



**METEO
FRANCE**

- **Outils :**
 - **CliMAF**
 - **ESMValTool / C-ESM-EP / Atlas**
 - **EcliS**
 - **dr2xml**
 - ...
 - **Compte-rendu de réunion :**
 - **Prospective INSU – données**
 - **Ressources, stockage et portage**
 - **CMIP6 et co :**
 - **Production**
 - **Publication**
 - **Site web**
-

Réunion GMGEC-Tech – 10 mars 2020

Outils

CliMAF

- Formation :
 - Poursuite des boccas CliMAF

- Contributions à CliMAF :

- Dernière version :
 - V1.2.12 (11/09/2019) :
 - ▶ Compatible avec la C-ESM-EP
 - ▶ Corrections de bogues
 - ▶ Ajout d'opérateurs cdo sans sélection cachée : `ccdo_fast`, `space_average_fast`, `time_average_fast`, `lbox_fast`, `annual_cycle_fast`, `clim_average_fast`

CliMAF : Prochaines versions

- Prochaines versions :
 - V1.2.13 (mars 2020) :
 - ▶ Mise à jour de la documentation
 - ▶ Mise à jour des exemples
 - ▶ Correction de bogues
 - ▶ Modification de l'appel à cdo
 - ▶ Amélioration des figures
 - ▶ Possibilité d'appel des opérateurs sans fichiers d'entrée
 - ▶ Modification du cache
 - V2.0 (pour l'été ?) :
 - ▶ Compatibilité python 3
 - ▶ Écriture de tests unitaires

CliMAF : Bogues connus

- Problème dans certains cas avec sellonlat : lié à cdo
- Lenteurs dans les cas où les fichiers sont gros : utilisation des fonctions *_fast à généraliser
- Problème dès que les calendriers utilisés ne sont pas standards : lié à la fonction check des dataset et aux comparaison de dates
- ~~Problème de lecture d'un champ fixe dans un projet (voir si cela fonctionne avec fds plutôt qu'avec la définition d'un projet)~~
- Débogage nécessaire de la fonction check des dataset
- Sélection multiple du même dataset
- ~~cdftools ne fonctionne plus → recompilation à faire sous mageia7~~

CliMAF : Et pour la suite...

- Homogénéisation de la récupération des données
- Extension des projets
- Remplacement de NCL
- Interrogation du service d'errata
- Récupération de données sur des sites distants
- Remplacement de CDO (?)

ESMValTool

- Discussions entre Jérôme et l'équipe ESMValTool en cours, pas de participation à l'EGU.

- Installation :
 - Environnement installé et disponible sur lustre via les commandes suivantes :
 - ▶ *unset PYTHONPATH*
 - ▶ *source*
/cnrm/est/USERS/rigoudyg/tools/ESMValTool/install_env/miniconda37/etc/profile.d/conda.sh
 - ▶ *conda activate ESMValTool*
 - Une fois l'environnement mis en place, la commande suivante doit fonctionner :
 - ▶ *esmvaltool --help*

C-ESM-EP

- Atlas réalisés par Laure Coquart sur les simulations CMIP6 disponibles sous : <http://cerfacs.fr/giec6/>

- Version installée (2.0) :
 - Disponible sur le lustre du CNRM sous /cnrm/est/USERS/rigoudyg/C-ESM-EP.
Utilisable après copie chez soi.

Les jobs sont lancés sur la partition P8HOST sur aneto (valable pour tous les serveurs). Attention, restriction de la possibilité d'utiliser cette partition, demander à CTI.

- Des évolutions sont en cours :
 - Modifications liées à la nouvelle version de CliMAF
 - Bugfixes diverses
 - Ajout d'une option pour indiquer sur quelle grille les projections sont faites

Atlas / Générateur automatique de diagnostics

- Développements en pause
 - Discussion en cours avec Jérôme, cahier des charges à faire avant réunion sur le sujet
-
- Marie-Pierre avait développé un atlas CliMAF il y a deux ans. C'est une base qui peut être récupérée si besoin.

