

Entrepôt de données géolocalisées

14^{ième} journée du projet CASCIMODOT
Marie-Dominique Van Damme
30 juin 2011



INVENTAIRE FORESTIER
NATIONAL



Missions de l'IFN



Établissement chargé de réaliser l'inventaire permanent du patrimoine forestier sur tout le territoire métropolitain indépendamment de toute question de propriété.

Parmi ces objectifs:

- Connaissance de la ressource forestière et de son évolution dans l'espace et dans le temps.
- Publication de résultats
- Indicateurs de gestion durable
- Études de ressources



Etude de faisabilité visant à mettre en place un entrepôt de données sur les données de l'IFN.

⇒ Produire des résultats statistiques

⇒ Analyser et Explorer avec une grande interactivité

⇒ Enrichir la modélisation dimensionnelle en tenant compte des axes géométriques, produire des cartes





1. De OLAP à SOLAP
2. Entrepôt de données à l'IFN
3. Résultats et Interface Web OLAP Cartographique



De OLAP à SOLAP

Entrepôts de données



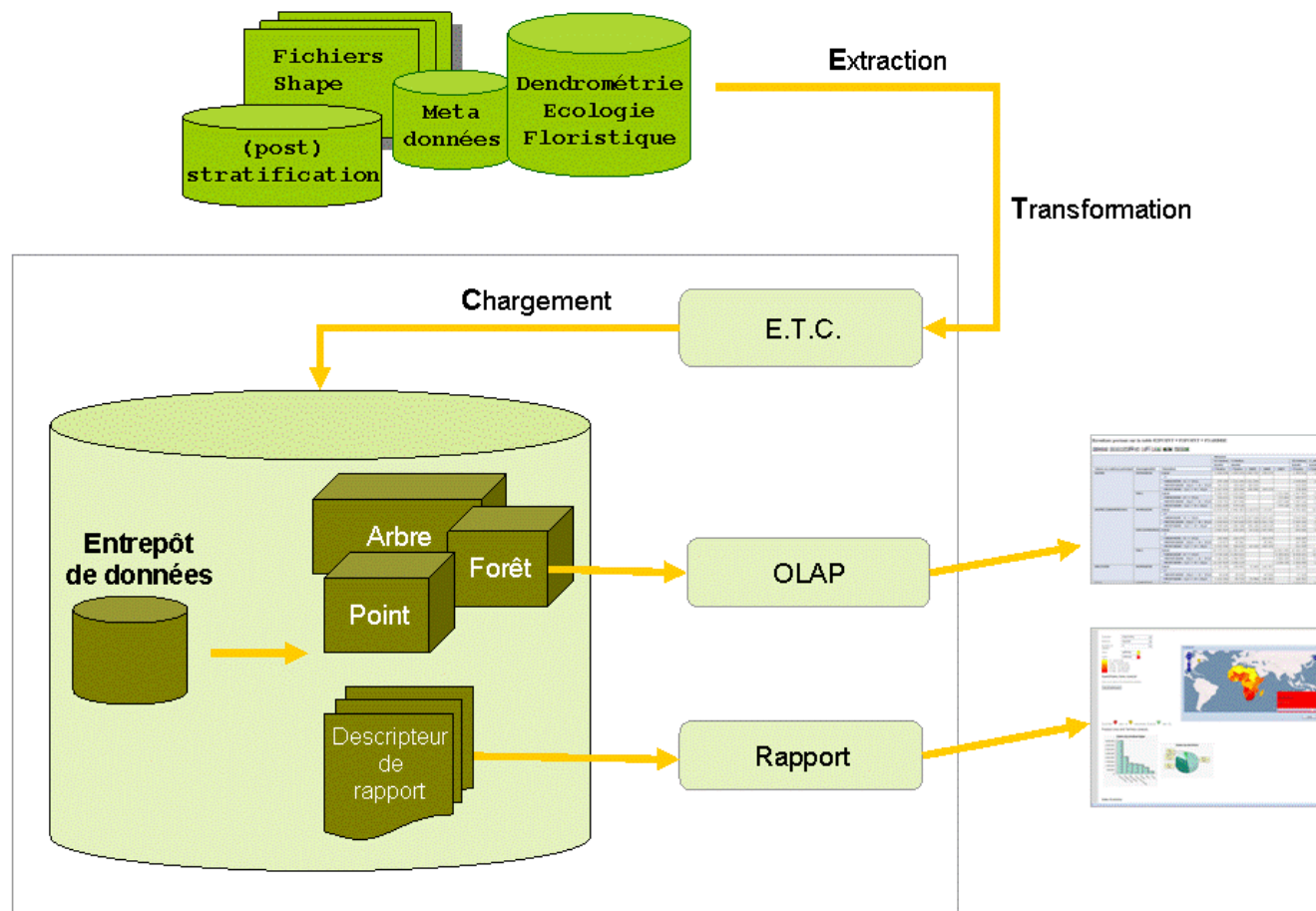
Un entrepôt de données est un ensemble de technologies destinées à permettre à une personne qui manipule des connaissances de prendre des décisions bonnes et rapides.

Ralph Kimball

- Utilisé dans le cadre de l'informatique décisionnel
- Supporte le traitement analytique en ligne : **OLAP**
- Organisation des données selon le modèle multidimensionnel
- Applications des entrepôts de données :
 - Rapports et outils d'interrogation
 - Fouille de données



Architecture des systèmes décisionnels



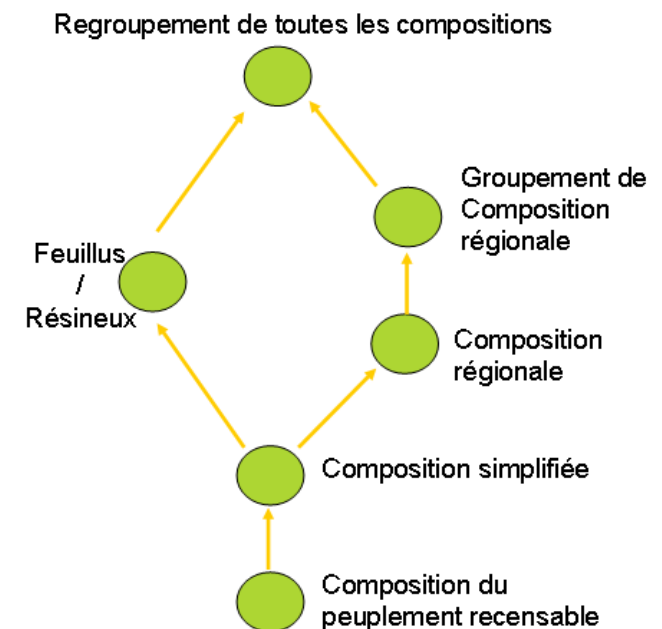
Approche multidimensionnelle 1/2



- La structure de base de toute application multidimensionnelle est la **dimension**.
- Une dimension est une liste complète d'éléments d'entrée et d'éléments calculés ou dérivés.

Elément = donnée qualitative

- Dans le cas d'une dimension hiérarchique, nous considérons la **hiérarchie** toute entière, et toutes les hiérarchies s'il y en a plusieurs, comme une seule et même dimension.



Approche multidimensionnelle 2/2



- Un **fait** est décrit par plusieurs mesures.
- Les **mesures** représentent usuellement des valeurs numériques qui fournissent une description quantitative du fait
- Certaines mesures peuvent être calculées à partir d'autres mesures ou propriétés de membres.
- Un fait est associé à une ou plusieurs combinaisons de membres des dimensions.

Arbre Forêt
VOL
NBTIGE
G
VOL_Q1
VOL_Q2
PV
BIOM

Table de fait





MOLAP : structure de stockage en cube :

- Les données pré-agrégées et pré-calculées sont stockées.
- Munie de techniques d'indexation et de hachage.
- Excellent temps de réponse.

