

Guillaume Cabanac, Max Chevalier, Franck Ravat et Olivier Teste. Méta-modélisation des bases de données multidimensionnelles annotées. Dans : *EDA 2006 : 2e journée francophone sur les Entrepôts de Données et l'Analyse en ligne*, Versailles, France, 19 juin 2006. ND, p. à paraître (16 pages).

## Méta-modélisation des bases de données multidimensionnelles annotées

Guillaume Cabanac\*, Max Chevalier\*\*,\*\*  
Franck Ravat\*, Olivier Teste\*

\*Institut de Recherche en Informatique de Toulouse (IRIT – UMR 5505)  
Université Paul Sabatier - 118 route de Narbonne  
F-31062 Toulouse cedex 9

\*\*Laboratoire de Gestion et Cognition (LGC – EA 2043)  
IUT "A" Paul Sabatier - 129 avenue de Ranguel - BP 67701  
F-31077 Toulouse cedex 4

{cabanac, chevalie, ravat, teste}@irit.fr

**Résumé.** Cet article présente le concept de Mémoire d'Expertises Décisionnelles (MED). Une MED comprend d'une part des fonctionnalités d'un système d'aide à la décision et d'autre part des fonctionnalités de mémorisation et d'exploitation du capital immatériel des décideurs. Notre système d'aide à la décision repose sur une Base de Données Multidimensionnelles (BDM) composée de faits, dimensions, mesures, hiérarchies et paramètres. Le capital immatériel des décideurs est articulé au travers d'annotations associées aux différents concepts et valeurs de la BDM. Nous définissons un modèle conceptuel de BDM annotées et le méta-modèle associé.

**Mots clés :** bases de données multidimensionnelles, annotations décisionnelles.

### 1 Contexte et problématique

D'une part, face à la mondialisation et à la concurrence exacerbée de ces dernières années les entreprises souhaitent disposer d'outils facilitant les prises de décisions rapides et fiables nécessaires à leur pérennisation et à leur expansion. De tels systèmes visent à transformer les données opérationnelles en informations pouvant être interprétées par les décideurs afin que ces derniers puissent effectuer des analyses complexes et prendre de meilleures décisions de façon toujours plus réactive.

D'autre part, afin d'être toujours plus performantes, les entreprises peuvent mettre en place des outils permettant de capitaliser les connaissances et les expériences de leurs collaborateurs. Cette capitalisation du savoir-faire (capital immatériel) consiste à créer une mémoire d'entreprise. Une mémoire d'entreprise est donc la « *matérialisation explicite et persistante des connaissances et informations cruciales d'une organisation pour faciliter leur accès, partage et réutilisation par les membres de l'organisation dans leurs tâches individuelles ou collectives* » (Dieng-Kuntz, 2005). Une telle mémoire évite notamment la perte de

savoir-faire lors du départ d'un collaborateur et facilite le transfert de connaissances entre collaborateurs afin que l'entreprise puisse atteindre ses objectifs.

Dans le cadre de cet article, nous souhaitons proposer un système d'aide à la décision intégrant également des fonctionnalités de mémorisation et d'exploitation du capital immatériel des décideurs. Nos travaux visent donc à proposer aux entreprises plus qu'un système d'aide à la décision : un véritable outil de **Mémoire d'Expertises Décisionnelles (MED)**.

## 1.1 Les données décisionnelles

Aussi bien au niveau de la recherche scientifique que dans le monde professionnel, il est reconnu que les bases de données multidimensionnelles (BDM) sont adaptées pour le stockage et la manipulation des données décisionnelles (Inmon, 1996). Les modèles conceptuels (Ravat *et al.*, 2005) des BDM organisent les données en sujets et axes d'analyses (Kimball, 1996). Tout sujet d'analyse est représenté par un **fait** composé de plusieurs **mesures** (indicateurs d'analyse). Les **dimensions** représentant les axes d'analyse sont formées de **paramètres** en fonction desquels les mesures sont étudiées. Les paramètres sont organisés en **hiérarchies**, de la granularité la plus fine (attribut racine servant d'identifiant à la dimension) à la plus générale. Dans le cadre des BDM, la structure de représentation des données la plus couramment utilisée est la Table Multidimensionnelle (TM) : tableau représentant les valeurs d'un fait en fonction de deux dimensions (Gyssens et Lakshmanan, 1997 ; Ravat *et al.*, 2006).

## 1.2 Limites des systèmes d'aide à la décision

Les tables multidimensionnelles, même si elles ont pour objectif de faciliter les prises de décisions, ne peuvent suffire. En effet, les prises de décisions reposent non seulement sur les données brutes mais également sur les réflexions, les commentaires des décideurs voire la confrontation de différentes interprétations. À notre connaissance, il n'existe pas d'outil logiciel permettant aux décideurs d'analyser les données décisionnelles en intégrant l'expertise des décideurs. La sauvegarde et la manipulation de l'expertise des décideurs permettront d'intervenir à deux niveaux :

- **Au niveau schéma.** La compréhension de la sémantique des composants et des instances d'une BDM est essentielle dans le processus de prise de décision. Cependant, les outils actuels sont insuffisants car ils se limitent à la description d'attributs dans des dictionnaires de données qui sont, au mieux, structurés de manière arborescente.
- **Au niveau des analyses décisionnelles.** Nous souhaitons faciliter la tâche des décideurs (usage personnel) tout en permettant le partage d'expertises (usage collectif) :
  - **Pour un usage personnel.** La spécification d'analyses décisionnelles au travers d'une TM accompagnée d'une réflexion critique s'apparente au concept de « lecture active » (Adler et van Doren, 1972). Les annotations aident le lecteur à matérialiser sa réflexion. Or actuellement, la formulation, la sauvegarde et la restitution informatique d'une TM et de ses annotations au sein d'un même outil n'est pas possible. Par conséquent, les expertises ne peuvent pas être partagées, exploitées voire réutilisées.
  - **Pour un usage collectif.** De nos jours, un outil d'analyse mono-utilisateur n'est plus suffisant. En effet, lorsque l'analyse est complexe, l'avis d'un autre expert est souvent sollicité, ce qui peut éventuellement donner lieu à des débats argumentés visant à atteindre un consensus pour une prise de décision collégiale.

