

Intégration de versions fonctionnelles dans les entrepôts de données multimédias

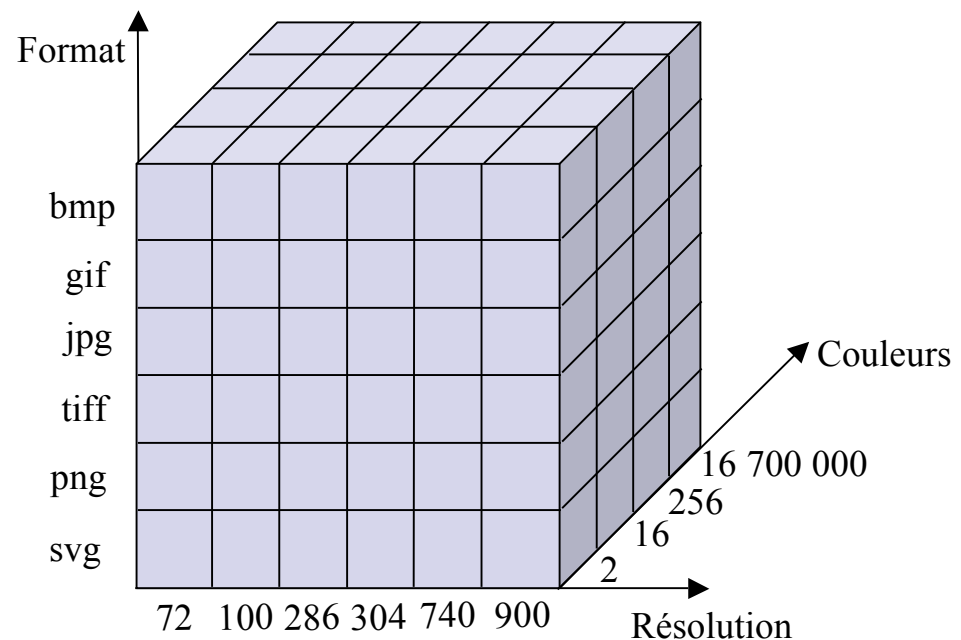
Anne-Muriel Arigon^{1,2}, Maryvonne Miquel¹, Anne Tchounikine¹

¹Laboratoire d'InfoRmatique en Images et Systèmes d'information
INSA de Lyon – UMR CNRS 5205

²Laboratoire de Biométrie et Biologie Evolutive
Université Claude Bernard Lyon1 – UMR CNRS 5558

Exemple d'entrepôt de données multimédias

- Entrepôt de données regroupant des images
- Les faits: les images
- Les dimensions:
 - le format
 - la résolution (dpi)
 - les couleurs (palette)



Problématique liée aux entrepôts de données multimédias

- **Stockage de données volumineuses**

- différents types de données complexes : textes, vidéos, images, signaux, sons,...

- **Fonctions d'agrégat à redéfinir**

- fonctions d'agrégat spécifiques : liste, fusion de données multimédias,...

- **Utilisation des descripteurs pour former les dimensions du modèle**

- donnée multimédia définie par :

- des descripteurs textuels

âge du patient
technologie
date



Électrocardiogramme

- des descripteurs sur le contenu

niveau de bruit
durée du QT
segment ST

- plusieurs modes de calcul pour un descripteur :

- algorithmes

durée du QT → intervalle QT par algorithme1 de traitement
 → intervalle QT par algorithme2 de traitement

- classifications

âge du patient → âgeParTranche
 → âgeParCatégorie

⇒ versions fonctionnelles du descripteur

Solutions

- **Système d'analyse de données multimédias : MultiMediaMiner [O.R.Zaïane98]**
 - faits : liens vers les données multimédias
 - dimensions : descripteurs des données

- **Recherche de données multimédias : MédiaHouse [J.You01]**
 - faits : contenus des données multimédias
 - dimensions : différents types possibles de médias associés chacun à leurs descripteurs

- **Implémentations d'entrepôts de données multimédias dans le domaine médical pour :**
 - l'aide au diagnostique (détection du cancer du sein) [H.Zhang01]
 - le stockage et la restitution d'images médicales [R.V.Tikekar95]

Bilan

- Modèles statiques
 - descripteurs fixés à la modélisation
 - descripteurs calculés d'une manière unique

- Donner la possibilité de choisir le mode de calcul pour sélectionner les descripteurs
 - ⇒ intégration de versions fonctionnelles de descripteurs aux entrepôts de données

