

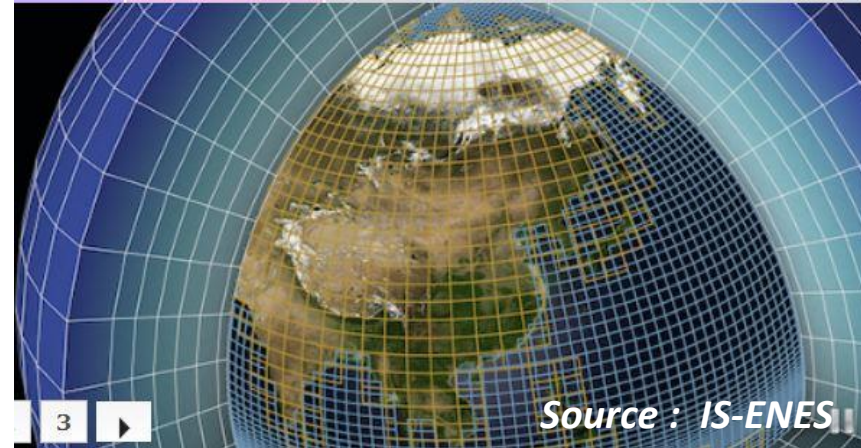
Quand la climatologie utilise les services ESGF et EUDAT pour la gestion de ses données.

Marie-Pierre Moine

Journée utilisateurs du CINES – 9 juin 2015

1. Introduction : présentation de la donnée « climat »
2. La climatologie « data provider »
3. La climatologie (*et les impacts*) « data user »
4. Conclusions

- **sorties de modèles** de climat global ou régional
- **données maillées** (*grille régulière, gaussienne, curvilinéaire, non-structurée, etc...*)
- coordonnées (latitude, longitude) + axe temporel
- 2 grilles différentes : atmosphère / océan
- grilles modèles = 3D
- 1 variable = 1 champ 2D (*sur 1 niveau du modèle*) ou 3D
- des centaines de variables pronostiques ou diagnostiques à des fréquences variées (*mensuelle, journalière, tri-horaire, horaire, pas de temps*)



La donnée climat : formats de fichier & métadonnées

CIM & CV
(METAFOR EU
project, an
ISENES
initiative)

➤ **Chaîne et contexte de production :**

- Modèle (dynamique, physique, paramétrisations, ...)
- Expérience et simulations (méthode d'initialisation, jeux de données utilisée forçages externes, etc..)

CMOR
Climate Model Output
Rewriter
(NOAA/BADC/PCMDI)

➤ **Attributs globaux additionnels** (centre producteur, modèle, expérience, ...)

➤ **Nomenclature & arborescence DRS** (Data Reference Syntax)

CF-NetCDF
Climate Forecasting conventions
for NetCDF (PCMDI/NCAR-UCAR)