Malheureusement, le support de cet échange n'est pas assuré par l'application informatique : l'analyste doit utiliser un moyen de communication classique (réunion, appel téléphonique, courrier électronique, etc.) pour établir le contact. Il doit ensuite restituer le contexte complet de l'analyse et formuler sa question à son interlocuteur. Cet échange gagnerait à être informatisé : une discussion en contexte pourrait accélérer la prise de décision. De plus, les arguments de la discussion pourraient être sauvegardés et réutilisés lors d'une expertise future.

### 1.3 Vers une mémoire d'expertises décisionnelles

Pour répondre à ce besoin, notre objectif est de proposer un cadre informatique permettant d'exploiter et de partager les données multidimensionnelles tout en supportant des fonctionnalités d'annotation pouvant impliquer des fils de discussions (principe de communication asynchrone employé notamment par les forums Usenet et Web). Ainsi, en permettant d'enrichir interactivement les composants des BDM et des TM, les décideurs deviendront des **lecteurs actifs**, créant leur propre système de repérage au travers de signes graphiques (surlignage, cerclage...), de commentaires, des réponses (affirmation, réfutation)...

Nous souhaitons proposer un outil informatique permettant :

- de visualiser les données décisionnelles sous un format de TM ;
- de pouvoir manipuler une telle TM au travers d'opérations multidimensionnelles ;
- d'annoter les schémas de BDM afin d'explicitier les composants décisionnels ;
- d'annoter les analyses décisionnelles dans les TM (ces annotations pouvant susciter des échanges et des débats au travers de fils de discussion) ;
- d'exploiter ces annotations au travers de fonctions d'exploration, d'édition, etc.

En considérant les deux premiers objectifs, de nombreux travaux ont apporté une réponse. Ainsi des algèbres pour la définition et la manipulation de BDM ont été proposées (Marcel 1999), notamment avec l'opérateur cube (Gray *et al.*, 1996). De plus, (Agrawal *et al.*, 1997) propose différents opérateurs algébriques dans le contexte relationnel tandis que (Abelló *et al.*, 2003 ; Cabibbo et Torlone, 1998) définissent plusieurs opérateurs pour un langage algébrique et/ou graphique. Certains travaux proposent également des opérateurs algébriques relatifs à la fusion de TM facilitant ainsi les corrélations entre sujets d'analyse (Benitez-Guerrero *et al.*, 2003 ; Ravat *et al.*, 2005).

Aussi, dans le cadre de cet article, nous nous centrons sur la définition d'un modèle multidimensionnel conceptuel de BDM intégrant un système d'annotations.

### 1.4 Existant : l'activité d'annotation

À notre connaissance, les systèmes d'annotation couplés aux BDM n'ont pas fait l'objet d'étude. Une première solution à notre problématique consiste à transposer le principe des commentaires associés aux schémas de BD transactionnelles au BD décisionnelles. Par exemple, la commande COMMENT d'Oracle permet d'associer un commentaire à une table, une vue ou une colonne et ce dernier est stocké dans le dictionnaire de données d'Oracle. Cette proposition est insuffisante pour répondre à notre problématique car elle reste difficilement exploitable et très limitée.

Par ailleurs, les annotations ont largement été étudiées dans le contexte de la gestion électronique de documents. Elles sont qualifiées « *[d'] informations ou marques supplémentaires apportées au document pour l'enrichir avec des explications brèves et utiles afin de permettre au lecteur de conserver une trace de ses réactions et par la suite marquer les passages*

*mis en valeur* »<sup>1</sup>. Ces annotations sont dites **informelles** (Marshall, 1998), contrairement aux annotations formelles qui sont employées pour cataloguer et indexer les documents dans des langages formels, avec un vocabulaire strict qui peut être par exemple issu d'une ontologie dans le cadre du Web Sémantique. Dans cet article, nous considérons uniquement les annotations informelles car nous ne souhaitons pas contraindre les décideurs à l'utilisation d'un vocabulaire normé et limité.

La majorité des systèmes d'annotation informatisés (SAI) opèrent sur des documents électroniques de type textuel. Dans ce cadre, les annotations sont matérialisées sous différentes formes : contenus textuels, marques libres (astérisques, accolades, flèches, soulignement, etc.). Elles permettent principalement de mettre en valeur des passages textuels en y associant éventuellement un commentaire (Marshall, 1998). À ce jour, on dénombre plus de vingt SAI : ce sont des prototypes de recherche *e.g.* Annotea/Amaya du W3C (Kahan *et al.*, 2002) comme des produits commerciaux tels qu'iMarkup Client ou Microsoft Office Web Discussions (Brush, 2002). Fondamentalement, un SAI permet de visualiser des documents, d'en consulter toutes les annotations et d'en créer de nouvelles.

Au regard de nos problématiques, nous proposons d'adapter les principes des SAI à l'analyse décisionnelle. Dans la section suivante, nous présentons les concepts de notre modèle de base de données multidimensionnelles annotées. Dans la section 3, nous étudions le méta-modèle servant de support à la mémoire d'expertise décisionnelle.

## 2 Modélisation multidimensionnelle annotée

Notre modèle, comme la plupart des modèles de représentation des données multidimensionnelles, repose sur une modélisation des sujets d'analyse (faits) en fonction d'axes d'analyse (dimensions).

### 2.1 Dimension et Hiérarchie

Les dimensions représentent les axes d'analyse. Une dimension est formée d'attributs exprimant les caractéristiques en fonction desquelles sont analysées les mesures d'activité.

**Définition.** Une dimension  $D$  est définie par  $(N^D, A^D, H^D, I^D)$  où

- $N^D$  est le nom de la dimension,
- $A^D = \{a_1, a_2, \dots, a_u\}$  est un ensemble d'attributs,
- $H^D = \{h^D_1, h^D_2, \dots, h^D_v\}$  est un ensemble de hiérarchies,
- $I^D = \{I^D_1, I^D_2, \dots\}$  est l'ensemble des instances de  $D$ .

On pose IDE un ensemble d'identifiants. Toute dimension possède les attributs All et Id tels que  $\text{dom}(\text{All}) = \{\text{'all'}\}$  et  $\text{dom}(\text{Id}) \in \text{IDE}$  ; All désigne la granularité de plus haut niveau tandis que Id représente la granularité la plus fine.

Les hiérarchies permettent de déterminer les niveaux de granularité auxquels peuvent être manipulées les mesures des faits. Chaque hiérarchie propose un ordonnancement des attributs d'une dimension afin de spécifier ces différents niveaux de granularité.

