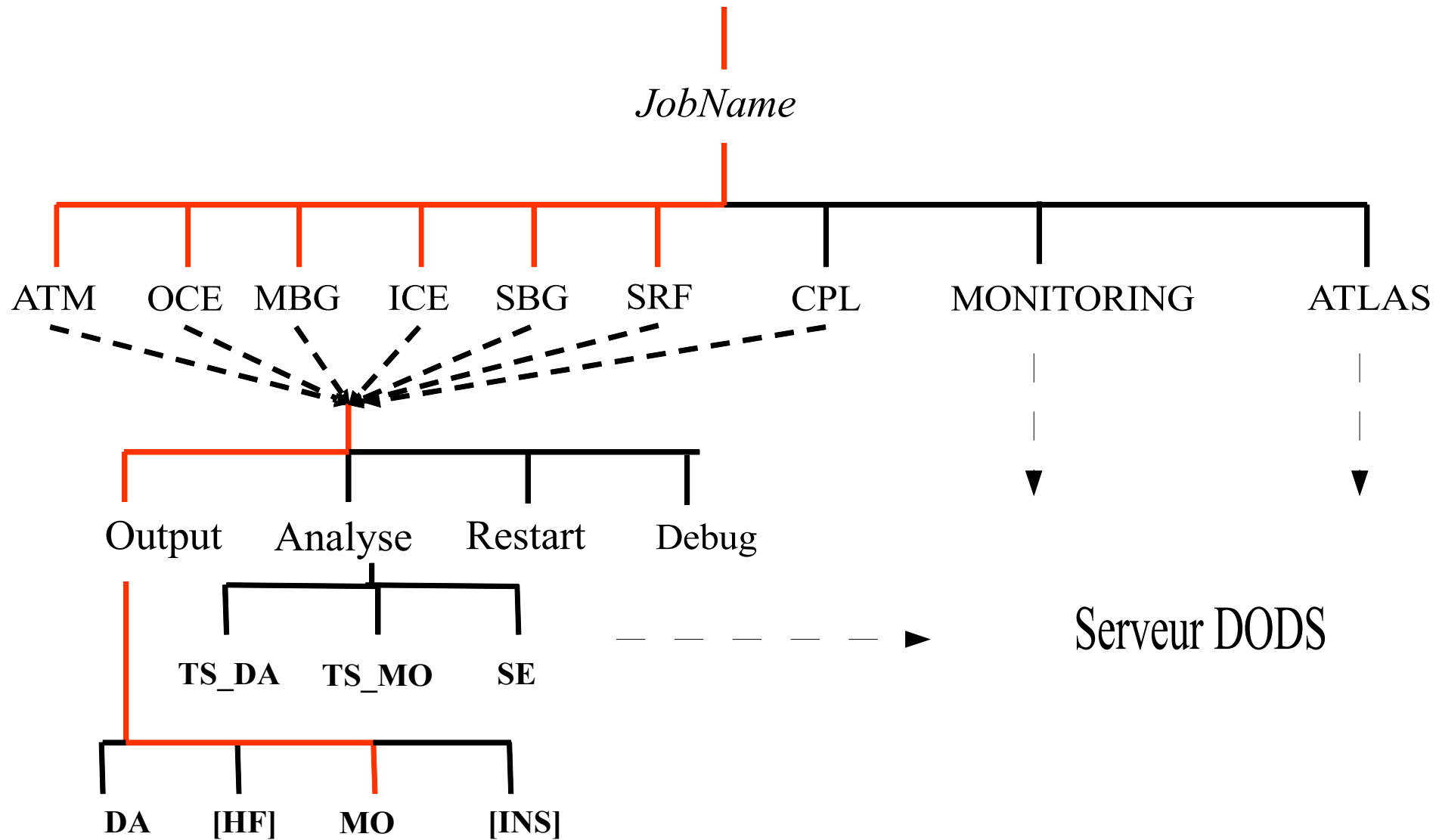


Types de fichiers à migrer

- Simulations IPSL :
 - Restart : petits fichiers netcdf ou texte (**tar sur STORE**)
 - Debug : petits fichiers texte (**tar sur STORE**)
 - Output : petits et gros fichiers netcdf (**ncrcat ou tar sur STORE**)
 - Analyse : gros fichiers netcdf (**cp sur STORE**)
 - Atlas et monitoring : petits fichiers (**cp sur WORK**)
 - Fichiers sur serveur DODS (**cp sur STORE ou WORK + lien**)
- Hors simulations IPSL :
 - ccc_archive
 - tar CCRT-TGCC

