Modeling groups and Data Center Requirements. Session's Keynote.

Sébastien Denvil,

CNRS, Institut Pierre Simon Laplace (IPSL)

Outline

- Major constraints (requirements' DNA)
- Modeling center requirements/constraints
- Data center requirements/constraints
- Take home slide

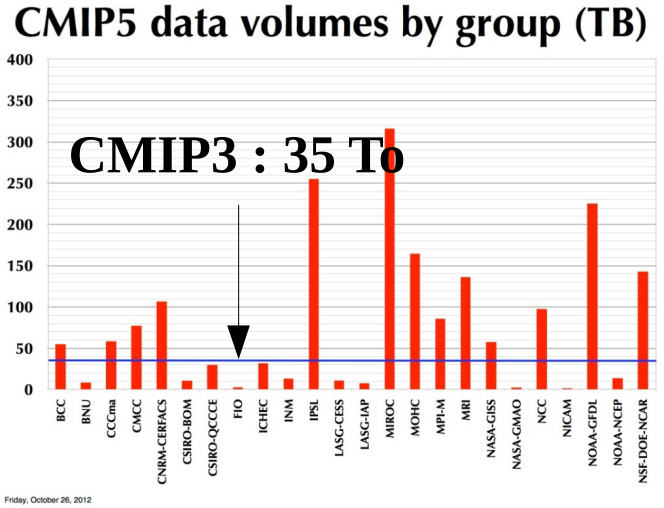
A one slide guide to CMIP5 from a data perspective

Fifth Climate Model Intercomparison Project (CMIP5)

World Climate Research Programme WCRP- WGCM Involves all the major climate modelling centres.

Original Timing:
o(2) PB of requested output from 20+ modelling centres finished early 2010!
Actual Timing?
Years late.

101 experiments
61 model variants
59,000 datasets!
4.5 million files
2 PB in global archive.
Unknown PB locally!



PCMDI-led,
Community developed (GO-ESSP)
s/w infrastructure for data delivery:
Earth System Grid Federation

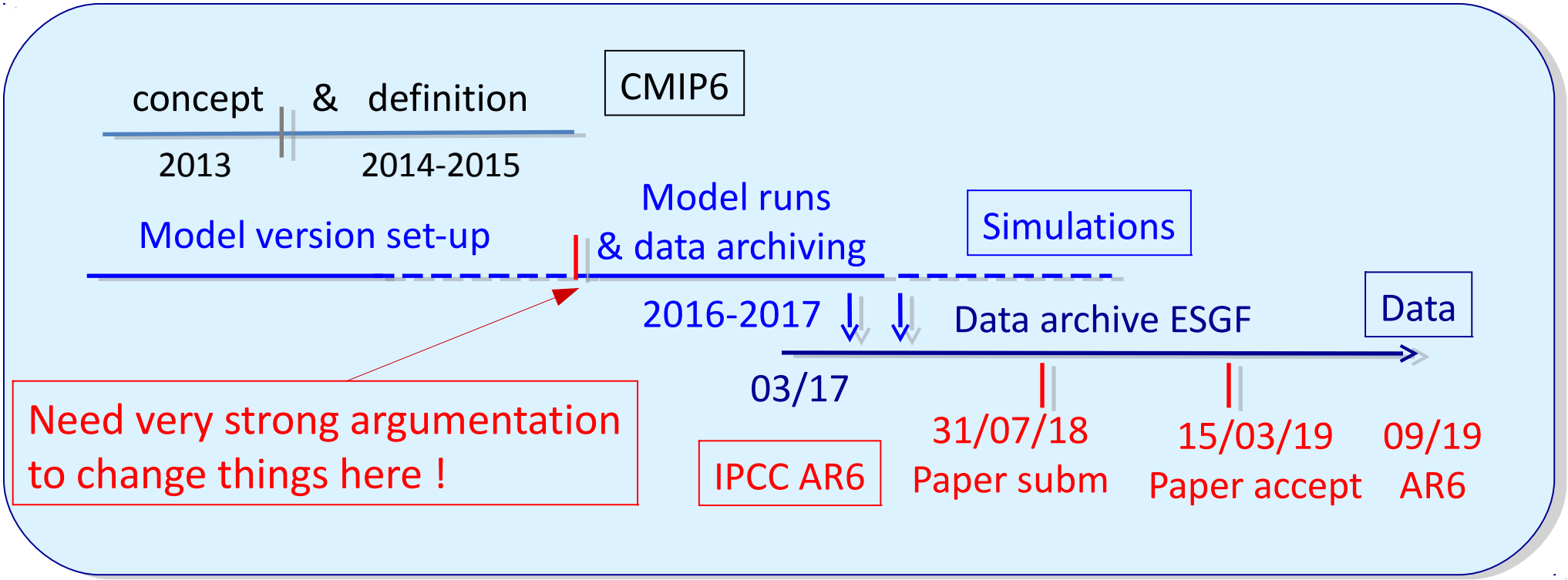
Sciences de l'environnement
Institut Pierre Simon Laplace

