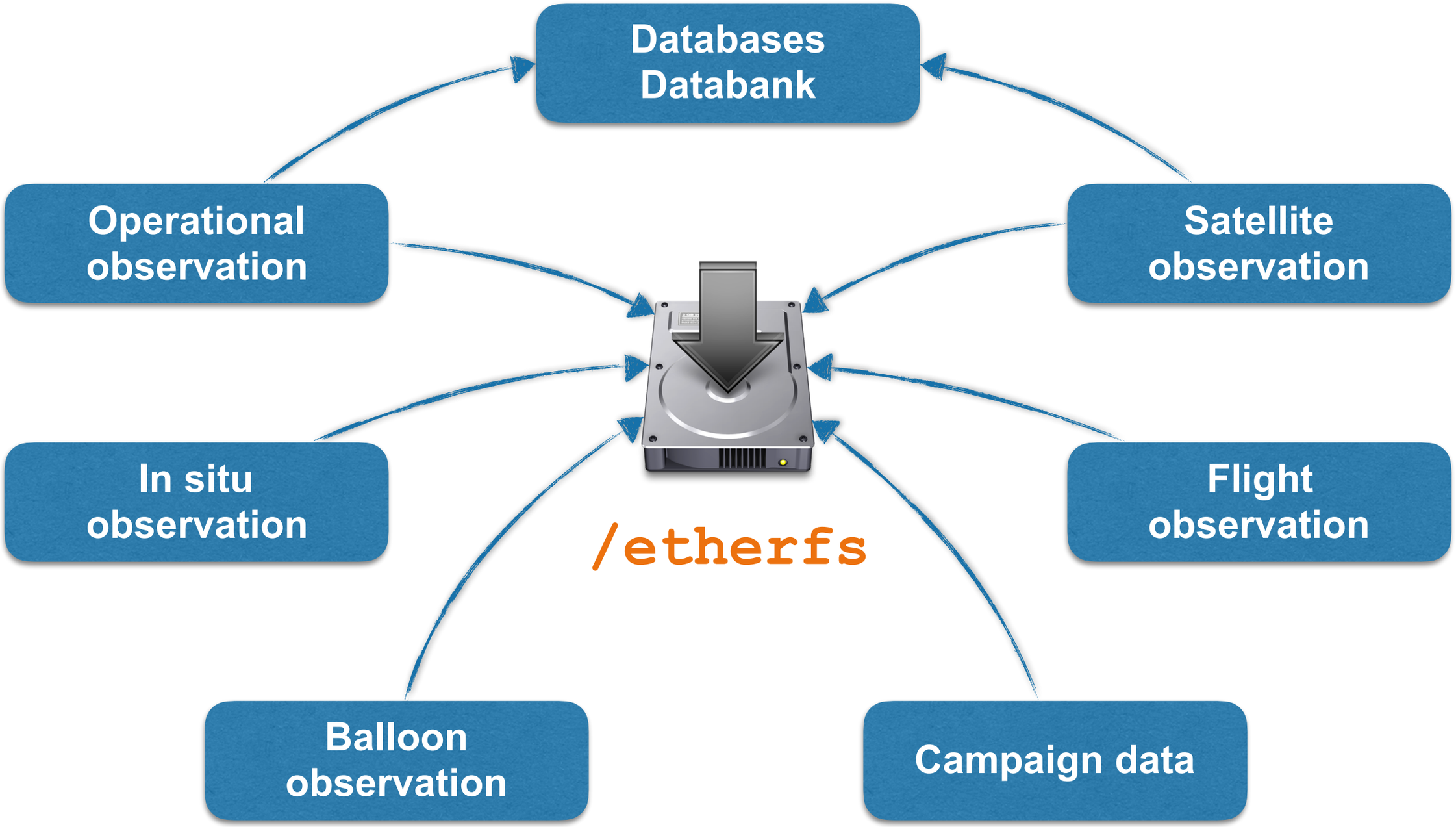
- On the IPSL mesocentre (and on IDRIS and TGCC)
- In NETCDF format (+ GRIB at IDRIS)
- Available through ssh, FTP or OpenDAP

(Re)Analyses
ECMWF *(protected by MeteoFrance convention)*: • ERA40 • ERA-Interim • ERA20C • Analyses 1.125° COPERNICUS C3S: • ERA5
MERCATOR-OCEAN: • GLORYS2V3 • MEDRYS1V1 • OSTIA_SST
NASA: • MERRA1/2
NCEP1/2 *NCEP/CFSR v1* *NCEP FNL*
NOAA CIRES 20th century reanalysis v2 (20CRv2)

FCDR
Archive AMSU (2000- présent)
Archive SSMI (1987-présent)
GridSat Géostationnaires (1980-présent)

Observation Data

Examples of observation data on `/etherfs`



Volume ≈ 470To

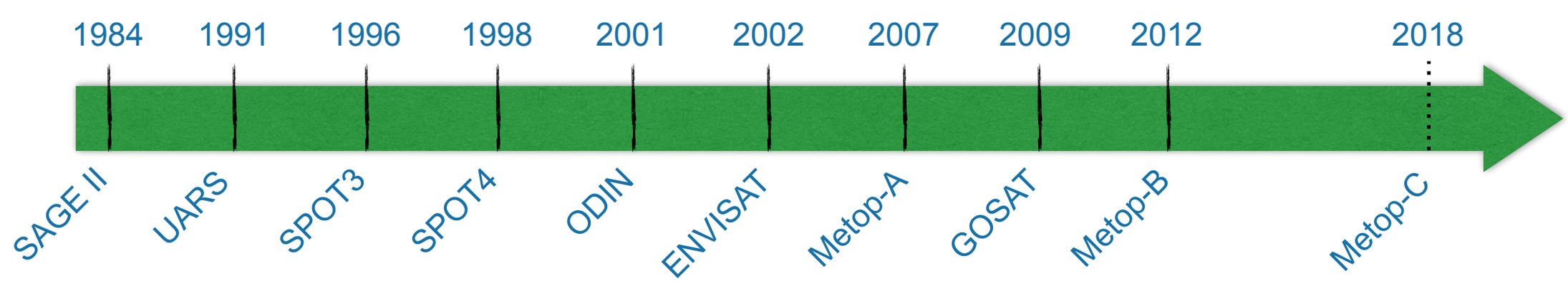
OBSERVATION DATA

Which observation data on `/etherfs` ?

`/etherfs` data management context: 470To

A huge variety of data sets: 250 from level 1 to 4

- Earth satellite data from UARS, ADEOS1, ERBS, POAMII, POAMIII, ODIN, ENVISAT, METOP/IASI – GOME2 and GOSAT missions