But : diminuer le nombre de fichiers et augmenter leur taille

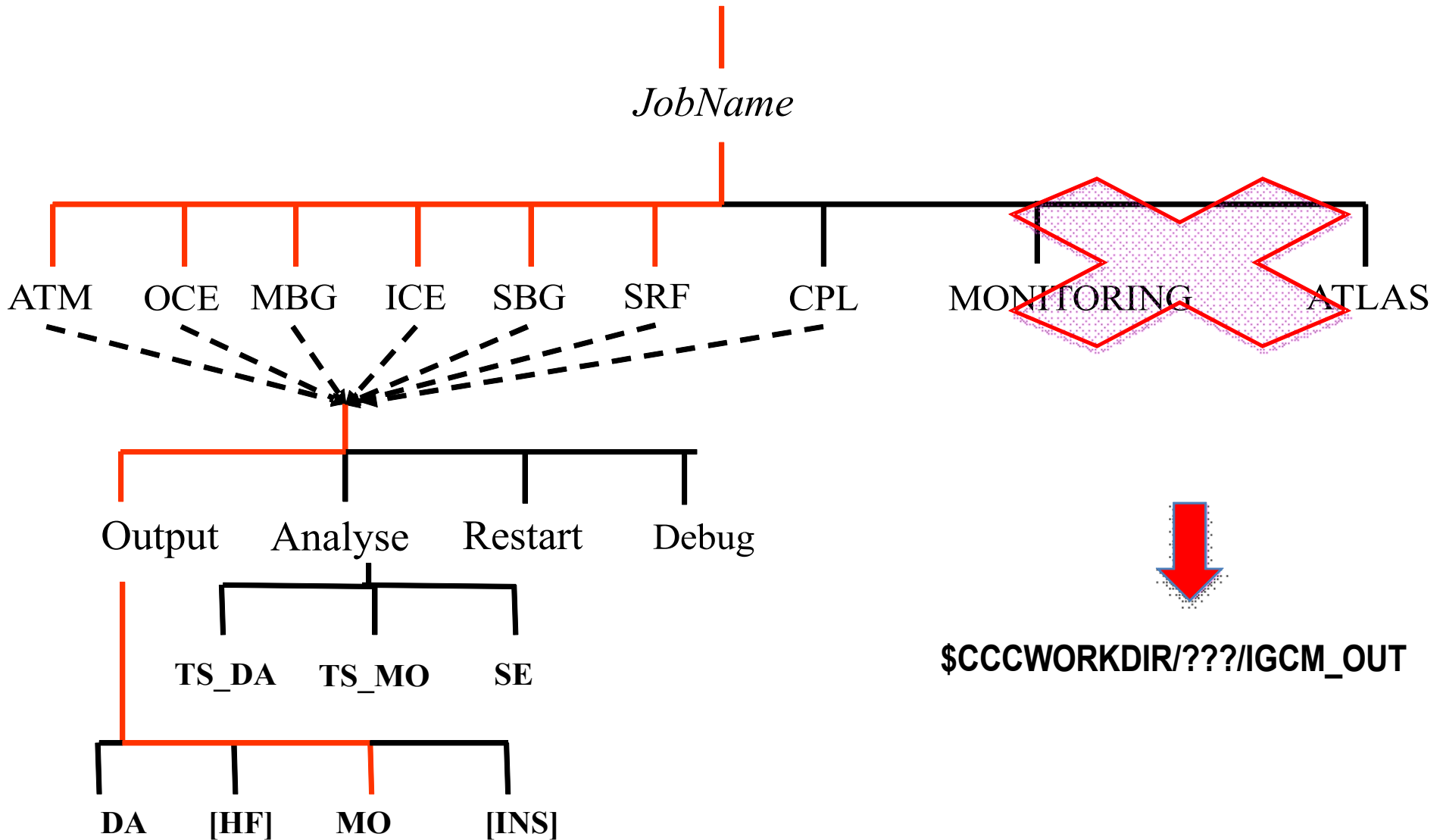
\$DMFDIR/IGCM_OUT/IPSLCM5A/DEVT/pdControl