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Coupled Model Intercomparison Project – CMIP6

International community under **strong** pressure

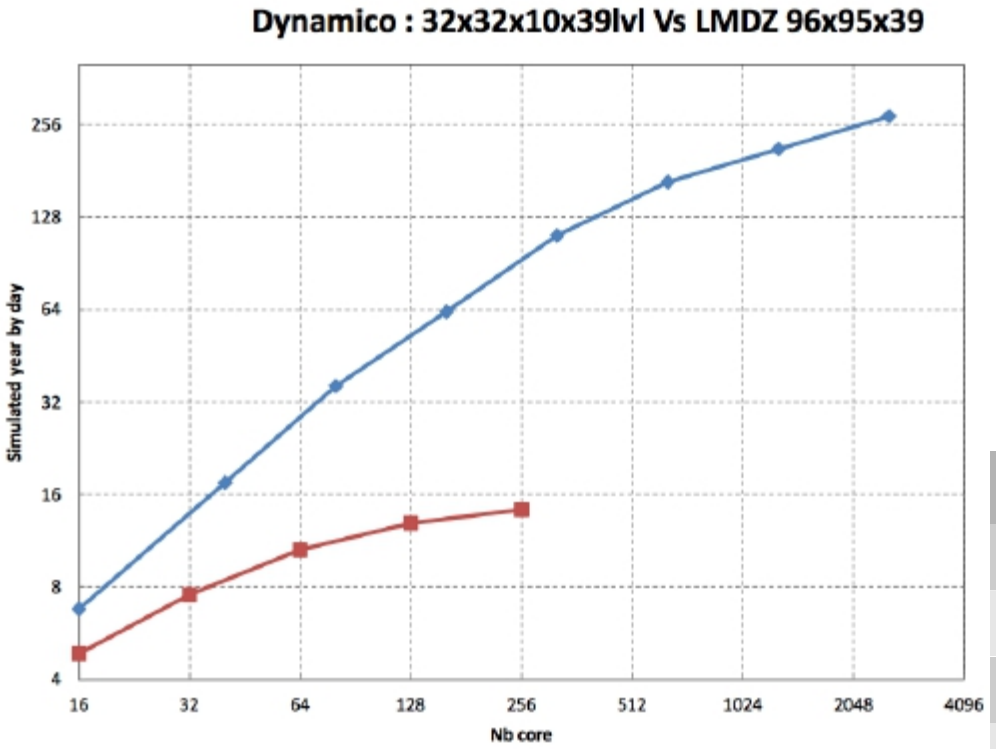
CMIP6/AR6 cycle

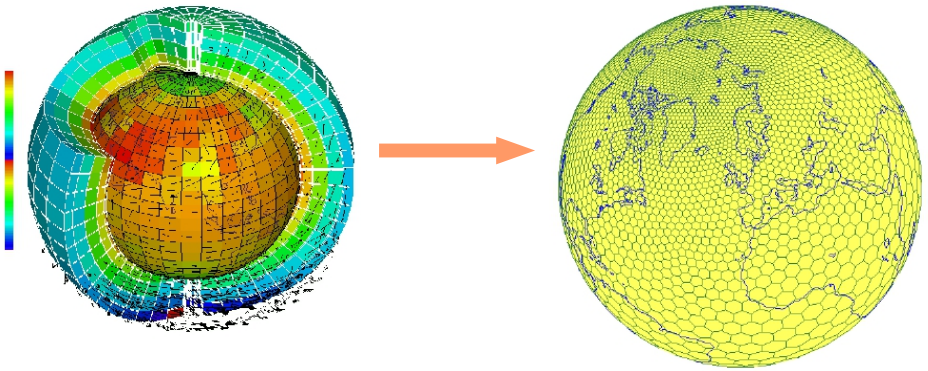


Outline

- Major constraints (requirements' DNA)
- Modeling center requirements/constraints
- Data center requirements/constraints
- Take home slide

Next generation model performance





degree	cores	Year/day	Mh/century	
3	320	**110**	0,0077	*Measured*
1	1280	**20**	0,15	
1	5120	**55**	0,22	
½	5120	*10*	1,2	*Extrapolated*
½	11520	*18*	1,5	
½	20480	*28*	1,8	
¼	20480	*5*	10	
¼	46080	*8*	14	