ROLAP : structure dans des SGBDR avec un moteur supplémentaire OLAP :

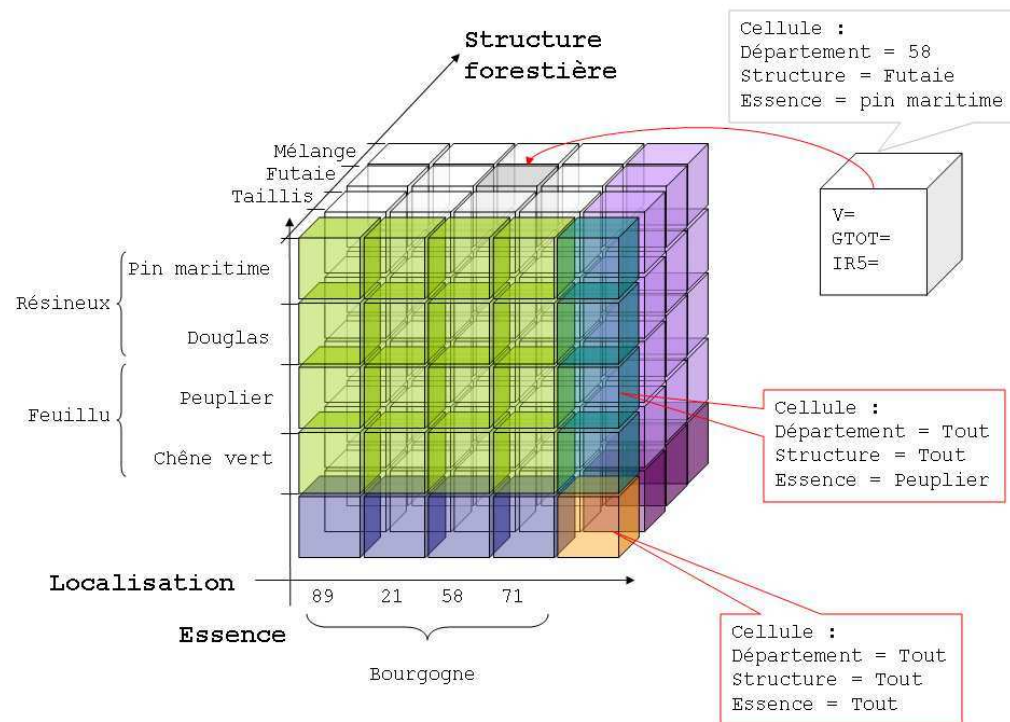
- Fournit une vision multidimensionnelle
- Des calculs dérivés et des agrégations à différents niveaux.
- Génère les requêtes SQL mieux adaptés au schéma de l'entrepôt par une indexation spécifique et des vues matérialisées.



Hypercube



La représentation physique d'un modèle multidimensionnel s'appelle un **hypercube**, on parle de **cube**.



Les opérations sur les données multidimensionnelles les plus fréquentes sont :

- Agrégation ou Roll-Up
- Désagrégation ou Drill-Down
- Projection ou Slicing
- Sélection ou Dicing
- Réorientation ou Rotate (Pivot)



Alliance des applications décisionnelles et des outils SIG

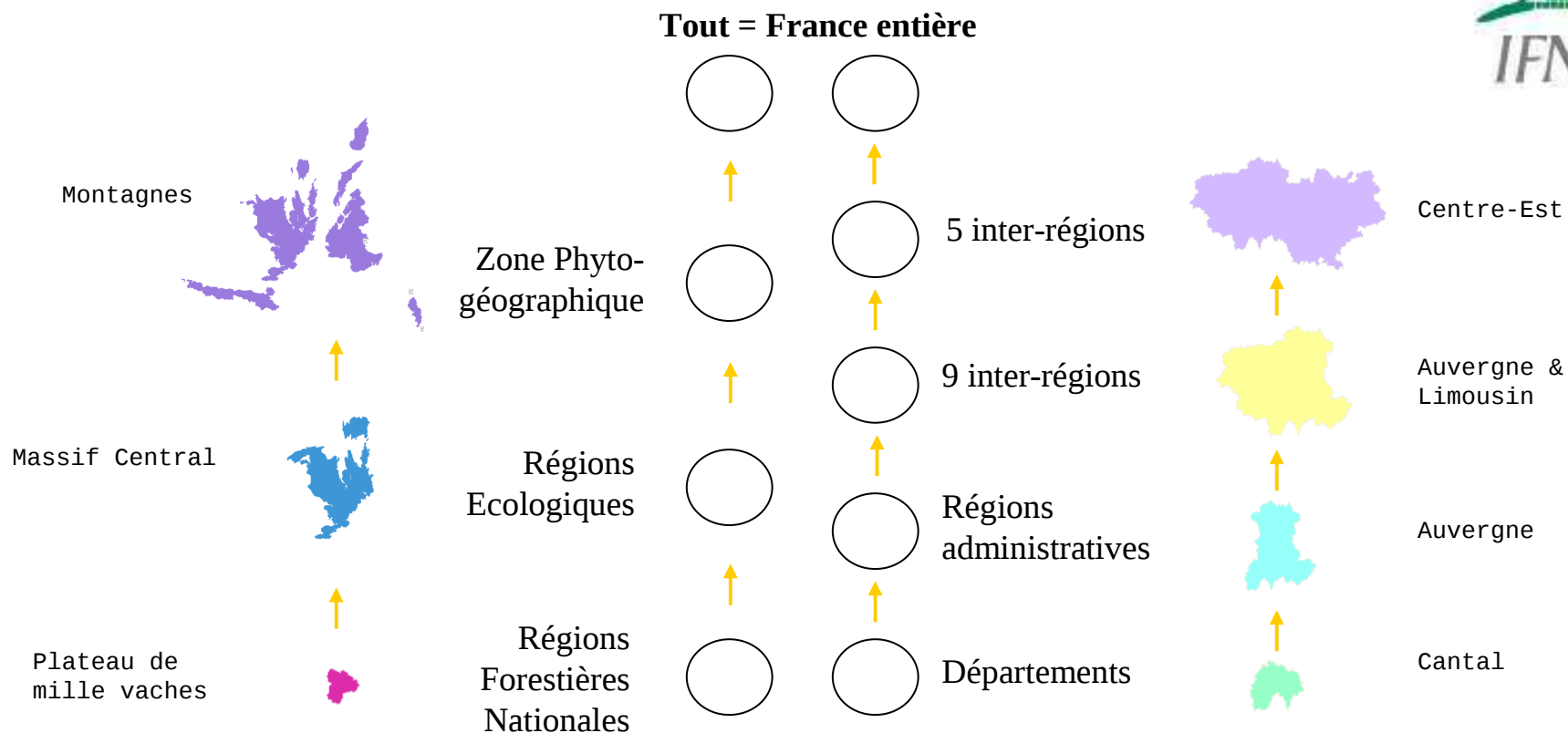
⇒ exploiter la dimension spatiale

⇒ associer la navigation OLAP à la représentation cartographique

Le concept SOLAP s'appuie sur l'ajout de la dimension spatiale au concept de l'OLAP avec adaptation des opérations d'analyses, enrichissement de l'interface et affichage cartographique.