➤ **Variables standardisées :**

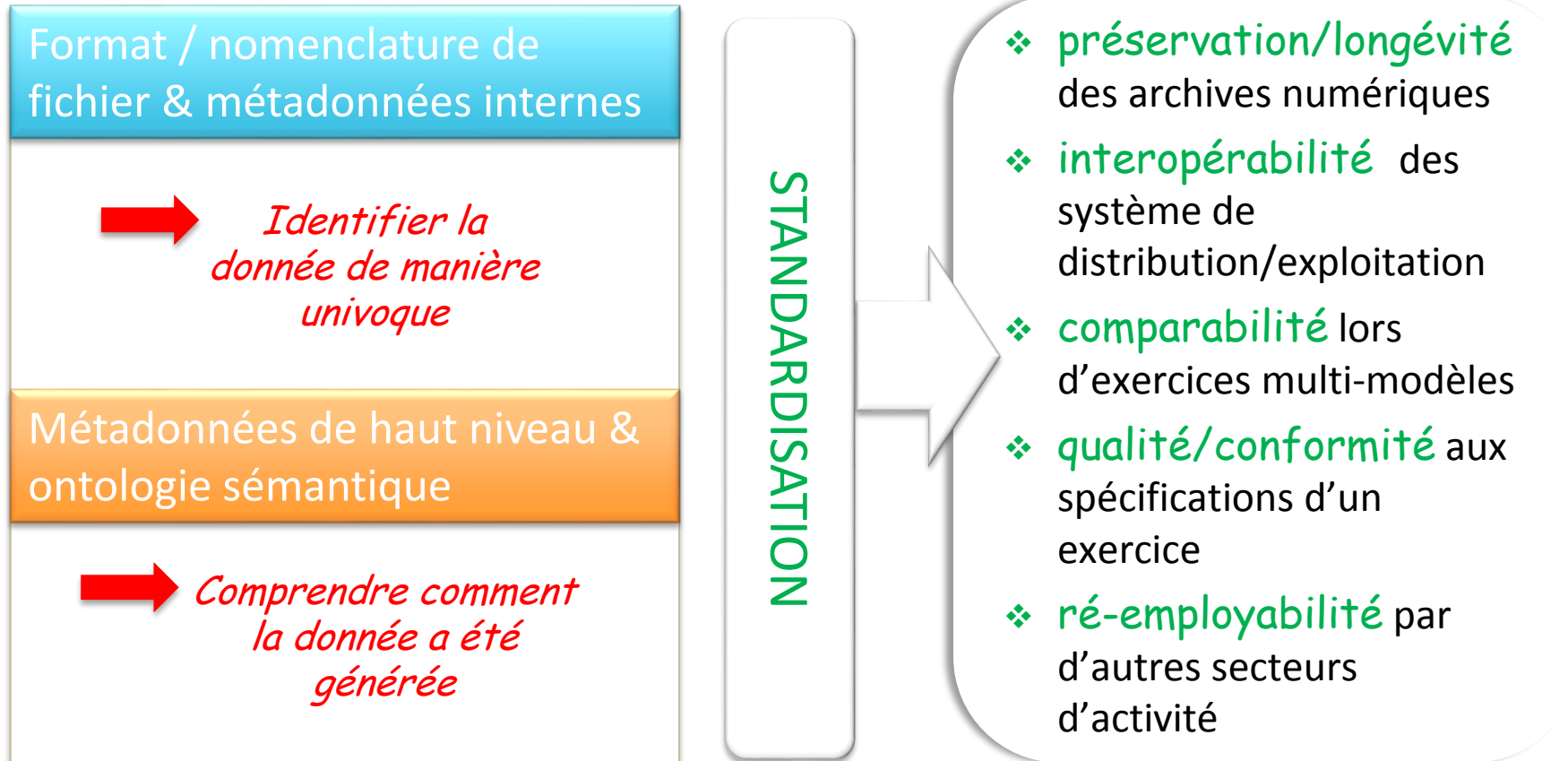
- noms
- unités

NetCDF
Network Common Data Format
(unidata, UCAR)

➤ **Format auto-descriptif :**

- Dimensions (2D ou 3D + temps)
- Coordonnées géographiques
- Champ(s) physique(s)

Les métadonnées, une préoccupation majeure ces dernières années





Earth System Grid Federation :
archivage & distribution des données



Earth System DOCumentation :
documentation des modèles et expériences (métadonnées)



Earth System COG :
environnement collaboratif; interface Web vers les services
ESGF et ES-DOC



**Infrastructure for the European Network of Earth System
Modelling :**

cadre d'initiative et de coordination de l'e-infrastructure
européenne en modélisation climatique (calcul, données,
métadonnées)



« data provider »

Les différent types de données

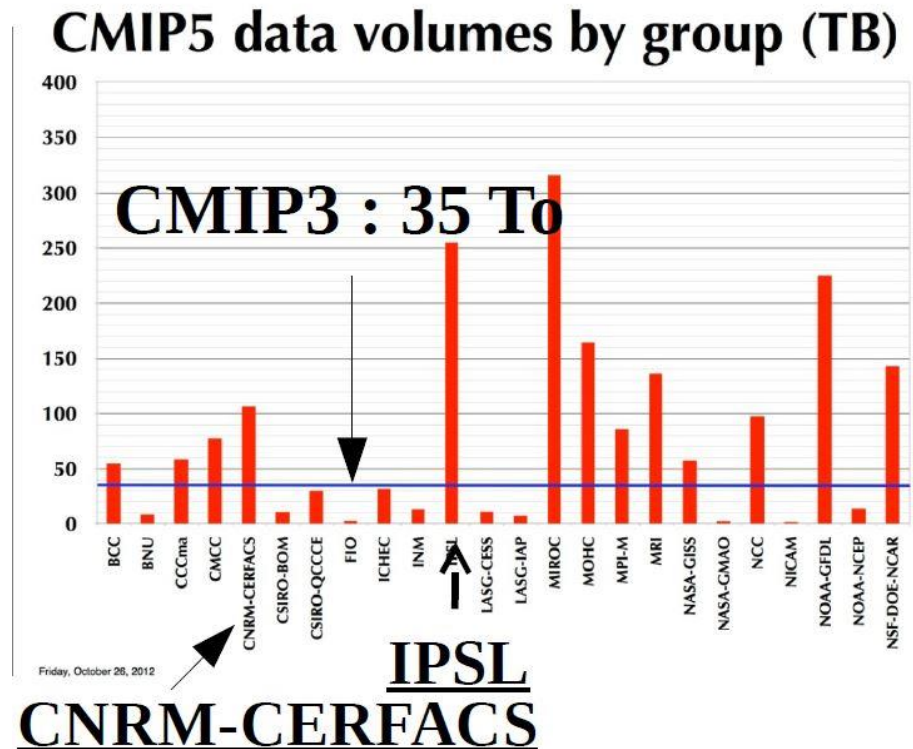
Type de données	Archivage	Sécurisation	Durée préservation	Niveau métadonnées (standard)	Distribution
Etude hors projet	local	pas de copie	1 -2 ans	minimal	non distribué
Projet ANR	local, chez les partenaires	pas de copie	2 -3 ans	minimal à renforcé	labos partenaires
Projet européen	EU data center / ESGF data nodes	dépôt central < base locale	5 ans... et plus	renforcé à élevé	labos partenaires ou public
Projet international	ESGF data node	réplicat(s) du "core" < base locale	ad-vitam (?)	élevé	public