**Définition.** Une hiérarchie  $h^D$  est un chemin élémentaire acyclique débutant par Id et se terminant par All. Elle est définie par  $(N^{\text{hd}}, \text{Param}^{\text{hd}}, \text{Suppl}^{\text{hd}})$  où

- $N^{\text{hd}}$  est le nom de la hiérarchie,

---

<sup>1</sup> *cf.* le site Web du projet Annotea initié par le W3C : <http://www.w3.org/2001/Annotea/>

- $\text{Param}^{\text{hd}} : P \rightarrow P (P \subseteq A^D)$  est une fonction décrivant la hiérarchie des attributs, appelés paramètres de la hiérarchie,
- $\text{Suppl}^{\text{hd}} : P \rightarrow 2^{(A^D-P)}$  est une fonction spécifiant les attributs faibles qui complètent les paramètres (un ensemble d'attributs faibles est associé à un paramètre).

## 2.2 Fait

Chaque sujet d'analyse est représenté par un fait caractérisé par une ou plusieurs mesures représentant les indicateurs analysés.

**Définition.** Un fait  $F$  est défini par le quadruplet  $(N^F, M^F, I^F, \text{IStar}^F)$  où

- $N^F$  est le nom du fait,
- $M^F = \{f(m_1), f(m_2), \dots, f(m_w)\}$  est un ensemble de mesures (ou indicateurs) agrégées selon une fonction  $f \in \{\text{AVG}, \text{SUM}, \text{MAX}, \text{MIN}, \text{COUNT}, \dots\}$ ,
- $I^F = \{I_1^F, I_2^F, \dots\}$  est l'ensemble des instances de  $F$ ,
- $\text{IStar}^F$  est une fonction associant les instances de  $I^F$  aux instances des dimensions liées au fait.

## 2.3 Constellation

Une base de données multidimensionnelles est modélisée au sein d'une constellation. Le schéma en constellation est une généralisation du schéma en étoile (Kimball et Ross, 2002) dans laquelle plusieurs sujets d'analyse (faits) sont associés à différents axes d'analyse (dimensions) potentiellement partagés.

**Définition.** Une constellation  $C$  est définie par  $(N^C, F^C, D^C, \text{Star}^C, \text{ANNOTE}^C)$  où :

- $N^C$  est le nom de la constellation,
- $F^C = \{F_1, F_2, \dots, F_p\}$  est un ensemble de faits,
- $D^C = \{D_1, D_2, \dots, D_q\}$  est un ensemble de dimensions,
- $\text{Star}^C : F^C \rightarrow 2^{D^C}$  est une fonction associant les faits aux dimensions afin de spécifier les sujets d'analyses et les axes d'étude associés,
- $\text{ANNOTE}^C = \{AD_1^C, AD_2^C, \dots\}$  est l'ensemble des annotations décisionnelles globales (cf. section 2.5) définies sur la constellation et les éléments de cette dernière.

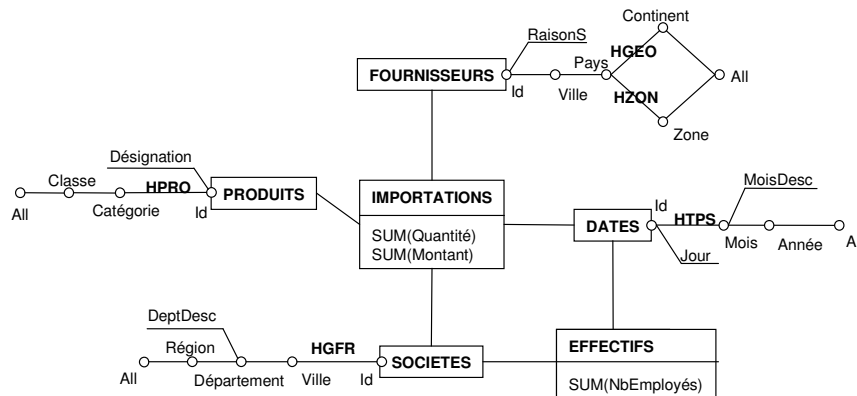


FIG. 1 – Schéma d'une base de données multidimensionnelles.

**Exemple.** Les décideurs d'un ministère désirent analyser les quantités de produits importés en France, le montant de ces importations ainsi que les effectifs des sociétés françaises en fonction des produits importés, des sociétés fournissant ces produits et du temps. Pour ce faire une base de données multidimensionnelles est modélisée conformément au schéma :

```
('SH_IMPORT',
 {IMPORTATIONS, EFFECTIFS},
 {PRODUITS, DATES, SOCIETES, FOURNISSEURS},
 {IMPORTATIONS→{PRODUITS, DATES, SOCIETES, FOURNISSEURS},
 EFFECTIFS→{DATES, SOCIETES}},
 ANNOTEC).
```

Pour représenter le schéma d'une constellation, nous adoptons des notations graphiques proches de (Golfarelli *et al.*, 1998).

## 2.4 Table Multidimensionnelle

La visualisation d'une constellation est centrée sur un seul fait au travers d'une table multidimensionnelle. Le choix de visualiser en détail les données d'un seul fait réduit la complexité de l'information visualisée et facilite donc l'interprétation et l'analyse des données. La représentation sous forme de tableau à double entrées hiérarchisées est motivée par sa simplicité et sa précision ; il s'agit d'une représentation répandue à laquelle les décideurs sont habitués (Gyssens et Lakshmanan, 1997) notamment au travers des outils du marché.

**Définition.** Une table multidimensionnelle TM est définie par (S, L, C, R ANNOTE<sup>TM</sup>)

- S = (F, { $f_1(m_1), \dots, f_p(m_p)$ }) est le sujet analysé au travers d'un fait et d'un ensemble de p mesures  $m_1, \dots, m_p$  agrégées à l'aide de fonctions d'agrégations  $f_1, \dots, f_p$ ,
- L = (DL, HL,  $\langle p_{1, \dots, p}^{DL} \rangle$ ) est l'axe d'analyse en ligne contenant la dimension DL, une hiérarchie courante HL et une liste de paramètres affichés dans la table,
- C = (DC, HC,  $\langle p_{1, \dots, p}^{DC} \rangle$ ) est l'axe d'analyse en colonne contenant la dimension DL, une hiérarchie courante HL et une liste de paramètres affichés dans la table,
- R =  $\text{pred}_1 \wedge \dots \wedge \text{pred}_s$  est une restriction des données analysées exprimée par prédicats en forme normale disjunctive portant sur les dimensions et le fait,
- ANNOTE<sup>TM</sup> = {AD<sup>TM</sup><sub>1</sub>, AD<sup>TM</sup><sub>2</sub>, ...} est l'ensemble des annotations locales (*cf.* section 2.5) définies sur la table et les éléments de cette dernière.