Dimension spatiale



Dimension : Localisation écologique

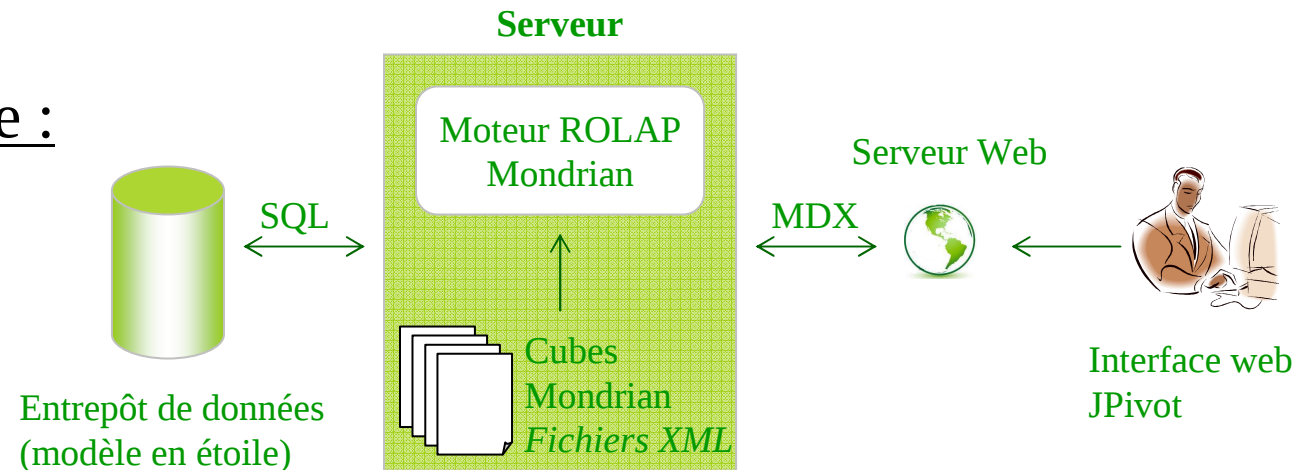
Dimension : Localisation administrative

Mondrian ou Pentaho Analysis



- Serveur web J2EE
- Moteur OLAP qui permet de concevoir des cubes multidimensionnels.
- Serveur ROLAP
- Référence open source
- Composant indépendant

Architecture :



GeoMondrian



- Basé sur Mondrian, réalisation de l'équipe GeoSOA
- GeoMondrian est le premier serveur open source spatial Olap
- Il implémente les types de données géométriques

Exemple d'une requête avec un filtre spatial sur les membres d'une dimension basée sur la fonction distance :

```
Select {[Measures].[Population]} on columns,  
      Filter (  
          {[Unite géographique].[Region économique].members},  
          ST_Distance([Unite géographique].CurrentMember.Properties("geom"),  
                    [Unite géographique].[Province].[Ontario].Properties("geom")) < 2.0  
      ) on rows  
From [Recensements]  
Where [Temps].[Recensement 2001 (2001-2003)].[2001]
```





Entrepôt de données à l'IFN

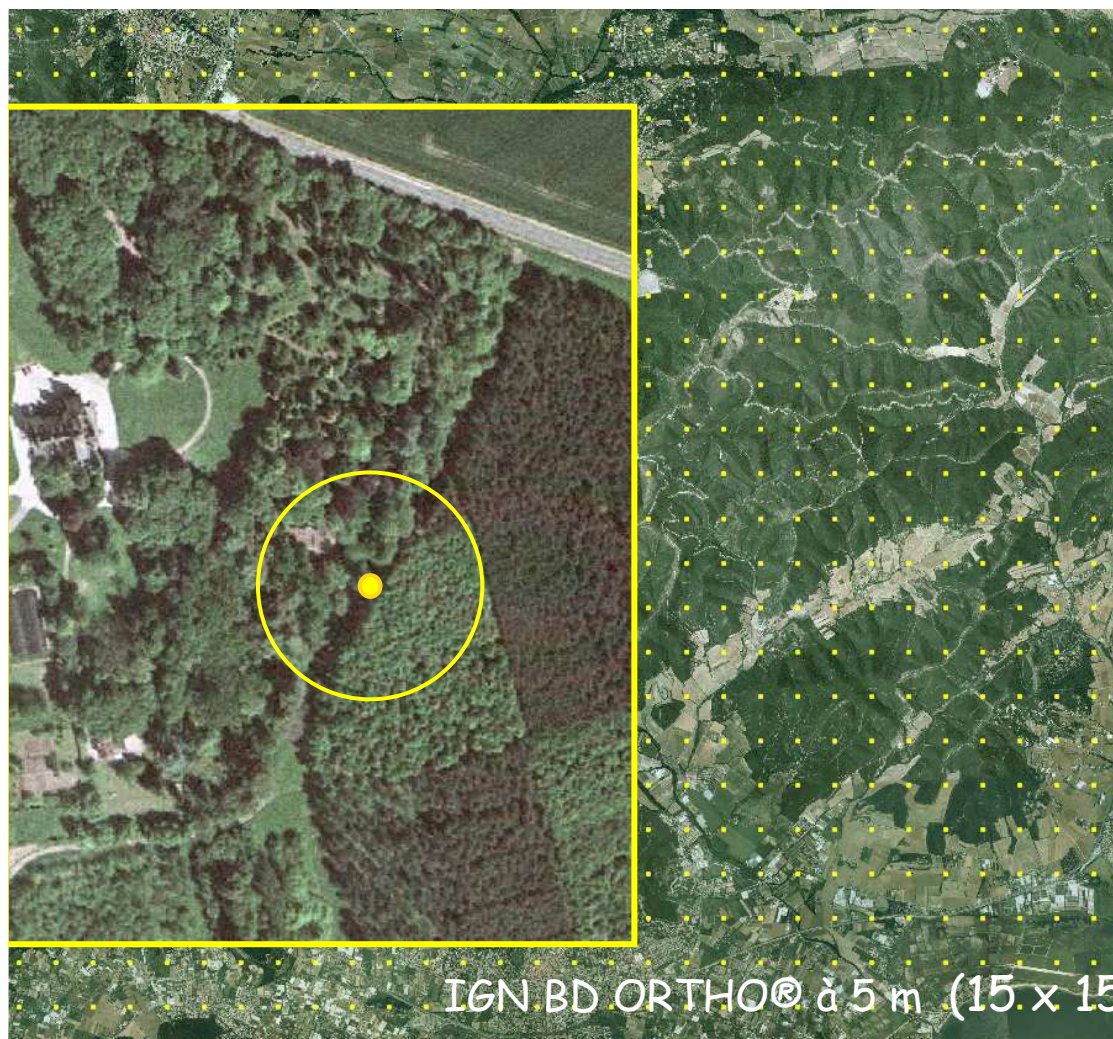
Méthode statistique : photo-interprétation ponctuelle



- Un échantillon sur l'ensemble du territoire métropolitain
- 1^{ière} subdivision : points photo-interprétés (couverture du sol, utilisation, ...)

**Échantillonnage
systématique
sur BD ORTHO®**

Echantillon de phase 1 complet :
80 000 points d'inventaire par an



IGN BD ORTHO® à 5 m (15 x 15 km)

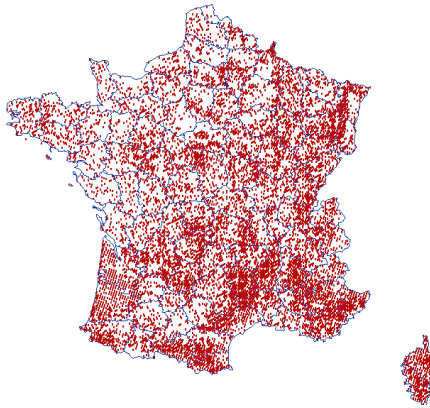
[Revenir au sommaire](#)



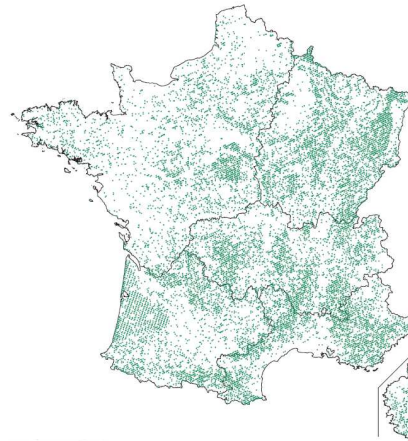
Méthode d'échantillonnage



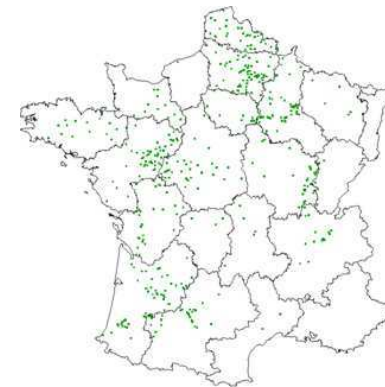
- 2^{ième} subdivision : points levés (observations, mesures sur des arbres, ...)



