

Données modèles sur le mésocentre IPSL

Guillaume LEVAVASSEUR
Sébastien DENVIL
Atef BENNASSER



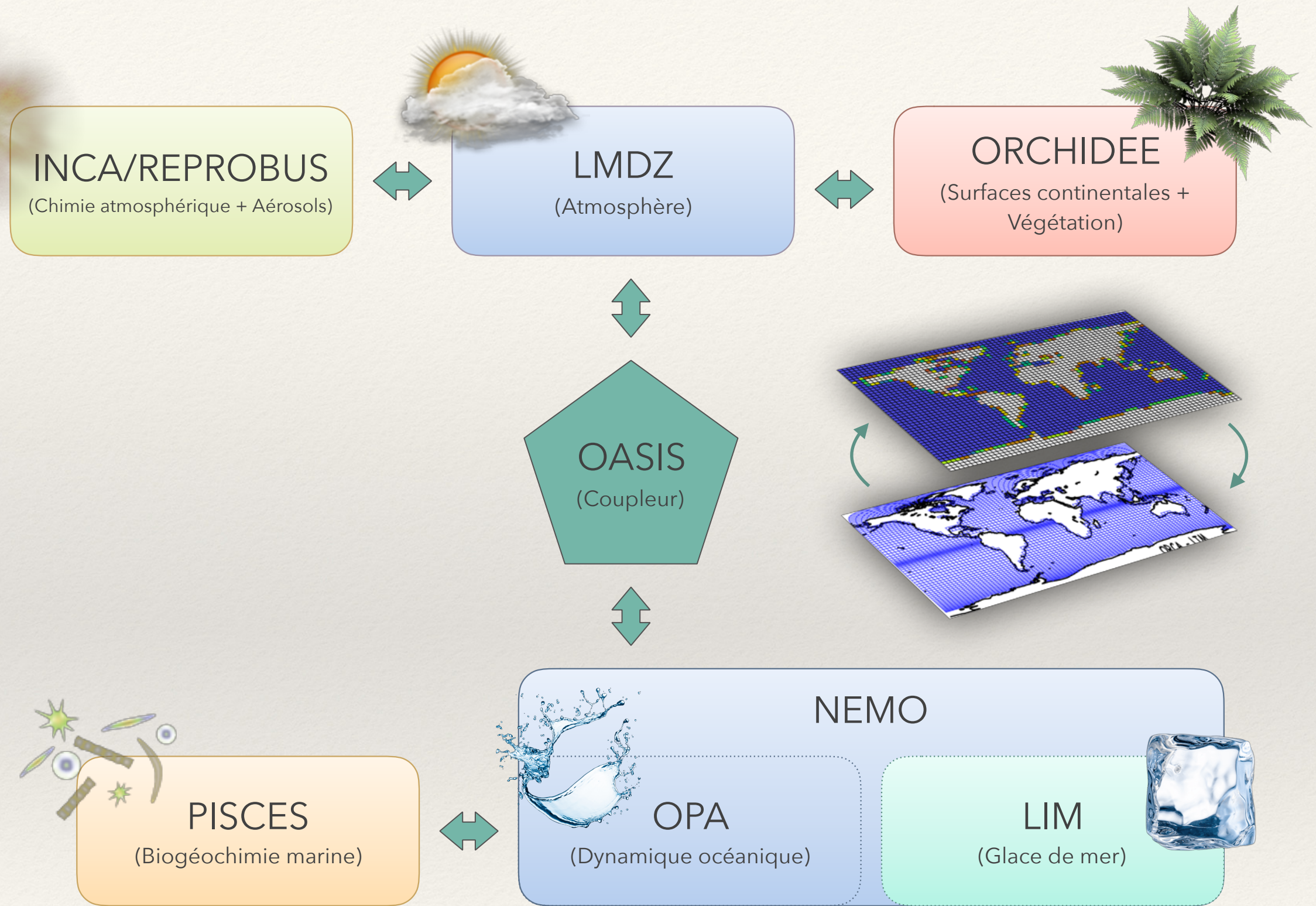
Institut
**Pierre
Simon**
Laplace

19 Novembre 2018

Plan:

- Introduction
- Workflow
- Données globales et régionales
- Standardisation et organisation
- Une aiguille dans une botte de foin ?
- J'en veux plus !