« data provider » :
L'exemple d'un projet international (passé), CMIP5

CMIP5 (IPCC-AR5, production 2010-2012)

Coupled Model Intercomparison Project Phase 5 (WCRP-WGCM)

- **+20 groupes** de modélisation du climat à travers le monde
- **~60 configurations de modèles**
- **~100 types d'expérience**
- **59 000 datasets**
- **4.5 millions de fichiers**
- **~11 000 ans** de simulation
- **o(2) PB outputs imposés**
- **o(1) PB données répliquées**



(dont ~20 To exercice décennal CERFACS)

L'exemple d'un projet européen (en cours), SPECS



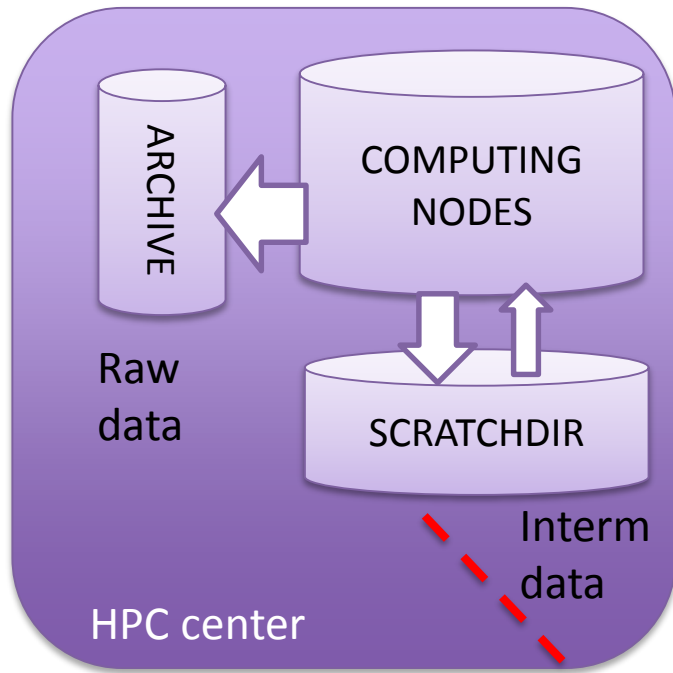
SPECS : « Seasonal-to-decadal climate Prediction for the Improvement of European Climate Services »

- 20 laboratoires européens (IC3, KNMI, MetOffice, ECMWF, CERFACS, IPSL, Meteo-France, ...)
- **200 To** de données prévues à la production – **80 To** à achever (compression netcdf4)
- **107 variables** à produire pour chaque simulation

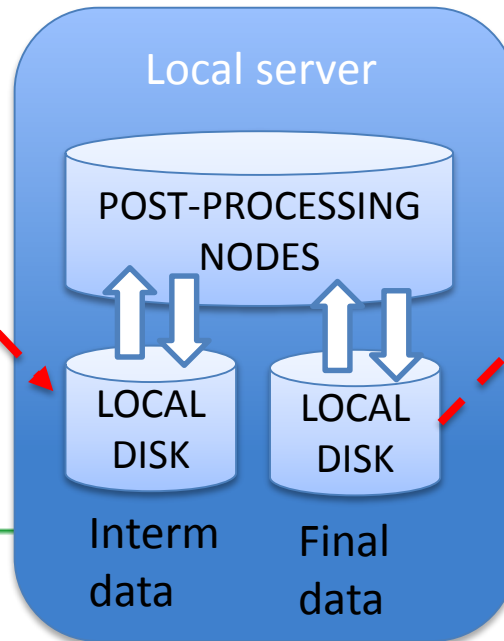
Contribution CERFACS à SPECS :

- Modèle couplé océan-atmosphère haute résolution ARPEGE-NEMO
- Super-calculateur Marenostrum (BSC, Barcelone) : IBM 48 000 cœurs, 1.1 Petaflops)
- **800 ans** de simulations
- 17.6 Mh de calcul consommées
- **30 To** produits – **16 To** à archiver (**+ 50 To** de données brutes - temporaire)
- **~85 000 fichiers**

