

Pourquoi cette réflexion ?

- 1999 : Rapport Waldteufel (« Bases de données pour les géosciences »)

Référence pour la mise en place de pôles thématiques OT

- Depuis 13 ans, de nombreuses évolutions

- scientifiques :

- approches de plus en plus intégrées (système Terre, compartiments)
- nouveaux systèmes d'observation, besoin de continuité
- utilisation des observations dans les systèmes de surveillance environnementale

- techniques

- volumes & diversité des données, traitement, archivage, accès
- partage d'expertise

- institutionnelles



- national : nouvelles logiques programmatiques
- européen : services Copernicus, infrastructures de recherche, ESA, EUMETSAT
- international : coordinations GEO, OMM, etc...

Besoin de lignes directrices à moyen / long terme

Le mandat du groupe

- **Groupe multi-organismes**
 - mandat défini par CNES et CNRS/INSU
 - recommandations sur actions à mettre en place pour répondre aux besoins des communautés nationales dans un contexte national, européen & international

- **Le mandat inclut**
 - **une analyse :**
 - état des lieux national, européen, international
 - besoins/demandes à long terme des utilisateurs nationaux
 - des propositions sur
 - les schémas de coopération inter-organismes
 - une approche d'intégration européenne, voire internationale
 - un plan d'actions : priorités, moyens, acteurs

Le groupe de réflexion

■ Membres

- désignés intuitu personae
- couvrent un ensemble de compétences thématiques & techniques

■ Composition

- G. Bergametti, président TOSCA
- Ph. Bertrand, CNRS/INSU
- M. Diament, CNRS/INSU
- J.-P. Gleyzes, CNES
- S. Hosford, CNES
- N. Papineau, IPSL
- M. Pontaud, Météo-France
- D. Roumigières, CNES
- F. Vial, CNRS/INSU
- J.-P. Wigneron, INRA
- A. Podaire, CNES – animateur

Autre participant : F. Genova, INSU

Périmètre & méthode de travail

■ Périmètre

- les observations concernées couvrent tous les compartiments terrestres & sont acquises depuis l'espace, *in-situ*, en avion, ballon, etc...
- ce sont des données numériques...
- les questions liées aux infrastructures d'observation ne sont pas traitées : on considère que les données sont acquises
- l'utilisation des produits en aval des centres de données (modélisation, surveillance, valeur ajoutée) est prise en compte en termes de besoins – pas de moyens...



■ Méthode de travail :

- ◆ le groupe privilégie la consultation - **plus de 60 auditions depuis décembre 2011**
 - des parties prenantes : communautés de recherche, « grands comptes », organismes, producteurs de données, secteur privé...
 - avec parfois nécessité d'échantillonner age pour la partie « analyse »
- ◆ le rapport comprendra des propositions & **une feuille de route**
- ◆ il sera délivré à nos mandataires mi-2013



Qu'est-ce qu'un pôle de données ?