Evolutions temporelles dans les entrepôts

■ Problématique

- donner la possibilité de choix de versions temporelles pour la représentation des données

■ Modèles intéressants

- modèles prenant en compte l'historique des évolutions

■ Modèle multidimensionnel multiversion [M.Body02] :

- une table de fait multiversion
- une dimension représentant les versions temporelles

⇒ navigation assez limitée car choix unique de la version temporelle

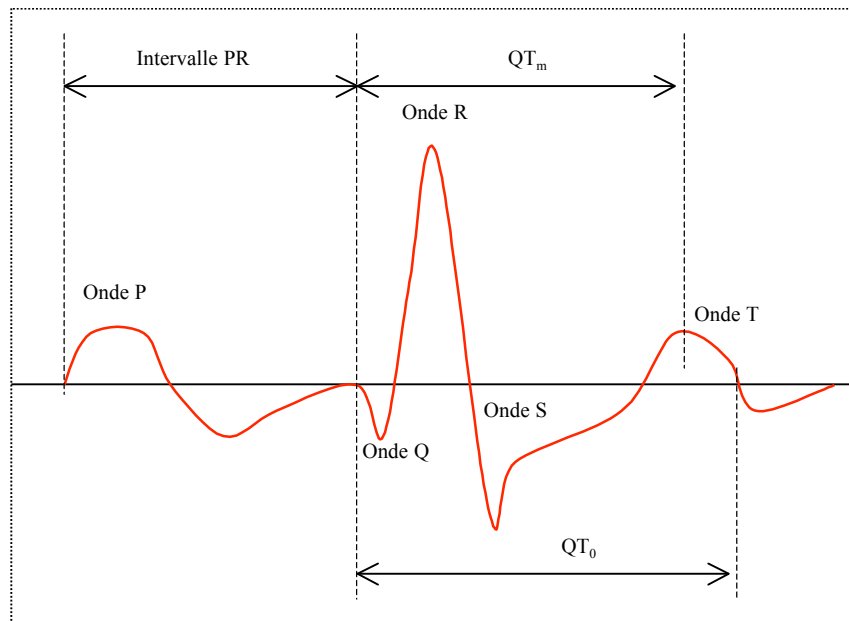
L'essai thérapeutique (EMIAT)

■ Le but

- évaluer les effets de l'amiodarone comparée à un placebo chez des patients ayant survécu à un infarctus du myocarde

■ Résultats

- électrocardiogrammes (ECGs) et autres données médicales des patients

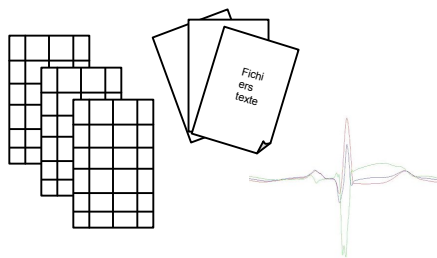


■ ECGs caractérisés par :

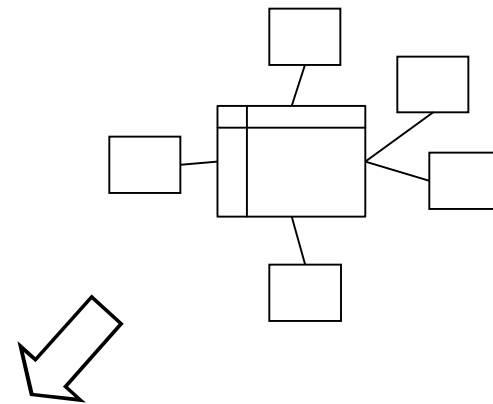
- descripteurs textuels
 - pathologie principale
 - âge
 - tranche horaire d'acquisition
 - date
 - ...
- descripteurs de contenu
 - durée du QT
 - niveau de bruit
 - ...

