

# Mise à disposition des simulations climatiques via l'ESGF

---

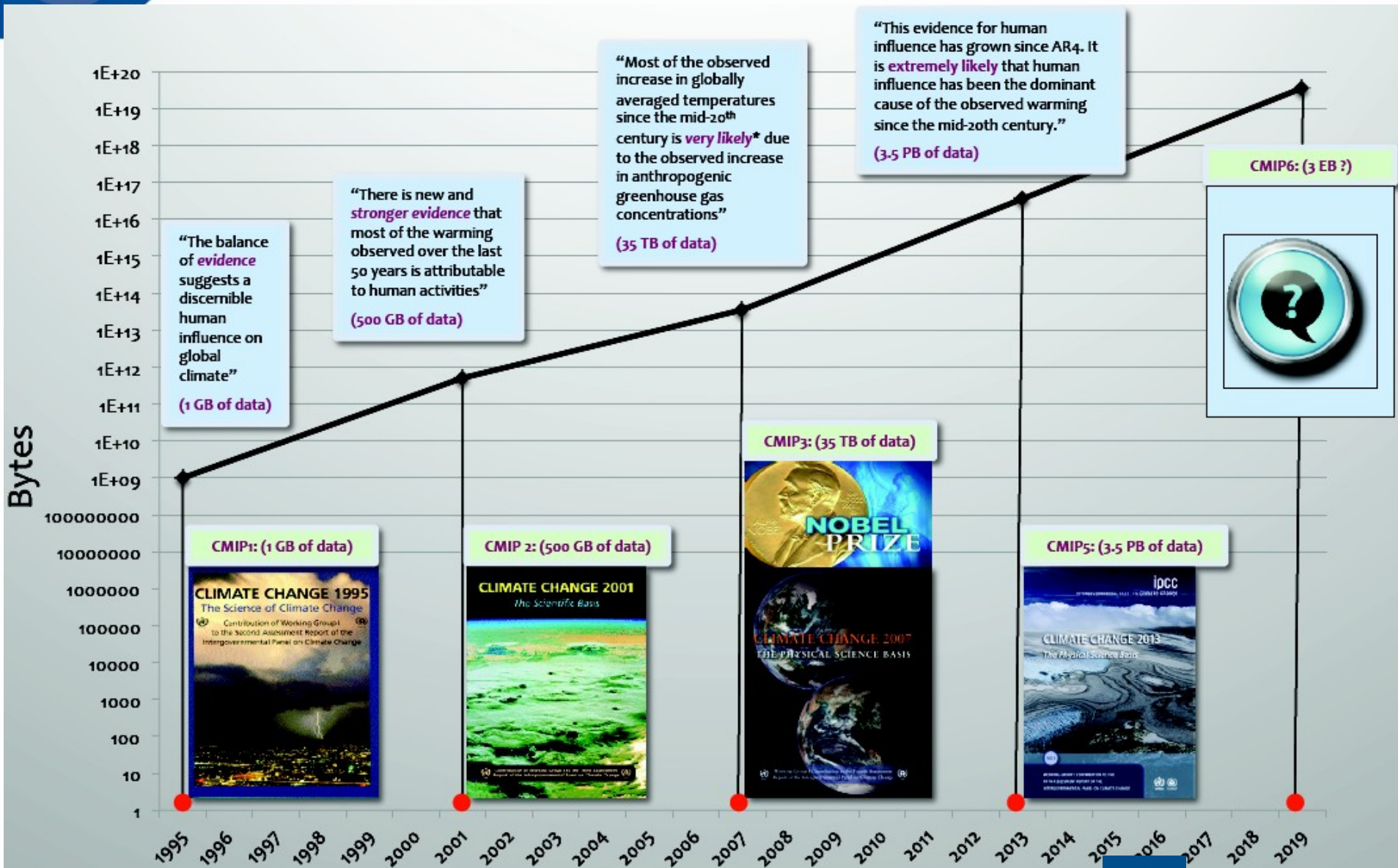
**Eric Gerbier, CNRM/CTI**  
**Laurent Franchistéguy, CNRM/GMGEC**