**Exemple.** La figure suivante présente l'exemple d'une table multidimensionnelle obtenue à partir du schéma de la constellation décrite en figure 1. Cette table multidimensionnelle représente la somme des importations annuelles de produits de la classe 'Electronique' par continent et par pays d'origine des fournisseurs de produits importés.

IMPORTATIONS			DATES   HTPS			
SUM(Montant)			Année	2003	2004	2005
FOURNISSEURS	Continent	Pays				
HGEO	Asie	Chine		100	120	230
		Inde		50	80	160
		Thaïlande		90	110	220
	Amérique	Etats-Unis		160	120	100
		Brésil		20	40	100
PRODUITS.Classe='Electronique'						

FIG. 2 – Analyse du montant des importations annuelles par continent et par pays pour les produits de la classe 'Electronique'.

## 2.5 Annotations

Comme indiqué précédemment, notre MED permet de mémoriser et d'utiliser non seulement les données décisionnelles (décrites au travers des concepts définis précédemment) mais également l'ensemble du capital immatériel mentalement associé à ces données par les décideurs. Notre proposition consiste à modéliser ce savoir-faire au travers d'annotations. Les annotations visent à conserver les commentaires formulés lors des analyses et les décisions prises. L'expertise que véhiculent ces annotations est utilisée à des fins personnelles ou collectives et elles contribuent à améliorer les analyses futures.

Les annotations contiennent des informations de plusieurs natures :

- les informations subjectives regroupant un contenu et le type de l'annotation, et
- les données objectives comportant un identifiant, une date de création, le créateur, des références à une annotation père, un point d'ancrage.

**Définition.** Une annotation décisionnelle  $AD_g$  est définie par le couple  $(IS_g, DO_g)$

- $IS_g$  = un ensemble d'informations subjectives regroupant
  - le contenu textuel formulé par le décideur qui annoté,
  - le type de l'annotation caractérisant son contenu (commentaire, question, réponse à une annotation, conclusion).
- $DO_g$  = un ensemble de données objectives comportant
  - son identifiant,
  - sa date de création permettant de caractériser sa position dans le fil de discussions ordonnées chronologiquement,
  - son créateur (décideur) avec sa fonction et ses coordonnées,
  - une référence éventuelle à une annotation père,
  - son point d'ancrage spécifiant la localisation précise de l'annotation.

Il est important de noter que le point d'ancrage peut être défini

- pour désigner un ancrage global sur un concept dans une constellation (l'annotation sera présente dans toutes les TM intégrant le concept annoté globalement), ou bien,
- pour désigner un ancrage local sur un élément dans une table multidimensionnelle (l'annotation n'est présente que dans le contexte local de la TM).

**Exemple.** La définition des annotations est réalisée par une simple action, soit sur le schéma de la constellation, soit sur une TM (comme l'indique la figure 3).

L'ensemble des données objectives est produit automatiquement par le système ; le point d'ancrage est calculé automatiquement en fonction de l'élément désigné par l'annotateur. Remarquons que plusieurs ancrages au sein d'une table multidimensionnelle peuvent être produits, l'utilisateur choisit alors celui qui sera conservé (*cf.* section 2.5.2). Les informations subjectives sont spécifiées par l'annotateur au travers d'un formulaire qui lui est proposé.

La définition du point d'ancrage ne peut se faire de manière simple car les annotations portent sur deux niveaux dans la base multidimensionnelle : global ou local. Le point d'ancrage doit pouvoir être défini suivant la même spécification pour désigner un ancrage global sur un concept d'une constellation, ou bien, pour désigner un ancrage local sur un élément d'une table multidimensionnelle.

Dans la suite de l'article,  $[elt]?$  désigne  $elt$  comme optionnel dans la spécification,  $[elt]^*$  désigne une répétition 0 à  $n$  de  $elt$ ,  $\{ elt1 \mid elt2 \}$  désigne l'alternative entre  $elt1$  et  $elt2$ , et  $\lambda$  désigne une expression de chemin vide.

## Méta-modélisation des bases de données multidimensionnelles annotées

IMPORTATIONS			DATES   HTPS			
SUM(Montant)			Année	2003	2004	2005
FOURNISSEURS HGEO	Asie	Chine		100	120	230
		Inde		50	80	160
		Thaïlande				
	Amérique	Etats-Unis				
		Bésil				

PRODUITS.Classe='Electronique'

Création d'une annotation (identifiant : @25)

Point d'ancrage défini par la sélection

TM.IMPORTATIONS.SUM(Montant)=230.DATES.  
HTPS/Année=2005

Informations subjectives

Contenu textuel :

Doublément des importations : Ouverture des  
quotas Européens d'importation le 01/01/2005

Type de l'annotation

commentaire     conclusion     question

Portée de l'annotation

globale     locale

Autoriser des réponses    Références..

OK    Annuler

Enregistre l'annotation

Auteur : Guillaume Cabanac (cabanac@irit.fr)    Jeudi 28 sept. 2005 13:15

FIG. 3 – Formulaire de saisie d'une annotation.

**Définition.** Un point d'ancrage  $\alpha$  est défini par le triplet  $(S, D_1, D_2)$  où :

- $S = \{C \mid TM\}.[F.[f(m)[=val]?]??$  désigne un ancrage relatif au fait F,
- $D_1 = \lambda \mid D.[H[/p[=pos]?]*]??$  désigne un ancrage relatif à une dimension,
- $D_2 = \lambda \mid D.[H[/p[=pos]?]*]??$  désigne un ancrage relatif à une dimension.

Notons que C désigne une constellation, TM désigne une table multidimensionnelle, F est un fait,  $f(m)$  est une mesure associée à une fonction d'agrégation, val représente une valeur prise par la mesure, D désigne une dimension, H désigne une hiérarchie, p désigne un paramètre, pos représente une valeur prise par le paramètre.