« data provider »
cycle de vie : de la production à la publication ESGF

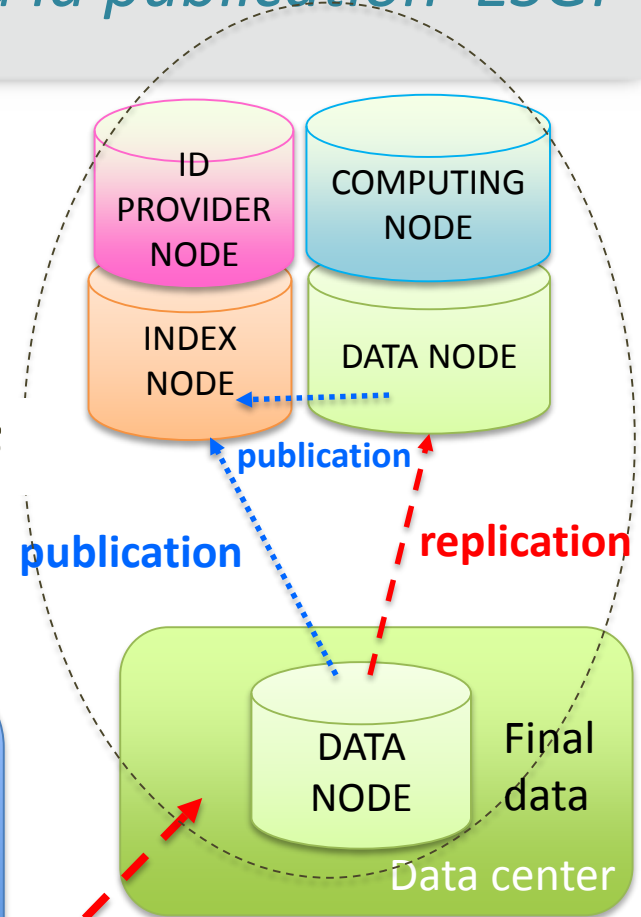


download



upload

ESGF



- Recherche par catégories / mots-clés (vocabulaire contrôlé)
- Sélection des jeux de données
- Script Wget auto généré
- Rapatriement (intégralité !) par exécution du Wget

(ou FTP pour un utilisateur averti)

ESGF Earth System Grid Federation

is-enes Institut Pierre Simon Laplace Sciences de l'environnement

Home Search Tools Login Help

Current Selections

- remove all
- (x) project:CMIP5
- (x) institute:CNRM-CERFACS
- (x) experiment family:Decadal
- (x) data node:esgf.cines.fr
- (x) variable:clt
- (x) experiment:decadal1975

Search Categories

- Project
- Institute
- Model
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name
- CMIP Table
- CF Standard Name
- Ensemble
- Domain
- Driving Model
- Downscaling realisation
- Data Node

Search

Examples: temperature, "surface temperature", climate AND project:CMIP5 AND variable:hus.
To download data: add datasets to your Data Cart, then click on Expand or wget.

☒ Search All Sites ☒ Show All Replicas ☒ Show All Versions

< 1 > displaying 1 to 10 of 10 search results

Display 10 datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Results Data Cart

project=CMIP5, model=CNRM-CM5, Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique (CNRM/CERFACS), experiment=10- or 30-year run initialized in year 1975, time frequency=mon, modeling realm=atmos, ensemble=r10i1p1, version=20150423
Data Node: esgf.cines.fr
Version: 20150423
Description: CNRM-CM5 model output prepared for CMIP5 10- or 30-year run initialized in year 1975
Further options: [Add To Cart](#) [Visualize and Analyze](#) [Model Documentation](#)

project=CMIP5, model=CNRM-CM5, Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique (CNRM/CERFACS), experiment=10- or 30-year run initialized in year 1975, time frequency=mon, modeling realm=atmos, ensemble=r1i1p1, version=20150423
Data Node: esgf.cines.fr
Version: 20150423
Description: CNRM-CM5 model output prepared for CMIP5 10- or 30-year run initialized in year 1975
Further options: [Add To Cart](#) [Visualize and Analyze](#) [Model Documentation](#)

project=CMIP5, model=CNRM-CM5, Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique (CNRM/CERFACS), experiment=10- or 30-year run initialized in year 1975, time frequency=mon, modeling realm=atmos, ensemble=r2i1p1, version=20150423
Data Node: esgf.cines.fr
Version: 20150423
Description: CNRM-CM5 model output prepared for CMIP5 10- or 30-year run initialized in year 1975
Further options: [Add To Cart](#) [Visualize and Analyze](#) [Model Documentation](#)

project=CMIP5, model=CNRM-CM5, Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique (CNRM/CERFACS), experiment=10- or 30-year run initialized in year 1975, time frequency=mon, modeling realm=atmos, ensemble=r2i1p1, version=20150423
Data Node: esgf.cines.fr
Version: 20150423
Description: CNRM-CM5 model output prepared for CMIP5 10- or 30-year run initialized in year 1975
Further options: [Add To Cart](#) [Visualize and Analyze](#) [Model Documentation](#)

Motivation : des volumes en inflation !

- augmentation de la résolution des modèles
- augmentation du nombre de variables haute fréquence (tri-horaire, horaire, ...)
- des dizaines d'années à analyser simultanément (30, 50, 100 ans)
- de plus en plus de « candidats » pour le multi-modèles
- au total, des To de données à manipuler... !

➡ Le téléchargement en local pour analyse n'est plus une solution viable.