Utilisation d'un entrepôt de données dans l'étude EMIAT

Données de l'étude EMIAT



Entrepôt de données



Constitution de populations d'étude à partir de requêtes

- femmes entre 40 et 50 ans
- visite médicale en mars 1998 durant la nuit
- tachycardie
- durée de QT de 258
- niveau de bruit de 20

Les descripteurs des données

■ Descripteurs organisés en hiérarchies complexes

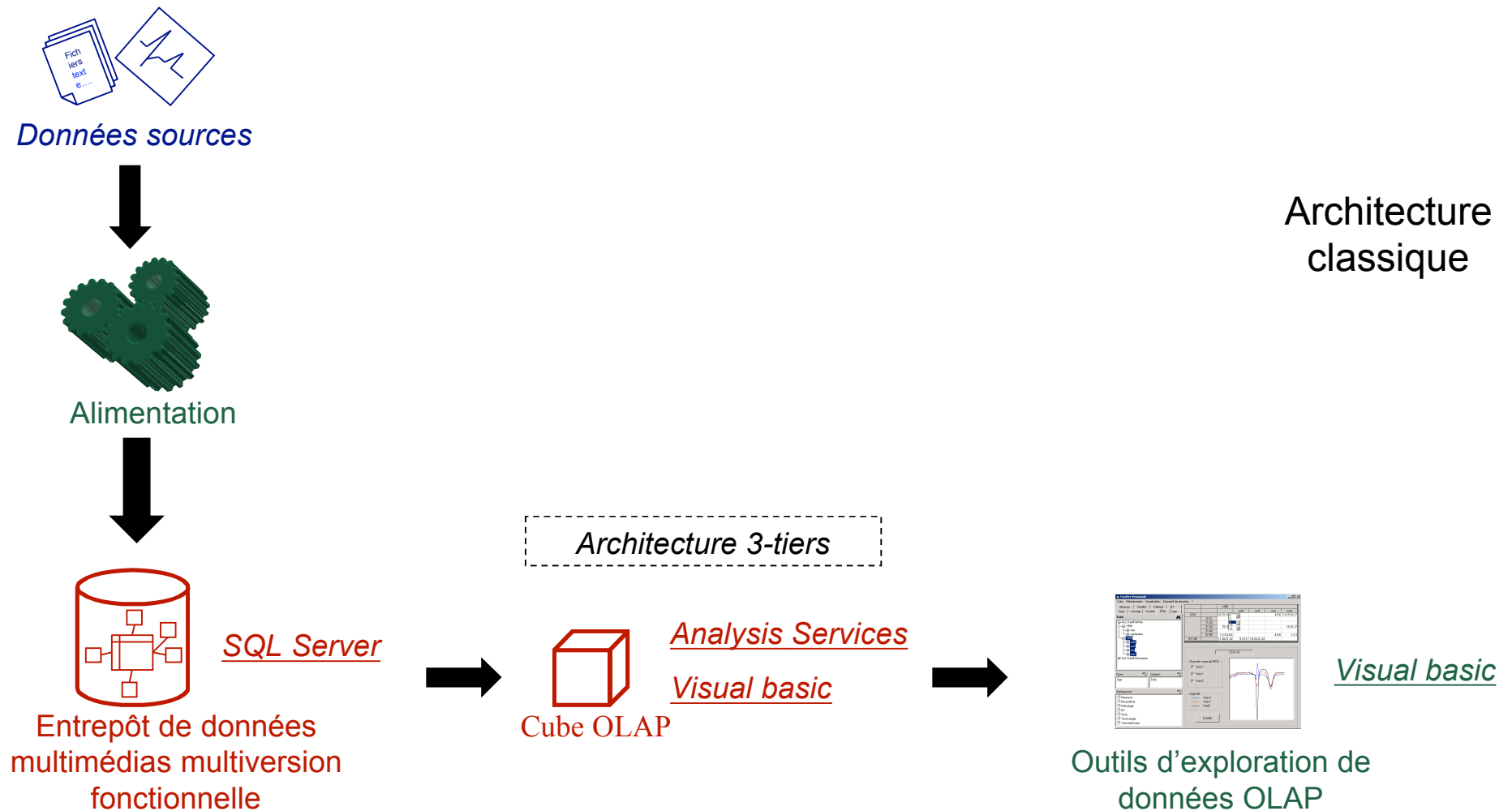
- ⇒ hiérarchies complexes dans les dimensions
- hiérarchie multiple
 - hiérarchie non-onto
 - hiérarchie non-stricte
 - hiérarchie non-couvrante