# Contexte

---

- Rapports d'évaluation du GIEC (Groupe d'Experts Intergouvernemental sur l'évolution du Climat, IPCC) tous les 6 ans. Volumes (4) du 5ème et dernier rapport parus sur la période septembre 2013 – novembre 2014
- CMIP (Coupled Model Intercomparison Project) : une base de données de résultats de simulations climatiques essentielle pour les rapports du GIEC :
- activité de recherche coordonnée par le World Climate Research Program (WCRP, sponsors : WMO, ICSU et IOC/UNESCO)
- Visibilité du laboratoire sur la thématique du climat
- mise à disposition des données : standardisation des formats de données et métadonnées (NetCDF / CMOR)

# Contexte : exercices CMIPs et volume de données



# historique

	date	Volume global	Volume cnrm-cerfacs
CMIP1	1995	1 Go	
CMIP2	2001	500 Go	
CMIP3	2007	35 To	2 To

Modèle : publication des données au plus près de la production

1) Site ftp dans une DMZ sur le réseau RENATER, débit 100 Mb/s

2) Envoi des données par la poste sur des disques USB (« véloprocessing »)



**METEO FRANCE**

# CMIP5

	date	Volume global	Volume cnrm-cerfacs
CMIP5	2013	3,5 Po	52 +20 To

réflexion dès 2009 sur les améliorations à apporter

- infrastructure réseau en 1 Gb/s (firewall, switch ...)

- faire mieux que ftp :

- identification des utilisateurs
- sélection des données via une interface indépendante du stockage
- Sous-grilles
- ... etc

