



*Retour d'expérience
sur HDF
au CGTD ICARE*

HDF au CGTD ICARE

1. HDF4 : le choix du CGTD ICARE
2. Exemples d'utilisation
3. Conclusions

1. HDF4 : le choix du CGTD ICARE

Dès 2003, orientation naturelle vers un FDS pour tous les produits ICARE générés au CGTD, permettant de :

- accéder de manière uniforme aux différents produits
- s'affranchir d'une documentation spécifique de description de format
- manipuler et visualiser aisément tous types de produits
- fournir aux utilisateurs des outils génériques de manipulation des données et produits distribués par le CGTD

HDF4, HDF5, NetCDF, ... ?

Choix de HDF4 / HDF5 ou NetCDF

- HDF4 très largement répandu dans la communauté scientifique de la télédétection
- très gros volume de données utilisées et/ou redistribuées en HDF4 et HDF4-EOS (NASA : Modis, Calipso, ...)
- souci d'homogénéité dans le format des données/produits
- beaucoup plus riche que NetCDF supporté en interne
- beaucoup plus d'outils de manipulation et de visualisation que pour HDF5 (encore un peu « exotique » à l'époque...)
- peu concernés par les limitations de HDF4
- API HDF5 plus difficile à maîtriser

Les problèmes rencontrés (1)

Obstacle « culturel » :

- difficulté de s'affranchir des habitudes dans le choix d'une technologie nouvelle (ici un FDS)

Compatibilité ascendante pas toujours assurée :

- certains fichiers produits avec la version 4.2r1 ne peuvent pas être relus avec la version 4.2r2
- effets de bords indésirables rencontrés lors de l'utilisation d'outils intégrant des versions différentes (MATLAB, HDFView, Python, IDL)

Les problèmes rencontrés (2)

Difficultés liées à la compression des SDS :

- impossibilité de compresser un SDS déjà existant
- impossibilité d'étendre un SDS comportant une dimension illimitée s'il est compressé
- accès direct inefficace avec des SDS compressés (bufferisation des données)
- le « chunking » avec compression nécessite une conception d'écriture non triviale

Les problèmes rencontrés (3)

Suppression et/ou modification d'objets :

- impossibilité de modifier la taille d'un objet (sauf SDS défini avec une dimension illimitée)
- impossibilité de modifier la structure d'un objet (modification du nombre ou de la valeur de ses dimensions)
- impossibilité de supprimer un objet

Les problèmes rencontrés (4)

Visualisation :

- impossibilité d'attacher une palette de couleurs à un SDS

Support de NetCDF :

- NetCDF, qui évolue encore, est embarqué dans HDF qui n'évolue plus ; conflits possibles en utilisant les API NetCDF et HDF en même temps (création NetCDF)

2. Exemples d'utilisation

Quelques exemples

Développement d'outils et génération de produits grandement facilité par les aspects d'abstraction, d'auto-description et d'auto-documentation de HDF :

- fonctions de lecture générique d'un SDS d'un fichier et/ou d'attributs pour les langages les plus utilisés (C, C++, Fortran 77, Fortran 95, Python, IDL)
exemple : `sds = lire_sds(fichier, nom_sds)`
- `ncf2hdf`, `hdf2raw`, `raw2hdf` : conversions de/vers HDF
- `Hdf2plot`, `Hdf2img` représentation graphique de SDS
- `HDFCompressor` : compression des SDS d'un fichier
- `Geo2Hdf`

Geo2Hdf (1)