■ Descripteurs calculés par divers modes de calcul

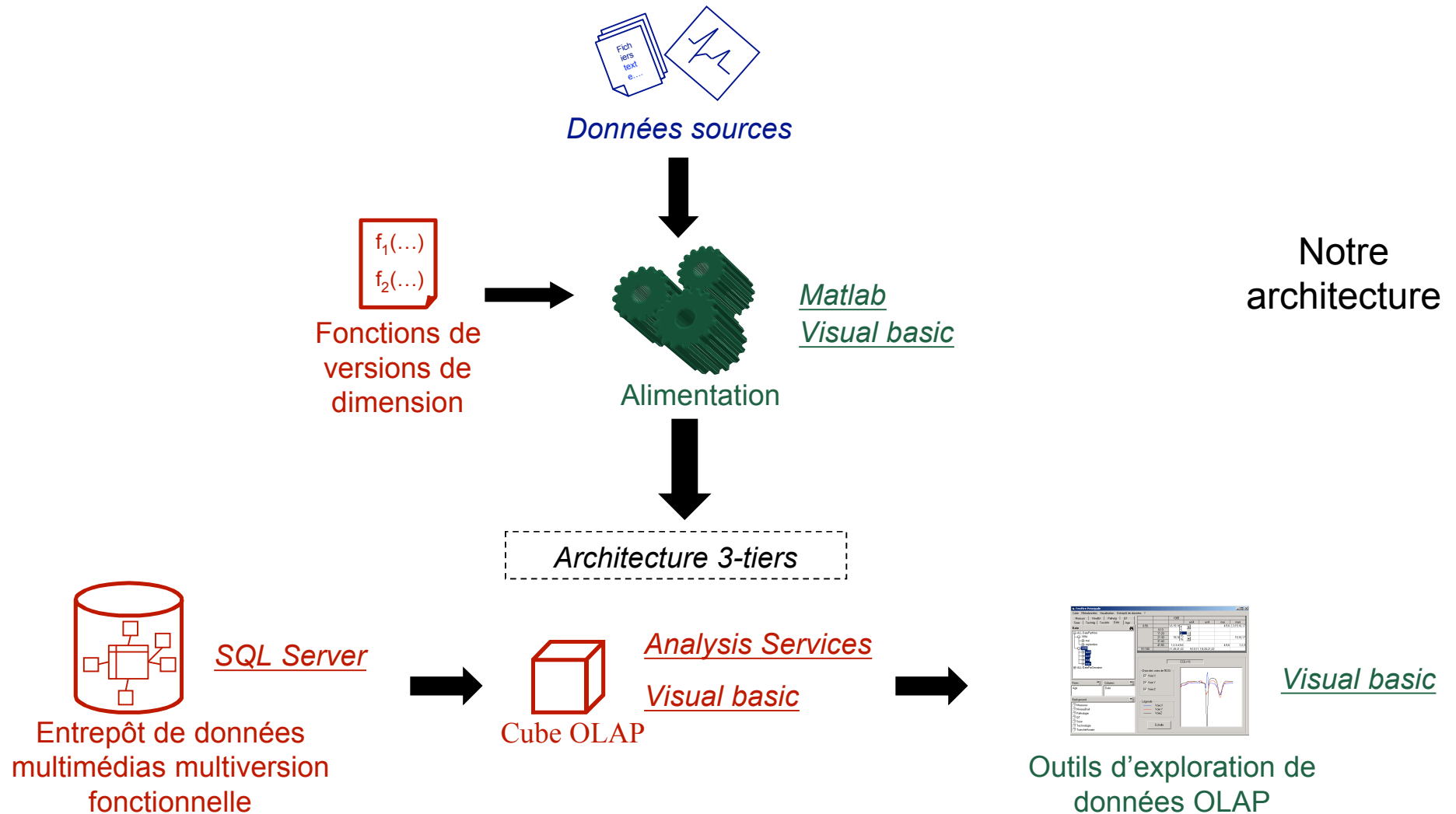


⇒ **version de dimension**

Architecture de notre solution



Architecture de notre solution



Modèle conceptuel

- **Schéma de version de dimension**

= ensemble des niveaux d'une dimension de version donnée, liés par des relations de hiérarchie

$$S_{VD} = \langle idVD, N, \langle idVD \rangle \rangle$$

- **Version de dimension**

= ensemble de membres d'une dimension de version donnée, liés par des relations de hiérarchie

$$VD = \langle idVD, nomVD, M, \langle VD \rangle, [descriptionVD] \rangle$$

- **Dimension multiversion**

= ensemble de versions de dimension

$$DMV = \langle idDMV, nomDMV, VD, [descriptionDMV] \rangle$$

Modèle conceptuel

- **Table de fait multiversion fonctionnelle**

= ensemble des mesures selon différentes combinaisons de versions

$$tf : DMV_1 \times DMV_2 \times \dots \times DMV_n \rightarrow dom(\mu_1) \times \dots \times dom(\mu_m)$$