Eclis : Versions en cours

- Versions :
 - V6.44 (mars 2020) :
 - ▶ Ajout de la configuration pour Epona
 - ▶ Réécriture de la partie post-traitement
 - ▶ Ajout d'un warning dans relan si jamais il y a déjà eu un post-traitement pour éviter d'avoir des données plusieurs fois sur Hendrix
 - ▶ Ajout de la variable PROJECT pour indiquer le projet auquel appartient la simulation
 - V7.0 (en développement) :
 - ▶ Permet de faire des run annuels
 - ▶ Travail sur le nudging
 - ▶ Adaptation pour les run aladin et ESM
 - V6.45 (dès que la V6.44 sera prête) :
 - ▶ Modification du répertoire de relance et de l'archivage des fichiers intermédiaires pour s'adapter aux changements d'architecture
 - ▶ [Vérifier les runs sur aneto](#)

Eclis : Développements prévus

- Pause dans les gros développements pour le moment. A suivre dans les prochaines versions :
 - Corrections de bogues (si besoin)
 - Adaptations mineures (si besoin)
 - Mise à jour de la documentation
 - Mise à jour des params de référence
 - Interfaçage de xiofilechecker (vérification cohérence xml / fichiers produits)
 - Interfaçage d'un débogueur
 - Suppression de LDR_EXPID

Eclis : Et pour la suite...

- Vérifier quelles variables devraient être modifiées par Eclis dans les settings de dr2xml (en lien avec les développements dr2xml)
- Intégration d'un équivalent de verifrelan dans Eclis ? Voir si les fonctionnalités de slurm le permette comme on ne peut pas faire de cron sur le calculateur. A discuter en fonction.
- Les post-traitement peuvent se faire au fil de l'eau, voir s'il faut le généraliser au vu de la charge de lustre sur les calculateurs.
- Continuer à suivre le développement de simu-lanceurs pour voir si d'autres outils pourraient être à même de remplacer EcliS. En cas de refonte d'Olive côté GMAP, suivre pour voir ce qui sera développé (même démarche à mener éventuellement avec l'IPSL et l'outil Cylc).

dr2xml

- Versions disponibles :
 - V2.1 (27/11/2019) :
 - ▶ Adaptations pour Cordex
 - ▶ Début de documentation du code
 - ▶ Utilisation d'accents possible dans les fichiers de settings (encodage utf-8 requis)
 - ▶ Ajout de l'attribut `branch_method` (y compris quand il n'y a pas d'expérience parent)
 - V1_CMIP6_dev :
 - ▶ Version proche de celle utilisée pour CMIP6 (pour le rendu des fichiers de sorties)
 - ▶ Embarque des nouveautés (variables dev...)
 - Base de tests :
 - ▶ Possibilité d'inclure des configurations particulières pour les tests de non régression.
- Pause dans les gros développements pour le moment. A suivre dans les prochaines versions (2.1.*) :
 - Poursuite de la documentation
 - Écriture des tests unitaires
 - Corrections de bogues (si besoin)
 - Adaptations mineures (si besoin)

dr2xml : Et pour la suite...

- Généralisation de dr2xml à d'autres projets :
 - Actuellement adaptation CMIP6 (défaut) et CORDEX (via tests dans le code)
 - Demandes pour une utilisation hors CMIP6 (IPSL et Cerfacs)
 - Réunion à organiser pour écrire les spécifications pour les usages IPSL, Cordex, prévision saisonnière... puis développement quand cela sera fait
- Utilisation d'un fichier settings propre aux variables dev ?
 - Ajout des axes personnalisés
 - Gestion des grilles natives pour les variables dev
- Poursuite de la rédaction de la doc
- Simplification de l'installation (pip, ...)

Autres outils

- Nouveau package pour créer les conditions aux limites SIC/SST pour Arpège :
 - Pré-requis :
 - cdo 1.9.1 (interpolation spatiale)
 - Utilitaire nctofiee
 - Pas de dépendance avec des bibliothèque non disponibles en python 3
 - Au CNRM, disponible sur lustre sous :
/cnrm/amacs/TOOL_PACKAGES/PACKAGE_Make_SST_SIC_ARPV6.tar.gz
 - A tester au Cerfacs (Marie-Pierre, Rym)
 - Dès que les tests seront terminés, faire un mail à GMGEC avec les informations.

- Get :
 - L'option « -n » ne fonctionne plus (voir pour l'enlever ou pour corriger le bogue)

- XIOS :
 - Problèmes de compilation sur mageia7.
 - David a réussi à faire fonctionner la 1442 en ajoutant des modifications issues de versions ultérieures (uniquement en mode debug).
 - Reprendre les différents éléments.