Sciences de l'environnement — Institut Pierre Simon Laplace

Courtesy Thomas Dubos LMD/IPSL and Yann Meurdesoif LSCE/IPSL

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

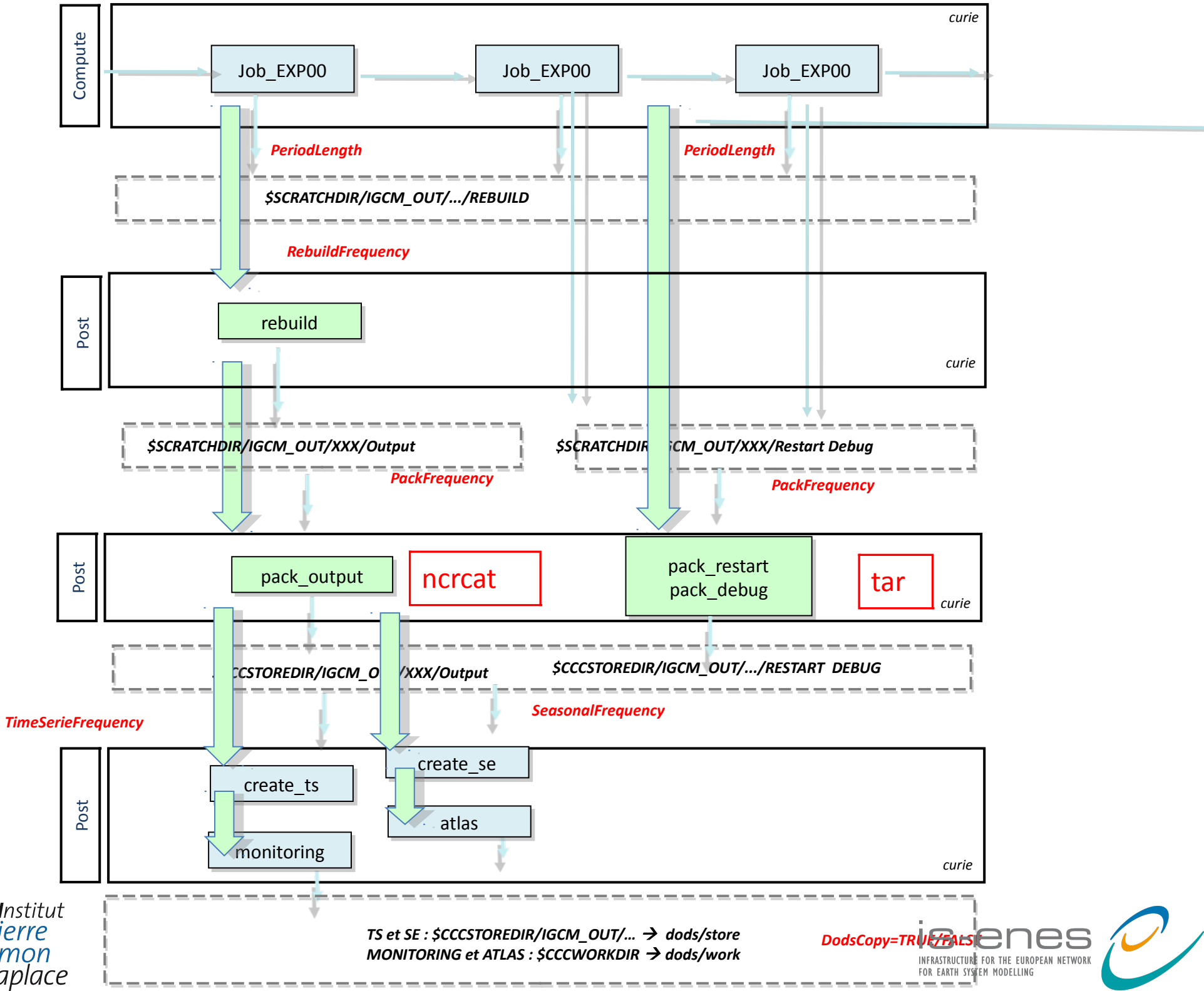
To keep in mind

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

But metadata alone won't get us there !