### 2.5.1 Annotations globales

Les annotations globales sont définies sur un élément conceptuel d'une constellation : fait, mesure, dimension, hiérarchie, paramètre ou attribut faible. Ces annotations sont indépendantes de tout contexte d'analyse, et apparaissent dans toutes les tables multidimensionnelles construites à partir de la constellation.

Le point d'ancrage  $\alpha^C$  d'une annotation globale est défini par le triplet  $(C.[F.[f(m)[=val]?]??, \lambda \mid D.[H[/p[=pos]?]*]?, \lambda)$ . Le point d'ancrage  $\alpha^C$  a pour racine une constellation C, et peut intégrer le fait, les mesures, les dimensions, les hiérarchies et les paramètres dans la spécification du chemin. Il faut noter qu'un point d'ancrage global n'intègre ni les valeurs des mesures, ni les positions (valeurs) des paramètres.



**Exemple.** L'exemple suivant présente la constellation relative aux importations. Le concepteur du schéma a placé diverses annotations globales,

- au niveau du fait IMPORTATIONS,
- au niveau de la dimension SOCIETES,
- au niveau du paramètre Département (le concepteur annotateur précise le nom du fait dans le point d'ancrage car il souhaite rendre l'annotation disponible uniquement à partir du fait IMPORTATION, et non pas du fait EFFECTIFS),
- au niveau du paramètre Année (le fait n'étant pas spécifié dans le point d'ancrage, l'annotation concerne tous les faits liés à la dimension DATES).

Ces annotations sont visibles dans la constellation, mais également dans toutes les tables multidimensionnelles produites à partir de la constellation.

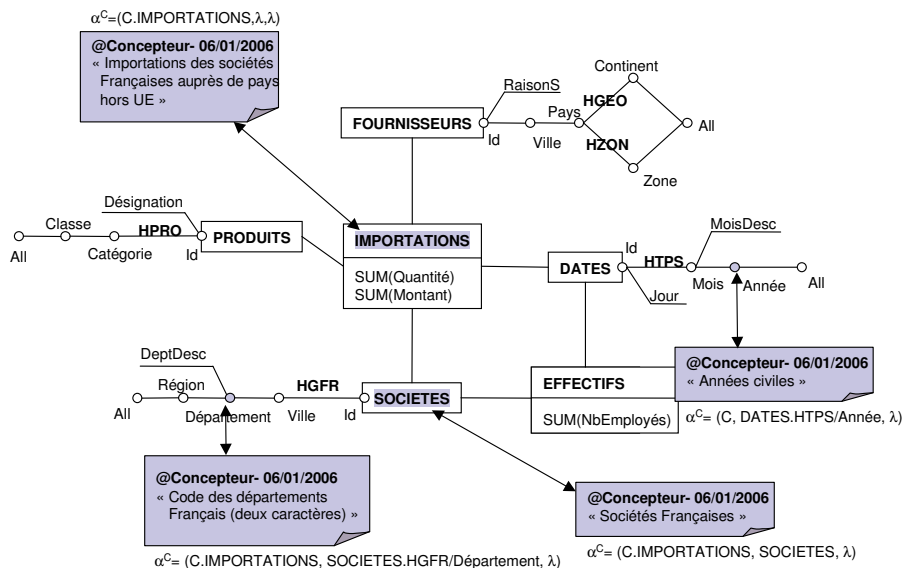


FIG. 4 – Exemple d'annotations globales avec leur point d'ancrage.

### 2.5.2 Annotations locales

Les annotations locales sont définies au sein d'une table multidimensionnelle ; elles sont par conséquent spécifiques au contexte de l'analyse que représente la table.

Le point d'ancrage  $\alpha^C$  d'une annotation globale est défini par le triplet  $(C[F.f(m)[=val]?]?], \lambda \mid DL[.HL[/p[=pos]?]*]?], \lambda \mid DC[.HC[/p[=pos]?]*]?]$ . Elles peuvent être attachées localement au fait, aux mesures, à l'une des dimensions ou des hiérarchies courantes (en ligne ou colonne), aux paramètres affichés, aux positions de ces derniers et enfin, aux valeurs observées contenues dans les cellules de la table multidimensionnelle.

**Exemple.** L'exemple suivant présente la table multidimensionnelle des importations annuelles de produits de la classe 'Electronique' par pays d'origine des fournisseurs. Trois anno-

## Méta-modélisation des bases de données multidimensionnelles annotées

tations locales à la TM sont disponibles, écrites par le Concepteur, le Directeur des Importations et le Directeur Général.

- L'annotation du Concepteur, portant sur le paramètre Année, est une annotation globale spécifiée sur la constellation, et visible sur toutes les tables multidimensionnelles intégrant le paramètre Année.
- L'annotation du Directeur Général porte sur toute la ligne correspondant aux importations de produits provenant des Etats-Unis (TM.IMPORTATIONS.SUM(Montant), FOURNISSEURS.HGEO/Continent='Amérique'/Pays='Etats-Unis', λ).
- L'annotation du Directeur des Achats porte sur la valeur 230 de la somme des montants des importations (TM.IMPORTATIONS.SUM(Montant)=230) de l'année 2005 (DATES.HTPS/Année=2005) et des fournisseurs chinois (FOURNISSEURS.HGEO/Continent='Asie'/Pays='Chine'). On peut constater que cette annotation est ancrée dans la TM à partir de la dimension des Fournisseurs et de la dimension des Dates.

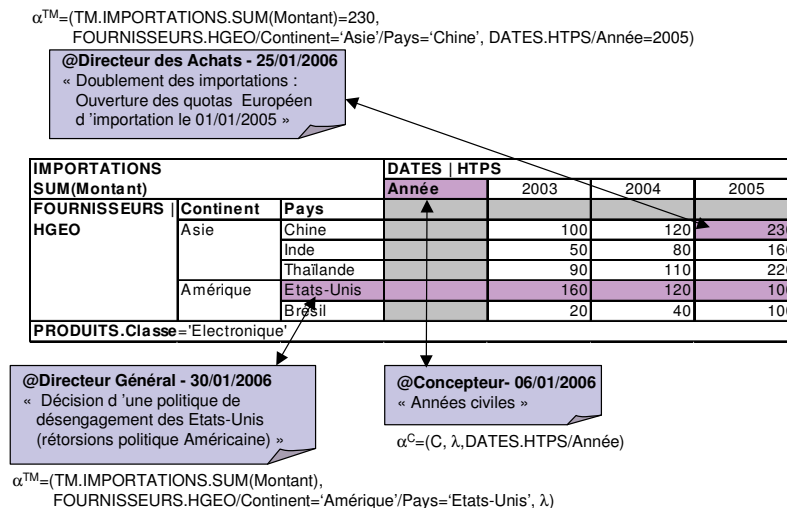


FIG. 5 – Exemple d'annotations locales avec leurs points d'ancrage.