Compte-rendu de réunions



Compte-rendu : Prospective INSU – données

- Une trentaine de participants (Astrophysique, Télédéttection, Géologie, Sismologie, Génomique, Environnement, Océanographie, Géographie, Climat, ...)
- Des observatoires INSU (Obs Paris, OMP, et Infras de Recherche (DataTerra, OZCAR, ...) et GDR (MADICS)
- Modélisation climatique : présentations de Guillaume Levavasseur (DMP CMIP6) et Gaëlle (workflow de production ECLIS)
- Des données hétérogènes (numériques ou matérielles, petites ou volumineuses, issues de divers capteurs ou modèles numériques, échantillons, ...)
- Des besoins/problèmes semblables : maîtrise de la production, qualification de la donnée, contrôle qualité, documentation, stockage, archivage, distribution (principe FAIR) => des workflows à toutes ces étapes + DMP
- Données massives ou complexes à analyser : rôle de l'IA (ex. approches hybrides physique/Deep-Learning pour l'analyse d'images, PANGEO -[environnement et communautés de développeurs autour de Python et des bibliothèques xarray, dask et pandas](#)- pour l'analyse de données massives,...)
- Des communautés diversement avancées dans la FAIRisation : Sismo, Astro et Climat bien structurées, les autres moins
- Philo : « repenser » les pratiques à tous les étages (les chaînes de traitement, la conservation de données, la façon de les traiter, les analyser...)
- RH : besoin de profils hybrides à l'interface DataScience / Thématique. Constat classique : pas de postes, métier non reconnu, attractivité des salaires CNRS, gros effort de formation...
- Réflexion sur mutualisation moyens (infra de type mésocentres, outils). Construction de l'IR DataTerra (vocation ?, périmètre?). [Pour faciliter l'accès aux données climatiques pour le CNRM \(actuellement très coûteuse via l'IPSL\), demande de financement sur le projet PIA3 dans le cadre de l'IR Data Terra \(qui inclura l'infrastructure Climeri\). Data Terra permettra d'accéder à des données de disciplines très diverses.](#)
- Gouvernance (transdisciplinarité, infrastructure, réseau de travail, formation) : il existe une gouvernance réseau (RENATER) et calcul (GENCI) mais rien pour la donnée => besoin d'une division technique Data à INSU (+ agir au sein d'ANITI)
- [Question des liens avec l'industrie : craintes des labos vis-à-vis des géants de l'industrie, dont GAFA. Questions d'éthique, de marchandisation des données. Pas du tout la même approche au USA.](#)

Ressources calcul, stockage, portage

Portage Calcul 2020

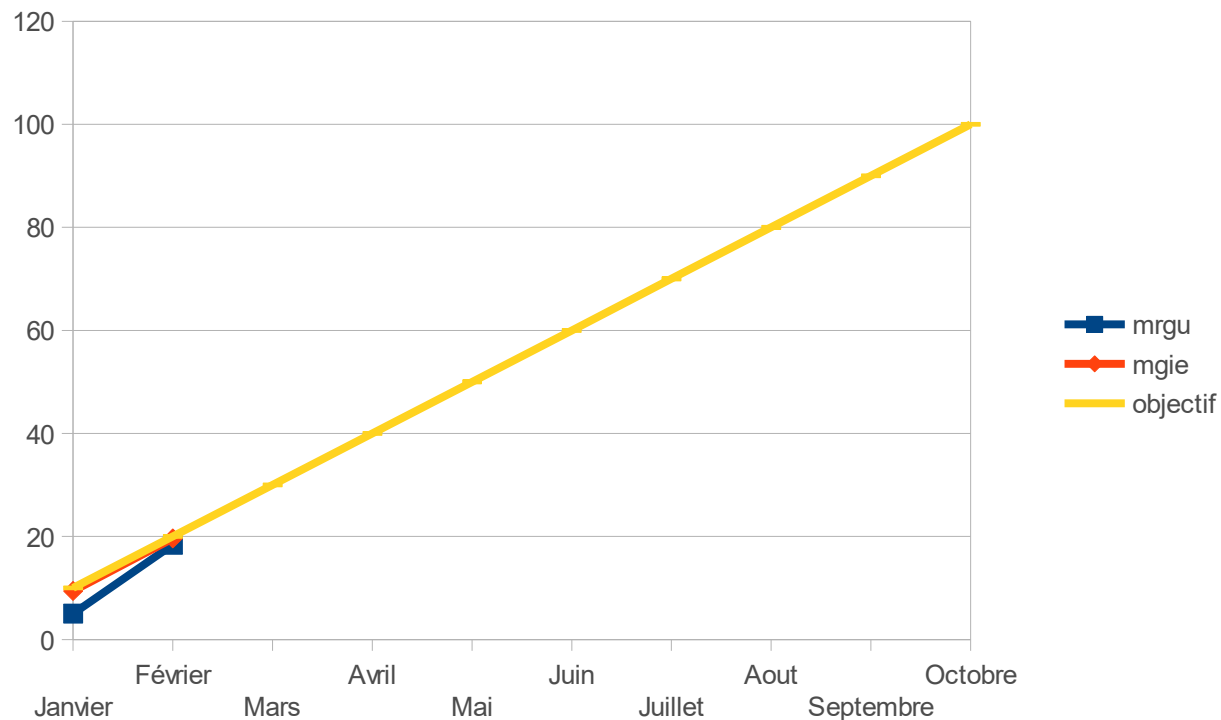
- A noter :
 - Le volume des \$HOME sera limité à 50Gb/utilisateur sur les prochains calculateurs
 - Enquêtes sur les besoins
 - Chaque utilisateur aura la responsabilité du transfert de ses propres données entre anciennes et nouvelles machines Bull
 - Les répertoires sous /scratch seront nettoyés tous les deux à trois mois
 - Cartographie des flux entrant et sortant des calculateurs en cours (réunion 4 mars)