Computation useless if results cannot be stored/distributed/read

TGCC

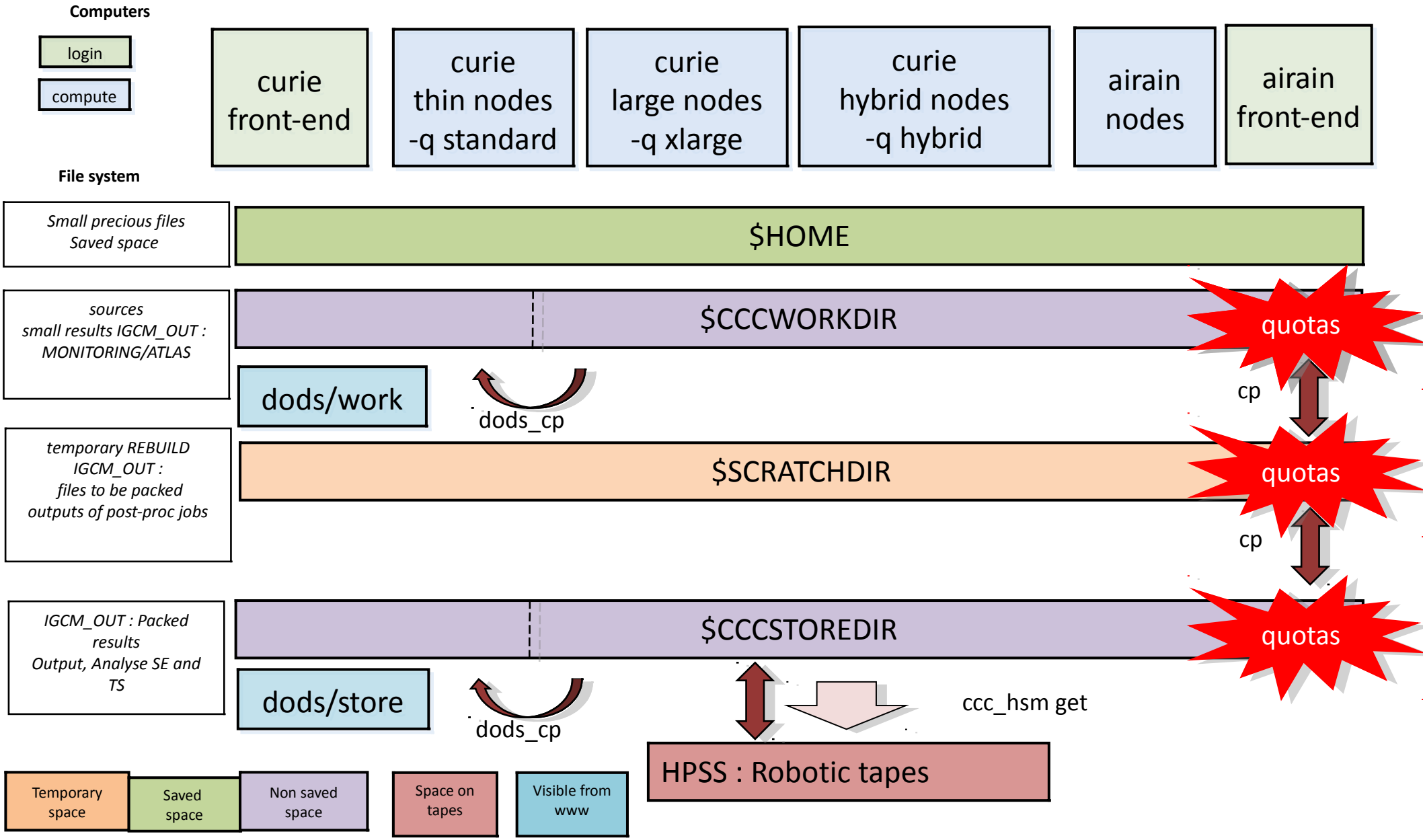


Sciences de l'environnement

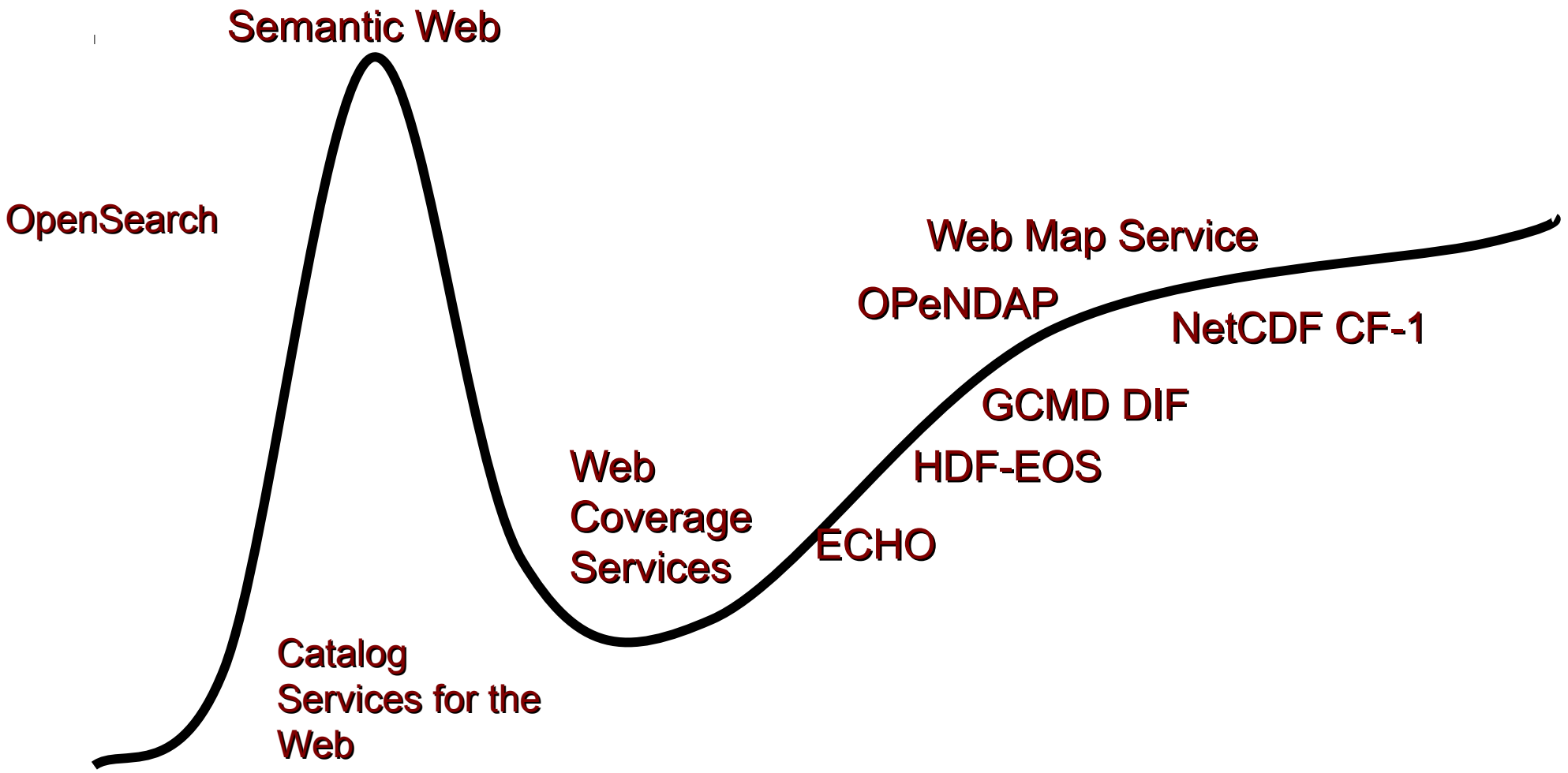
Institut Pierre Simon Laplace

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

TGCC computers and file system in a nutshell



Hype .vs. Reality



Sciences de l'environnement
Institut Pierre Simon Laplace

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Why is it good to log « all around » ?

THEN...