### 3 Méta-modélisation d'une constellation annotée

Dans cette section, nous présentons le méta-modèle supportant les concepts utilisés dans la spécification d'une Mémoire d'Expertises Décisionnelles (MED). Ce méta-modèle est représenté à l'aide d'un diagramme de classes UML. Par souci de simplification, nous présentons les différents concepts au travers de trois schémas. Le premier permet de spécifier les composants du modèle multidimensionnel, le second se focalise sur la manipulation multidimensionnelle au travers de TM et enfin, le dernier présente les concepts relatifs à l'annotation d'une BDM. Pour ne pas alourdir le schéma, nous n'avons pas fait figurer la partie comportementale des classes.

### 3.1 Composants du modèle multidimensionnel

Tous les concepts de ce modèle possèdent un nom unique, une date de création et une date de dernière modification. Lors de la création d'une BDM, le décideur crée un schéma et détermine la liste des décideurs autorisés à manipuler ce schéma. Un schéma de BDM est composé d'un ou plusieurs faits comprenant une ou plusieurs mesures dont on conservera le nom, le domaine et la fonction d'agrégation associée. Chaque fait est analysable selon différents axes d'analyse. Tout axe d'analyse, également appelé dimension, est composé de hiérarchies. Une hiérarchie permet de définir les différents niveaux d'agrégation des attributs d'une dimension (attribut Position de la classe Niveau). Chaque niveau d'agrégation comprend un paramètre éventuellement complété d'un ou plusieurs attributs faibles ; l'attribut booléen `Attrib_Faible` de la classe Niveau permet de préciser si l'attribut de la dimension joue ou non le rôle de l'attribut faible.

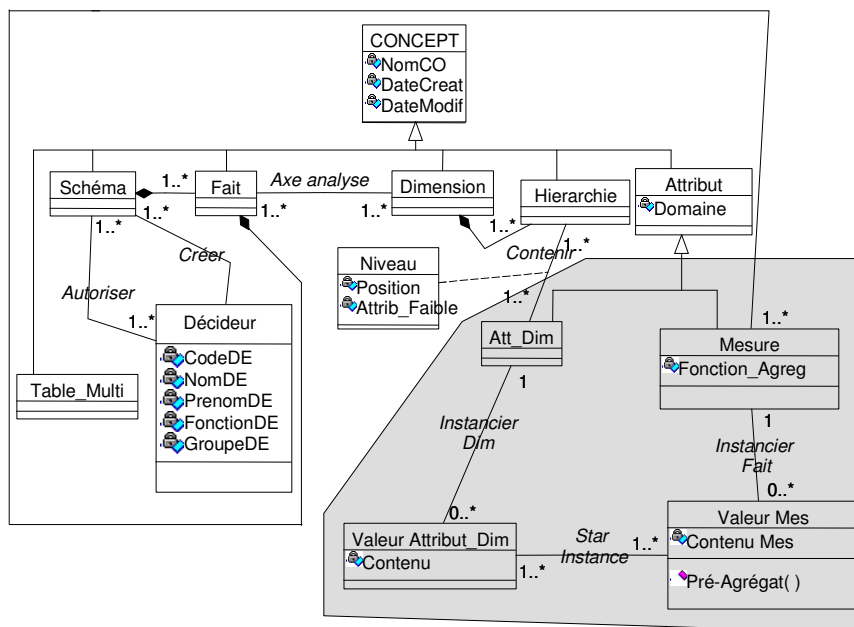


FIG. 6 – Méta-modélisation des concepts d'une constellation annotée.

La partie grisée correspond à l'instanciation d'une BDM. Tout attribut de dimension est instancié par un ensemble de Valeurs (classe "Valeur Attribut\_Dim" et association "Instancier Dim"). L'instanciation initiale d'une mesure consiste à lui affecter la valeur associée aux valeurs des paramètres de plus basse granularité (classe "Valeur Mes" et association labélisée "Star Instance"). Par contre, la valeur d'une mesure en corrélation avec des attributs de dimensions de granularité quelconque peut se calculer à l'aide de la fonction `Pré-Agrégat()`.

### 3.2 Manipulation multidimensionnelle

Cette seconde partie de notre méta-modèle permet de traduire la manipulation d'un schéma multidimensionnel au travers de Tables Multidimensionnelles (TM).