Portage Calcul 2020

- A faire :
 - Archivage des données de Beaufix et Prolix :
 - ▶ Définir les répertoires dont il faut connaître le contenu (/scratch/CMIP6/V2, ...)
 - ▶ Le ls récursif sera lancé par adminHPC (échéance à préciser)
 - ▶ Discussion sur la pertinence de réaliser ce ls récursif → on sait à peu près ce qu'il y a dans ces répertoires et on peut le faire nous
 - Lister les besoins en termes d'espace pérenne en particulier CMIP6, CEDRE, données du régional, packs arpège :
 - ▶ Un mail sera envoyé à la DSI pour indiquer et justifier ce besoin (évaluation volume disque /scratch GMGEC demandé à faire avant envoi du mail).
 - ▶ Réfléchir à l'organisation des données sur un espace partagé pérenne sur lustre (une seule architecture avec quelques versions utilisées)
 - ▶ Archiver ce qui est actuellement sur Beaufix/Prolix sur Hendrix (en particulier /scratch/CMIP6/V1 et V2)
 - Modifications d'EClIS nécessaires :
 - ▶ Ajout des fichiers de configuration pour Belenos à faire
 - ▶ Alléger le contenu du répertoire de relance sur le \$HOME (listings sur lustre, archivage de tous les éléments) → les listings sont déjà envoyés sur Hendrix dans un tar
 - ▶ Généraliser la récupération des forçage sur la machine d'archivage
 - ▶ Modification du post-traitement à faire pour éviter de perdre des fichiers sur le lustre en cas de nettoyage pendant la simulation. Les fichiers seront ensuite récupérés au moment du post-traitement s'ils ne sont plus disponibles sur le scratch
 - Run VSR :
 - ▶ Reproductibilité CM6-1 LR (200 ans) envisagée

Ressources calcul et stockage

- (Rappel renouvellement calculateurs : arrêt Beaufix/Prolix prévu le 19/10/2020 pour le moment, voir mail A. Beuraud 10/01)
- Consommation mgie 2020 sur Beaufix : 3,5 Mh consommées depuis début 2020 sur la queue Libre sur mgie alors que quota file classique non consommée → merci d'imputer sur la queue mgie classique !!
- Le quota pour 2020 comprenait l'OMIP-HR et quelques simulations LR.
- Liste comptes devant imputer sur mgie sur Belenos/Taranis communiquée à CTI pour diffusion à DSI (17 comptes)



CMIP 6 et co



Production CMIP6

- Liste des champs fixes manquants à finaliser et à générer si besoin.
- Variable vt100 :
 - Version 1.00.31 de la data request disponible avec des modifications sur la variable vt100 mais a priori pas finalisées
- Le parent_variant_label est erroné pour certaines simulations pour lesquelles $r > 1$ (liste des simulations concernées à faire encore)
- Fichier de recensement des publications par MIP utilisant les sorties des modèles du CNRM-Cerfacs pour CMIP6 :
<https://docs.google.com/document/d/1HZyJWuGMQxVonX5gQltD9wP2TRjHHkE2/edit#>
- Fichier de recensement des variables non produites qui manquent (si besoin, des reruns pourront être envisagés pour ces variables, pour l'instant, fait à titre d'information) :
<https://lite.framacalc.org/9dxk-ocyj7tesh5>