Modèle CMIP5
CNRM-CERFACS

Atmos : 1 million pts
Ocean : 4.4 millions pts

X 25 !

Modèle CMIP6
CNRM-CERFACS

Atmos : 24 millions pts
Ocean : 110 millions pts

**10 ans de données
(1 variable) :**

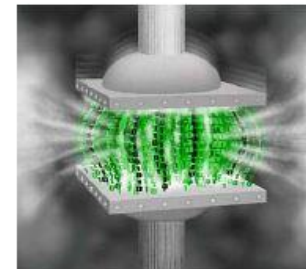
Atmos. 2D, 1m = 140 Mo
Atmos. 2D, 1d = 3.8 Go
Atmos. 2D, 6h = 15 Go
Ocean 2D, 1m = 820 Mo
Ocean 2D, 1d = 20 Go
Ocean 3D, 1m = 530 Go



« data user » : *nouveau besoins, nouveaux services (EUDAT, ISENES)*

Sélection & Réduction de données :

- « **data subsetting** » = extraction spatiale ou temporelle sur la région ou période d'intérêt
- « **regridding** » = interpolation à plus basse résolution
- « **coarsening/upscaling** » = traitement à haute résolution, résultat de l'analyse à résolution dégradée
- (« **data chunking** » = traitement opérés sur des sections de tableaux)



Data reduction...

Traitements près des données :

- Calcul statistiques simples (min/max, moyennes, écarts-types,...)
- Calcul d'indices climatiques (écart de température diurne, nbre de jours de sécheresse,)

Exemple du portail Climate4impact (IS-ENES)

- API (prototype) orientée vers la communauté impact & adaptation
- Accès aux données ESGF (CMIP5, CORDEX, obs4MIP,...)
- Sélection région
- Traitement à la volée (subsetting, regridding, climate indices,...)
- Résultat :
 - visualisable (plot)
 - téléchargeable (fichier)
 - accessible par URL openDAP pour d'autres séries de traitements

The screenshot shows the IS-ENES climate4impact portal. The header includes the logo 'is-enes' with the tagline 'Exploring climate model data' and navigation links for 'Home', 'Data discovery', 'Downscaling', 'Documentation', 'Help', 'About us', and 'Sign in'. Below the header, the page title is 'IS-ENES climate4impact portal'. The main content area welcomes users and provides information about the portal's purpose, data access, and documentation. It also lists recent events, such as a workshop on design of scientific portals. At the bottom, there is a grid of eight images representing different climate change impact and adaptation themes: Agriculture/Forestry, Energy, Health, Infrastructure/Urban, Marine/Coastal, Nature/Biodiversity, Tourism, and Water Management. A footer section mentions funding from the European Union's Seventh Framework Programme and includes a disclaimer.

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

Exploring climate model data

Home Data discovery Downscaling Documentation Help About us Sign in

IS-ENES climate4impact portal

Welcome to the IS-ENES climate4impact portal, oriented towards [climate](#) change impact modellers, impact and adaptation consultants, as well as other experts using climate change data.

Here you will find [access to data](#) and [quick looks](#) of global climate models (GCM) scenarios, as well as regional climate model (RCM) and downscaled higher resolution climate data. The portal provides data transformation tooling for tailoring data to your needs and [mapping & plotting capabilities](#).

[Guidance](#) on how to use climate scenarios, documentation on the climate system, [frequently asked questions](#) and examples in several impact and adaptation themes are presented and described, along with the steps required to go from [GCM](#) data to impact model input data.

Latest

- Workshop held on design of scientific portals (Nov 2014, KNMI (NL)) [download the presentations](#)
- The climate4impact portal is operational since 15 April 2014: [read more](#).

Agriculture/Forestry Energy Health Infrastructure/Urban

Marine/Coastal Nature/Biodiversity Tourism Water Management

Click on one of these images to go to a specific climate change impact and adaptation theme.

The IS-ENES project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration.

Disclaimer

Exemple du portail Climate4impact (IS-ENES)

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

Exploring climate model data

Data discovery

Home Data discovery Downscaling Documentation Help About us Account

Search Faceted Search Catalogs Explore your own catalogs or files Map & Plot Processing

Search

Project	☐ CMIP5	☐ CORDEX	☐ SPECS	+				
Variable	☐ Temperature	☐ Precipitation	☐ Windspeed	☐ Shortwave radiation down	☐ Surface specific humidity	+		
	☐ Min temperature	☐ Conv. precipitation	☐ Max windspeed	☐ Shortwave radiation up	☐ Surface relative humidity			
	☐ Max temperature	☐ Snow	☐ Eastward wind	☐ Longwave radiation down	☐ Specific humidity			
			☐ Northward wind	☐ Longwave radiation up	☐ Relative humidity			
	☐ Evaporation	☐ Surface pressure		☐ Diffuse radiation	☐ Surface relative humidity			
	☐ Potential evaporation	☐ Pressure		☐ Total cloud cover	☐ Max relative humidity			
					☐ Minimum relative humidity			
Frequency	☐ 3 hourly	☐ daily	☐ monthly	+				
Time frame	+							
Experiment	☐ Historical	☐ RCP26	☐ RCP45	☐ RCP60	☐ RCP85	☐ Evaluation	☐ 1pctCO2	+
Domain	Search domain (CORDEX)				+			
Models	Found 186 model(s)				+			