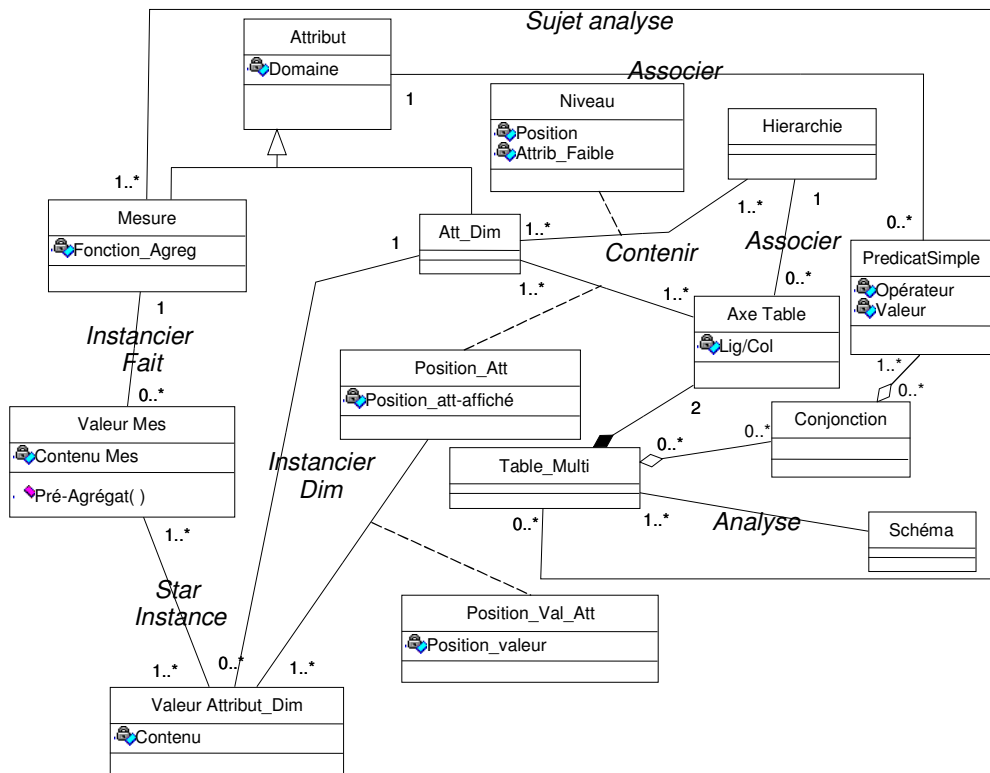


FIG. 7 – Méta-modélisation des tables multidimensionnelles annotées.

Toute TM est construite à partir d'un schéma. Comme indiqué dans la section 2, à l'exception des annotations, une TM comprend 4 composants : un sujet d'analyse matérialisé au travers de ces mesures (toute mesure appartenant à un seul sujet d'analyse), deux axes d'analyse (ligne et colonne) comprenant une hiérarchie d'une dimension (association labellisée Associer), et éventuellement un prédicat de restriction. Le prédicat de restriction d'une TM est exprimé au travers d'une conjonction de prédicats (classe Conjonction). Chaque prédicat est composé d'une disjonction de prédicats simples. Un prédicat simple comprend un attribut (paramètre, mesure ou attribut faible), un opérateur et une valeur.

Pour chaque axe de la table, en plus de la hiérarchie affichée, nous conservons la position de chaque attribut de dimension affichée au sein de cet axe (classe Position\_Attr permettant de modéliser la liste des attributs affichés). De plus, pour conserver l'ordre d'affichage des valeurs d'un attribut de dimension, le système instancie la classe "Pos\_val\_Attr".

### 3.3 Annotation des concepts

Cette dernière partie du méta-schéma permet de mettre en avant le capital immatériel d'une MED. Cette partie permet de gérer les annotations. Une annotation est un concept de notre méta-modèle (le nom du concept étant unique, il joue le rôle d'identifiant d'annotation comme défini dans la section 2.5). Une annotation possède un contenu un type dont les valeurs sont précisées dans la section précédente et éventuellement une annotation "père" lors de la réponse à une annotation précédemment définie.

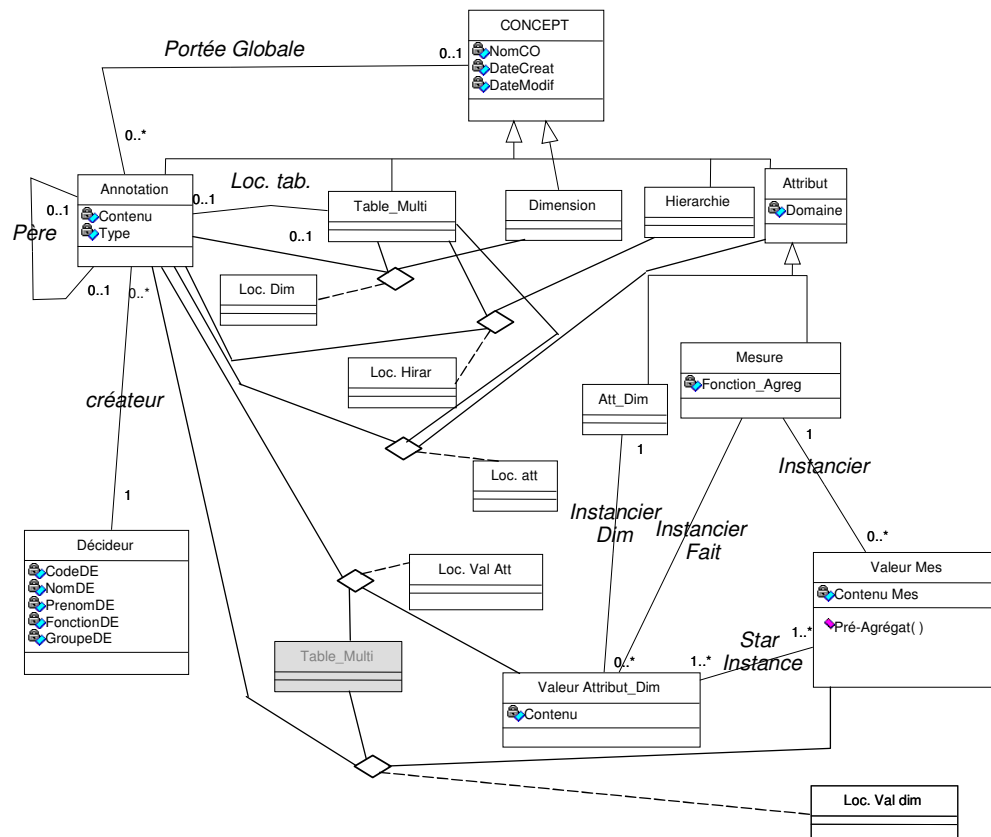


FIG. 8 – Méta-modélisation des annotations.

Toute annotation possède une portée globale ou locale. Une annotation globale est relative à un concept du modèle multidimensionnel tandis qu'une annotation locale permet de commenter un des composants précis d'une TM. Pour traduire ces annotations locales, nous avons défini des classes d'association ternaires liant l'annotation la table concernée et les composants de la table (dimension, hiérarchie, mesure, attribut de dimension, valeur d'un attribut de dimension ou valeur d'une mesure). Pour faciliter la lecture de ce méta-schéma, nous avons dupliqué la classe "Table Multi" ; cette copie est grisée.