Un modèle, plusieurs composantes



Workflow



Données globale, régionales, observations reformatter et forçages

Coupled Model Intercomparison Project (CMIP)

- Plusieurs orientations scientifiques,
- Plusieurs processus à étudier,
- Plusieurs communautés (interconnectées)
- Plusieurs projets associés

ISIMIP	Inter-Sectoral Impact Model Intercomparison Project
GeoMIP	Geoengineering Model Intercomparison Project
PMIP3	Paleoclimate Modelling Intercomparison Project
TAMIP	Transpose-Atmospheric Model Intercomparison Project
LUCID	Land-Use and Climate, IDentification of robust impacts

Coordinated Regional Climate Downscaling Experiment (CORDEX)

- Focus sur la variabilité climatique régionale ou locale,
- Plusieurs domaines géographiques,
- Plusieurs modèles régionaux,
- Plusieurs techniques de correction de biais.

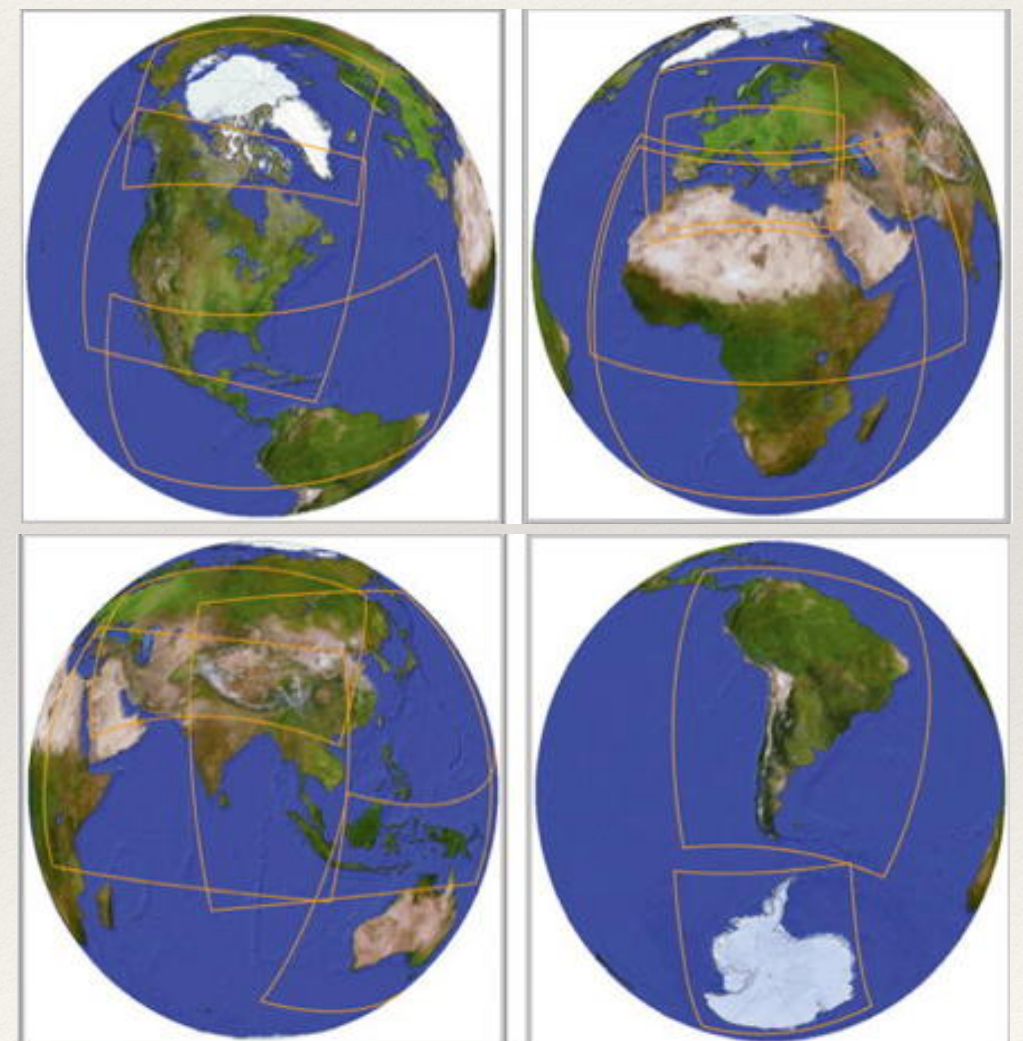
Observations for Model Intercomparison Project (obs4MIP)