Échantillon terrain 2005



Points levés en forêt de production
en 2005 et 2006



500 points en peupleraies
2005, 2006 et 2007

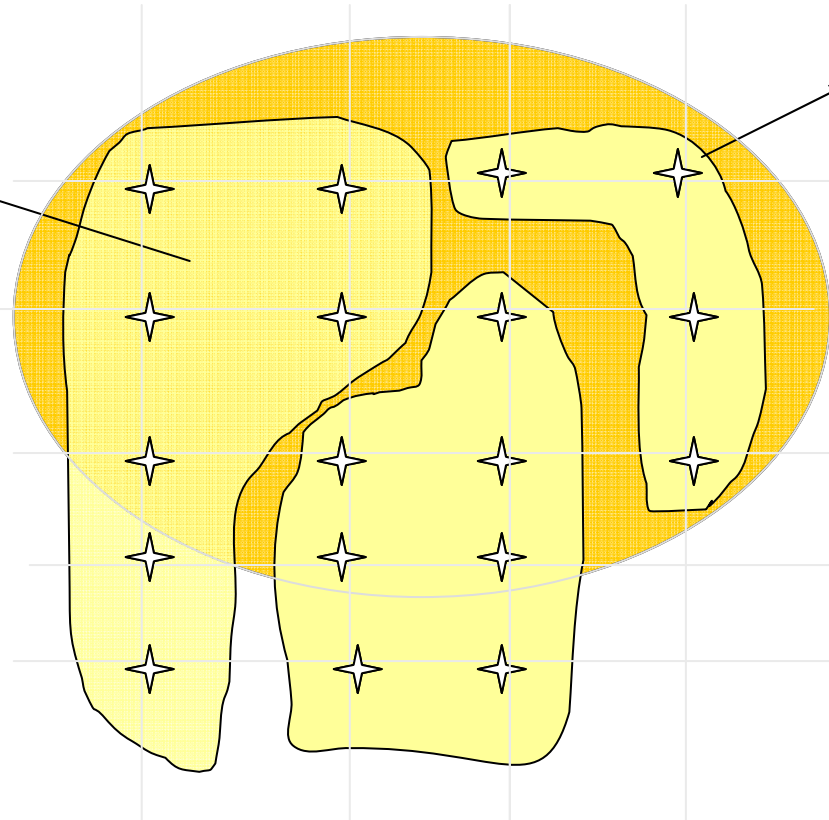
- Domaines d'études différents : Forêt, Landes, Peupleraies, Ecologie, Haies-alignements.
- 8000 points levés par an, 65 000 arbres vifs
- ~ 400 variables saisies par point d'inventaire



Estimation des résultats



Strate = k
de surface S_k



i : point de l'échantillon de poids w_i

Sous domaine = f

$$S_f = \sum_{k \in K} \left(S_k * \frac{\sum_{i \in f \cap k} w_i}{\sum_{j \in k} w_j} \right)$$

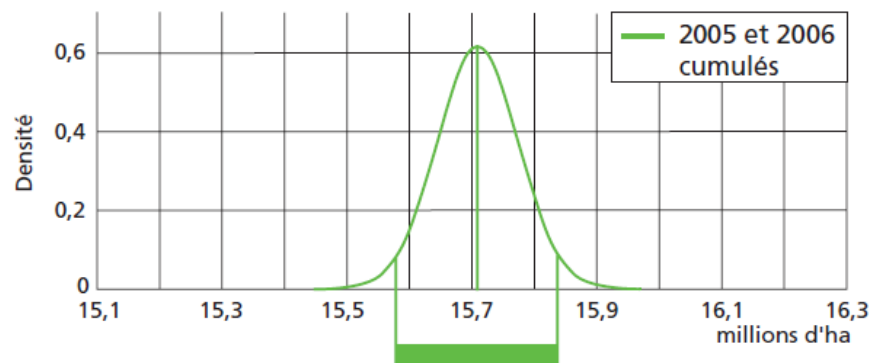
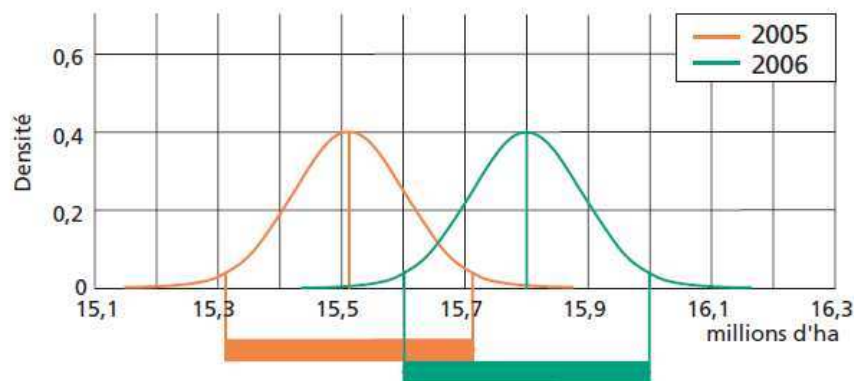
Pour une année et une stratification.



Fenêtre glissante



Le principe de la nouvelle méthode est de produire des résultats sur la base d'un échantillon pluriannuel.



Moyenne simple sur x années de campagne :

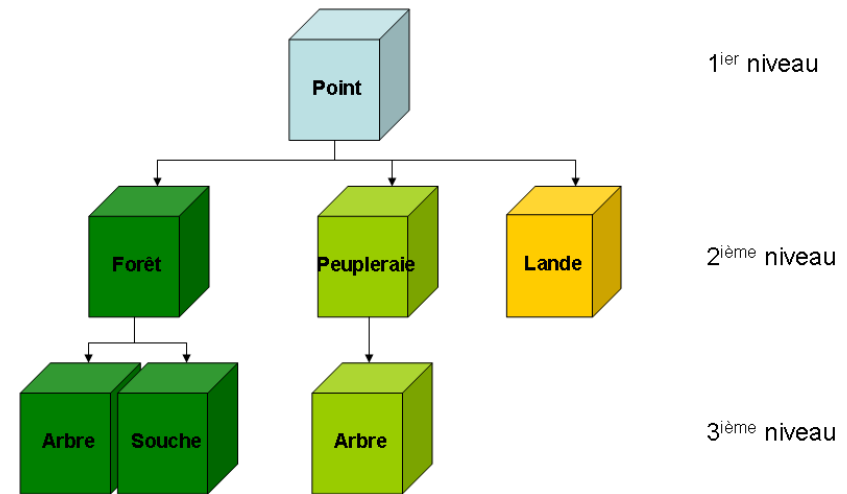
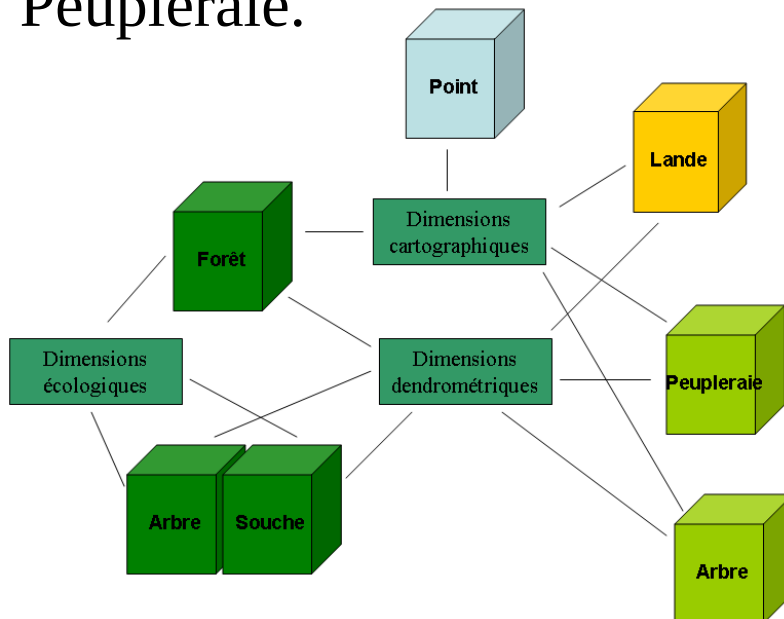
$$S_{2007=[2005-2009]} = \frac{S_{2005} + S_{2006} + S_{2007} + S_{2008} + S_{2009}}{5}$$



Constellation IFN



- Chaque domaines d'études est candidat pour un modèle dimensionnel.
- Fusion des modèles Forêt et Peupleraie.