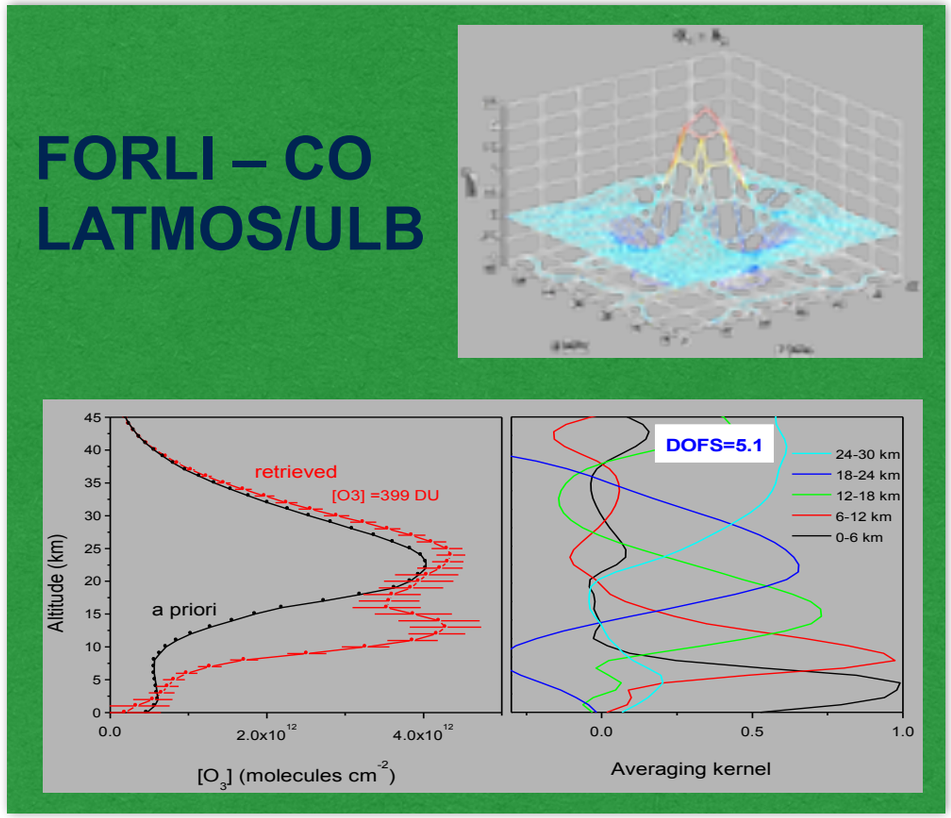


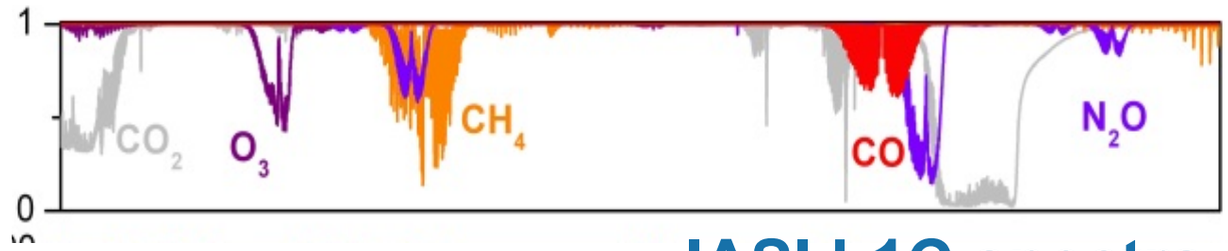
- Ground-based and in-situ data : Lidar, Radar, Balloon, gondolas, aircraft data, radio sounding, etc.
- Modelling data from REPROBUS, MIMOSA, etc.
- Ancillary data like spectroscopic data (GEISA), kinetic data, etc.

Observation Data

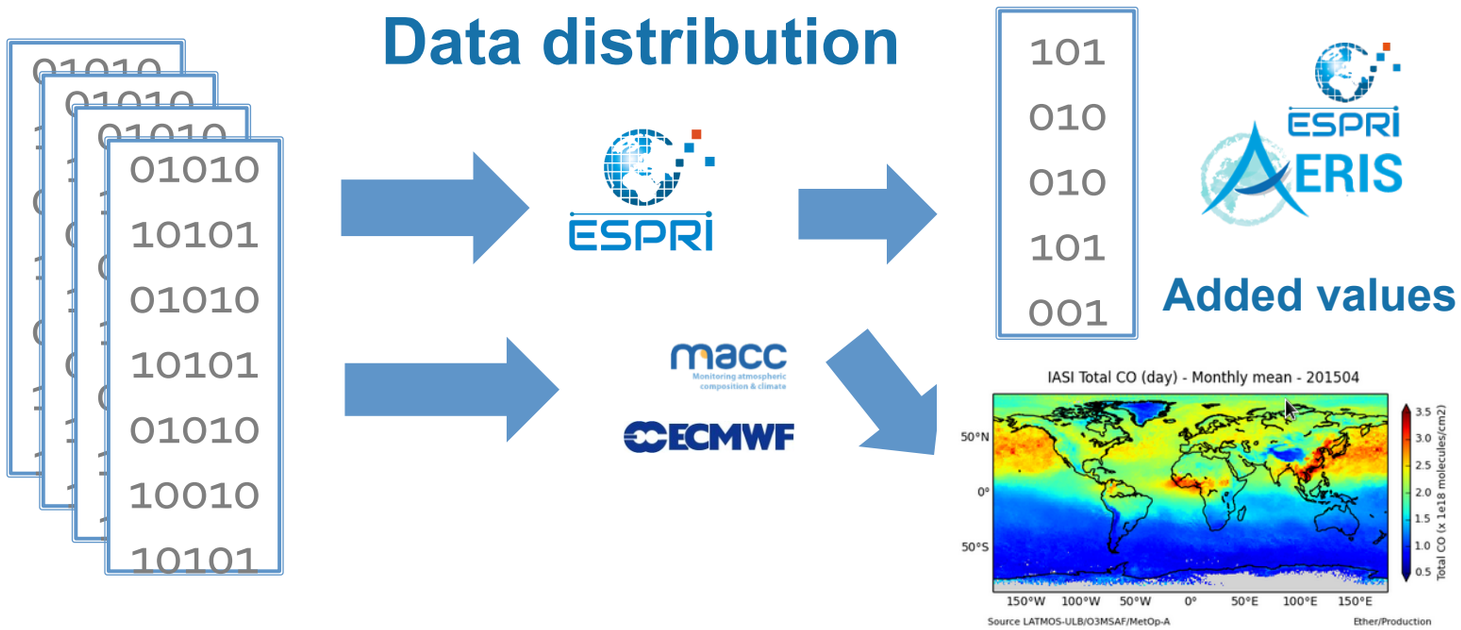
Satellite data management and distribution : IASI example

CO data treatment





IASI L1C spectra



CO data retrieval

Production level 2 data and QL on the fly

Currently, IASI O3, CO, SO2, HCOOH data retrieval and production uses IPSL Mesocentre.

OBSERVATION DATA

Satellite data management and distribution

Satellite data and quick looks distribution:

- On the IPSL mesocentre
- In native format
- Available through SSH or HTTP

Satellite data	Format	Distribution
IASI level 1C (METOP-A-B)	BUFR	SSH
IASI level 2 (O3, CO, SO2, CH4, HCOOH, NH3)	ASCII,	HTTP
AMSUA-MHS-HIRS4 level 1C (METOP-A-B)	BUFR	SSH
GOME2 level 1B (METOP-A-B)	HDF	SSH
GOME2 level 2 (METOP-A-B)	HDF, NetCDF	SSH
GOSAT level 1B / FTS/CAI	HDF	SSH
GOSAT level 2 / FTS/CAI	HDF	SSH
SAGE II, UARS, SPOT3, SPOT4, ODIN, ENVISAT (level 2)	Native	HTTP

Non public data

Observation Data

Satellite data management and distribution

Distribution through cds-espri.ipsl.fr

CO total column from IASI

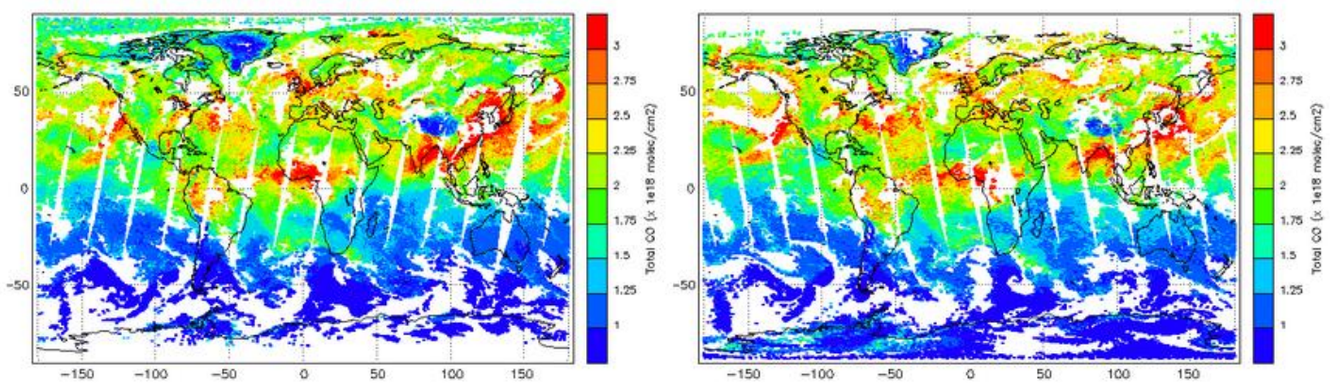
Daily maps | Monthly maps

How to read the data
Contact & publications

Satellite : Metop-A
Projection : Global
Date : < 2017-04-05 >

Metop-A

IASI Total CO (day) 2017/04/05 | IASI Total CO (night) 2017/04/05



Metop-A CO data file for this day
Data file archive (via calendar)

Carbon monoxide (CO)

How to read the data

Download this Readme LATMOS-ULB FORLI-CO data retrieved from IASI **here**.