- Collection limitée d'ensembles de données d'observation bien établis et documentés,
- Organisés selon les exigences CMIP

Input Datasets for Model Intercomparison Project (input4MIP)

- Conditions aux limites et forçages pour CMIP6
- Données centralisées et distribuées par le LLNL (US)

Domaines géographiques CORDEX



Climate "big" data

Archive CMIP3 : 24 modèles + 12 expériences = **39To** (82 340 fichiers)

Archive CMIP5 : 63 modèles + 101 expériences = **1,8Po** (4,3 millions de fichiers)



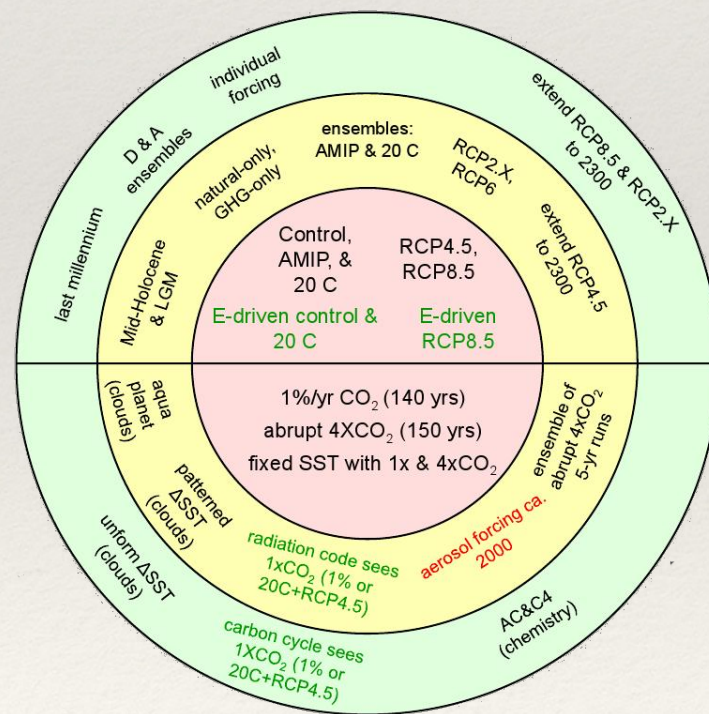
Extrapolation à CMIP6:

- Protocole plus complexe de 287 expériences
- 101 modèles aux résolutions spatiales plus fines
- Des ensembles plus larges