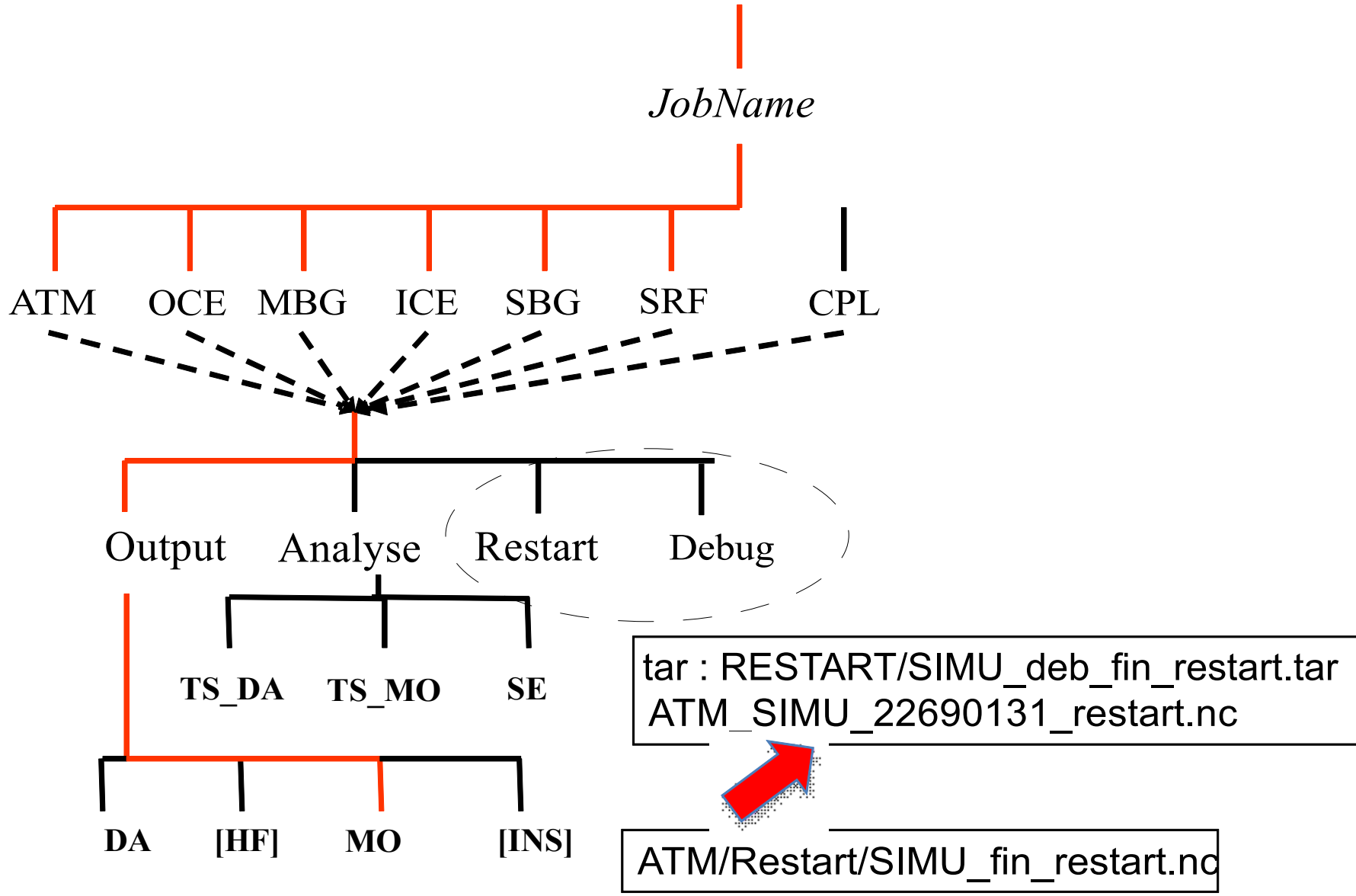
Outil : Pack_IPSL

- Traite les fichiers issus des simulations IPSL
- Regroupe les différents types de fichiers par « pack »
- Une période de pack correspond à un nombre de fichiers représentant des années de simulation.
- Une même période de pack pour tous les types de fichiers d'une même simulation
- Restart et Debug : archivage via tar
- Output : archivage via nrcat ou tar suivant les cas
- Analyse : copie telle quelle