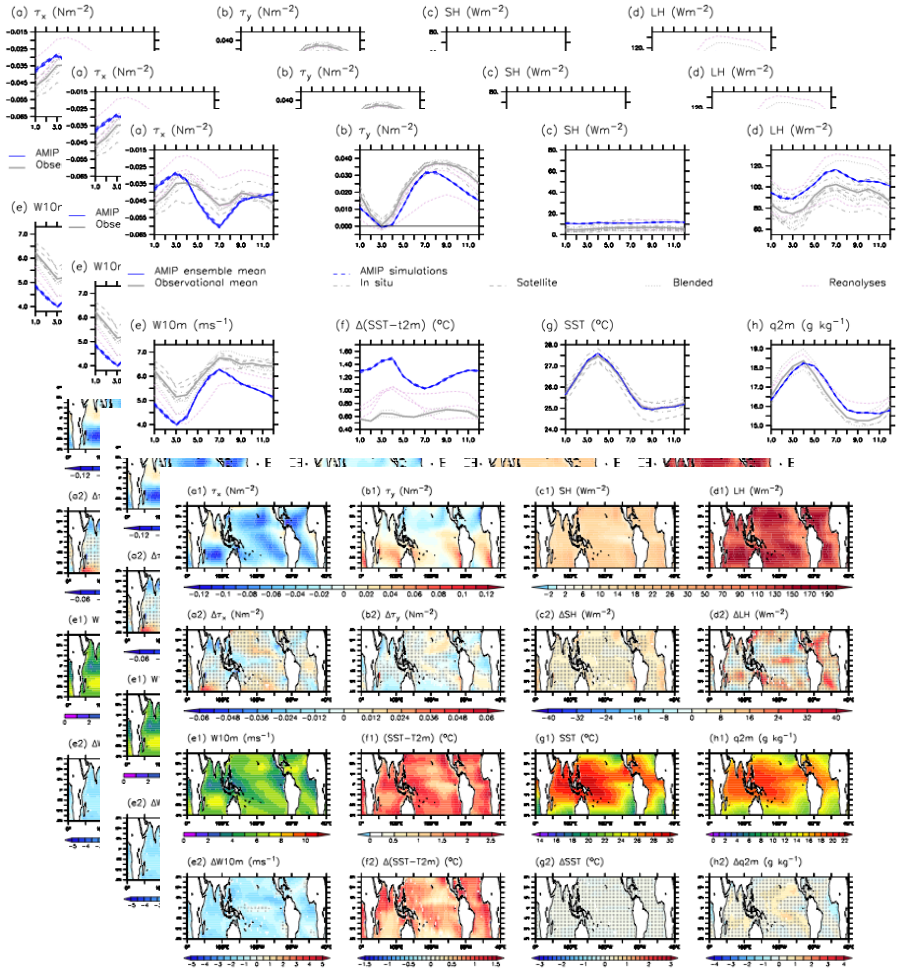
Log files

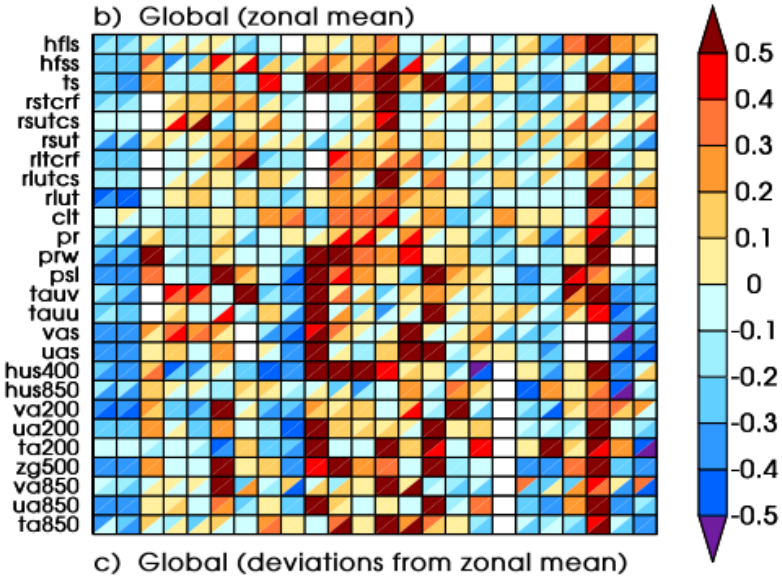
```
[Sun Dec 21 09:17:09 2008] [error]
[Sun Dec 21 10:04:53 2008] [error]
[Sun Dec 21 10:45:50 2008] [error]
[Sun Dec 21 11:14:09 2008] [error]
[Sun Dec 21 12:26:04 2008] [error]
[Sun Dec 21 13:41:05 2008] [error]
[Sun Dec 21 14:37:16 2008] [error]
[Sun Dec 21 15:19:39 2008] [error]
[Sun Dec 21 15:26:05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

NOW...

Operational intelligence

Synthesis is so important here also





Metrics Garden

Metrics Garden User Web Interface

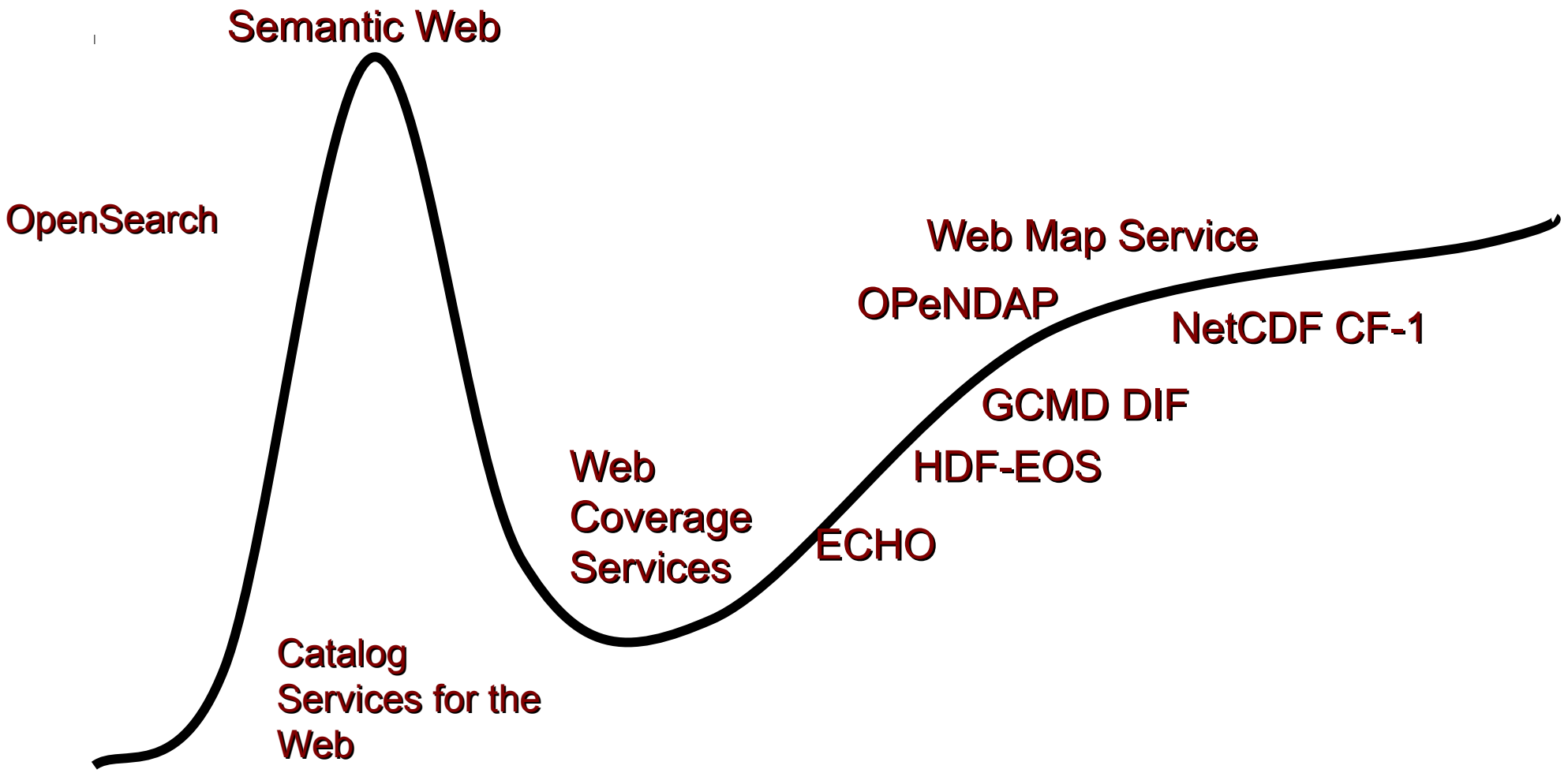
Test Glecker like metrics on CMIP5 version of IPSL models

			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.

Outline

- Major constraints (requirements' DNA)
- Modeling center requirements/constraints
- Data center requirements/constraints
- Take home slide

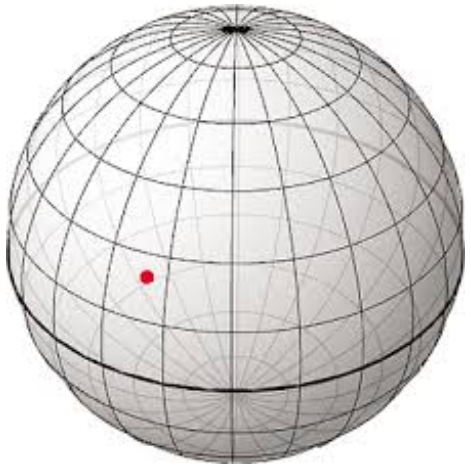
Hype .vs. Reality

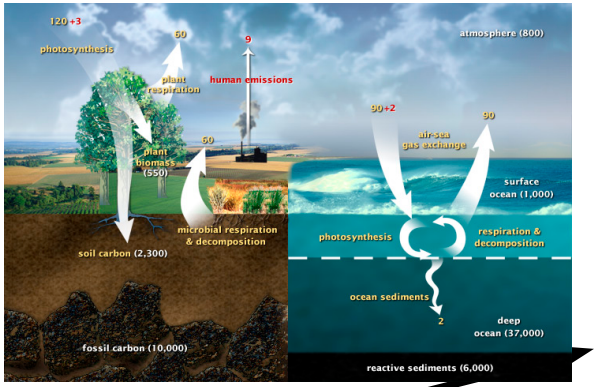


Sciences de l'environnement
Institut Pierre Simon Laplace

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING



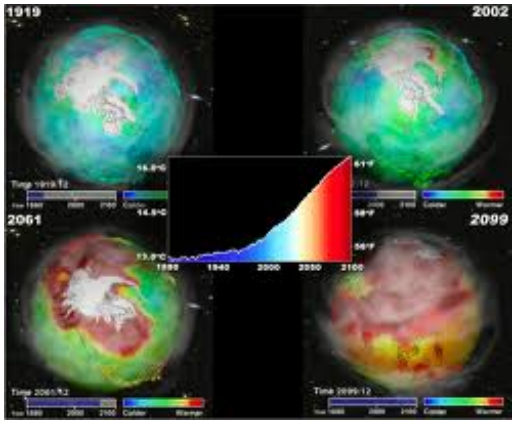