=> ESGF

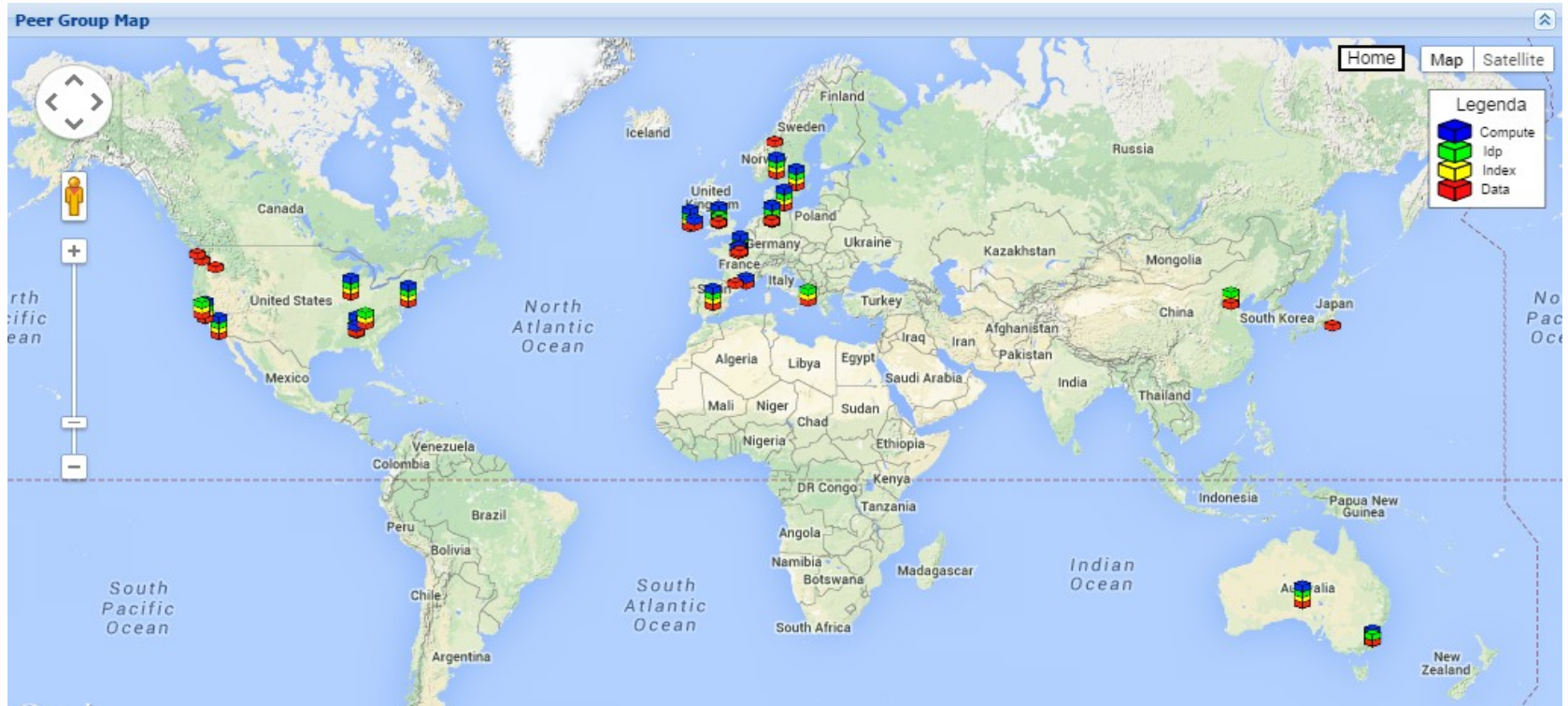


METEO FRANCE

# ESGF

- ESGF (Earth System Grid Federation) : consortium ouvert, sponsors : DOE, NASA, NOAA, NSF, IS-ENES (Europe) et NCI (Australie). <http://esgf.llnl.gov>

Structuration en portails (indexNode), nœuds de données, nœuds d'identification



(juin 2015)

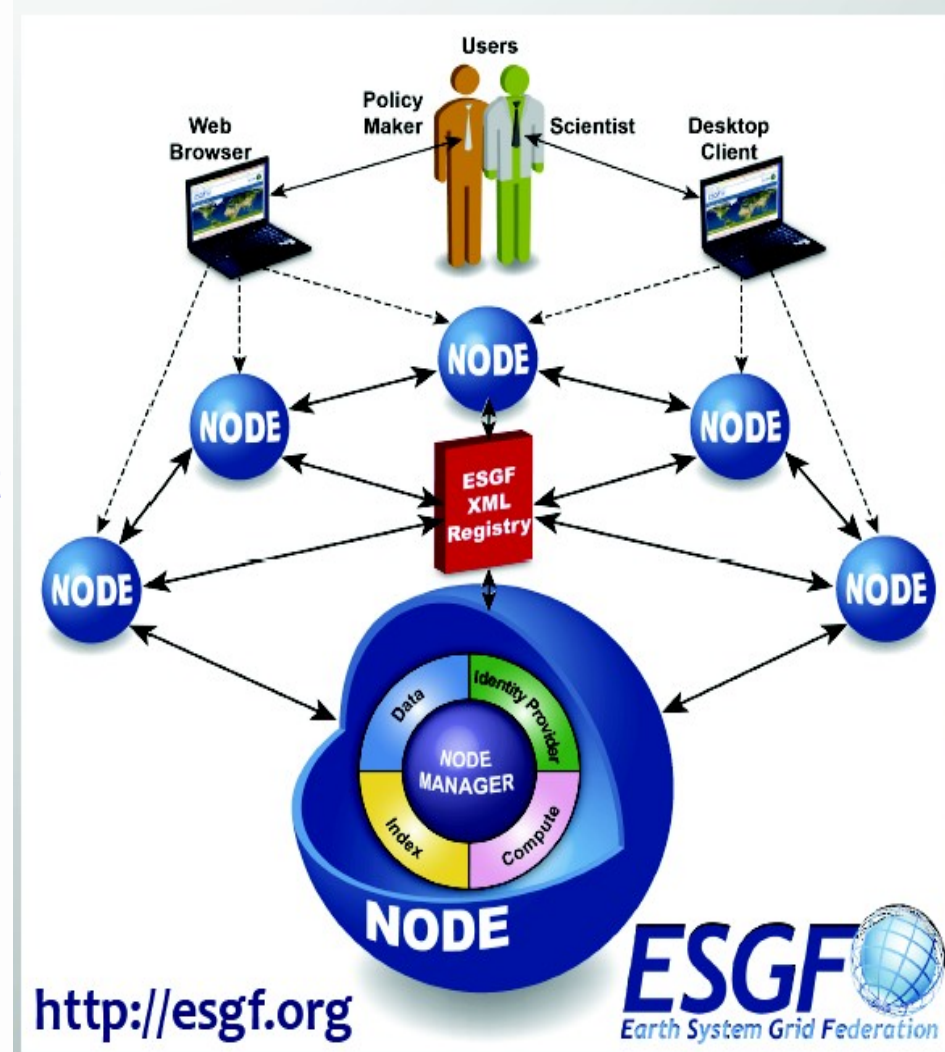
40 datanode ESGF dans le monde (dont 15 Index Node)  
7 datanode français : IPSL(3), TGCC, IDRIS, CINES, CNRM  
➔ réseau ESGF-France (animation IPSL)