How to read the data since December 2, 2010 :

Each file contains data for one day of observation. In each file, each line corresponds to one observation.
File names include date of observation. Their structure is :
iasi_CO_LATMOS_ULB_YYYYMMDD_vXXXXXXXX.txt
where : YYYY = year, MM = month, DD = day, XXXXXXXX = version number of the retrieval code.
The data are organized in 60 columns :

- columns 1 and 2 : latitude and longitude
- columns 3 and 4 : date (yyyymmdd) and time (hhmmss)
- columns 5 : solar zenith angle
- columns 6 : IASI field of view (0, 1, 2 or 3)
- columns 7 : information flag about Eumetsat IASI L2 temperature profiles retrieval method (see

Carbon monoxide (CO)

For any use in publication please contact Cathy Clerbaux (cathy.clerbaux@latmos.ipsl.fr)

Contact & publications

References :

Algorithm description

FORLI radiative transfer and retrieval code for IASI

D. Hurtmans, P.-F. Coheur, C. Wespes, L. Clarisse, O. Scharf, C. Clerbaux, J. Hadji-Lazaro, M. George, S. Turquety J. Quant. Spectrosc. Radiat. Transf., 113, 11, 1391.1408, 2012

Validation

Carbon monoxide distributions from the IASI/METOP mission: evaluation with other space-borne remote sensors

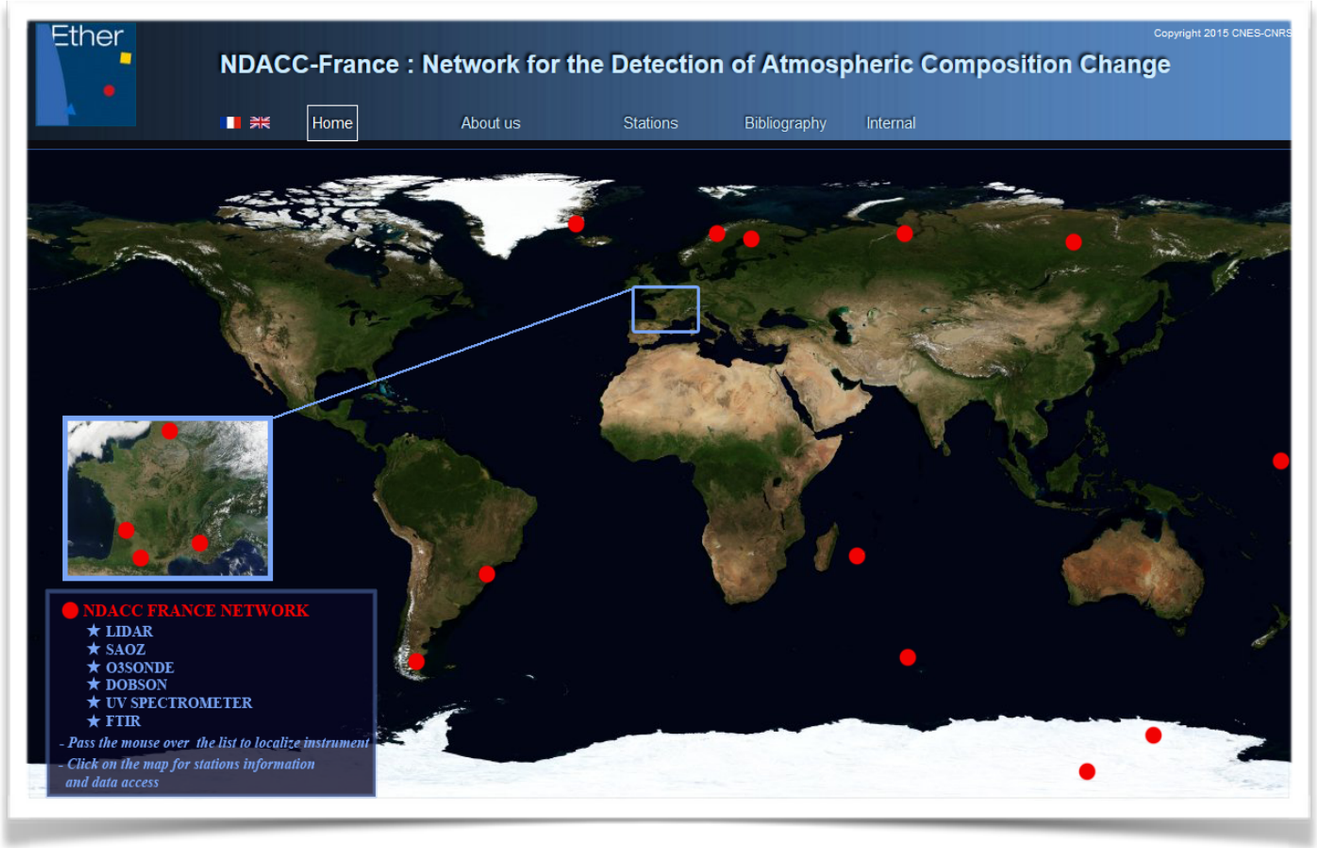
M. George, C. Clerbaux, D. Hurtmans, S. Turquety, P.-F. Coheur, M. Pommier, J. Hadji-Lazaro, D.P. Edwards, H. Worden, M. Luo, C. Rinsland, and W. McMillanAtmos. Chem. Phys., 9, 8317-8330, 2009

Observation Data

Ground based and in situ data

Ground based and in situ data and quick looks distribution (ACTRIS-FR)

- On the IPSL mesocentre
- In ames format
- Available through ftp or http



17 stations:

OHP, La Réunion, Dumont d'Urville, OMP, Villeneuve d'Ascq, Briançon, Scorebysund, Andoya, Sodankyla, Zhigansk, Salekhard, Bordeaux, Tarawa, Bauru, Kerguelen, Rio Gallegos, Concordia

Atmospheric components measured:

- O_3 tropo et strato,
- NO_2,
- Temperature,
- Aerosols

http://cds-espri.ipsl.fr/NDACC/

OBSERVATION DATA

Ground based and in situ data

Distribution data via FTP

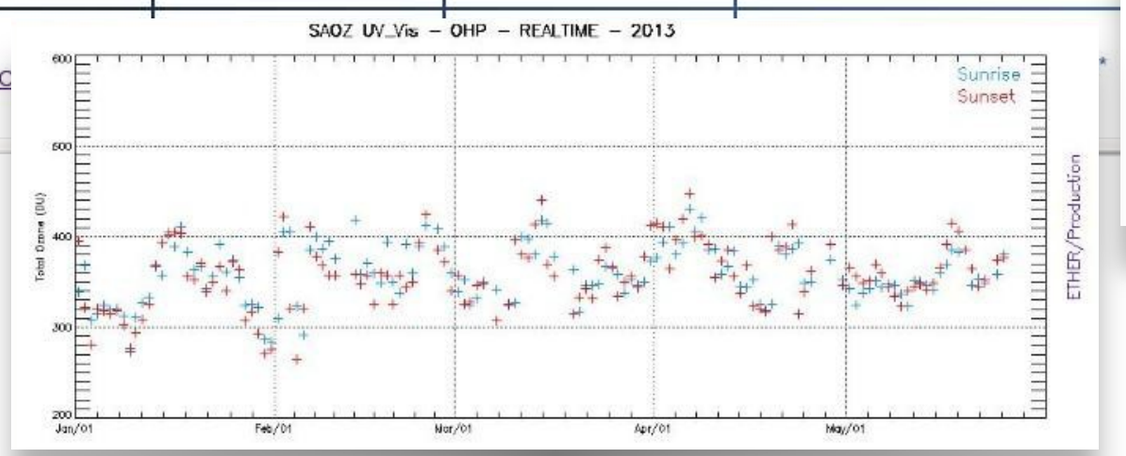
OHP | La Réunion | Dumont d'Urville | OMP | Villeneuve d'Ascq | Briançon | Scorebysund | Andoya | Sodankyla | Zhigansk | Salekhard | Bordeaux | Tarawa | Bauru | Kerguelen | Rio Gallegos | Concordia

Observatoire de Haute Provence

[Alpine station - Lat : 43°56'N, Lon : 05°42'E] Station data access *Scientist coordinator : Philippe Keckhut* *Station leader : Guy Tournois*

INSTRUMENT	MEASUREMENT	PI	NDACC STORE DATA	QUICKLOOK		STATUS
				Long term	Real Time	
SAOZ	NO2/O3 column	A. Pazmino, F. Goutail and J. P. Pommereau	1992-2015 *pattern files : ohtc[yy][mm].pzv*			Routine
LIDAR *(LIDAR DIAL)*	O3 tropospheric profiles *(3-14 km ASL)*	G. Ancellet	1990-2015 *pattern files : ohto[yy][mm].anl*			Routine
LIDAR *(LIDAR DIAL)*	O3 stratospheric profiles *(10-50 km)*	S. Godin-Beekmann	1985-2016 *pattern files : oho3[yy][mm].gol*			Routine
LIDAR	Temperature profiles *(30-80 km)*	P. Keckhut and A. Hauchecorne	1991-2009 2012-2015 *pattern files : ohte[yy][mm].hal*			
LIDAR *(LIDAR RETRODIFFUSION RAYLEIGH/MIE)*	Aerosols profiles	J. Jumelet and P. Keckhut	1991-1994, 1999-2010, 2012-2013/0 *pattern files : ohae[yy][mm].dal*			
EC						