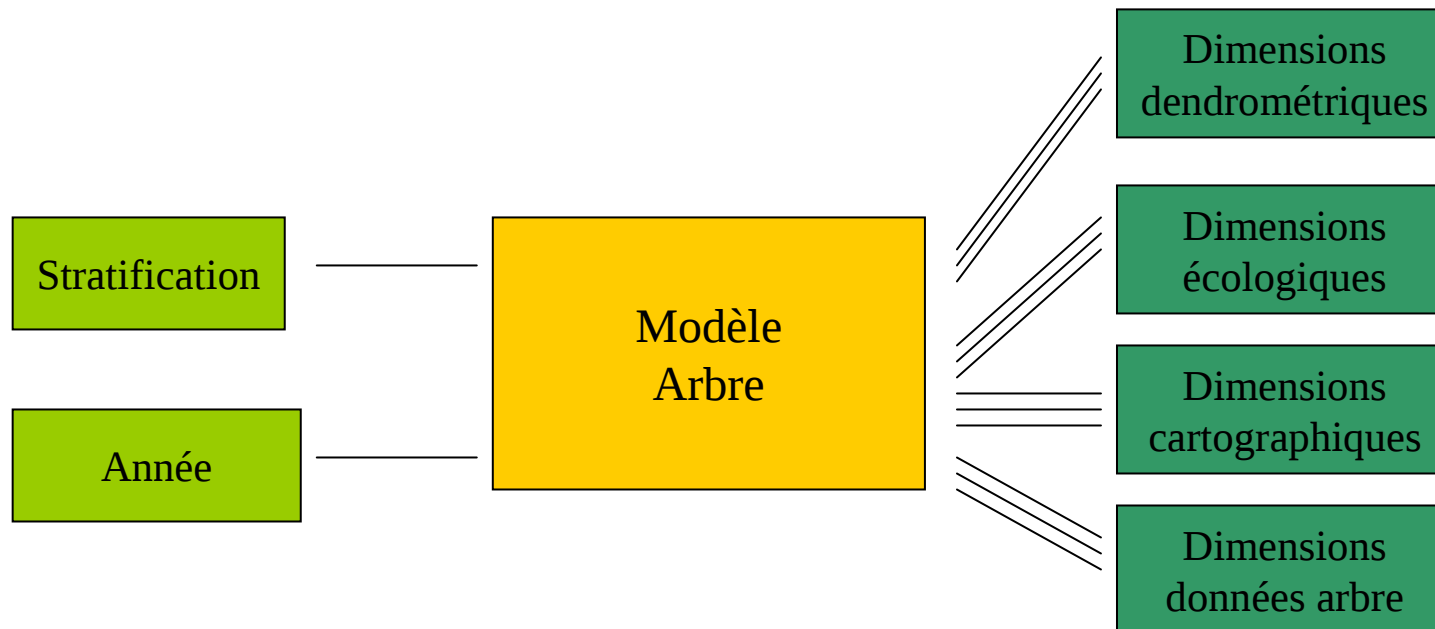
- Héritage de l'information.
- Partage des dimensions $\Rightarrow$ garder une similarité.



Granularité



Spécifier ce que représente une ligne individuelle de la table de fait.



Dimensions choisies



- Dimensions thématiques : caractérisées par les données issues des relevés, composées ou calculées, discrétisées ou regroupées.
- Dimension stratification-année : spécifique à l'IFN, liée au chargement d'une campagne d'inventaire et une stratification.
- Dimensions spatiales : caractérisées par la présence du contour géographique dans les membres des différents niveaux.
 - Dimension Administrative
 - Dimension Écologique



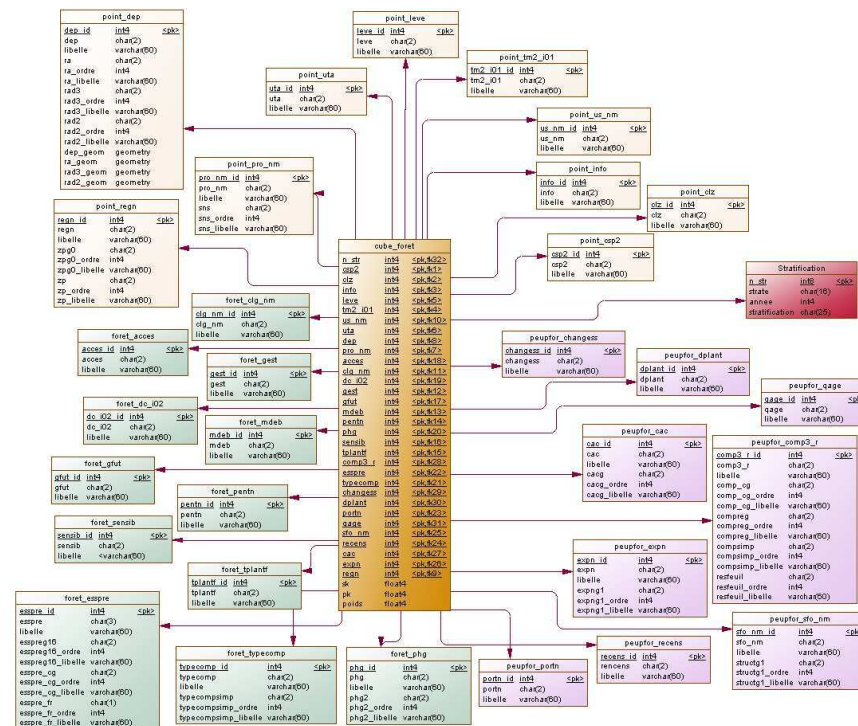
Architecture Mille-pattes



Les nombreuses jointures vont réduire la commodité d'utilisation et la performance.