Mission & fonctions

~~pôle thématique~~

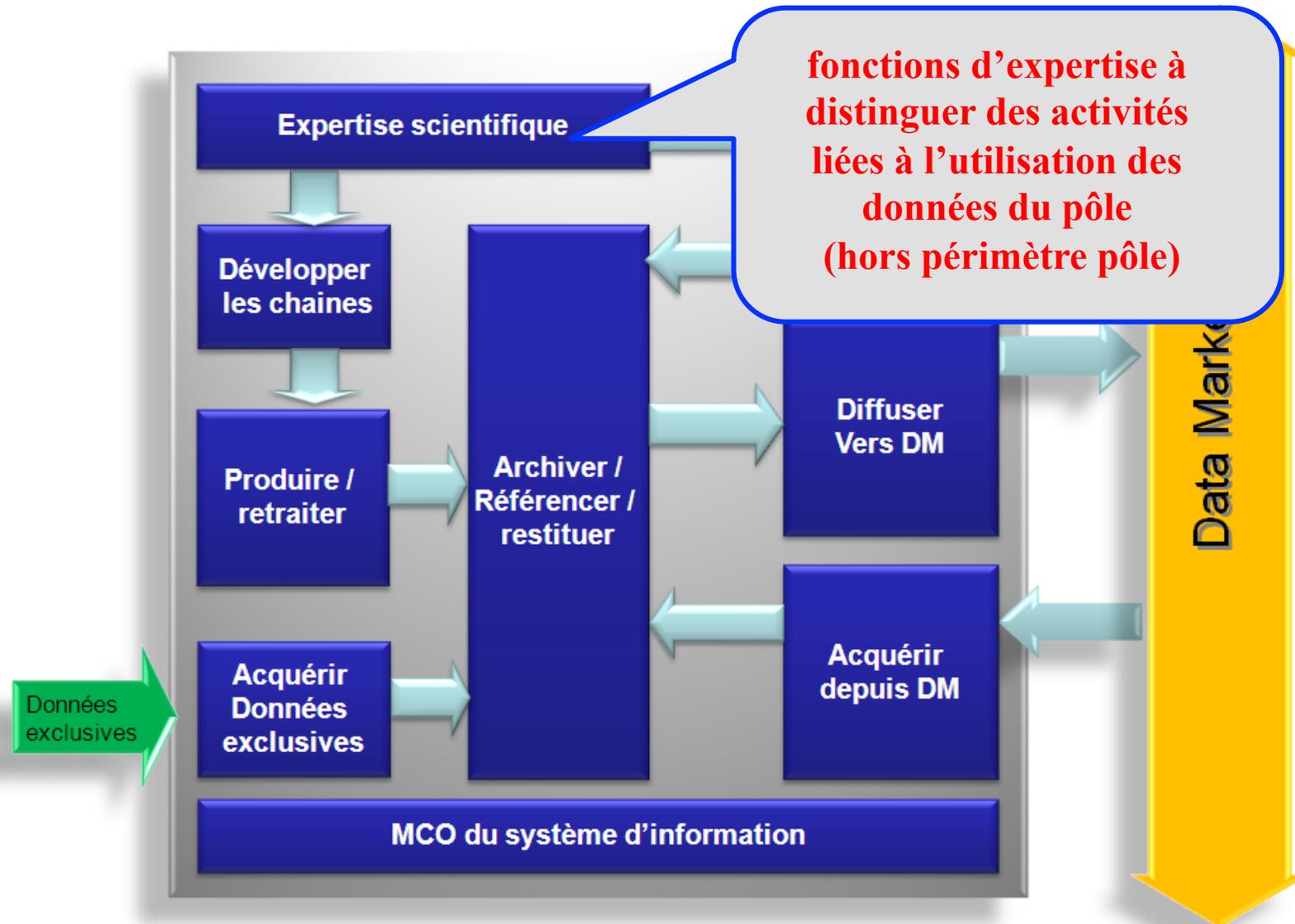


pôle de données

La mission d'un pôle de données...

- Fournir des services et des informations
 - pour la communauté scientifique française
 - pour d'autres utilisateurs : internationaux, publics, privés, éducation, ...
- Ces services incluent
 - l'accès à des jeux de données dont certains sont produits/ retraités / validés... par le pôle
 - l'aide à la découverte de données
 - la diffusion d'outils logiciels (e.g. visualisation ou intégration des données, changement de format, etc.) et de documents
 - un support d'expert sur les données et outils diffusés
 - la fourniture de ressources : hébergement et exécution de codes de calcul, archivage de données, publication de références, etc
 - l'hébergement d'une communication éditoriale
 - l'hébergement de plateformes collaboratives (Wiki), forum, blogs, partage et échange de documents...

Les fonctions d'un pôle de données



Architecture & organisation

■ Architecture

- besoin d'un ensemble de **ressources** :
 - personnels, infrastructures
 - centralisées ou distribuées
- un pôle peut intégrer **plusieurs centres de données**
 - liens, consignes & règles entre le pôle et les centres de données

■ Organisation

- Périmètres et fonctionnement du pôle bien identifiés
- Prise de décision par les institutions qui financent
- Réalisation des tâches par personnel dédié
- Fonctions expertise contractualisées

Les besoins

- Approches de plus en plus intégrées sur système Terre ou ses compartiments
- Données & produits d'observation
 - essentiels pour les recherches
 - multi-variables, multi-sources
- Maximiser l'utilisation des données pour les communautés travaillant
 - sur la donnée (e.g. physique de la mesure)
 - avec la donnée (les produits des pôles de données)ces besoins peuvent être très différents....