Nom	Taille	Dernière modification	
ohae0...	314 KB	03/12/2004	00:00:00
ohae0...	150 KB	21/04/2017	08:00:00
ohae0...	242 KB	03/12/2004	00:00:00
ohae0...	150 KB	21/04/2017	08:00:00
ohae0...	146 KB	03/12/2004	00:00:00
ohae0...	190 KB	21/04/2017	08:00:00
ohae0...	122 KB	14/12/2004	00:00:00



Observation Data

Model data distribution

Model data and quicklooks distribution

- On the IPSL mesocentre
- In native format
- Available through HTTP, FTP

Index de ftp://mimosa@ftp.ipsl.fr/Haute_resol/2010/

Vers un rép. de plus haut niveau

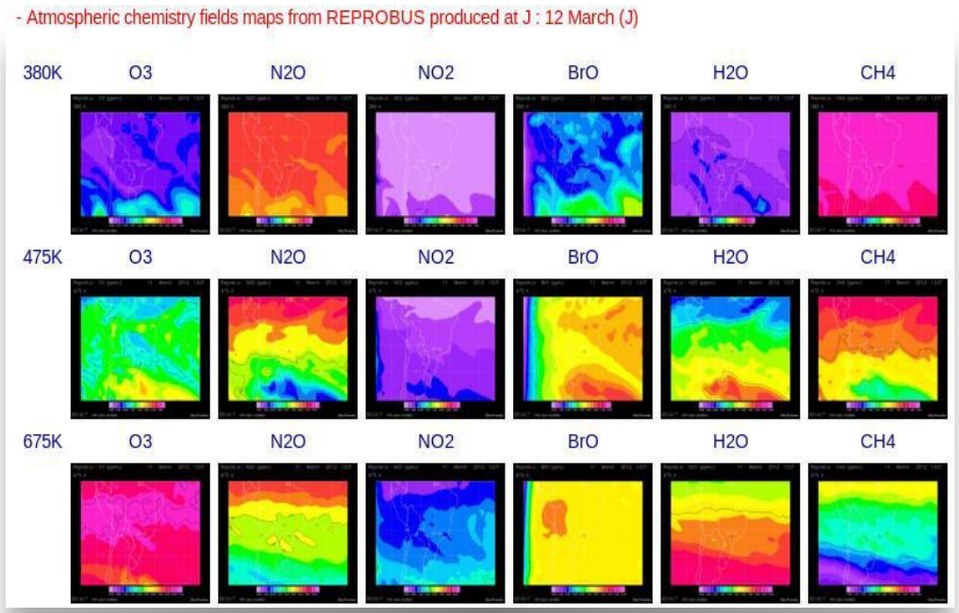
Nom	Taille	Dernière modification	
pv_t_ABE_201001_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_ABE_201002_profile.txt	147 KB	21/03/2011	00:00:00
pv_t_ABE_201003_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_ABE_201004_profile.txt	158 KB	21/03/2011	00:00:00
pv_t_ABE_201005_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_ABE_201006_profile.txt	158 KB	21/03/2011	00:00:00
pv_t_ABE_201007_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_ABE_201008_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_ABE_201009_profile.txt	158 KB	21/03/2011	00:00:00
pv_t_ABE_201010_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_ABE_201011_profile.txt	158 KB	21/03/2011	00:00:00
pv_t_ABE_201012_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_AIR_201001_profile.txt	163 KB	21/03/2011	00:00:00
pv_t_AIR_201002_profile.txt	147 KB	21/03/2011	00:00:00
pv_t_AIR_201003_profile.txt	163 KB	21/03/2011	00:00:00

MIMOSA :

Potential vorticity and temperature fields: daily analysis and forecasts production (Mimosa model, North and South Hemisphere and Tropic)

REPROBUS:

50 atmospheric constituent fields (O_3, ClO, NO_2, N_2O, BrO, etc.): daily analysis production (REPROBUS model)



OBSERVATION DATA

GEISA 2015 data distribution

GEISA data distribution:

- On the IPSL mesocentre
- In native format
- Available through HTTP only (cds-espri.ipsl.fr/GEISA)

Three SUB-DATABASES:

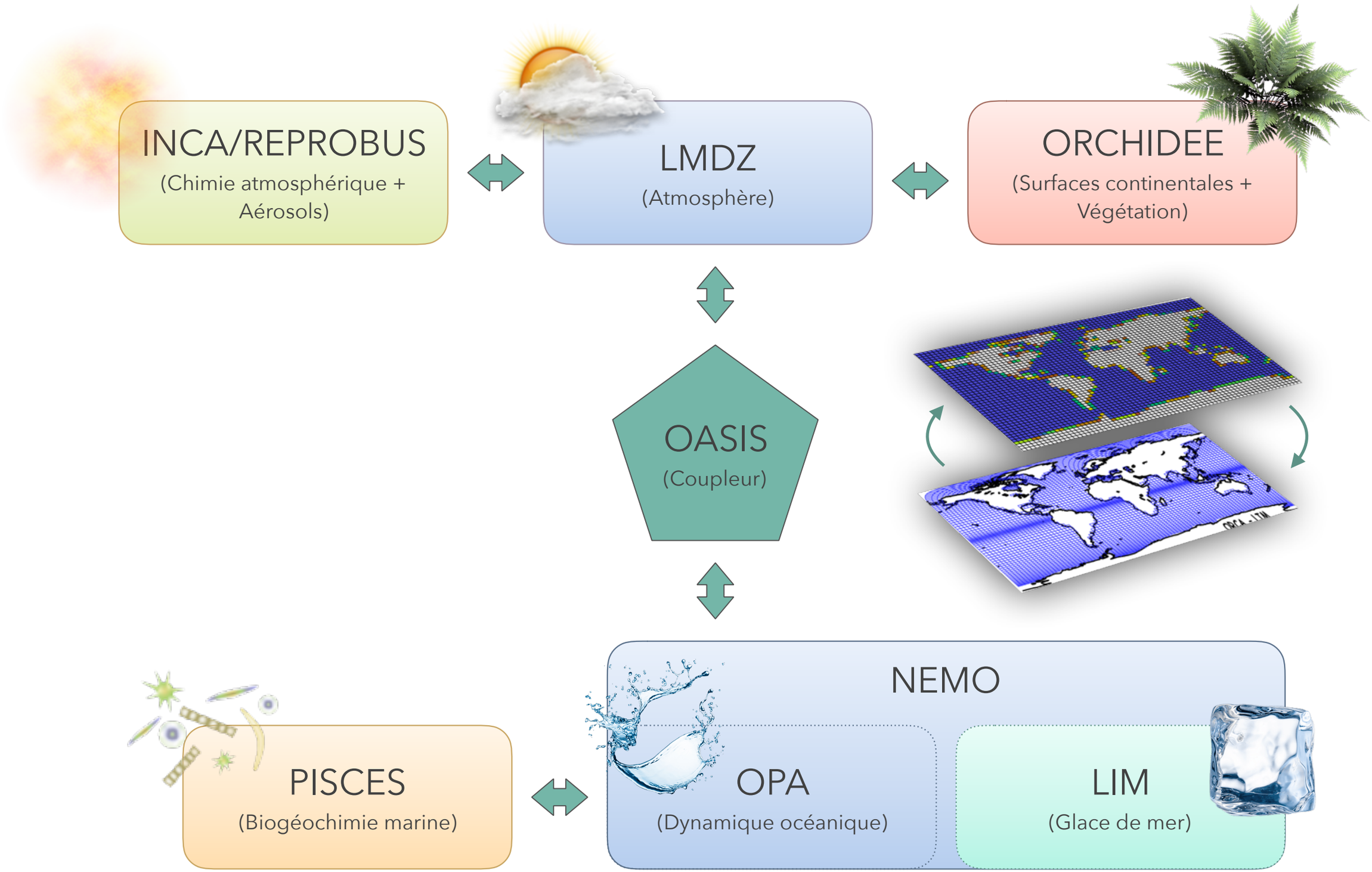
- Line parameters sub-database
  - 52 molecules (113 isotopic species)
  - Over 4,800,000 entries in the spectral range
    10-6 - 35,877 cm^{-1} (1010 - 0.28 µm)

(Major permanent constituants of EARTH atmosphere: O_2, H_2O, HDO, CO_2, etc. and trace molecules: CH_4, NO, SO_2, NO_2, NH_3, HNO_3, OH,HF, HCl, HBr, HI, ClO, OCS, H_2CO, PH_3)
(Molecules in atmospheres of JUPITER, SATURN, URANUS,TITAN, etc.: C_6H_6, CH_4, CH_3D, C_2H_2, C_2H_4, GeH_4, HCN, C_3H_8, C_3H_4)

- Absorption cross-sections sub-database (IR and UV/Visible)
- Microphysical and optical properties of atmospheric Aerosols sub-database

MODEL DATA

Climate models: example of IPSL-CM



MODEL DATA

IPCC Assessment Report cycle

The group's objective is the study of natural and anthropogenic variability in the global climate system. IPSL is also studying climate change impacts and usage of climate projections for adaptation to climate change related to industry. IPSL is one of the climate modelling centre of international repute contributing to the IPCC (Intergovernmental Panel on Climate Change).