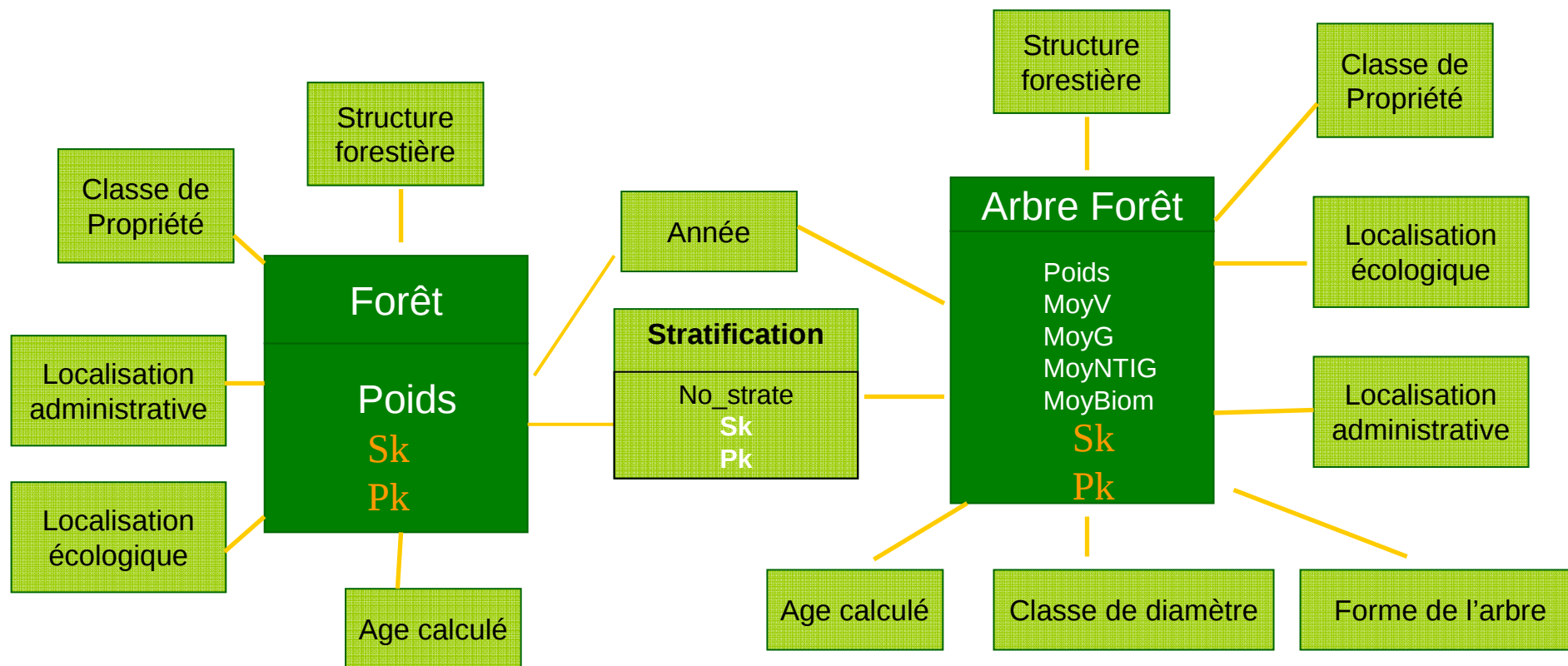
⇒ Combiner les dimensions corrélées en une seule dimension.

⇒ Créer des dimensions fourre-tout avec des données prenant un nombre restreint de valeur comme des indicateurs.



Trop de dimensions !

Faits et mesure



De nouveaux protocoles ont été mis en œuvre créant potentiellement une hétérogénéité sémantique, spatiale et temporelle :

- Nouvelles données dans un protocole.
- Relevés des données abandonnées.
- Changement de définitions des variables. Deux exemples :
Exploitabilité, Essence principale.
- Evolution du découpage des zones cartographiques.





Résultats et Interface Web OLAP Cartographique



[Revenir au sommaire](#)



Définition du schéma unique IFN dans Mondrian :

- 7 cubes + 4 cubes virtuels
- Dimensions et dimensions partagées
- Hiérarchies multiples
- Niveaux et propriétés
- Mesures et membres calculés

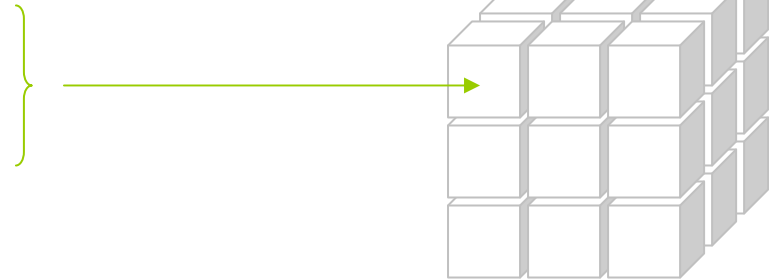


Exemple d'une mesure



On a dans chaque cellule de l'hypercube :
pour un sous-domaine, une année et une strate :

- nombre de point, poids du sous-domaine
- surface de la strate et poids de la strate
- variable moyenne



Exemple des mesures « Surf(A) Ha » :

```
<Measure name="Surf(A) Ha" aggregator="sum" visible="true">  
  <MeasureExpression>  
    <SQL dialect="generic">  
      cube_point.point_poids * cube_point.point_sk / cube_point.point_pk / 1E4  
    </SQL>  
  </MeasureExpression>  
</Measure>
```



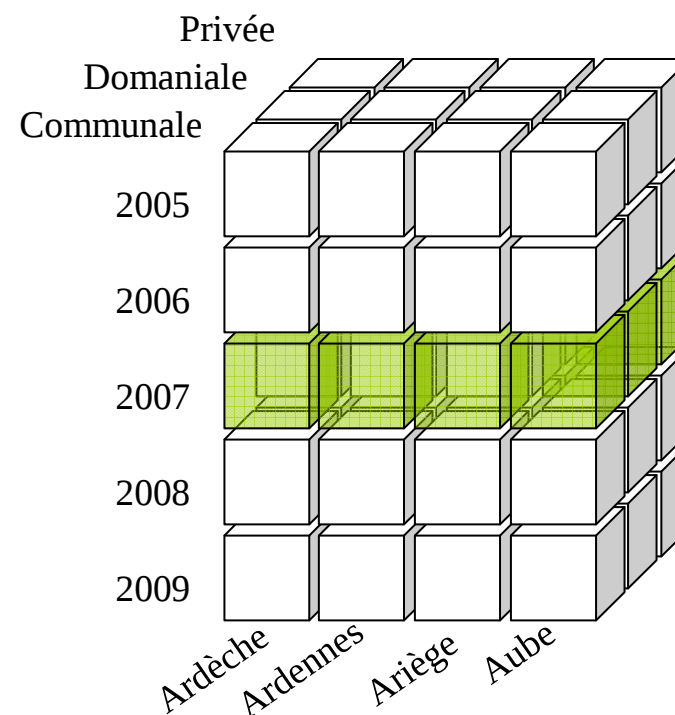
MDX : grammaire et exemple



SELECT *axis* {, *axis*}
FROM *cube name*
WHERE *slicer*

Exemple :

Select {[Measures].[Nb Point], [Measures].[Surf Ha] on Rows,
[Propriété].[Total].Children on Columns,
[Loc].[France entière].Children on pages
From Point



MDX : opérations



Slicer :

WHERE permet de sélectionner une partie du cube

Exemple : WHERE ([Loc].[Ariège])

Fonctions numériques :

SUM, MAX, AVG, MIN

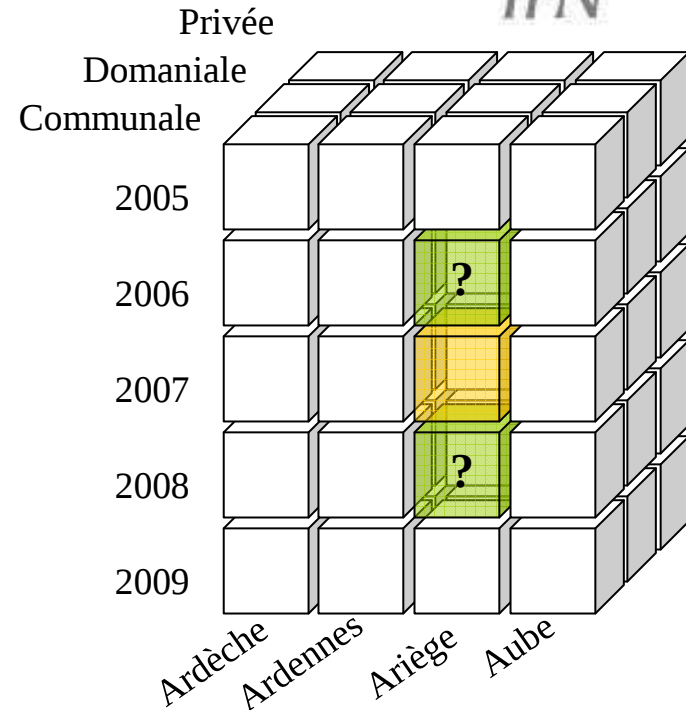
Exemple : AVG([Année].Members, [Mesures].[Nb point])