36 Po

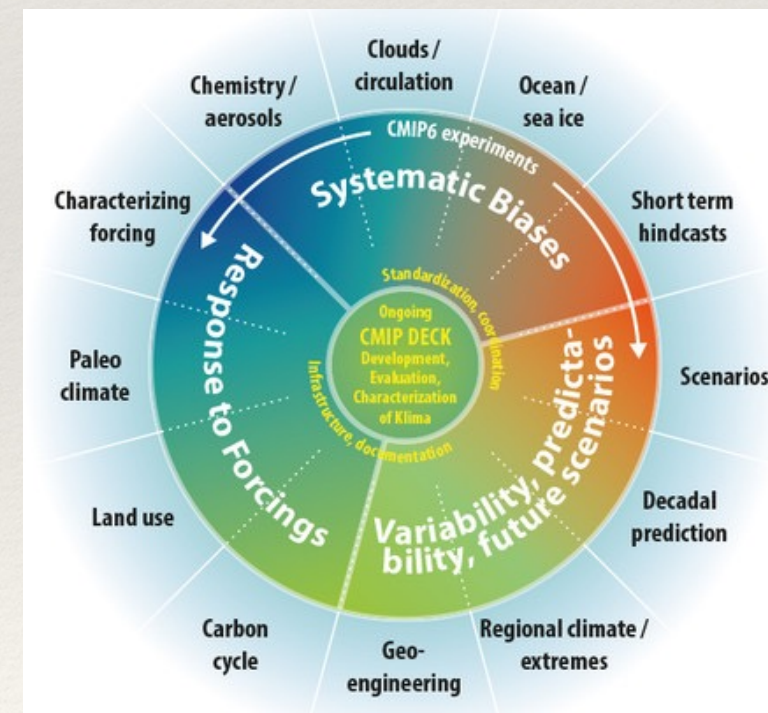
(86 millions de fichiers)



De la phase 5...

90 Po

(215 millions de fichiers)



... à la phase 6

Climate "big" data



**Vous
êtes ici**

Standardisation et organisation des données

La convention Climate and Forecast (CF)

Ce sont les règles qui régissent la descriptions des données et métadonnées de la communauté "climat". Elles sont conçues pour promouvoir le traitement et le partage des fichiers au format netCDF. Cette convention définit les métadonnées fournissant une description auto-suffisante de la donnée (e.g., propriétés spatiales et temporelles des données).



La Data Request (DR)

C'est un filtre "multidimensionnel" capable de lister les variables requises pour aborder les différentes thématiques scientifiques d'un projet. Les variables fournies par une simulation d'un modèle climatique dépendra donc de l'expérience, des fréquences, des composantes des modèles, des membres, etc.



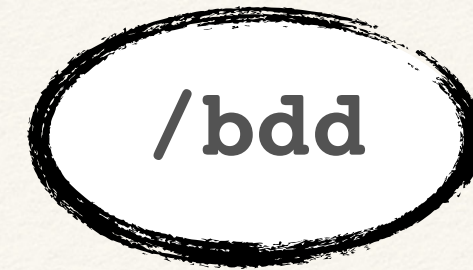
Le plan de gestion des données (ou "Data Reference Syntax" - DRS)

La DRS définit la structure de l'arborescence et le nommage des fichiers en stipulant toutes les valeurs possibles pour chaque paramètre ou "facette". Un DRS est crucial pour une gestion et analyse efficace des données.



Une aiguille dans une botte de foin ?

Où que vous soyez sur le mésocentre IPSL, une seule destination à retenir



```
# La base...
$> ls /bdd/CMIP3
$> ls /bdd/CMIP5
$> ls /bdd/CMIP6
$> ls /bdd/obs4MIPs
$> ls /bdd/input4MIPs

# WARNING : coming soon !
$> ls /bdd/CORDEX
# Pour patienter
$> ls /prodigfs/project/CORDEX

# Pour les plus téméraires...
$> ls /prodigfs/project/CMIP5-Adjust
$> ls /prodigfs/project/CORDEX-Adjust
```


Le cas CMIP6

Arborescence :

Pattern:

```

root
├── mip_era
│   └── activity
│       ├── institute
│           ├── source_id
│               ├── experiment
│                   ├── ensemble
│                       ├── cmor_table
│                           ├── variable
│                               ├── grid_label
│                                   └── version
└── /bdd
    ├── CMIP6
    │   ├── CMIP
    │   │   ├── CNRM-CERFACS
    │   │   │   ├── CNRM-CM6-1
    │   │   │   │   ├── historical
    │   │   │   │   │   ├── r1i1p1f2
    │   │   │   │   │   │   ├── 3hr
    │   │   │   │   │   │   │   ├── huss
    │   │   │   │   │   │   │   │   ├── gr
    │   │   │   │   │   │   │   │   └── v20171115
    │   │   │   │   │   │   │   └── huss
    │   │   │   │   │   │   │   └── gr
    │   │   │   │   │   │   └── huss
    │   │   │   │   │   └── huss
    │   │   │   │   └── huss
    │   │   │   └── huss
    │   │   └── huss
    │   └── huss
    └── huss