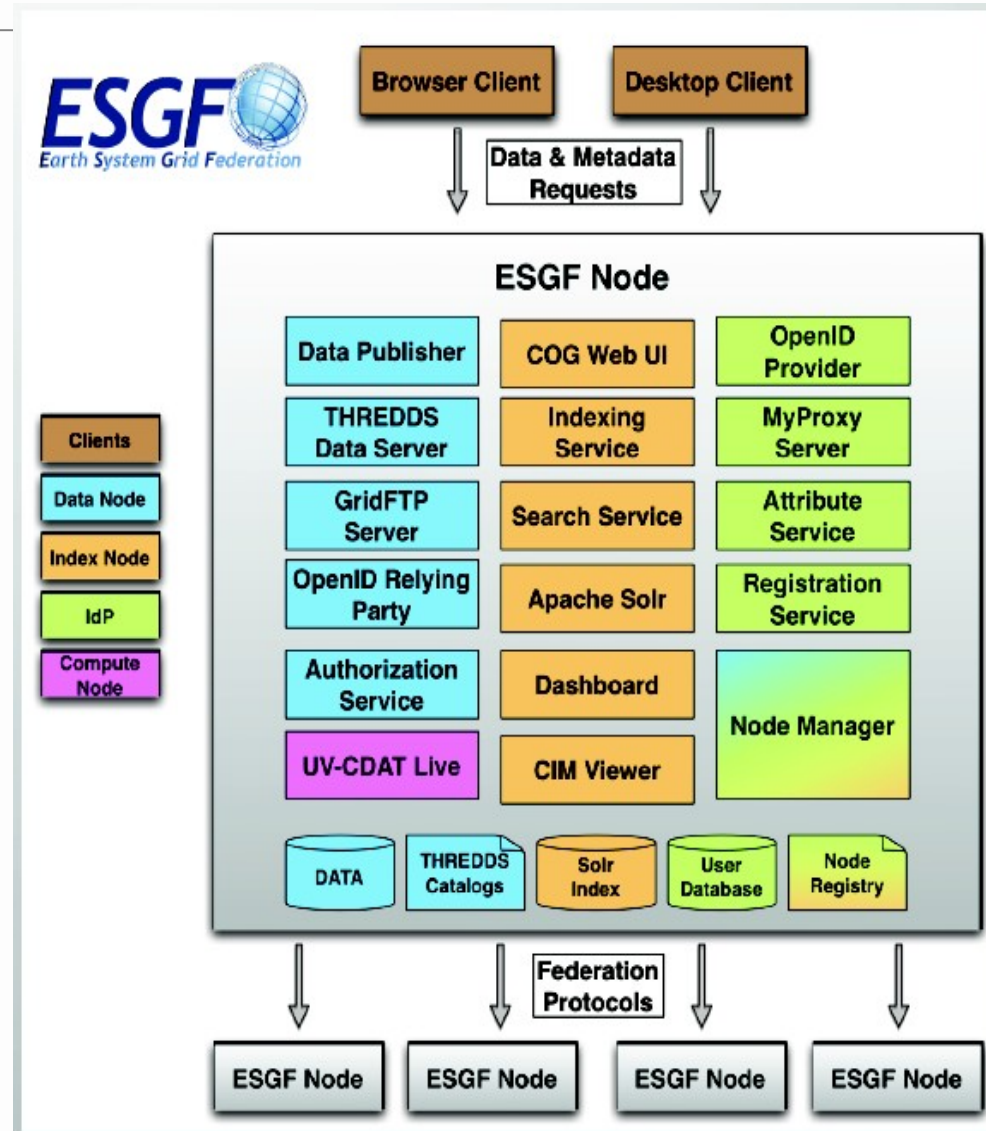
# ESGF : archive distribuée et système de téléchargement