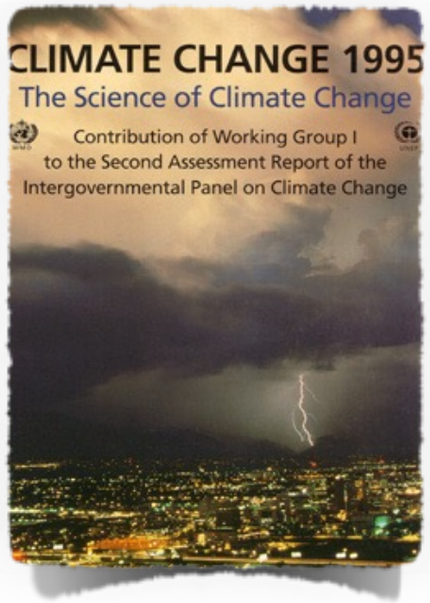
WCRP World Climate Research Programme

IPCC ARs
Collect the state of the art in understanding the climate system

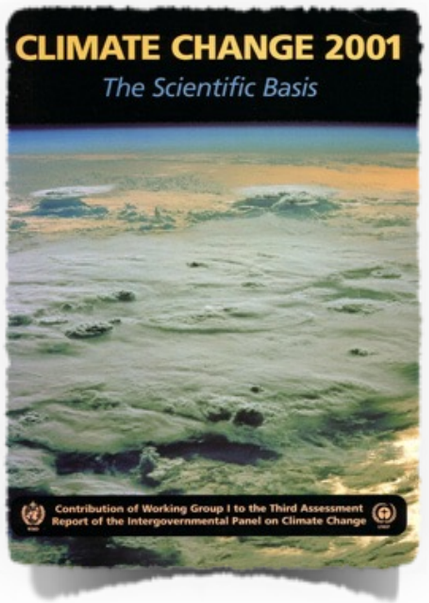
CMIP The World Climate Research Programme's Coupled Model Intercomparison Project

...

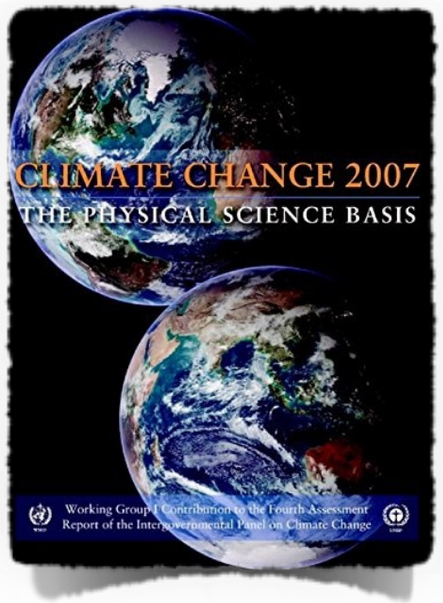
#2



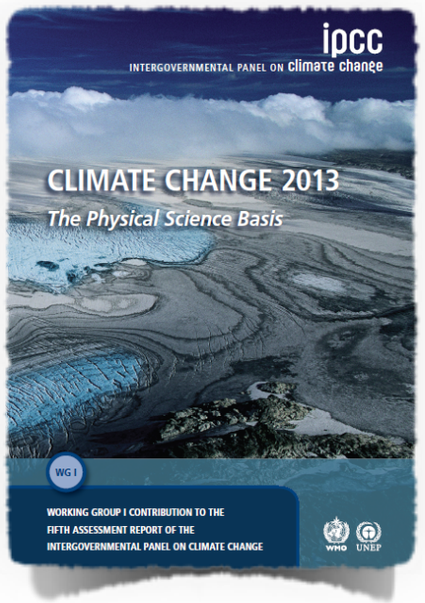
#3



#4



#5

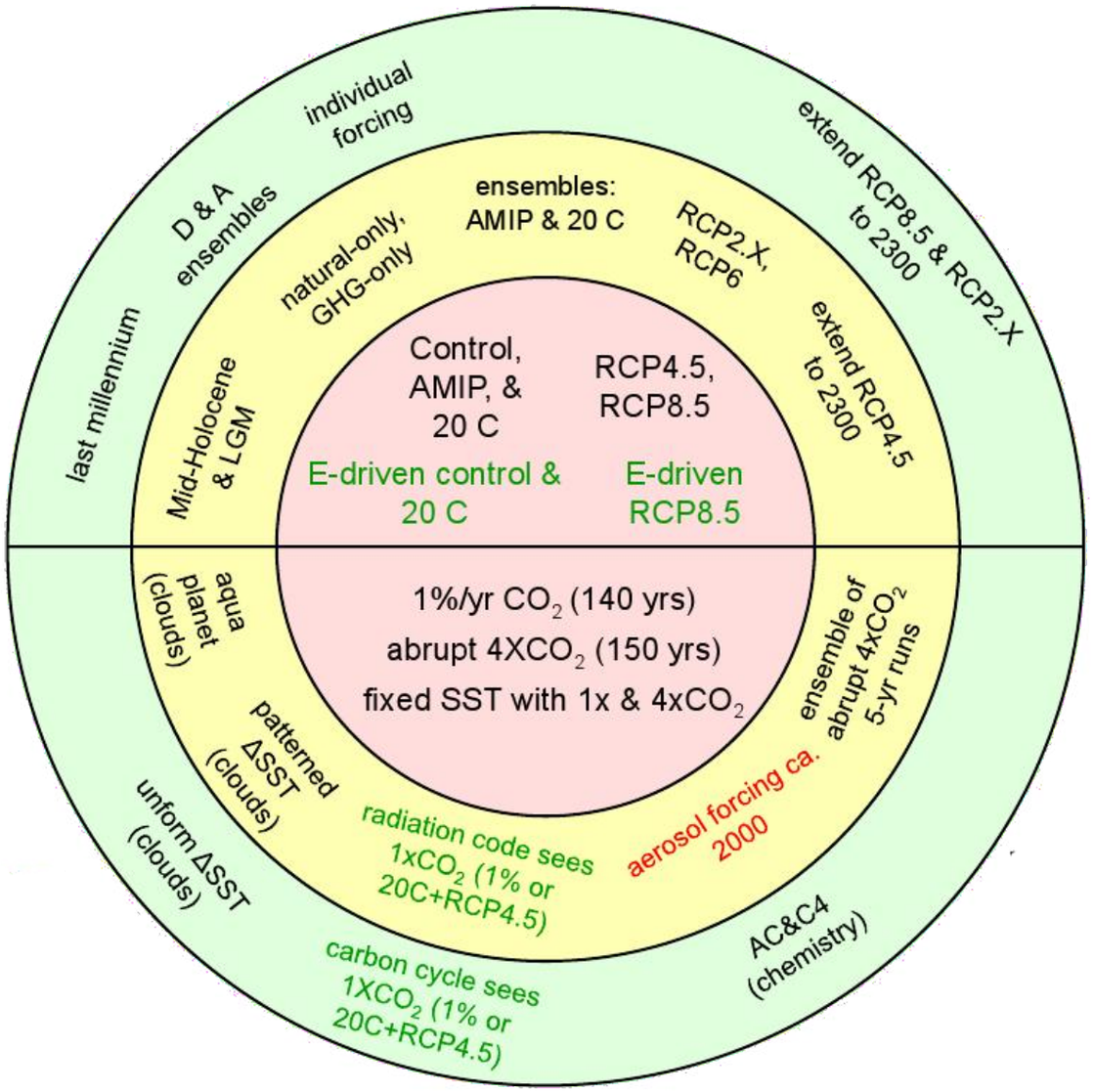


...