$$m_1, m_2, \dots, m_n \mapsto v_1, \dots, v_m$$

- **Fonction de version de dimension**

= mode de calcul pour obtenir les membres d'une version de dimension à partir des données de la base de données de production

$$f_{VD} : BD_f \rightarrow MF_{VD}$$

$$d \mapsto m$$

- **Structure multidimensionnelle multiversion fonctionnelle**

= structure formée par la table de fait, l'ensemble des dimensions multiversions et l'ensemble des fonctions de versions de dimension

$$M2F = \langle DMV, tf, F \rangle$$

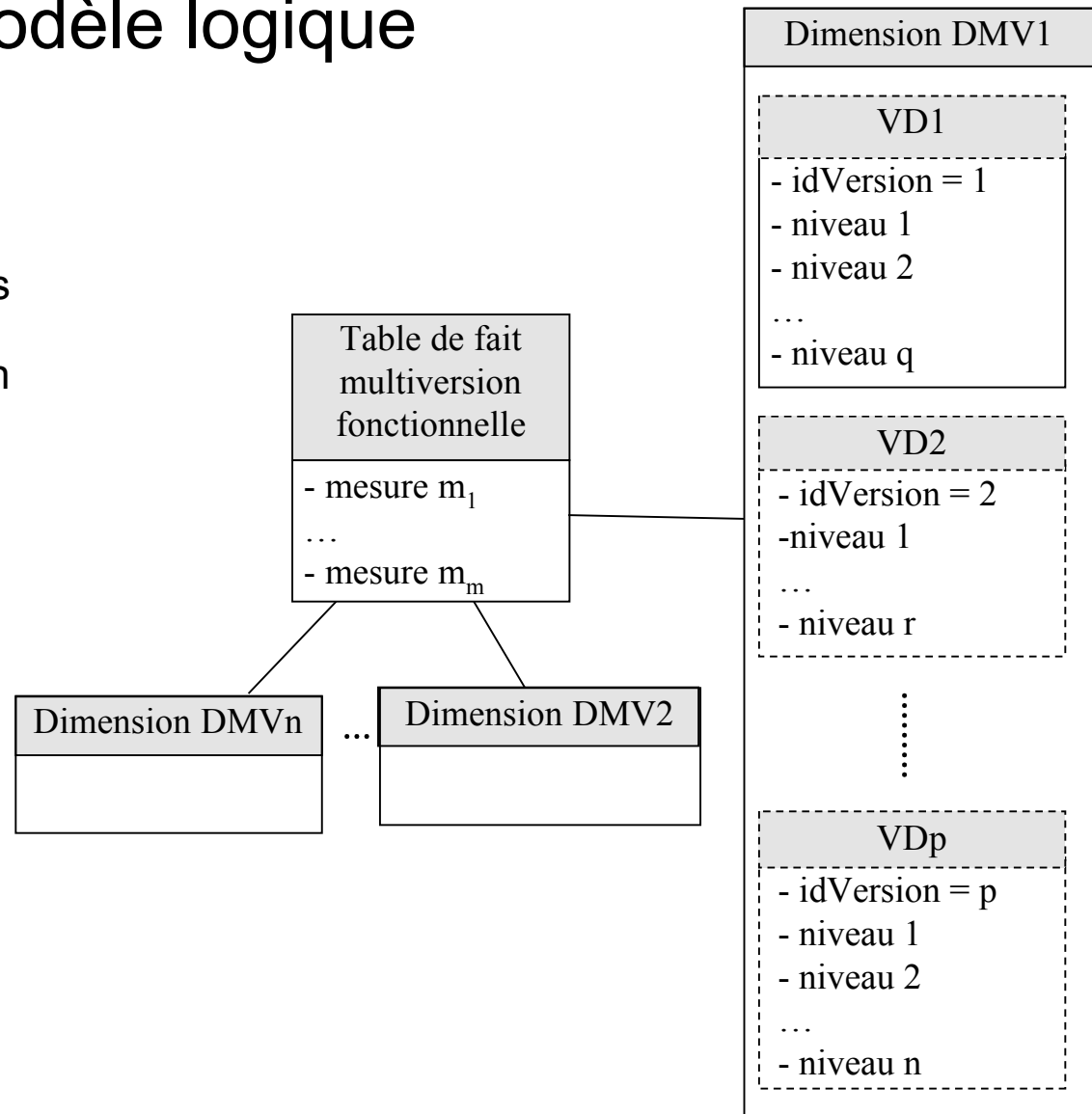
Modèle logique

■ Dimension multiversion

- regroupe les membres des versions de dimension qu'elle contient par version de dimension