- Approche DAR (Discover-Access-Retrieval) respectée :
  - découverte des données (catalogue de données)
  - authentification
  - accès aux données depuis les portails ou les noeuds de données (navigateur ou script wget)
- Découverte des données sur l'index node de son choix puis redirection automatique vers le datanode hébergeant les données (fédération de nœuds, archive distribuée)
- Système de réplication des données (BADC, DKRZ)
- Références :  
Williams et al, 2015 : A Global Repository for Planet-Sized Experiments and Observations. Bull. Americ. Meteor. Soc., doi:10.1175/BAMS-d-15-00132.1, in press : <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-15-00132.1>



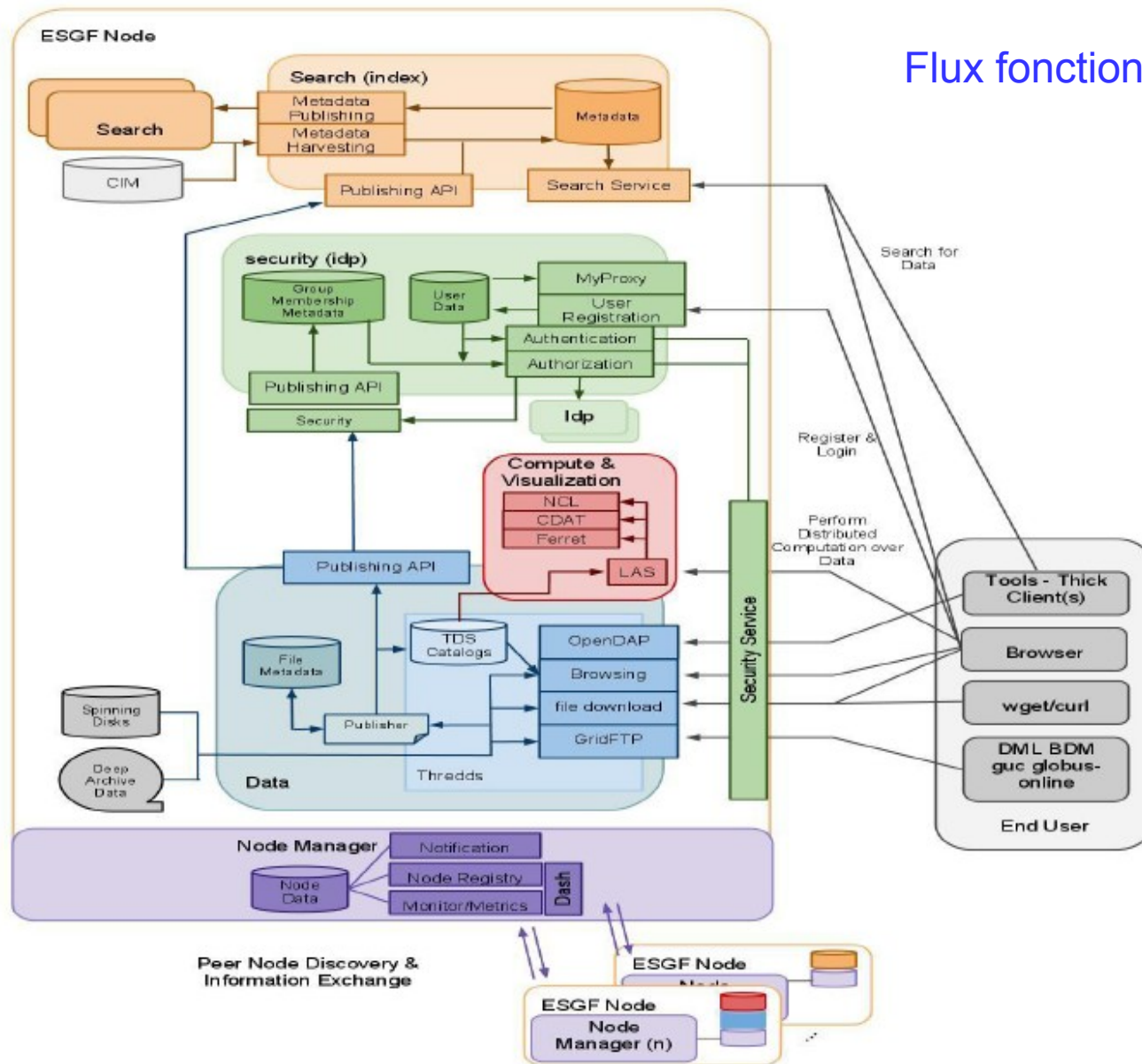
# ESGF : technologies déployées (1)