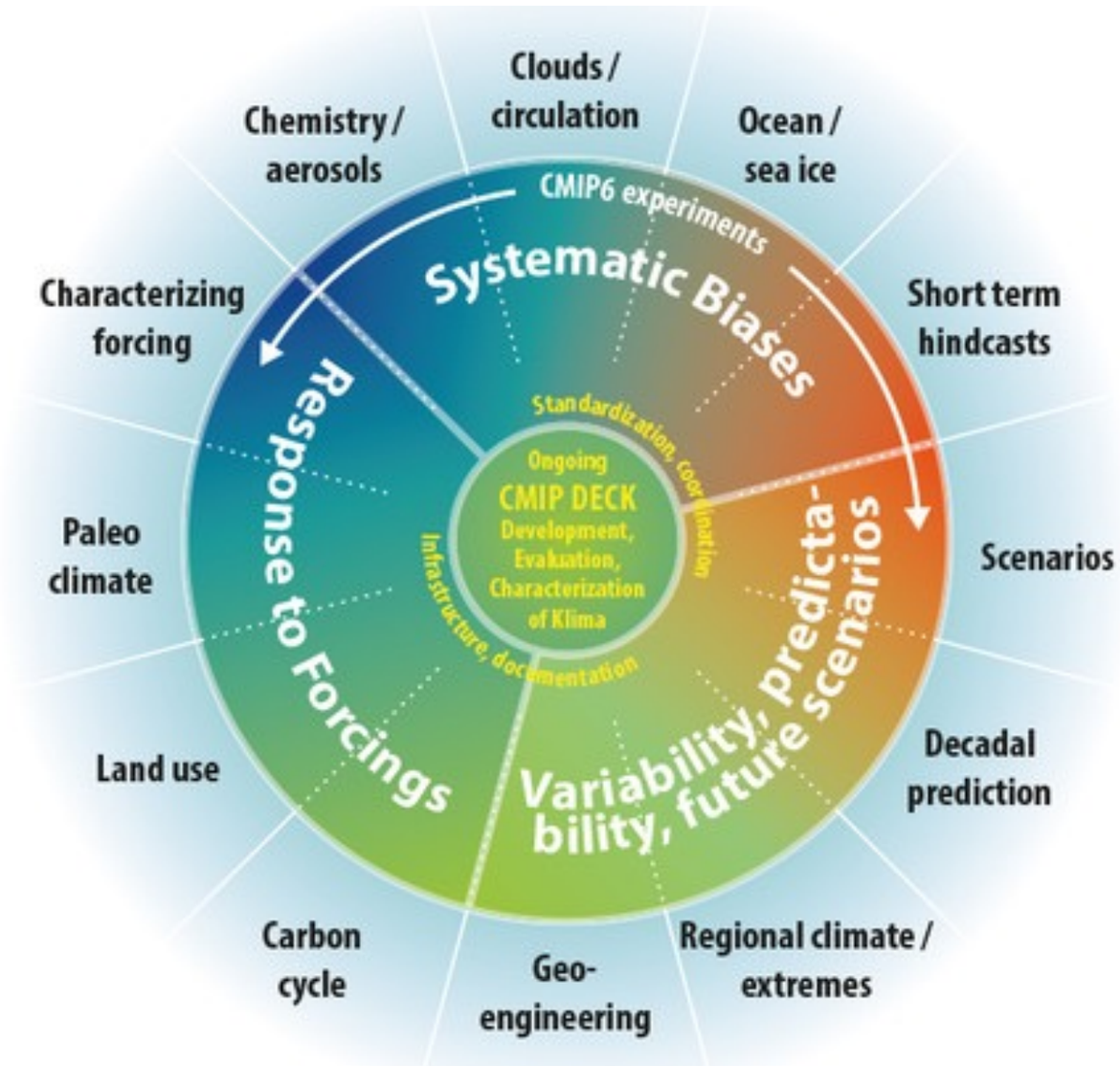
MODEL DATA

Couple Model Intercomparison Project

From phase 5 (CMIP5)



To phase 6 (CMIP6)

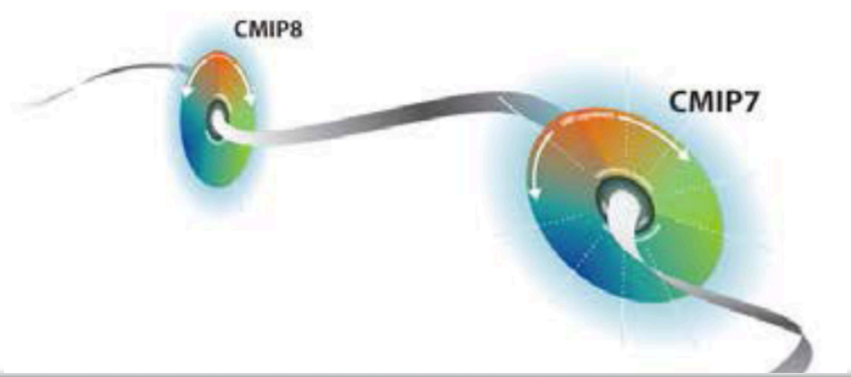


MODEL DATA

Couple Model Intercomparison Project

Many, many processes !
Many, many (interconnected) communities !
Many, many closed tied projects:

input4MIPs	Input datasets for Model Intercomparison Projects
ISIMIP	Inter-Sectoral Impact Model Intercomparison Project
GeoMIP	Geoengineering Model Intercomparison Project
PMIP3	Paleoclimate Modelling Intercomparison Project
TAMIP	Transpose-Atmospheric Model Intercomparison Project
EUCLIPSE	European Union Cloud Intercomparison, Process Study and
LUCID	Land-Use and Climate, IDentification of robust impacts

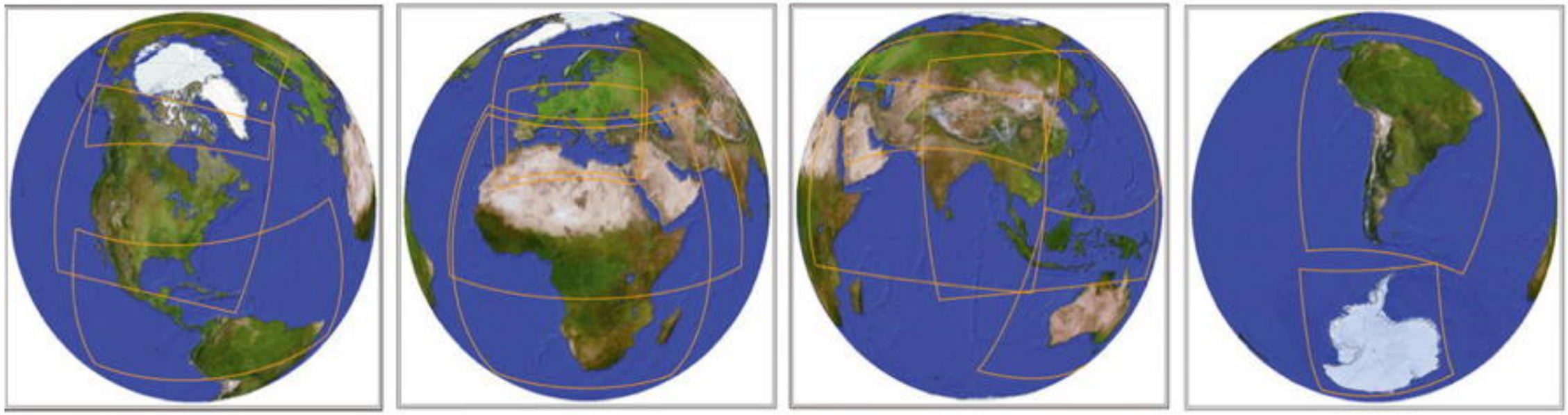


The Coordinated Regional Climate Downscaling Experiment (CORDEX)

The CORDEX vision is to advance and coordinate the science and application of regional climate downscaling through global partnerships. CORDEX goals are:

- ✓ To better understand relevant regional/local climate phenomena, their variability and changes, through downscaling,
- ✓ To evaluate and improve regional climate downscaling models and techniques,
- ✓ To produce coordinated sets of regional downscaled projections worldwide, and
- ✓ To foster communication and knowledge exchange with users of regional climate information.