Publications CMIP6 ESGF

- Publication CMIP6 réalisée à ce jour sur 28 nœuds (38 institutions, 86 modèles).
Détails sur l'ESGF et sur la page de monitoring :
https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/
- Nouvelle version de la pile logicielle ESGF (v4.0.4) : installée sur le nœud de test esg2 du CNRM et validée (publication ok pour tous projets). A déployer sur nœud de prod esg1
- CDNOT (CMIP Datanode Operations Team) : poursuite réunion bimensuelle. A noter : demande aux Index Node de retirer replica dans un délai d'une semaine si dépublication données primaires
- Meeting annuel ESGF F2F (Face2Face) 23 au 25/03 organisé pour la 1ère fois en Europe (CIC Toulouse) annulé (Covid)

Publication ESGF au CNRM

■ Bilan publication CNRM au 06/03 :

- détails sous /cnrm/est/USERS/franchisteguy/CMIP6/CNRM/Publication/SuiviPubli/SuiviPubliGlobal.ods

- **279** bons à publier émis au CNRM

- **276** simulations publiées soit **~94 %** du nombre total de simulations à publier sur l'exercice (**291** prévues à ce jour).
15 MIPs couverts côté CNRM-CERFACS, volume données CMIP6 d'ores et déjà publiées **502 To**

- Depuis la dernière réunion Tech du 23/01 :

* fin publication membres supplémentaires DECK CM6-1 historical (30 membres disponibles)

* fin publication membres supplémentaires DECK ESM2-1 historical (10 membres disponibles)

* publication 4 membres supplémentaires CM6-1 ssp245 (10 membres disponibles)

* publi ScenarioMIP HR ssp126 (sauf variables table 6hrLev hus, ua, va), ssp370, ssp585 (sauf variables table 6hrLev hus, ta, ua, va)

* simulations LS3MIP(3)

* simulation PMIP CM6-1 lig127k

■ A venir côté CNRM (par ordre de priorité décroissant) :

* dépublication simulations erronées (cf diapo Errata) puis publication reruns correspondants

* variables P1 tables Slday et SImon pour ESM2-1 (fait pour CM6-1)

* extension '2300' 3 membres r1 ScenarioMIP ESM2-1

* ESM2-1_esm-hist r1i1p1f3 (modèle en émission), ESM2-1_1pctCO2_r1i1p1f2_C1 (1pctCO2-cdr de CDRMIP)