- Services de recherche (basés sur les métadonnées)
- Services de sécurité
- Service de publication de données
- Live Acces Server (interface graphique)
- Transfert de données (Wget, GridFTP, Globus Online)
- Services DAP (Thredds Data Server, OpenDAP)
- UV-CDAT (outil d'analyse)
- Tableau de bord
- Replication et versionnement
- Fonction Compute Node (implémentation de WPS possible)
- Gestion politique d'accès (restricted/unrestricted)





# ESGF : technologies déployées (2)



Flux fonctionnels

# ESGF : technologies déployées (3)

Nouvelle pile logicielle ESGF v2 en 2016 :

1. Modernisation infrastructure sous-jacente :

- CentOS , Java 1.8, Tomcat 8, Postgres 9.4, OpenSSL 1.0, Python 2.7.9
- industrialisation : installation via des composants RPM
- exécution serveur Apache devant Tomcat (performance, flexibilité, sécurité)

2. Mise à jour de tous les services ESGF :

- services de recherche (Solr5), téléchargement (TDS5), transfert (Globus-Connect-Server), calcul (UV-CDAT), visualisation (LAS)
- refonte interface utilisateur (CoG)

## Le noeud ESGF du CNRM-CERFACS (1)

---

Architecture technique de CMIP5 (2009-2015):

- 1 serveur (Dell PowerEdge R610) sous Centos 5
- 6 baies de stockage avec raid matériel (MD1000, MD 1200) => 100 To utiles
- Réseau 1Gb/s, avec 2 parefeu Cisco en maître/esclave

Données du groupe de modélisation CNRM-CERFACS proposées au téléchargement ( 75 To ) :

- données CMIP5 du CNRM et du CERFACS
- données Cordex (domaines Europe EUR-11/EUR-44 et Afrique AFR-44)
- Données TAMIP