CORDEX domains



MODEL DATA

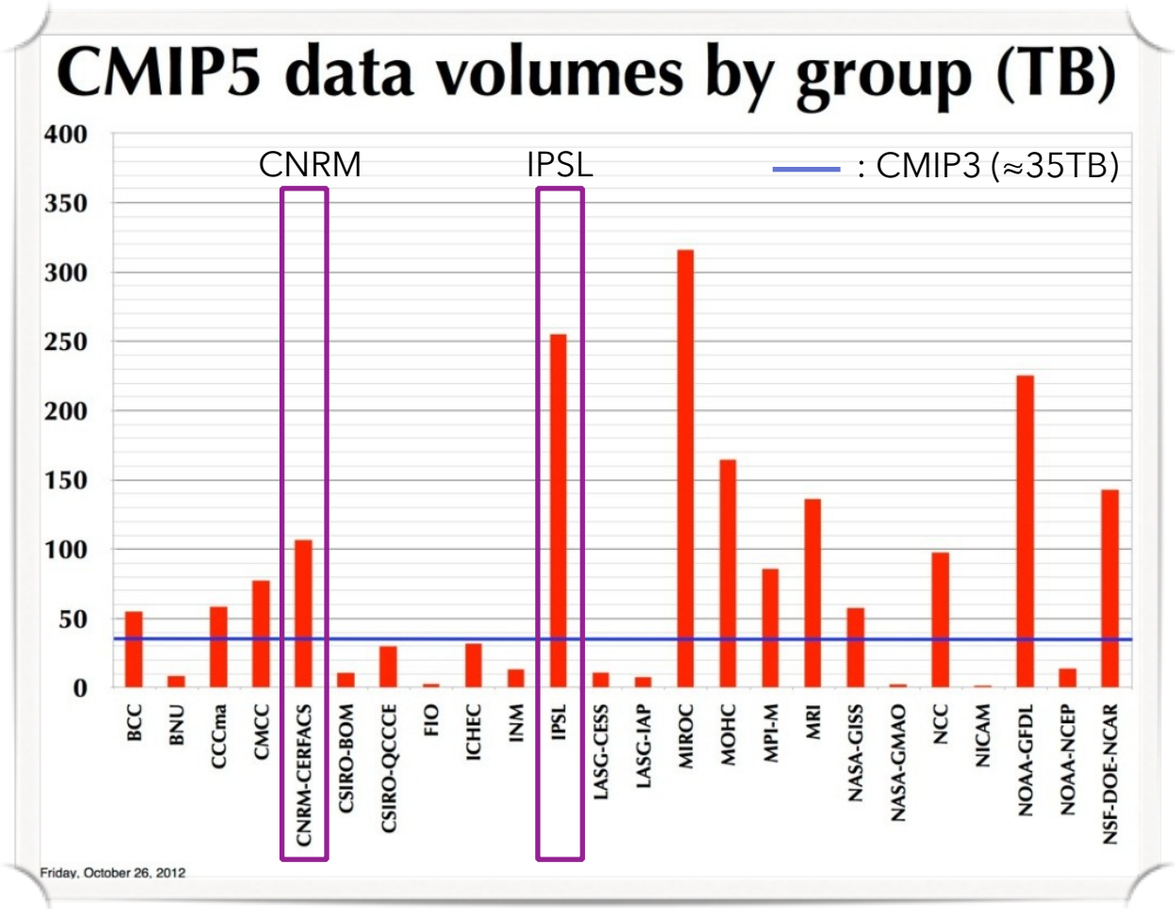
Climate « big » data

CMIP5 data archive final status

- 1.8 PB for 59,000 datasets, 4.3 Mio Files,
- 60 climate models
- CMIP5 data is about 50 times CMIP3.

Extrapolation to CMIP6

- CMIP6 has a more complex experiment structure than CMIP5,
- More models, finer spatial resolution and larger ensembles,
- **Factor of 20: 36 PB in 86 Mio Files**
- **Factor of 50: 90 PB in 215 Mio Files**



MODEL DATA

A world of 'facets'

Computation is useless if results cannot be stored, distributed or read.

The conventions for Climate and Forecast metadata (http://cfconventions.org/) are designed to promote the processing and sharing of files created with the Network Common Data Form (NetCDF). The conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data.

Data Request

The Data Request is coordinated by the WGCM Infrastructure Panel. The requested variables and their output frequencies can be filtered following the scientific backdrop of the WCRP Grand Challenges and additional scientific themes to investigate.

Data Reference Syntax defines the directory structure and the available values for each directory. The DRS also fixed the filename pattern. A DRS is crucial for a clear and efficient management of their publication.

MODEL DATA

The concept of 'dataset'

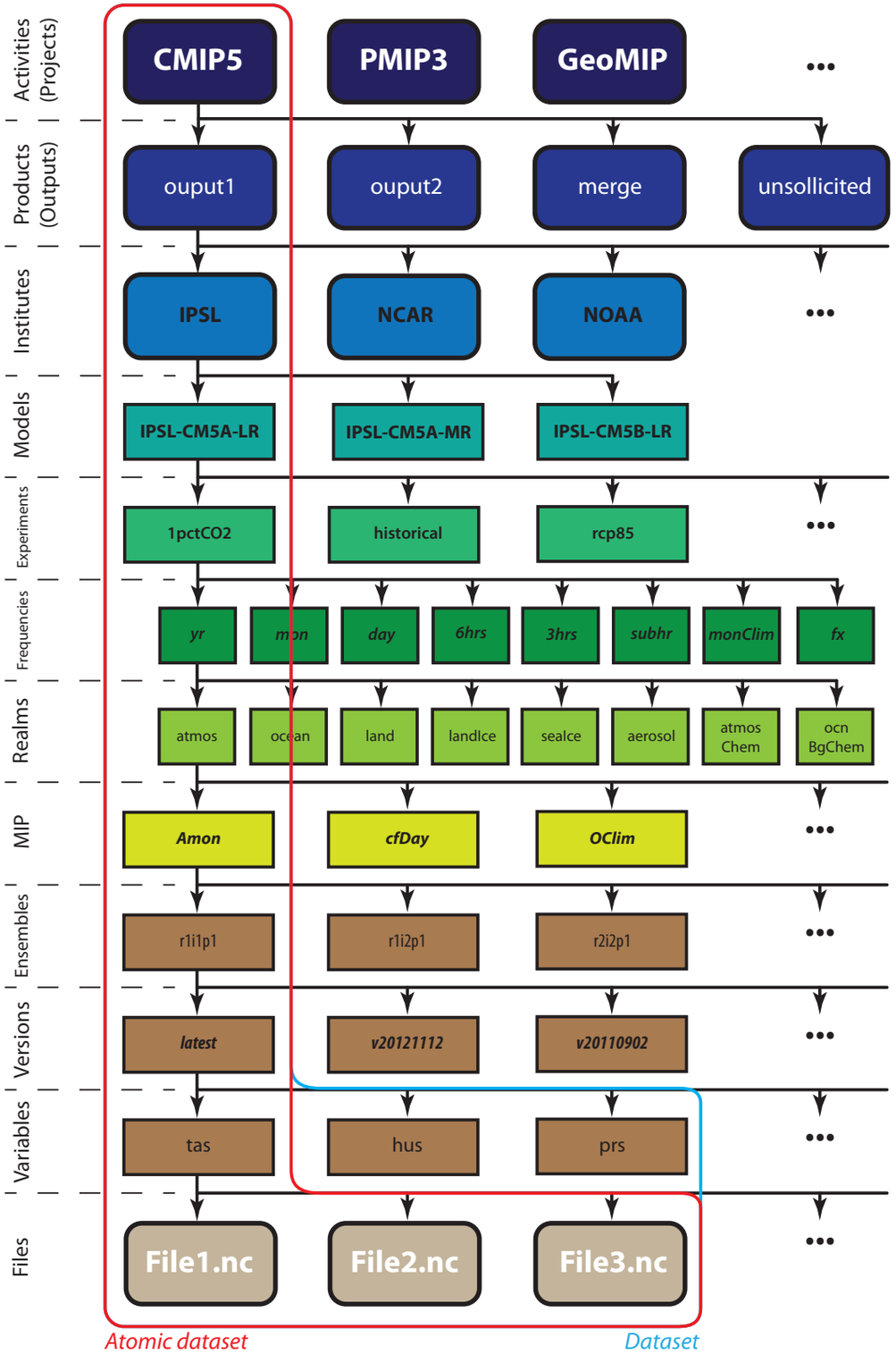
A ***'dataset'*** (as defined by ***publication***) is **ONE version** of a data set resulting from **a single simulation** (i.e., **characterised by a unique option of each attribue of the DRS before the version** such as the institute, the model, the domain, the experiment, the frequency, the ensemble, etc.).

Examples:

CMIP5 dataset: output1/IPSL/IPSL-CM5A-LR/1pctCO2/mon/atmos/Amon/r1i1p1/v20110427

CORDEX dataset: output/EUR-11/IPSL-INERIS/IPSL-IPSL-CM5A-MR/historical/r1i1p1/IPSL-INERIS-WRF331F/v1/mon/tas/v20140301

A dataset is the finest granularity for publication.