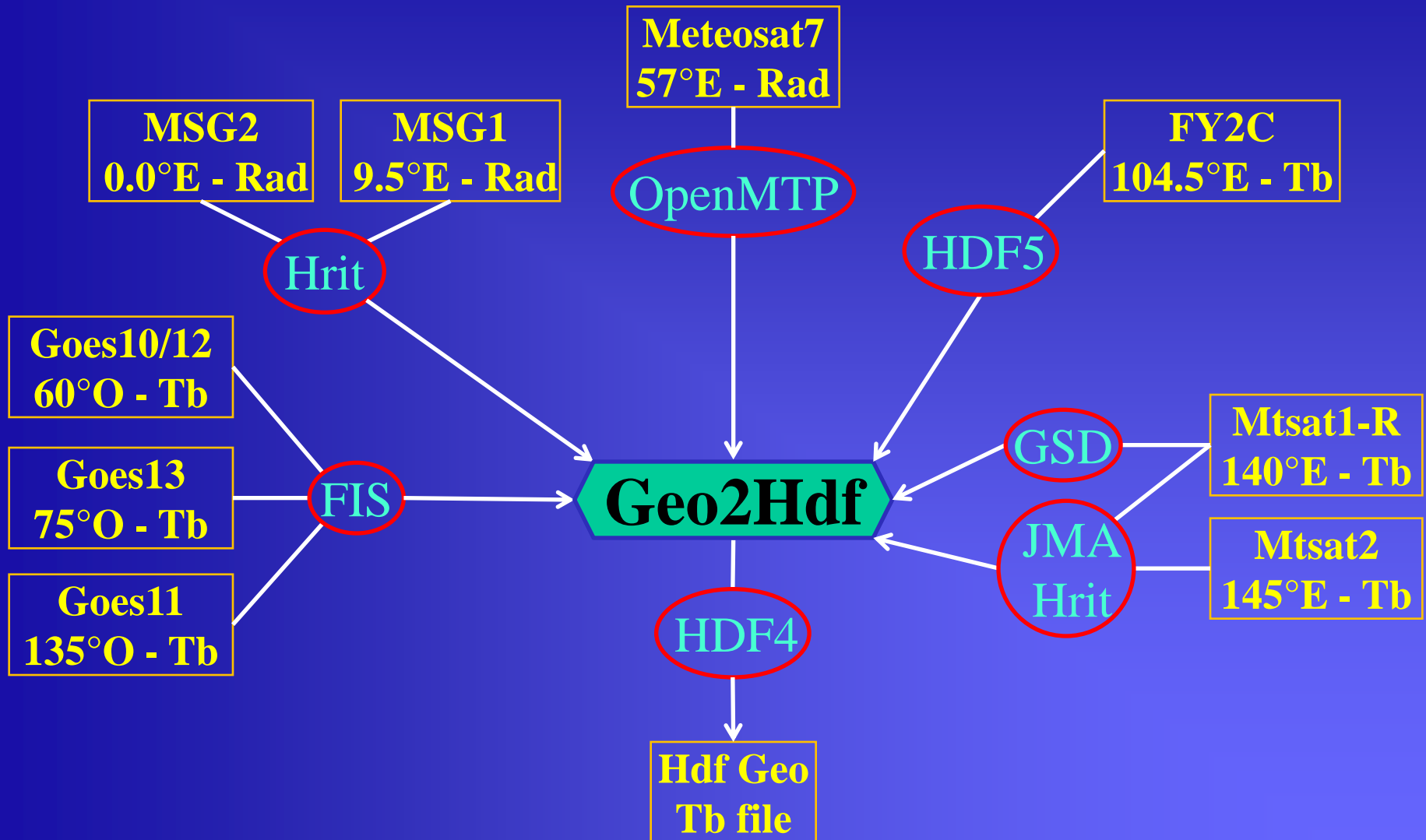
Le CGTD récupère les données de niveau 1 de nombreux satellites géostationnaires, les exploite et les redistribue.

Multiples différences :

- formats (FIS, GSD, Hrit, JMA-Hrit, HDF5, OpenMtp), mono ou multi-fichiers
- grandeurs physiques : radiance, température de brillance
- compression en entier (facteur, pente + offset, table)
- ...

Geo2Hdf : reproduction de ces données dans un unique modèle au format HDF, qui alimente ensuite de manière uniforme redistribution et différentes chaînes de traitement

Geo2Hdf (2)



Recommandations HDF au CGTD (1)

L'expérience montre que même avec un unique FDS, une certaine hétérogénéité persiste dans les produits générés par différentes personnes ou équipes.

Le CGTD a donc établi une série de recommandations pour la constitution des produits ICARE au format HDF.

- **attributs de fichier systématiques** :
 - caractéristiques de production : identification, version, description, date et lieu de production, version HDF, ...
 - caractéristiques spatiales, morphologiques, temporelles
 - informations sur les attributs « spécifiques »
- **attributs de fichier recommandés** :
 - capteurs/satellites sources
 - fichiers d'entrées, fichiers annexes

Recommandations HDF au CGTD (2)

- attributs de SDS prédéfinis (si pertinents) :
 - nom, description, unités, « valid range », « fill value »
 - coefficients de calibration HDF
- attributs systématiques (si pertinents) :
 - équation de la calibration HDF
 - attributs liés à la qualité
- jeux de données recommandés :
 - latitude et longitude

Document de recommandations en constante évolution

3. Conclusions



Après 7 ans, HDF4 continue de satisfaire nos besoins, en dépit de ses limites et des divers problèmes apparus

- plus d'efficacité dans le développement des interfaces de sortie (et d'entrée le cas échéant)
- contrôle des produits de sortie par le développeur puis par l'utilisateur beaucoup plus rapide et facile
- accès uniforme aux méta-données utiles à l'archivage (dates, géolocalisation, ...)
- produits composites en un seul fichier



Néanmoins, certaines des limitations de HDF4 deviennent maintenant un peu plus gênantes.

- l'orientation de plus en plus évidente vers le parallélisme risque de nous faire évoluer vers HDF5
- HDF5 est souvent choisi aujourd'hui pour les produits satellitaires (*Megha-Tropiques, NPOESS, NCMRWF, Ifremer, ...*)

Le *HDF Group* lui-même préconise fortement HDF5 pour les nouveaux utilisateurs et conseille la migration aux utilisateurs de HDF4



Quelques obstacles à une migration globale :

- volumétrie de l'archive HDF4 ICARE
- quantité de chaînes de traitements et outils liés à HDF4
- même problème « culturel » pour abandonner HDF4 que pour l'adopter

Moralité :

« HDF5 quand ça s'impose... ou quand c'est imposé ! »