Search datasets

Start search

Exemple du portail Climate4impact (IS-ENES)

Home Data discovery Downscaling Documentation Help About us Account

Account Basket Processing Monitor jobs (1)

Account » Processing » Wizard convert and subset

Processing

Convert and subset

Resource

Variable
tasmax

Projection
EPSG:4326

Bounding box
North:
West: East:
South:

Resolution
Horizontal: width: 73
Vertical: height: 124

Dates
Start date:
Stop date:
Time resolution:

Format
netcdf

Output file name

Start processing

Map
Daily Maximum Near-Surface Air Temperature (tasmax)
1990-12-31T12:00:00Z
1990-12-31T12:00:00Z

Projection: EPSG:4326
0 10 20 degrees



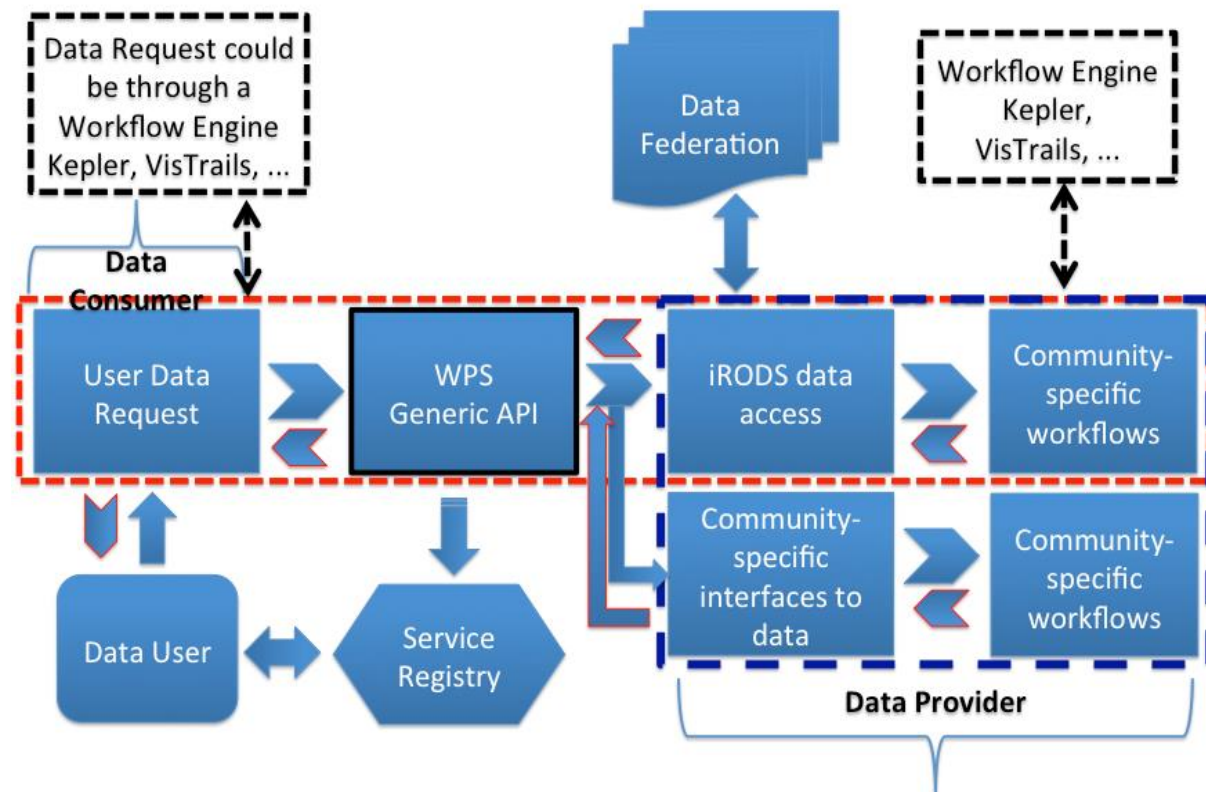
« data user » : Le développement du GEF (EUDAT)

Generic Executive Framework (EUDAT2020)

Implémentation des workflows sur des quantités massives de données

Implementation schematic

- gestion des workflows des communautés pour calcul près des données
- Protocole WPS de gestion de requête => interfaçable avec toute techno d'accessibilité: iRODS (EUDAT), P2P (ESGF)



- 
- La modélisation climatique est confrontée à la croissance exponentielle de ses données (volumes). Ce n'est plus le coût calcul qui devient rédhibitoire mais les **capacités de stockage**
 - Les migrations de données représentent une **part significative du processus de traitement**. Pour **limiter les déplacement** de To de données, le **rapprochement des calculs près des données** devient crucial (computing node)
 - Les volumes (#Tb, #Pb) soulèvent des problèmes de **sécurisation de la donnée** (copie intégrale difficile)
 - La **stratégie de production** « pléthorique » est certes problématique... mais les exercices coordonnés climat cherchent à répondre à de plus en plus de questions scientifiques, toujours plus spécifiques...