Opérations sur les membres avec

currentMember, prevMember, nextMember

Exemple :

```
([Propriété].[Communale], [Loc].[Ariège], [Mesures].[Nb point], [Année].[2007])  
=([Propriété].[Communale], [Loc].[Ariège], [Mesures].[Nb point], [Année].[2008].PrevMember)  
=([Propriété].[Communale], [Loc].[Ariège], [Mesures].[Nb point], [Année].[2006].NextMember)
```



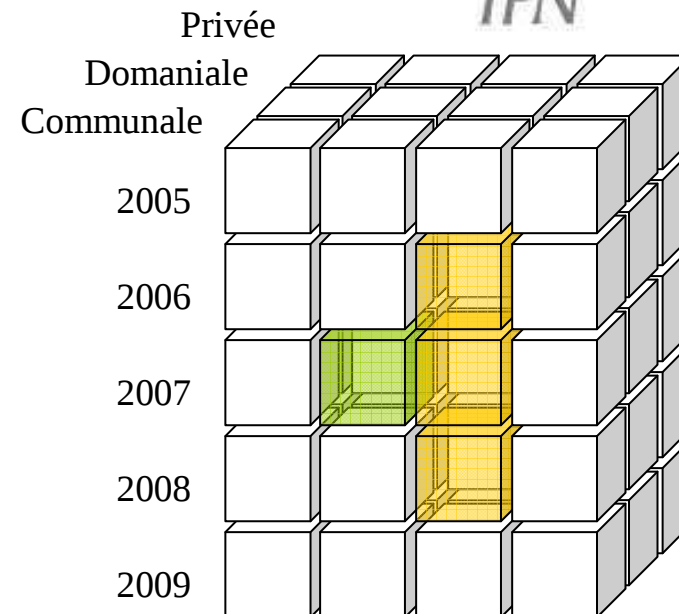
MDX : membres calculés



Pour :

- calculer des mesures
- faire des comparaisons
- recalculer suivant niveau d'agrégation choisi

```
<CalculatedMember name="tteSurf (A,A-1,A+1)"  
    dimension="Measures" visible="false">  
    <Formula>  
    ( ([Année].currentmember, [Measures].[Surf]) +  
      ([Année].currentmember.prevmember, [Measures].[Surf]) +  
      ([Année].currentmember.nextmember, [Measures].[Surf]) ) / 3  
    </Formula>  
</CalculatedMember>
```



Calcul de l'estimation de la surface



Pour les cubes « Point », « Forêt », « Peupleraie » et « Landes »

- On établit la surface à l'hectare pour une année et une stratification : $\text{Surf}(A)$ Ha
- On calcule la moyenne temporelle : $\text{tteSurf}(A-1, A, A+1)$.
- On retire les années extrêmes

[Croisement COUVERTURE x UTILISATION x TAILLE MASSIF] [total] [FORET DE PRODUCTION]	[Mesures] [Surf]	[Mesures] [tteSurf(A,A-1,A+1)]	[Mesures] [Surf(A,A-1,A+1)]
[Année] [2005]	15,257,800.717	10,073,587.1	
[Année] [2006]	14,962,960.584	15,079,770.618	15,079,770.618
[Année] [2007]	15,018,550.553	15,068,186.984	15,068,186.984
[Année] [2008]	15,223,049.816	15,176,661.846	15,176,661.846
[Année] [2009]	15,288,385.17	10,170,478.329	

- On ajoute la couleur afin de prévenir sur l'incertitude.

	Mesures					
	Surf (A-1,A,A+1) Ha					
	Année					
Localisation Administrative	2005	2006	2007	2008	2009	
-RHONE-ALPES		7 772	9 004	9 224		
AIN		3 215	3 531	4 030		
ARDECHE		698				
DROME		165	342	1 112		
HAUTE-SAVOIE						
ISERE		2 934	3 115	2 466		
LOIRE						
RHONE			123	322		
SAVOIE		760	1 894	1 293		

Sous-domaine : [us_nm=PEUPLERAIE]



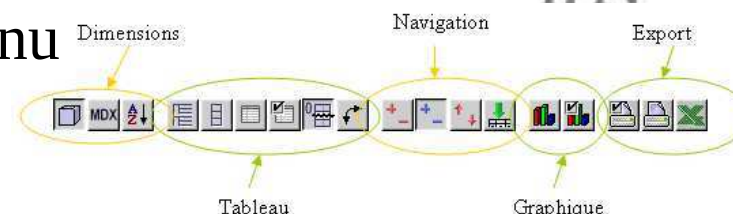
Résultats dans JPivot



Barre de Menu

Tableau résultat multidimensionnel

	Mesures	
	V (A,A-1,A+1) m³	BIOM_AR (A,A-1,A+1) tms
	Année	Année
Diamètre	2006	2006
-total	24 169 582,223	14 707 927,751
+GROS BOIS - D >= 37,5	12 084 648,978	6 977 931,139
+MOYEN BOIS - 22,5 <= D < 37,5	9 018 527,278	5 344 226,184
+PETIT BOIS - 7,5 <= D < 22,5	3 066 405,967	2 385 770,428



Mesures	
☒	V (A,A-1,A+1) m³
☒	V1 (A,A-1,A+1) m³
☒	V2 (A,A-1,A+1) m³
☒	V3 (A,A-1,A+1) m³
☒	VR (A,A-1,A+1) m³
☐	NT (A,A-1,A+1)
☐	V0 (A,A-1,A+1) m³
☐	V_AR (A,A-1,A+1) m³
☐	BIOM_AR (A,A-1,A+1) tms
☐	CARB_AR (A,A-1,A+1) tc
Aucun Grouper OK Annuler	

Mesures

Colonnes	
Mesures	
Lignes	
☒	Essence principale recensable
☒	Clone ou cultivar principal
Filtres	
☒	Age calculé
☒	Altitude
☒	Changement d'essence principale
☒	Composition de phase 3 du peuplement recensable
☒	Composition du peuplement recensable G1
☒	Coupe rase entre prise de vues et levé
☒	Couverture du sol
☒	Couverture du sol au-delà de la lisière
☒	Croisement COUVERTURE x UTILISATION x TAILLE MASSIF
☒	Densité initiale de plantation
☒	Diamètre quadratique moyen
☒	Distance de débardage
☒	Diversité du peuplement recensable
☒	Entretien de peupleraie
☒	Exploitabilité
☒	Homogénéité
☒	Identifiant de placette de peuplement
☒	Indicateur d'arbre courbe

Choix des dimensions : axe, ordre

Problématiques de l'axe « Année »



- Attention à la ligne « Hors limite temporelle ».
- Résultats sont des estimations !

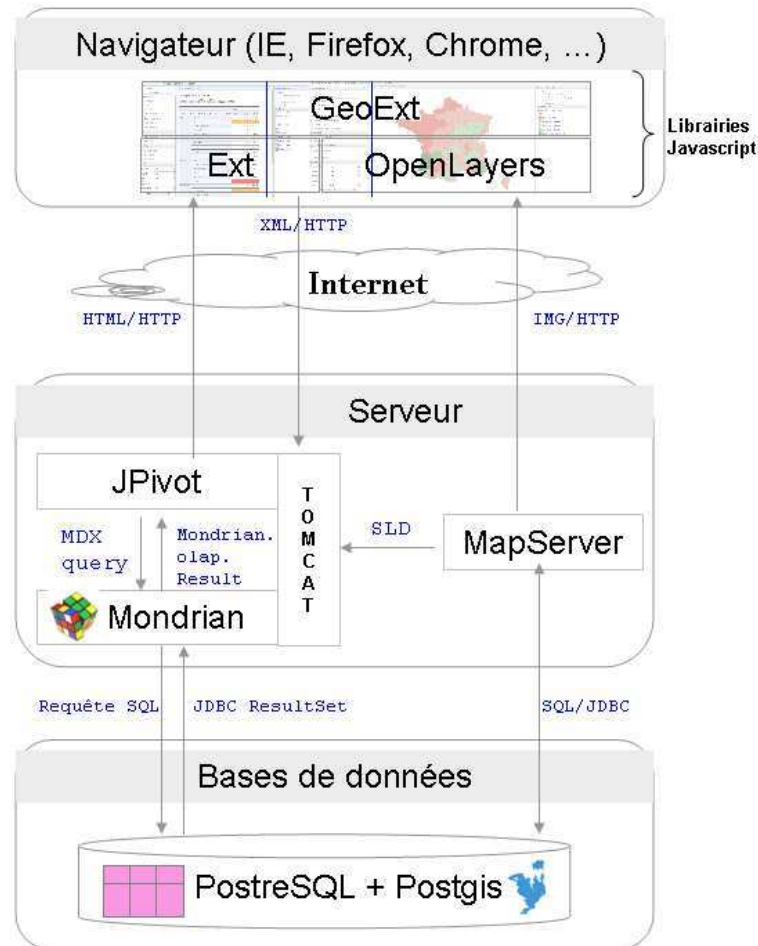
Évolution significative ?