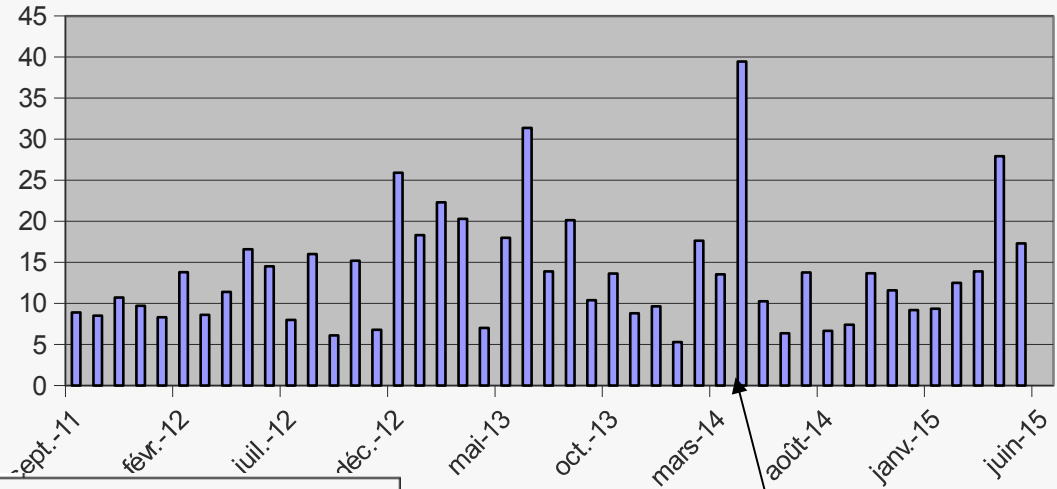
visibilité du laboratoire : 300 publications scientifiques utilisant nos données



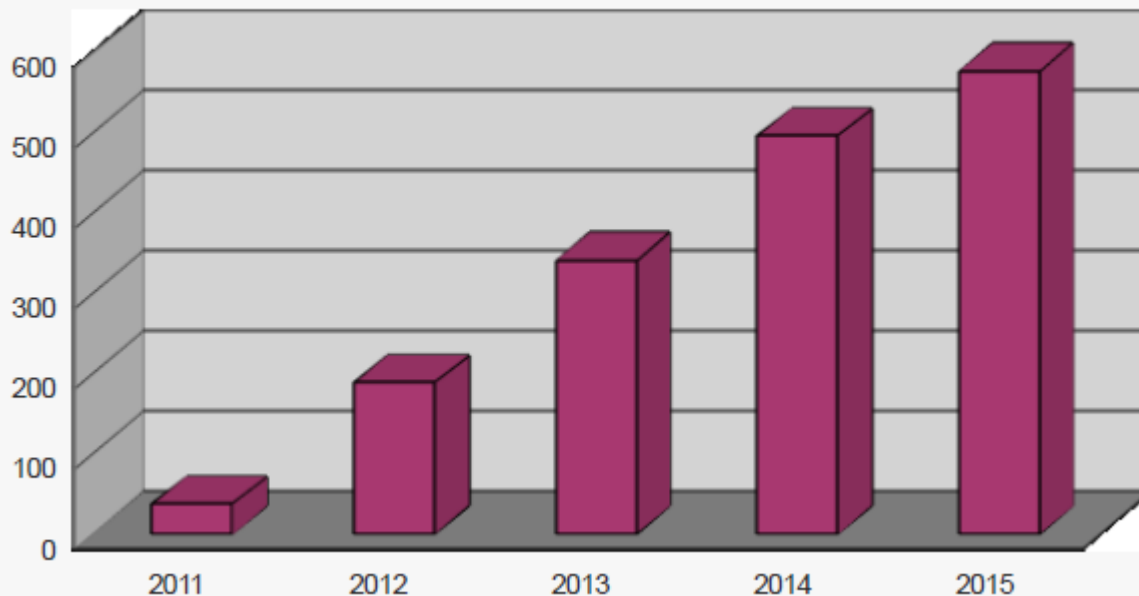
## Le noeud ESGF du CNRM-CERFACS (2)