■ Metadonnées

- dimensions multiversions
- versions de dimension
- niveaux
- fonction de version de dimension
- hiérarchie multiple et non-strict

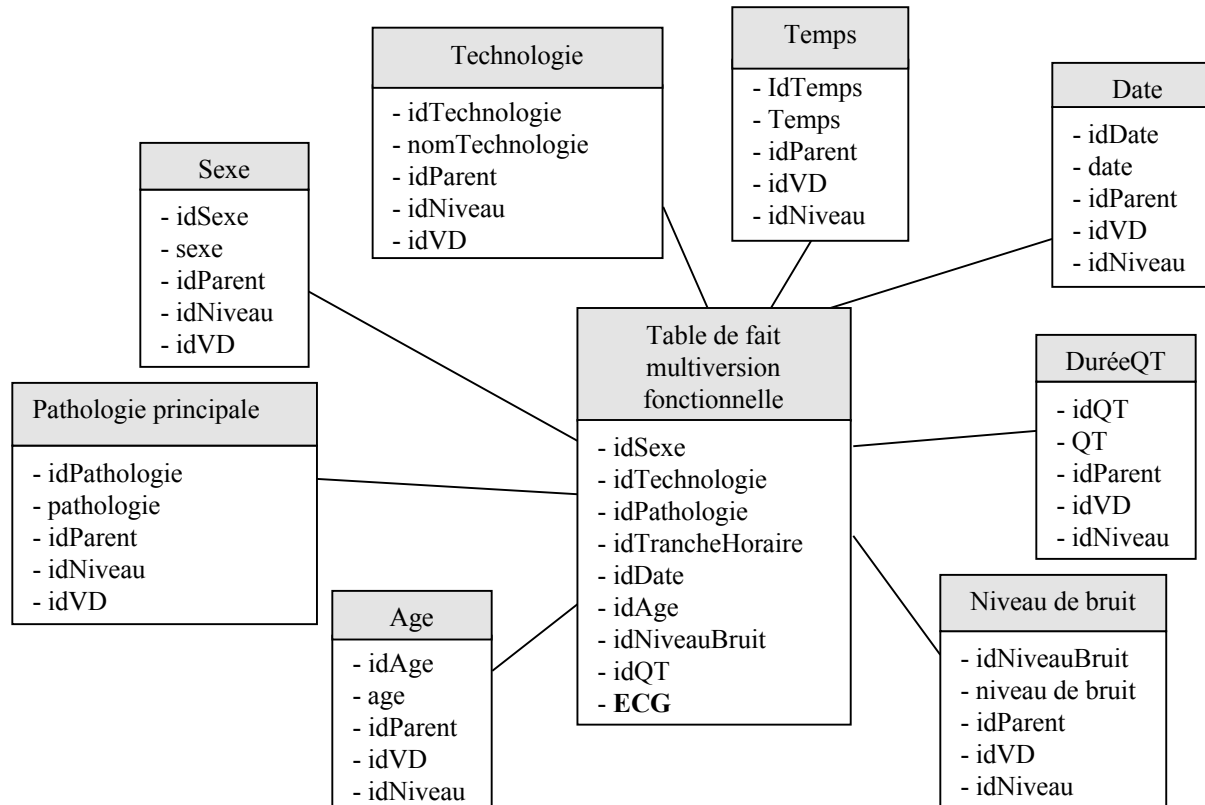


Modèle physique

- **Table de fait multiversion fonctionnelle**
 - un fait associé à une valeur de chaque dimension multiversion
 - un fait par combinaison de versions de dimension

- **Dimension multiversion**
 - ⇒ représentation parent-enfant [T.B.Pédersen01]
 - intégrer des versions de dimension
 - gérer les hiérarchies complexes