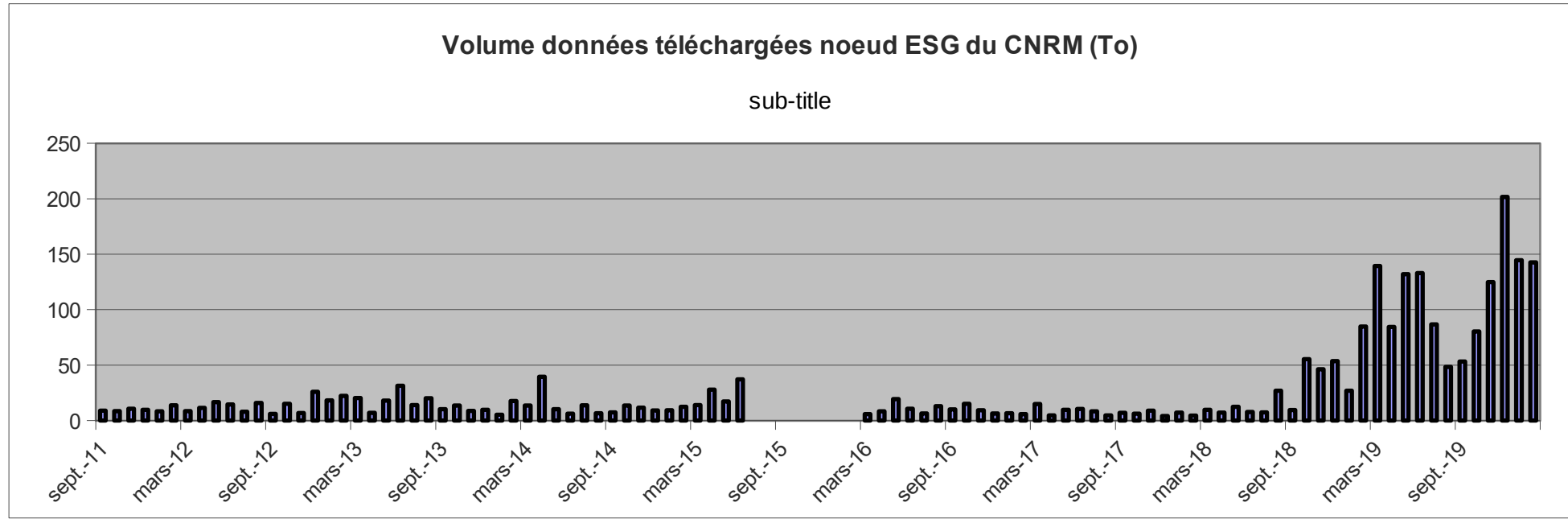
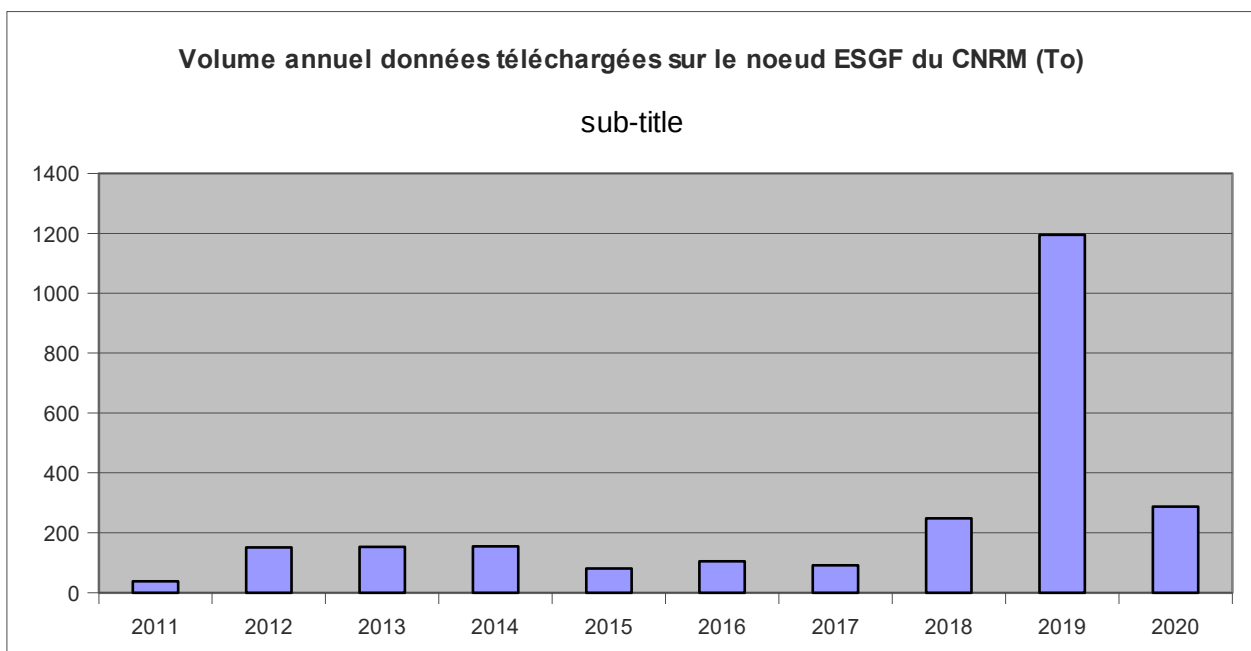
* complément DECK piControl LR (500 premières années seulement publiées) après émission bon à publier correspondant

* puis ordre chronologique d'arrivée des futurs bons à publier....