	Mesures									
	Surf (A-2,A-1,A,A+1,A+2) Ha					Surf (A-1,A,A+1) Ha				
	Année					Année				
Structure forestière	2005	2006	2007	2008	2009	2005	2006	2007	2008	2009
-total			15 375 229				14 994 944	15 284 306	15 636 622	
+FUTAIE			6 836 429				5 401 826	8 354 168	8 694 624	
+HORS LIMITE TEMPORELLE			3 000 664				5 001 107			
+MELANGE FUTAIE TAILLIS			3 342 789				2 708 626	4 153 128	4 229 468	
+NULL										
+PAS DE STRUCTURE			803 676				671 256	1 011 279	1 011 638	
+TAILLIS			1 391 671				1 212 129	1 765 731	1 700 891	

Sous-domaine : [us_nm=FORET DE PRODUCTION] [Leve=LEVE
REALISE]



Architecture



Architecture 3-tiers :

- PostgreSQL + Postgis
- 2 serveurs :
 - Mondrian
 - MapServer
- 2 clients web :
 - GeoExt
 - OLAP JPivot

Interface



Changer de domaines d'études ▾

- ☐ GArbre
- ☐ PArbre
- ☐ Tige
- ☐ Placette
- ☐ Peuplier
- ☐ Arbre
- ☒ Peuplier et Forêt
- ☐ Lande
- ☐ Point
- ☐ Forêt
- ☐ Souche
- ☐ Soucharbre

Menu Panel

DEP	Mesures
RAD2 - NORD-EST	3 461 429
RAD2 - CENTRE-EST	2 784 788
RAD2 - SUD-EST	2 660 429
RAD2 - SUD-OUEST	2 975 412
RAD2 - NORD-OUEST	2 948 746

Découpage

☐ France entière

Dimension administrative

☐ 5 interrégions

☐ 9 interrégions

☐ Région

☐ Département

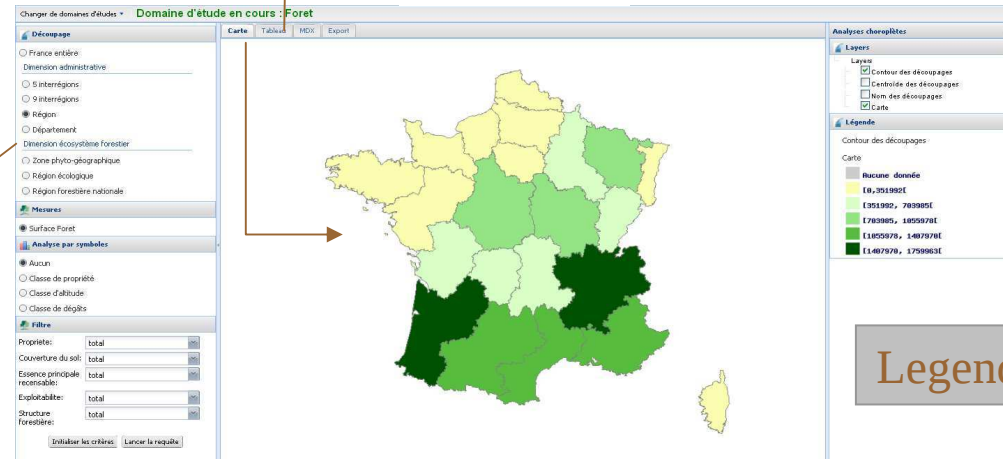
Dimension écosystème forestier

☐ Zone phyto-géographique

☒ Région écologique

☐ Région forestière nationale

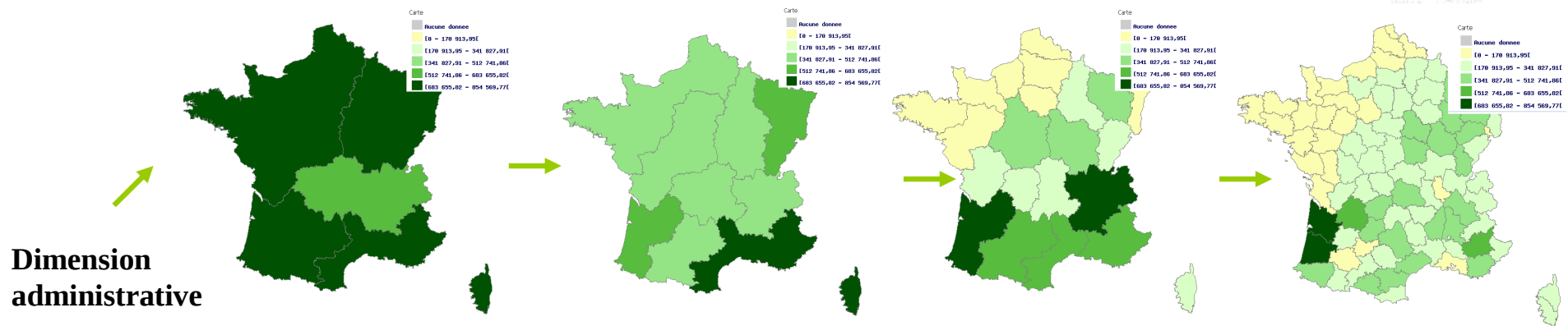
Cube Panel



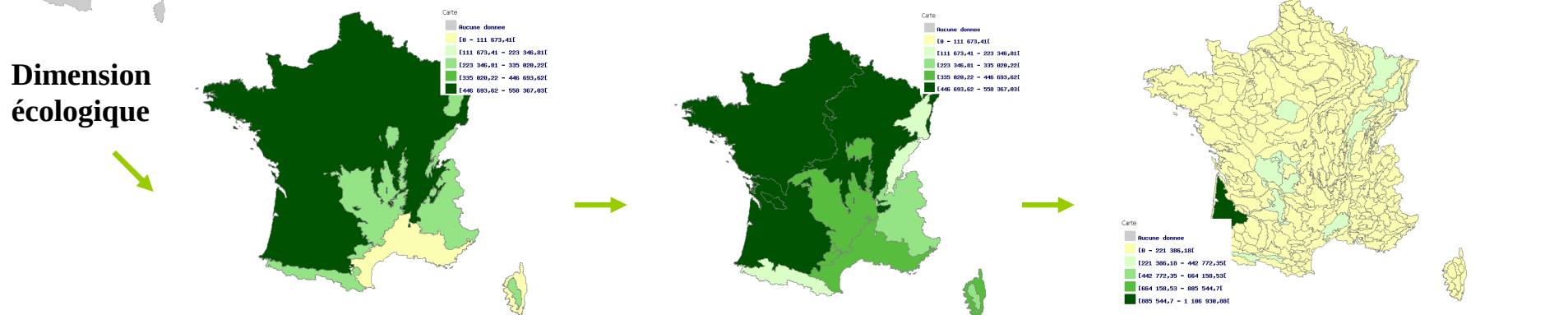
Legende Panel

Resultat Panel

Spatial Drill-Down et spatial Roll-Up



Surface boisée en forêt de production en ha calculée à partir des 5 dernières campagnes



Merci de votre attention



*INVENTAIRE FORESTIER
NATIONAL*