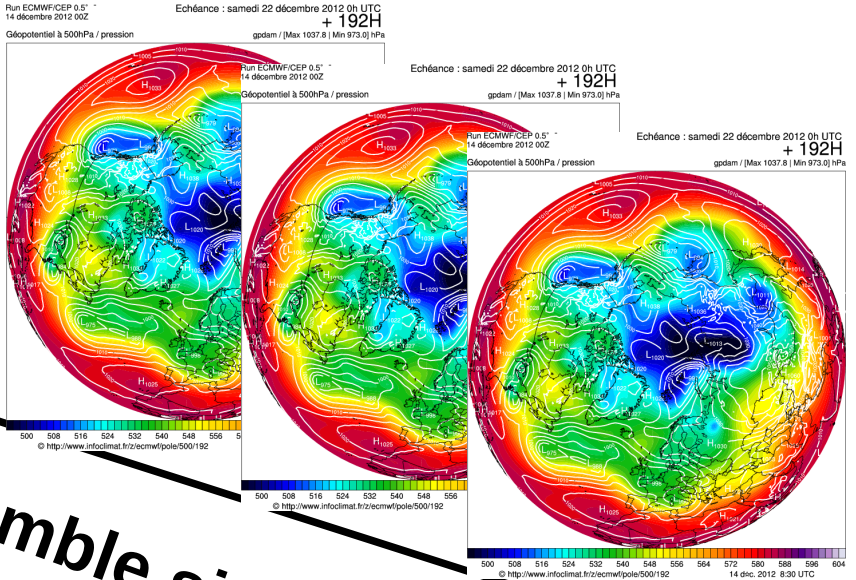
Earth Observations

Complexity

Resolution

Ehanced computing resources produce MORE DATA





Duration and ensemble size

My/Your Data Environment

At your home institution, you:

→ Have (some) control over your software environment

- Favourite packages, e.g. IDL
- Familiar Linux

We all like control!

→ Can buy/arrange more storage / compute on varying time-scales … can optimise ...

We all like the (illusion?) that we can scale our resources as necessary.

→ Are responsible for deleting / preserving your own data

We all lose/destroy/duplicate data.

→ Are likely to be duplicating data others have already downloaded *in your own institution* … let alone within a larger collaboration.

Most of us do our HPC remotely. Some of us do our analysis remotely. Why not more of us?

High Performance Data (HPD) Analysis Environment

Mutualized

Jointly *delivered* by
→IPSL laboratories.
Joint *users* (initially):
→IPSL community
Joint *users* (target):
→French Academic community

Analysis capabilities

Environmental Data
Compute Service
Web Service Provision for :
→Climate Science
→Earth Observation
→Environmental studies

Access services to ESGF System

Synchro-data find, download, and keep up to date the data users need.

CMIP5, CORDEX
Obs4MIPs
And *MIP
...

Big DATA Platform

Collaboration Environment
→ Access to Curated Archive.
→ Large shared "Group Work Spaces"
→ climate analysis enabled system
→ + 1 PB of high performance disk coupled to hundreds of cores configured for analysis

Sciences de l'environnement
Institut Pierre Simon Laplace

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Outline

- Major constraints (requirements' DNA)
- Modeling center requirements/constraints
- Data center requirements/constraints
- Take home slide

Take home slide

- Be ready for CMIP6 to **streamline** "production phase"!
- Be in a position to **make good decision** from torrent of data. Turn data into **information**. We need **operational** intelligence.
- **Roadmap, timeline, minute, milestone, deliverable, responsabilities ... publicly available.**
- The High Performance Data (HPD) Analysis Environment approach
  - We will need **not only** to **move** computation to data, **but aggregate** our data collections

Thank you for your attention

sebastien.denvil@ipsl.jussieu.fr