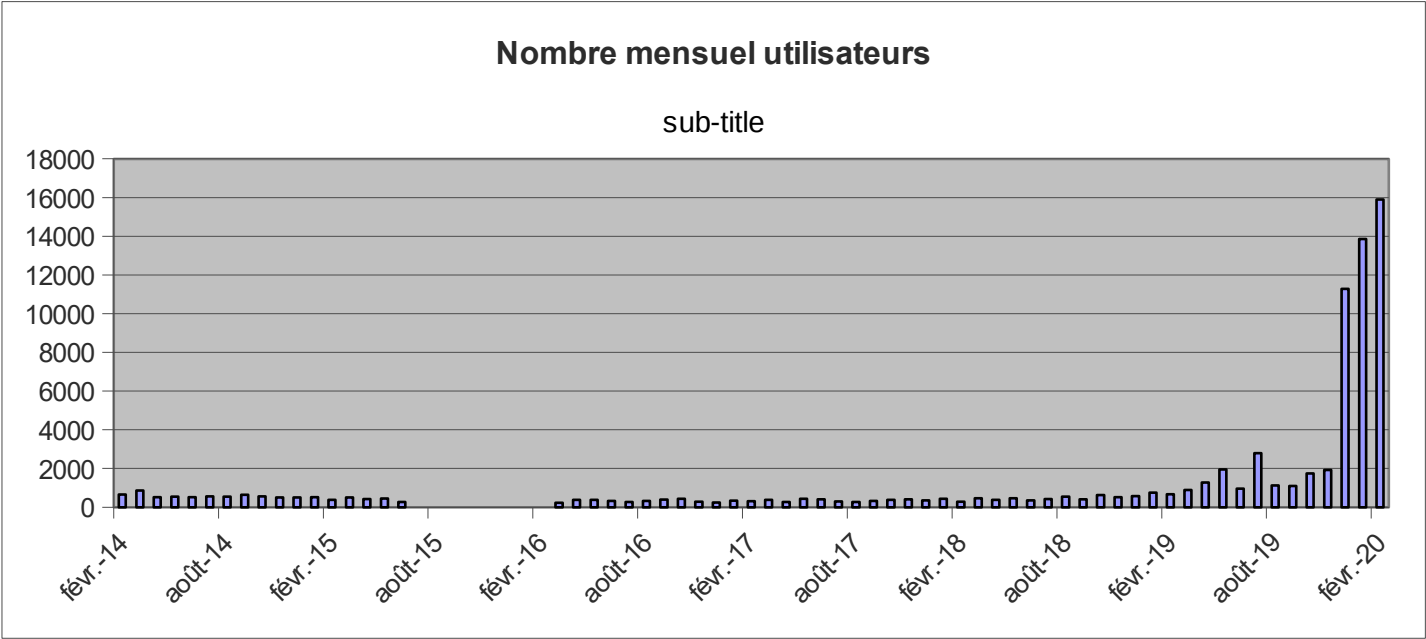
Publication ESGF au Cerfacs

- La publication de DCPD avance bien.
- Publication de HighResMIP :
 - avance doucement au MetOffice (en charge de la publication des membres 1 dans le cadre de Primavera ; publication sur le noeud CEDA)
 - publication des membres additionnels Primavera au Cerfacs : des difficultés liées à la volumétrie et à la publication "en masse"
- Quelques difficultés lors du lancement de publications en parallèle en cas de plantage

Trafic noeud ESGF au CNRM (volume téléchargé)



Trafic noeud ESGF au CNRM (nombre utilisateurs)



Errata Service, known issues

- champs ***4co2 ESM-2-1** erronés (calculés 1xCO2) : déclaration faite sur l'ErrataService (456 datasets, 38 simulations concernées) et info en ligne sur site web CMIP6. Dépublication en cours
- problème membre 29 historical CM6-1 et membre 3 de hist-GHG CM6-1 (continuité chaleur de l'océan) : info en ligne sur site web CMIP6, à dépublier et à déclarer sur Errata-Service. Reruns à publier quand dispos.
- variables **seaice** mal pondérées : feu vert David pour publication variables P1 (siu, siv, sithick, siconc, sisnthick et sivol) avec message de warning pour siu et siv. Fait pour CM6-1, à faire encore pour ESM2-1
- (stand by) attribut **parent_variant_label** erroné pour certaines simulations ('variant_label' au lieu de 'parent_variant_label'). A faire encore : liste des simulations concernées, texte pour site web (rubrique 'known issues')
- (stand by) Instruction à terminer suite à un retour utilisateur sur l'absence des champs lat_bnds / lon_bnds dans nos fichiers (attributs a priori présents pour grille native seulement, à confirmer encore)
- Rappel : tableau (usage interne) recensement erreurs sous /cnrm/est/USERS/franchisteguy/CMIP6/CNRM/Production/Issues (InfoIssuesInterne.odt)