Ne pas dupliquer ce qui est fait ailleurs
par opérateurs nationaux & internationaux
→ les communautés de recherche ont besoin
des données de ces opérateurs

Les besoins

- Accès à des données multi-variables et multi-sources, bien qualifiées et documentées,
- Du produit basique au produit élaboré
- Des longues séries temporelles avec cohérence et qualité quantifiées

- Benchmarking des produits nationaux
- Lien entre observations et modèles : analyses d'impact, simulateurs d'observation

- Harmonisation des standards (formats de données, métadonnées)
- Mutualisation des catalogues
- Traçabilité des données : production et expertise
- Besoin d'archivage de données : croissance des volumes, diversité des variables et des sources

Spécificité de l'imagerie haute résolution spatiale : licences, redistribution, synergie possible entre différentes utilisations,

L'existant...

- **De grandes différences d'approches et de structuration**
 - **des centres de données ou « pôles thématiques » fonctionnent déjà**
 - différences d'approches constatées entre services d'observation (SO) & centre de données spatiales
 - **des centres de données ou « pôles thématiques » sont en cours de mise en place**
- **Ce qui existe fonctionne, répond à des besoins & s'appuie sur des équipes motivées**



Une règle : partir des structures & moyens existants pour répondre à l'évolution des besoins – et aux améliorations nécessaires

Améliorer l'existant...

- Prendre en compte l'évolution des besoins : produits, contrôle qualité, retraitement, mise en cohérence
- Plus d'intégration des données et produits
- Simplifier l'accès aux données

- Traiter le problème d'archivage des données d'observation et produits dérivés
- Revoir certains processus de développement des chaînes de traitement : plus d'adaptabilité

- Simplifier l'organisation des pôles/centres de données & les mécanismes de décision
- Optimiser les moyens (masses critiques) affectés aux pôles & centres de données
- Améliorer visibilité / représentation externe
- Améliorer l'intégration européenne (et internationale) des pôles

Perspectives & stratégie



L'Europe

une évolution nécessaire

- un accès coordonné aux données acquises sur financement public (portails)
- **c'est une opportunité**
 - être acteurs de l'espace européen de la recherche
 - préparer les futurs services de surveillance de la Terre
 - mettre en valeur les compétences – renforcer les infras & l'expertise nationales
- cela induit des contraintes : standards, réciprocité (travailler avec les autres), non duplication



La stratégie nationale

- Des contributions « naturelles » & des positions européennes à prendre
- S'intégrer au mieux en Europe : définir des trajectoires pour la convergence / compatibilité
- Faire porter l'effort sur certaines fonctions des pôles de données (accès)
- Consolider les points forts
- Préparer l'accès libre & gratuit aux données
- Utiliser les nouvelles technologies



Propositions d'organisation

Périmètre – architecture - gouvernance

P1 – La stratégie européenne est un préalable...

■ Définir et délimiter :

- ce qui est **fait au niveau national pour des utilisations limitées au national**
- ce qui est **fait au niveau national pour des utilisations nationales et européennes**
- ce qui est **fait au niveau européen et est utilisé au niveau national** – donc pas dupliqué au niveau national

■ Une stratégie d'ensemble déclinée pour chacun des compartiments du système Terre

- définir les positions (fonctions) nationales relevant du niveau européen, en prenant en compte les contraintes (notamment de pérennisation de moyens) et les engagements institutionnels associés.
- ne pas remettre en cause les positions acquises au niveau européen : s'appuyer sur cette base

Stratégie nationale à définir avant mi-2014
Prise en compte des articulations existantes ou à établir
avec le niveau international, au delà de l'Europe

P2 – Vers 4 pôles de données nationaux

- Mettre en place **quatre pôles de données nationaux couvrant l'ensemble des observations (spatiales & *in-situ*) pour l'atmosphère, l'océan, les surfaces continentales et la terre solide**
- Cette organisation répond à :
 - **un besoin des utilisateurs recherche (et application)** : un portail unique par pôle pour les données gérées par le pôle & les liens externes
 - **une meilleure gestion des capacités & ressources** : mutualisation, masses critiques, utilisation de nouvelles technologies (cloud, etc)
 - **une gouvernance & un exécutif simplifiés** : moins de comités, plus de responsabilités
 - **une meilleure configuration pour traiter l'intégration européenne & internationale** : visibilité, représentation, cohérence avec les autres structures nationales (par exemple Royaume Uni)

Proposition = objectif à moyen/long terme

Logique d'intégration de l'existant

Mise en œuvre dépendante de chaque pôle : histoire, spécificités..