- Volume total données téléchargées depuis septembre 2011 : 577 To
- Moyenne annuelle : 154 To/an
- Moyenne mensuelle : 14 To/mois
- Volume maximal quotidien:  
7.4 To/jour

Volume mensuel données téléchargées nœud ESG du CNRM (To)



Volume cumulé annuel données téléchargées sur le noeud du CNRM (To)

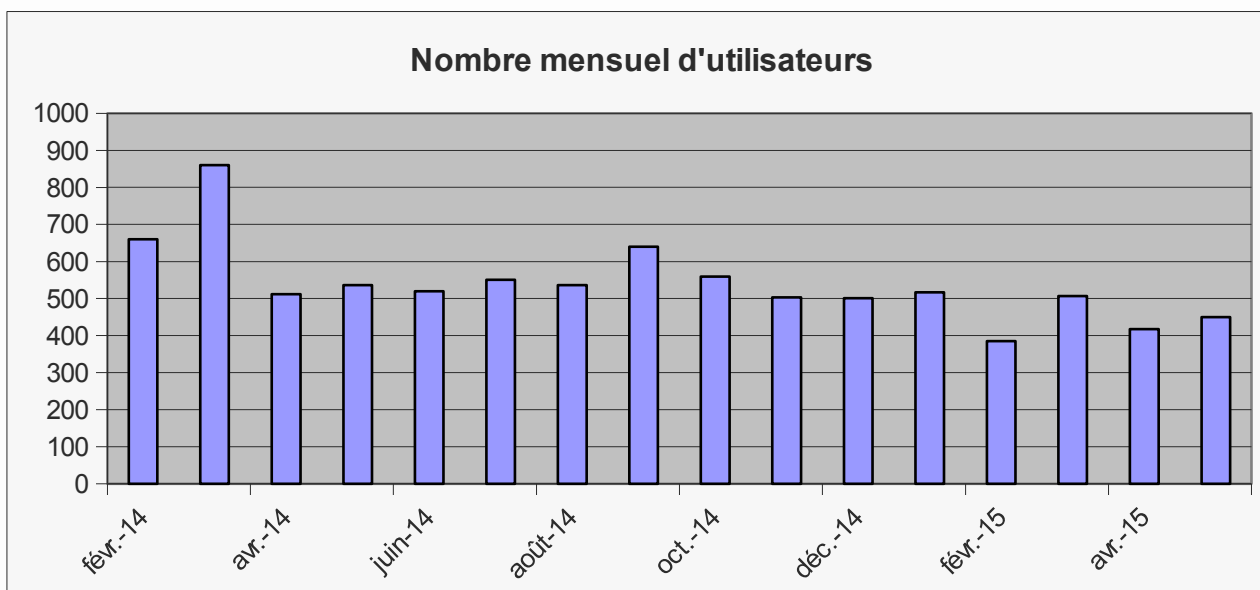


Parution volumes 3 et 4 du 5ème rapport du GIEC

(Statistiques Juin 2015)

## Le noeud ESGF du CNRM-CERFACS (3)

- période février 2014 – mai 2015
- moyenne d'environ 540 utilisateurs/mois
- pic de 860 utilisateurs en mars 2014





# CMIP6

	date	Volume global	Volume cnrm-cerfacs
CMIP6	2020	3 Eo	2 + 0,5 Po

réflexion dès 2015 sur les améliorations à apporter

- infrastructure réseau en 10 Gb/s
- amélioration de la sécurité : POC sur les parefeu => Palo Alto
- achats échelonnés sur 3 ans => matériel hétérogène
- extension de garantie sur 7 ans
- séparation stockage / production
- refonte de l'architecture de stockage : choix d'un système de fichier distribué simple et robuste (plutôt que performant) : glusterfs

# Avenir

---

- Janvier 2016 : installation de la pile logicielle ESGF V2
- Février/mars : republication des données CMIP5
- Nouveaux projets : CCMI, ANR
- Fonctions « compute » : le calcul se fera près de la donnée

**Merci pour votre attention !!**