ESDOC

- ESDOC alimenté à ce jour par 10 centres, cf <https://explore.es-doc.org/cmip6/models>
- Coté CNRM, pour ESM2-1 :

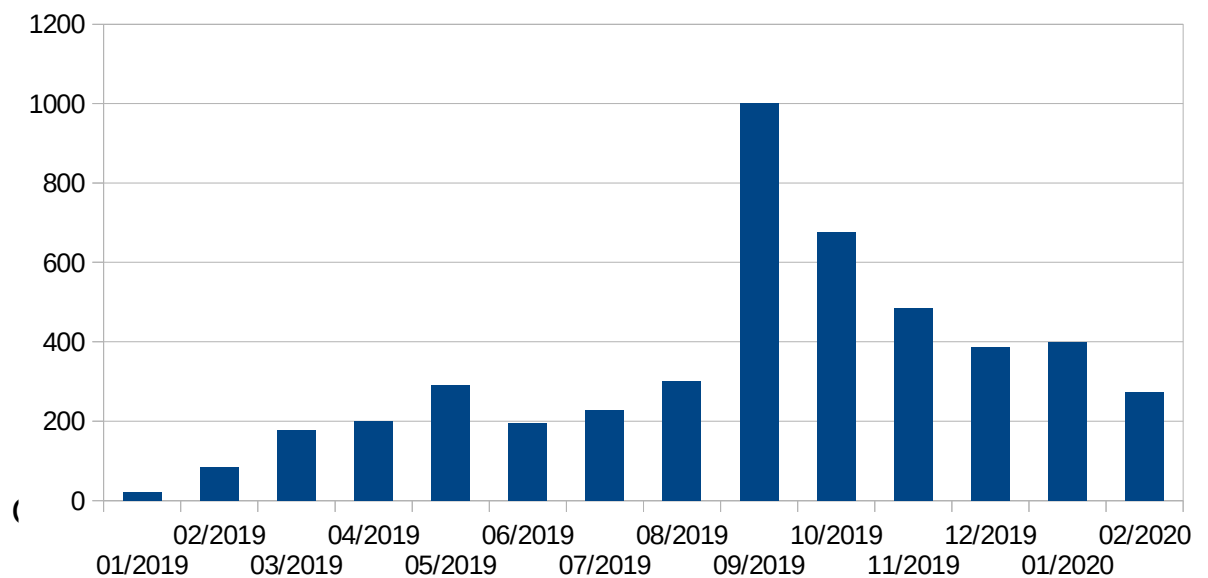
	landice	aerosol	atmoschem	land	coupling	atmos	ocean	ocnbgchem	seaice	toplevel	citation
Fichier xls	-	Ok	Ok	Ok	Ok	Ok	Ok	Ok			
EsDoc	-	Ok	Ok	Ok	Dde publi faite	Dde publi faite	Dde publi faite	Dde publi faite			

- Demandes de publication (cf ci-dessus) faites le 31/01 mais pas de prise en compte par le portail (malgré 2 relances par mail auprès de l'équipe ESDOC....)
- A suivre :
 - 1) récolte puis vérification des fichiers restants pour esm2-1 (seaice, toplevel et citation), dépôt sur GitHub ESDOC et envoi à l'équipe ESDOC de vos remarques éventuelles sur les fichiers à renseigner
 - 2) extension à CM6-1 (initialisation fichiers CM6-1 par contenus fichiers ESM2-1 faite) puis cm6-1-hr

Site web CMIP6 CNRM-CERFACS

- Site web ouvert fin janvier 2019
- Quelques statistiques (au 06/03/2020) :
 - 4872 visites
 - rubriques les plus visitées : Models and contribution to CMIP6 (84.9%), Data access and information (12%), About CNRM-CERFACS (2.3%), Useful links (0.5%), Publications (0.3%)
 - articles les plus visités : CNRM-CM6-1 model, CNRM-ESM2-1 model, Description of CNRM-CERFACS models and contributions to CMIP6, Data information and known issues, Data access, Contact
 - origine principale des visites : 55 % depuis www.umr-cnrm.fr, 28.2 % depuis Google
 - [le tableau de bord est beaucoup téléchargé](#)

Nombre de visites mensuelles site web CMIP6 CNRM



Suite...

- Prochaine réunion : ~~mardi 21 avril à 14h30 en salle 151~~ (à reprogrammer après le déconfinement)