```

```

00 000000
0000110101000100
01110111011001100000
010011111100100 111010100000 0
001 00111110101010110001001000 0 000
00010100001000001010011011010 0000 0010
00111011010 1010101001001010 0110100010 0
00 01000 010010 0100 1010110 111000 0101000
00 00110000101100 011 01011 11000 0001100 000
0000110 001011100 1 1011111100 000110 000000
0001 1 00001110001 11011 101000011 000 010 0
0001 0000111 10 1111 1101100 00010 100100 0
00 000011000100 001 10 1110 00 010 10 00 0000
011 00110 1 010 11 11010101000 0010
01100 11 11110 11010 10 10100
0000000 100 1111 1110 000110 001 00
000 100000000111 1101011000
011110010 110
111110
11010
1101
1000
11010
01101110110
011100000111001101110000100

```

Noms de fichiers :

Pattern:

<variable>_<cmor_table>_<source_id>_<experiment>_<ensemble>_<grid_label>[_<period_start>-<period_end>].nc

Exemple:

huss_3hr_CNRM-CM6-1_historical_r1i1p1f2_gr_185001010300-185002010000.nc

Versionnement :

```

# Quel est la dernière version disponible ?
$> ls -l /bdd/CMIP6/CMIP/CNRM-CERFACS/CNRM-CM6-1/1pctCO2/r1i1p1f2/Amon/pr/gr/
total 512
lrwxrwxrwx 1 444 9 12 nov. 16:54 latest -> v20180626/
drwxrwsr-x 2 444 0 12 oct. 11:44 v20180626/

```


J'en veux plus !

À la demande de ses utilisateurs, le mésocentre IPSL réplique un sous-ensemble relativement important de données ESGF (actuellement plus de 500To hors CMIP6) sur ses systèmes de fichiers locaux. Pour certains projets, un pipeline de post-traitement a été mis en place pour appliquer un contrôle qualité minimum sur les fichiers téléchargés et permettre ainsi les meilleures conditions d'analyses (e.g., suppression des overlaps, vérification de l'axe des temps).



J'en veux plus !

Comment formuler sa requête ?

1. Explorer la base de données existantes dans **/bdd**,
2. Si vous ne trouvez pas votre bonheur, explorer le catalogue ESGF : <https://esgf-node.ipsl.upmc.fr>
 - A. Votre bonheur ne se trouve pas si ESGF..... **GAME OVER**.
 - B. Votre bonheur existe sur ESGF : créer un fichier **.txt** avec la syntaxe suivante

```
$> vi my_template.txt
```

```
# My request
project=CMIP5
experiments=historical amip
models=IPSL-CM5A-LR CNRM-CM5
ensembles=all
variables[atmos][3hr]=clt tas
variables[seaIce][mon]=sic evap
```

3. Envoyer votre fichier à Guillaume Levavasseur (glipsl@ipsl.fr)
4. Votre demande est examinée (espace requis vs. espace libre)
 - A. Nous validons votre demande,
 - B. Nous revoyons ensemble vos attentes pour satisfaire aux espaces de stockage du mésocentre.
5. Nous vous notifions dès que vos données sont disponibles.

Merci !

► *Administrateurs données modèles :*

- Guillaume LEVAVASSEUR (SU/IPSL - glipsl@ipsl.fr)
- Sébastien DENVIL (CNRS/IPSL - sdipsl@ipsl.fr)

► *Développeur SYNDA :*

- Atef BEN NASSER (CNRS/IPSL - abennasser@ipsl.fr)

► *Administrateurs mésocentre :*

- Philippe WEILL (LATMOS/IPSL)
- Karim RAMAGE (CNRS/IPSL)

Guillaume LEVAVASSEUR



Sébastien DENVIL



Atef BEN NASSER