## 4 Conclusion et perspectives de recherche

Dans cet article, nous avons proposé un modèle et un méta-modèle pour décrire une Mémoire d'Expertises Décisionnelles (MED). Une MED permet de conserver dans un même environnement les données décisionnelles et les commentaires, remarques, questions et/ou réponses formulés par les décideurs lors de la définition des schémas ou des analyses décisionnelles. Plus précisément, les données décisionnelles sont modélisées au travers de schémas en constellation, faits, dimensions, hiérarchies, mesures, attributs faibles, paramètres et tables multidimensionnelles. Lors de la définition d'un schéma en constellation, le concepteur peut annoter ses composants. Ces annotations, dites globales, seront visibles à tout moment : lors de la définition ou de la modification d'un schéma ainsi que lors des analyses décisionnelles. Ces analyses s'effectuent au travers de Tables Multidimensionnelles (TM) annotées comportant les données brutes et les annotations locales associées. Ces annotations servent aussi bien à une lecture active des données décisionnelles (usage personnel) qu'à un usage collectif facilitant les prises de décisions (fil de discussion liant des annotations).

Le méta-modèle permet de représenter l'ensemble des concepts nécessaire à la mise en place et à l'exploitation d'une BDM Annotée.

Nous poursuivons ces travaux par l'étude de techniques permettant l'exploitation des annotations. Principalement, nous envisageons de développer un environnement décisionnel adaptatif qui pourrait lors de processus d'analyses décisionnelles similaires (à des analyses antérieures) anticiper les besoins des décideurs par des propositions d'expertises passées, stockées au sein des annotations. Ceci nécessite le développement d'une mesure de similarité entre deux contextes d'analyse. Nous pensons également proposer de nouveaux algorithmes pour déclencher préventivement des calculs de pré-agrégats en fonction des usages avérés détectés comme probables.

## Références

- Abelló, A., J. Samos, et F. Saltor (2003). Implementing Operations to Navigate Semantic Star Schemas. In *DOLAP '03 : Proc. 6<sup>th</sup> ACM Int. Workshop on Datawarehousing and OLAP*, New York, NY, USA, pp. 56–62. ACM Press.
- Adler, M.J. et C. van Doren, (1972). *How to Read a Book*. Simon & Shuster, NY, 1972.
- Agrawal, R., A. Gupta, et S. Sarawagi (1997). Modeling Multidimensional Databases. In A. Gray et P.-Å. Larson (Eds.), *Proc. 13<sup>th</sup> Int. Conf. Data Engineering, ICDE*, pp. 232–243. IEEE Computer Society.
- Benítez-Guerrero, E., Collet, C., Adiba, M. (2003), Le système WHES pour l'évolution des entrepôts de données, *BDA 2003, Lyon*.
- Brush, A. J. B. (2002). Annotating Digital Documents for Asynchronous Collaboration. Technical report 02-09-02, Department of Computer Science and Engineering, University of Washington, USA.
- Cabibbo, L. et R. Torlone (1998). From a Procedural to a Visual Query Language for OLAP. In *SSDBM '98 : Proc. 10<sup>th</sup> Int. Conf. on Scientific and Statistical Database Management*, Washington, DC, USA, pp. 74–83. IEEE Computer Society.
- Dieng-Kuntz, R. (2005). Le Web Sémantique pour la Gestion des Connaissances. *Conférence invité à EGC'2005 - Extraction et Gestion des Connaissances, Paris*

(<http://www.polytech.univ-nantes.fr/associationEGC/presentations/CI-RDK-transp-EGC2005-dieng.pdf>).

- Golfarelli, M., Maio, D. et Rizzi, S. (1998). Conceptual design of data warehouses from E/R schemes, *31<sup>st</sup> Hawaii International Conference on System Sciences, 1998*.
- Gray, J., A. Bosworth, A. Layman, et H. Pirahesh (1996). Data Cube : A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Total. In S. Y. W. Su (Ed.), *Proc. 12<sup>th</sup> Int. Conf. on Data Engineering, February 26 - March 1, 1996, New Orleans, Louisiana*, pp. 152–159. IEEE Computer Society.
- Gyssens, M. et L. V. S. Lakshmanan (1997). A Foundation for Multi-Dimensional Databases. In M. Jarke, M. J. Carey, K. R. Dittrich, F. H. Lochovsky, P. Loucopoulos, et M. A. Jeusfeld (Eds.), *Proceedings of the 23<sup>rd</sup> International Conference on Very Large Data Bases*, pp. 106–115, August 25-29, 1997, Athens (Greece).
- Inmon, W. H. (1996). *Building the Data Warehouse (2nd ed.)*. New York, NY, USA : John Wiley & Sons, Inc.
- Kimball, R. (1996). *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. New York, NY, USA : John Wiley & Sons, Inc.
- Kimball, R., Ross, M. (2002). *The Data Warehouse Toolkit*, John Wiley & Sons, New York, deuxième édition, 2002.
- Kahan, J., Koivunen, M-R., Prud'Hommeaux, É., et Swick, R.R. (2002). Annotea: an open RDF infrastructure for shared Web annotations. In *Computer Networks*, vol. 2, n°5, pp. 589-608, 2002.
- Marcel, P. (1999). Modeling and Querying Multidimensional Databases: an Overview. *Networking and Information Systems* 2(5-6), 515–548.
- Marshall, C. (1998). Toward an ecology of hypertext annotation. In *Proceedings of the 9<sup>th</sup> ACM conference on Hypertext and hypermedia: links, objects, time and space*, pp. 40-49, ACM Press, Pittsburgh (USA).
- Ravat, F., Teste, O. et Zurfluh, G. (2005). Manipulation et fusion de données multidimensionnelles. In *Revue des Nouvelles Technologies de l'Information (RNTI-E-3) – numéro spécial des actes de la conférence Extraction et Gestion des Connaissances, Volume I, Cépaduès*, pp.349–354, Toulouse.
- Ravat, F., Teste, O. et Zurfluh, G. (2006). Constraint-Based Multi-Dimensional Databases. *Chapitre XI de l'ouvrage "Database Modeling for Industrial Data Management"*, sous la direction de Zongmin Ma, IDEA Group (ed.), pp.323-368.

## Summary

This paper deals with Decisional Expertise Memory (DEM). A DEM regroups both a decisional system and mechanisms for preserving managers' expertise. The decisional system we present is based on Multidimensional DataBases (MDB), which are composed of facts, dimensions, measures hierarchies and parameters. Managers' expertise is modelled through annotations. Each annotation may be related to multidimensional structures or values. We provide a conceptual model and a meta-model for annotated MDB.