P3 – Une gouvernance & un exécutif plus lisibles

Pour chaque pôle

Décision

- un **comité directeur**
 - les institutions contribuant aux ressources du Pôle
 - il décide
- un **conseil scientifique**
 - des experts & représentants des programmes nationaux
 - il s'occupe de la stratégie scientifique du pôle, garantit l'état de l'art des produits, traite les retours utilisateurs

Exécutif

- un directeur
 - c'est un scientifique nommé par le comité directeur
 - il dirige le pôle
 - il le représente
- des équipes techniques & scientifiques
 - elles réalisent
 - elles sont co- ou délocalisées

Inter-pôles

- un **comité de pilotage** – les organismes
 - Un **comité technique**

P4 – Différents niveaux de mutualisation...

■ Pourquoi mutualiser ?

- pour une meilleure utilisation des ressources
- pour distribuer des données non accessibles actuellement

■ Plusieurs niveaux de mutualisation

– niveau national

- coordination des fonctions du pôle
- structure & organisation spécifiques

– niveau décentralisé

- fonctions contribuant aux activités nationales
- rôle spécifique des fédérations de laboratoires (e.g. OSU)

Pôles = structures indépendantes de type UMS

Pas de confusion entre moyens affectés aux pôles de données & ceux – tout aussi nécessaires – pour l'utilisation des données, à affecter aux unités de recherche ou à mutualiser dans les fédérations de laboratoires

P7 – Formation & communication

- Valorisation des données / produits des pôles pour
 - la formation (secondaire ou universitaire)
 - la diffusion grand public
- Intervention possible des pôles en cas d'événement catastrophique ou de crise
- Prise en compte des capacités des réseaux sociaux

**La formation & la communication des pôles
font partie des fonctions des pôles
Des moyens spécifiques doivent être attribués**



Spécificités

Pour chaque pôle & inter-pôles

Éléments spécifiques

■ Atmosphère

- Egale importance des données d'observation spatiale & in-situ
- Le Pôle doit être bâti à partir de l'existant sur ETHER & ICARE
- Les différentes composantes du pôle doivent être définies autour d'une approche fonctions/métiers
- Le lien entre les bases de données du pôle & les autres bases de données, notamment opérationnelles, doit être clarifié

■ Océan

- Un seul pôle multi-organismes pour les données spatiales & in-situ
- Prise en compte des données des domaines hauturier, côtier et littoral, et les préoccupations de la recherche et de l'océanographie opérationnelle
- Prise en compte des centres de données primaires (traitement & expertise) pour les observations in-situ des services d'observation

**Pôles atmosphère & océan
pilotes pour la nouvelle configuration**

Éléments transverses

- **La coordination inter-pôles doit traiter :**
 - les interfaces entre les pôles, par exemple pour les jeux de données situés dans plusieurs pôles
 - la cohérence technique & les développements conjoints
- **Les données pour le suivi climatique**
 - Traiter dans chaque Pôle les séries temporelles de données d'observation (dont les ECV)
 - Définir et dimensionner dans le cadre de la reconfiguration des Pôles les retraitements à appliquer aux séries temporelles, et les analyses de qualité associées (étalonnage / validation & leur documentation)
 - Analyser le lien entre les séries climatiques et les services climatiques en cours de mise en place

**Les observations pour le suivi climatique
sont un élément essentiel
de la coordination inter-pôles & de la stratégie européenne**



Propositions techniques

Ce qu'il faut améliorer et anticiper...

L'accès aux données : une fonction à améliorer

- Une **interface stratégique** avec les utilisateurs des pôles (les identifier, les caractériser)
 - caractérise la facilité d'accès aux données des pôles
- Une fonction **souvent mal structurée ou négligée au niveau français** – e.g. multiplication des portails

Il faut

- développer des **systèmes dynamiques d'accès aux données** (mise à jour des catalogues, informations)
- en **conformité avec des standards** (INSPIRE, ...), **principes** (GEOSS) ou **usages de facto**
- utiliser les **Digital Object Identifiers (DOI)**.
- **simplifier et harmoniser** – quand c'est nécessaire – **l'identification des utilisateurs...**

L'archivage : une fonction cruciale à gérer au niveau national

- **Des besoins croissants :**
 - les volumes & la diversité des données d'observation croissent de manière très importante – avec effet cumulatif
- **Il faut distinguer l'archivage**
 - à court / moyen terme : **stockage – toutes les données doivent être archivées**
 - à long terme : **pérennisation**
 - **demande des moyens spécifiques**
 - **doit être instruite au niveau national**

**Vers un centre national en charge de
l'archivage pérenne des données**