Merci de votre attention...des questions ?



« data provider » : Les différent types de données

« types » de données

- **Etudes hors cadre projets** (thèses, ...)
- **Projet nationaux** (ANR)
- **Projet européens**
- **Projets internationaux coordonnés**

*Finalité
(usage
scientifique,
public visé)*

« solutions » de stockage :

- **Lieu d'archivage**
- **Niveau de sécurisation** (copies, réplicas distants)
- **Durée de préservation**

« niveaux » de métadonnées :

- **Minimal** (CF-Netcdf)
- **Renforcé** (CMOR)
- **Elevé** (CMOR+CIM)

Motivations / Objectifs

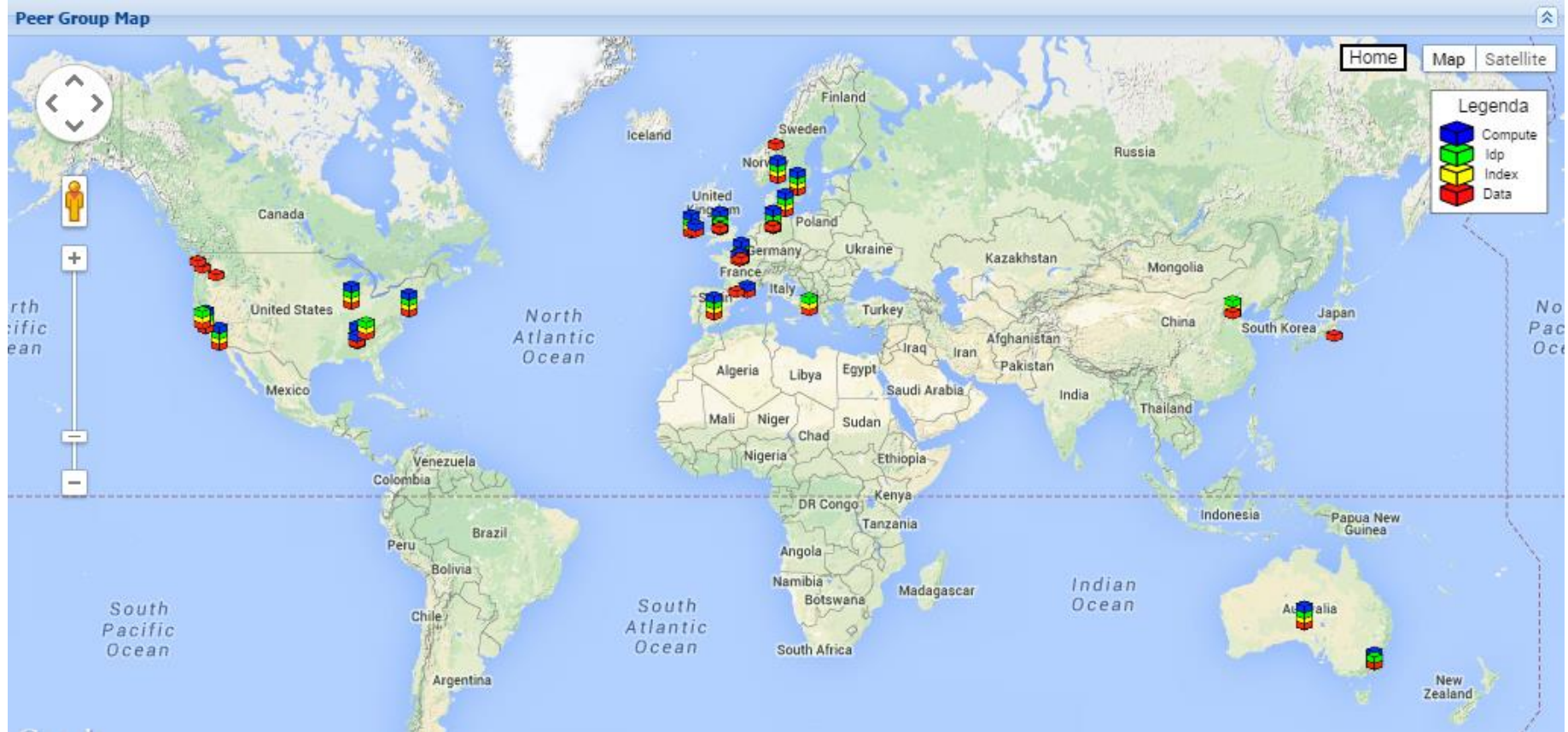
- Améliorer l'infrastructure logicielle dédiée à la recherche en climatologie et au nécessaire partage de données au sein de la communauté climat.
- Répondre au challenge du Big(Scientific)Data : des Terabytes (2007) aux Petabytes (2013) et maintenant vers l'Exabytes (2019).