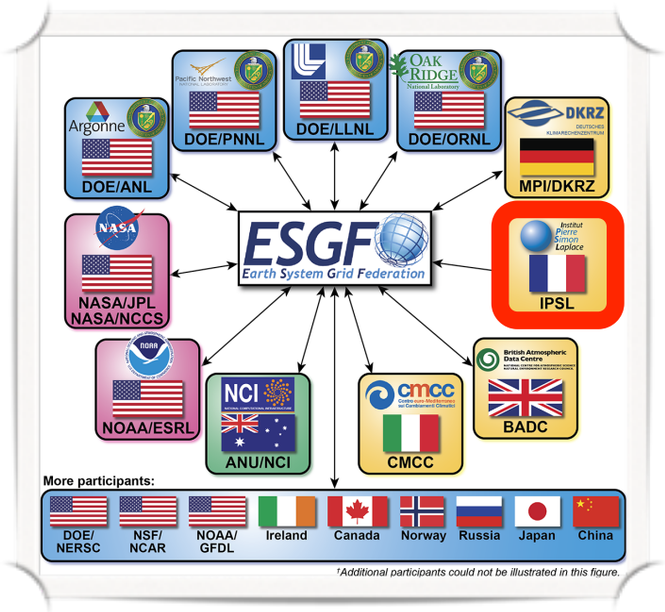


MODEL DATA

ESGF integration and international broadcasting

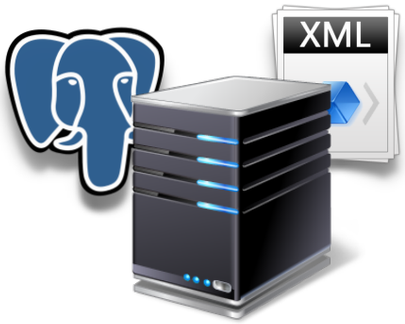
A **decentralised** and **federated** nodes network with international collaboration. Each ESGF Node is composed of services and applications that collectively enable metadata discovery, data access, and user management.

- ✓ Operational since 2011
- ✓ Hundreds of users per month
- ✓ Hundreds of TB per month
- ✓ About 10,000 registered users



IPSL climate models get involved in several projects

IPSL-CM results



ESGF datanode

- Metadata catalogs,
- Access protocols to data endpoints.

Scientific community

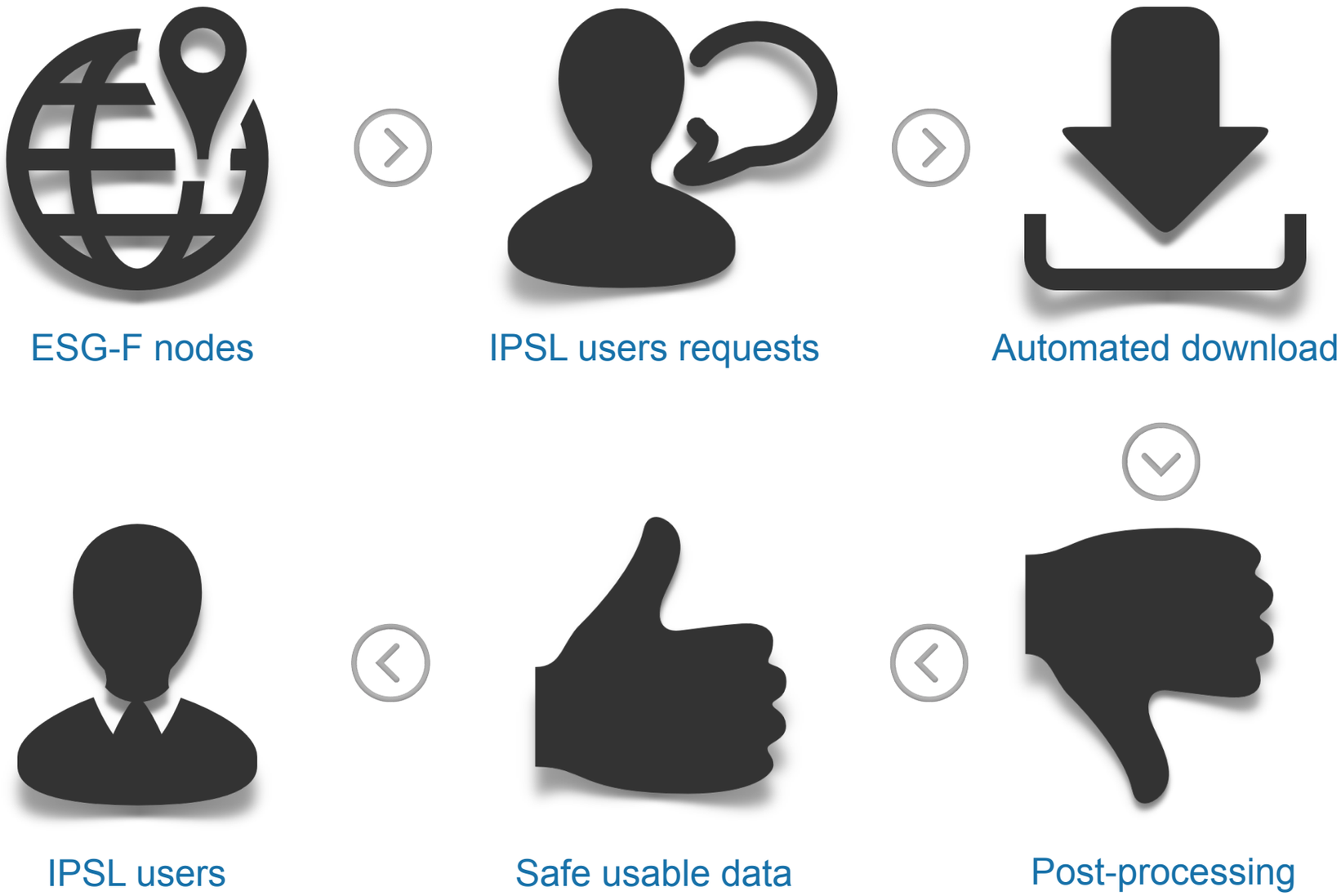
ESGF indexnode

- Indexes and broadcasts available data from all datanodes,
- Front-end with search UI,
- Links with models documentation,
- Redirects users requests to the appropriate datanode.

MODEL DATA

How ask for new data?

The IPSL mirrors at the request of its researchers a relatively large subset of ESG-F data (currently 420To) on its local filesystem. A post-processing pipeline is available to **analyse** climate simulations in the **best conditions**.



MODEL DATA

Edit a SYNDA template

```
# Edit a template
> vi my_template.txt
```

```
# My SYNDA template
project=CMIP5
experiments=historical amip
models=IPSL-CM5A-LR CNRM-CM5
ensembles=all
variables[atmos][3hr]=cltc tas
variables[land][fx]=sftgif
variables[seaIce][mon]=sic evap
```

Send your request to:

Guillaume Levavasseur: glipsl@ipsl.fr

MODEL DATA

Where on IPSL mesocentre?

```
# Location on CICLAD filesystem
> ls -l /prodigfs/project/
total 64
drwxr-sr-x  7 tnoel    ciclad 4096 Jul 11  2016 BCCORDEX
drwxrwsr-x  7 brocksce carbon 4096 Feb 24 12:25 CARBON
drwxrwsr-x  6 sdipsl   cmip5  4096 Mar 30 15:35 CLIM4ENERGY
drwxrwsr-x  6 jllod    cmip5  4096 May  5  2012 CMIP3
drwxr-xr-x  4 sdipsl   cmip5  4096 Jan 25 07:47 CMIP5
drwxr-xr-x  3 sdipsl   cmip5  4096 Feb 27 14:23 CMIP5-Adjust
drwxr-xr-x  4 sdipsl   cmip5  4096 Feb 27 14:27 CORDEX
drwxr-xr-x  3 sdipsl   cmip5  4096 Feb 27 14:23 CORDEX-Adjust
drwxr-xr-x  4 tnoel    ciclad 4096 Feb 24 12:20 DRIAS
drwxr-xr-x  3 sdipsl   cmip5  4096 Sep 12  2015 EMBRACE
drwxr-xr-x  7 brocksce ipsl   4096 Feb 19  2016 EUROCORDEX
drwxr-sr-x  3 glipsl   cmip5  4096 Apr 21 12:01 input4MIPs
drwxrwxr-x  4 brocksce ipsl   4096 Jul 25  2014 MAREMIP
drwxrwxr-x 23 brocksce ipsl   4096 Jan 31 18:21 OCMIP5
drwxr-xr-x  4 sdipsl   ciclad 4096 Nov  4  2014 ORCA025-LIM2-PISCES
drwxr-sr-x  3 sdipsl   cmip5  4096 Jul 31  2014 SPOOKIE

> ls -l /ccc/work/cont003
total 0
drwxr-x--- 1 apache ccc 0 Jan 26 15:45 cmip5
drwxr-xr-x 1 apache ccc 0 Mar  8 21:16 cordex
drwxr-xr-x 1 apache ccc 0 Oct 25  2013 dods
drwxr-xr-x 1 apache ccc 0 Nov 19  2014 esgf
drwxr-x--- 1 apache ccc 0 May 12 15:07 thredds
```

MODEL DATA

Documentation and errata

The potential to interpret, compare and reuse climate information results is strongly related to the quality of their description.

ES-DOC

Standards based eco-system in support of earth system modelling documentation creation, analysis & dissemination.

ES-DOC errata

Web-service to provide timely information about known issues and enable users to query about modifications and/or corrections applied to the data.