Mise en œuvre sur l'étude EMIAT



- **Les faits**
 - les ECGs de l'étude EMIAT
- **8 dimensions multiversions**
 - 3 correspondent au patient
 - 3 correspondent à l'acquisition de l'ECG
 - 2 correspondent au contenu de l'ECG
- **3 fonctions d'agrégation**
 - nombreECG
 - listeECG
 - ECGmoyen

L'interface de navigation OLAP

■ Interface Visual Basic

- composants spécifiques pour la navigation
- cube OLAP

■ Les données multimédias

- exploration des données agrégées selon
 - les versions de dimension
 - leur granularité
- visualisation des données multimédias
 - fenêtre principale : sélection de l'ECG dans le tableau
 - fenêtre de visualisation : indication de l'ECG

Interface : exploration et visualisation des données multimédias agrégées

The interface is titled "Fenêtre Principale" and contains several components:

- Left Panel (Age Tree):** A tree view showing categories: ALL-Catégorie (Adolescent, Adulte, Enfant, Jeune), ALL-DateNaissance, and ALL-TrancheDe5.
- Table:** A data table with columns for Age, Pathologie auriculaire (Fibrillation auriculaire, Flutter auriculaire, Tachycardie auriculaire), and various algorithmic measures (ALL_QT_Algo1, ALL_QT_Algo2).
- ECG Visualization:** A plot titled "ECG n°16" showing three traces (Voie X, Voie Y, Voie Z) with a legend and a scale button.

Age	ALL_QT_Algo1	Pathologie auriculaire		
		Fibrillation auriculaire	Flutter auriculaire	Tachycardie auriculaire
Adulte	0	1,2,3,4,5,6	7,8,9,10,11	
	200-400	1,2,3,4,5,6	7,8,9,10,11	
	401-800	7,8,9,10,11	7,8,9,10,11	
	ALL_QT_Algo2	2,3,4,5,6,7,8,9,10,11	1,2,3,4,5,6	7,8,9,10,11
Jeune	0	15,16,17		15,16,17
	200-400	15,16,17		15,16,17
	401-800	15,16,17		15,16,17
	ALL_QT_Algo2	15,16,17		15,16,17

L'interface de navigation OLAP

■ Interface Visual Basic

- composants spécifiques pour la navigation
- cube OLAP

■ Les données multimédias

- exploration des données agrégées selon
 - les versions de dimension
 - leur granularité
- visualisation des données multimédias
 - fenêtre principale : sélection de l'ECG dans le tableau
 - fenêtre de visualisation : indication de l'ECG

■ Les métadonnées

- schémas et instances de chaque version de dimension
- fonction de version de dimension

Interface : visualisation des métadonnées

The main window, titled 'Fenêtre Principale', shows a data table with the following structure:

Mez	Sexe	Age	Pathologie auriculaire	Fibrillation auriculaire	Flutter auriculaire	Tachycardie auriculaire
ALL_QT_Algo1		Adulte	2;3;4;5;6;7;8;9;10;11	1;2;3;4;5;6	7;8;9;10;11	
		0				
		200-400	1;2;3;4;5;6	1;2;3;4;5;6		
		401-800	7;8;9;10;11		7;8;9;10;11	
ALL_QT_Algo2		Adulte	2;3;4;5;6;7;8;9;10;11	1;2;3;4;5;6	7;8;9;10;11	
		0				
		200-400	1;2;3;4;5;6	1;2;3;4;5;6		
		401-800	7;8;9;10;11		7;8;9;10;11	
ALL_QT_Algo1		Adulte	15;16;17		7;8;9;10;11	15;16;17
		0				

The two pop-up windows, 'Versions de dimensions', provide hierarchical views of the 'Age' dimension. The left window shows a tree structure with nodes 'Niveau-ALL-Age', 'Niveau-NomCatégorie', and 'Niveau-Age'. The right window shows a more detailed tree structure starting with 'ALL-Catégorie' and branching into 'Enfant', 'Adolescent', 'Jeune', 'Adulte', and 'PersonneAagée', with further sub-nodes for each category.



Conclusion

■ Apports

- proposer à l'utilisateur de choisir diverses versions fonctionnelles des descripteurs de données
 - ⇒ constitution de différentes populations d'étude à partir de requêtes
- approche bien adaptée aux données multimédias

■ Limites et perspectives

- pas de version de fait
 - ajout d'une dimension version de fait
- un schéma par version de dimension
 - évolution du modèle conceptuel
- stockage de données
 - fonction de mapping et stockage des version de dimension les plus utilisées