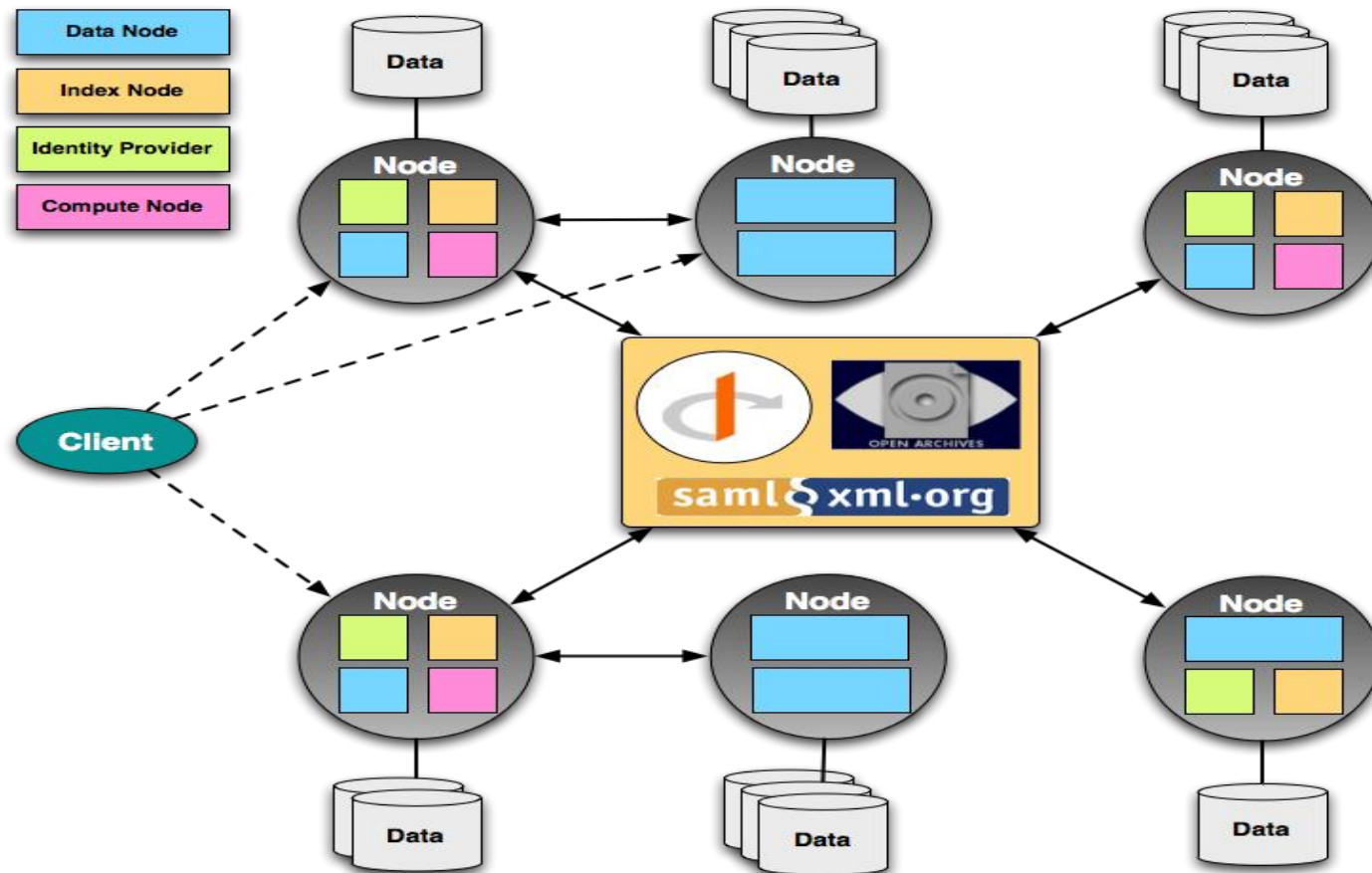
Missions / service

- Support aux activités associées aux exercices coordonnés internationaux d'intercomparaison de modèles climatiques globaux (MIPs) et exercice internationaux de modélisation régionale (CORDEX) + observations (obs4MIP).
- Infrastructure de données distribuées géographiquement et gérées de façon autonome.
- Outils de gestion des données (archivage, préservation, découverte, recherche, accessibilité) et des métadonnées.

Infrastructure de gestion & distribution de données climat : la fédération ESGF



Infrastructure de gestion & distribution de données climat : la fédération ESGF





Etapes de la publication des données auprès de la fédération ESGF :

- Obtention d'un OpenID depuis la Web-Front-End ESGF
- Demande d'autorisation de publication pour un projet
- Connexion SSH sur le nœud de données (archivage)
- Authentification sur le nœud Identity Provider
- Publication locale thredds (sur le data node)
- Publication distante sur l'index