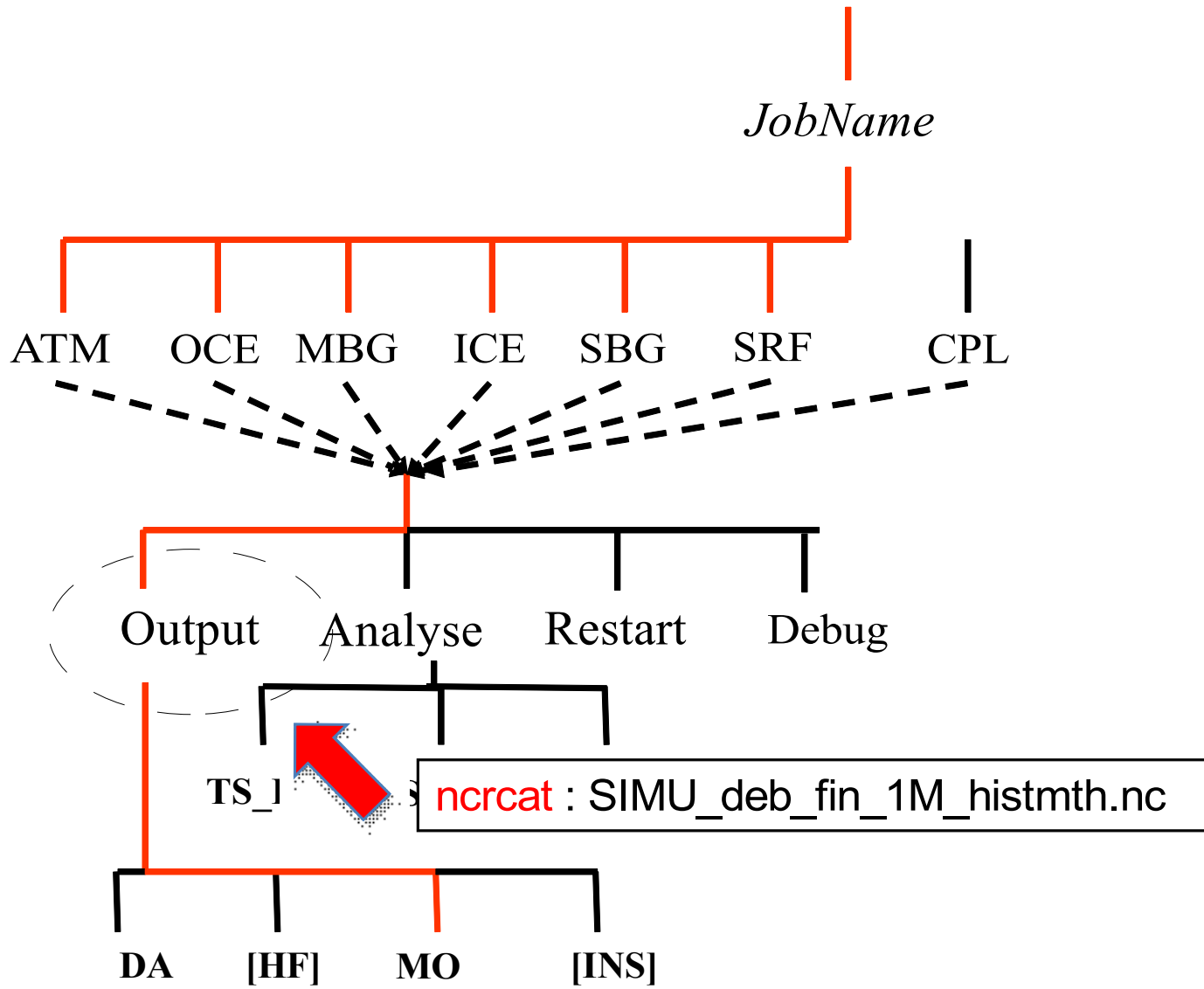
\$CCCSTOREDIR/??~/IGCM_OUT/IPSLCM5A/DEVT/pdControl



\$CCCSTOREDIR/??/IGCM_OUT/IPSLCM5A/DEVT/pdControl



\$CCCSTOREDIR/??~/IGCM_OUT/IPSLCM5A/DEVT/pdControl



Description du Pack_IPSL

- Deux parties :
 - Création des listes de fichiers à archiver
 - Création de l'archive : nrcrat (IPSL) – tar (?) - rien (cp ?)
- Déroulement du « package »:
 - Copie par le CCRT d'un dmnfs sur un espace tampon de Curie
 - Passage du pack_ipsl : création des listes de fichiers, création des différentes archives (vérifs diverses)
 - Copie par le CCRT (?) des archives sur le STORE ou WORK

Fonctions utilisées

- **find_directory_simul.sh (outil de création des listes)**
 - Objectif : pour chaque répertoire IGCM_OUT liste les simulations stockées
 - Output : créé un fichier texte liste_simul_*
- **create_config_card.sh (outil de création des listes)**
 - Objectif : pour chaque simulation listée dans liste_simul_* recherche ses caractéristiques principales :
 - Nom de la simulation (JobName)
 - Path de la simulation (PATH_SIMUL_FULL)
 - Date de début et de fin de la simulation
 - Path du répertoire de travail créé (IGCM_DEM_SIMU)
 - Output :
 - créé pour chaque simulation un fichier config_card_JobName stocké dans IGCM_DEM_SIMU
 - Un fichier texte regroupant les path de tous les config_card créés

- **find_size_pack.sh (outil de création des listes) :**
 - Objectif : pour chaque type de fichier d'Output trouver la période de pack idéale. Au final on conserve la plus petite période.
 - Output : créé dans IGCM_DEM_SIMU un fichier contenant la valeur de la période optimale
- **write_liste_pack.sh (outil de création des listes) :**
 - Objectif : créer les listes de fichiers d'Output à concaténer (ncrcat) par période de pack et par type de fichier. Si il manque un fichier dans la période alors la liste devient une liste à archiver via tar;
 - Output : créé des listes nrcrat_* et tar_* dans IGCM_DEM_SIMU
- **archive_restart.sh (outil de création des listes) :**
 - Objectif : créer les listes de fichiers Restart à archiver via tar en les renommant au préalable à la nouvelle norme de libIGCM. Puis tar ces listes.
 - Output : tar de types JobName_datebegin_dateend_restart.tar

- **Script de pack (outil de « package ») :**
 - Objectif : créer les archives à partir des listes fournies par `write_liste_pack.sh` et gérer les erreurs éventuelles des `ncrcat` (transfert des fichiers vers « liste à tarer »). Utilise `nco 4.10` pour la fonctionnalité `md5 (check)`.
 - Output : sur l'espace tampon au final
 - Test en interactif sur titane scratch : 7 ans de simu, pack de 5 ans, 2152 fichiers au départ, 50 à l'arrivée. Temps : 40 min sans check `md5`. En production : 120000 fichiers avant, 7000 (pack 1 an) et 3000 (pack 5 ans) maintenant.

Gestion de la chaîne - questions

- Création des listes :
 - Par défaut on travaille en considérant que les simulations sont stockées dans \$DMFDIR/IGCM_OUT/
 - Peut-on depuis l'espace tampon lire un fichier contenant un path alternatif pour chaque utilisateur n'ayant pas de répertoire IGCM_OUT (où ? comment?) ?
 - On utilise le listing fourni par le CCRT FROM_CCRT_INFORMATIONSDMNFS/information_dmnfs_2012-03-21
 - Impose la création de ce fichier avec la même méthode par le ccrt après migration de tout un dmnfs sur l'espace tampon (quelles infos on a ?)

Gestion de la chaîne - questions

- La chaîne : Pour une simulation
 - Création des listes
 - Pack[ncrcat | ccc_archive] ou Pack[ncrcat | tar]
 - Effacement des données traitées (Quand ?)
 - Ramasse miettes (fichiers restants sur le tampon ou fichiers restants dans la liste ?)
- Questions :
 - comment on boucle sur les différentes étapes (liste + pack) ?
 - Temps max pour un job/script (restartabilité)?
 - Gestion des plantages / surveillance ?
 - Conformance entre espace tampon et dmnsfs – comment c'est géré (liste à jour ?) ?

- Requête : avoir accès à un cas test grandeur nature (ex : p25luc) sur l'espace tampon.