

2 Modèle conceptuel des données

2.1 Constitution du dictionnaire des données

2.1.1 Etude des rubriques de documents: passer des rubriques aux données

Les informations recueillies se trouvent sur des documents variés et ne sont pas nécessairement prêtes pour un traitement systématique. Par exemple, il y a des quantités de manières de disposer une adresse... Il importe, à ce stade de l'étude, de se poser des questions du type suivant:

Telle donnée (par exemple une adresse) doit-elle être considérée comme un tout, ou au contraire résulte-t-elle de la juxtaposition d'éléments séparés?

Et, pour répondre de façon pertinente à une telle question, il faut se la poser sous la forme voisine suivante:

Y a-t-il éventuellement des traitements qui feront intervenir une partie de cette donnée (par exemple la partie "rue" ou la partie "ville" d'une adresse).

Autrement dit, les documents de base comportent des *rubriques*, et à partir de ces rubriques, il faut définir les *données élémentaires* relatives au cas étudié.

Par définition, une **donnée élémentaire** est la représentation d'informations élémentaires manipulées dans le domaine étudié.

Rubriques de documents et *données* sont deux notions différentes. Une donnée, comme une rubrique, peut être composée (ex. numéro de sécurité sociale). En revanche, le terme de *donnée élémentaire* indique qu'il est exclu de décomposer une telle donnée, dans le cadre des activités du domaine. Ce n'est pas nécessairement le cas pour les rubriques de documents. Par exemple, l'adresse du client peut ne constituer qu'une seule rubrique sur un document, et correspondre à trois ou quatre données: rue, code postal, ville, pays.

D'autre part, une donnée a un type bien déterminé (alphanumérique, numérique ...) alors qu'une rubrique n'est pas nécessairement normalisée (code numérique ou nom pour un département, par exemple).

2.1.2 Les types de données couramment acceptés par un SGBD

Pour chaque donnée identifiée, on lui attribue un nom et on en précise le type et éventuellement, le domaine de valeurs. Les principaux types de données sont décrits ci-dessous.

Type numérique

Par exemple, la population d'un département, le solde d'un compte sont des données numériques. Une donnée de ce type représente un nombre. Mais tout ensemble de chiffres n'est pas forcément une donnée numérique; ainsi, un numéro de sécurité sociale, un code postal ne sont pas des données numériques car les traitements qu'on leur fait subir ne sont pas ceux que l'on pratique sur les nombres.

Une donnée est numérique si elle peut, de manière naturelle, faire l'objet d'opérations mathématiques.

Type alphabétique

Par exemple, un nom, un prénom sont des données de type alphabétique.

Le plus souvent on généralise aux données contenant à la fois des lettres, des chiffres, certains symboles de ponctuation, etc. Ce sont alors des données *alphanumériques* dites encore *chaînes de caractères*;

Certaines données, bien qu'elles soient écrites uniquement à l'aide de chiffres, sont des données alphanumériques. Ainsi, par exemple, un code postal est une chaîne de caractères, car on ne fait pas d'opérations numériques sur une telle donnée.

Pour distinguer une chaîne de chiffres d'un nombre, de nombreux logiciels utilisent des guillemets :

- * "1 40 02 06 032 025" désigne la chaîne *un-quarante-zéro-deux-zéro-six-zéro-trente-deux-zéro-vingt-cinq*;
- * 1 400 206 032 025 désigne le nombre *mille quatre cent milliards deux cent six millions trente deux mille vingt cinq*.

Type texte

Il s'agit de données de nature alphanumérique, mais pouvant atteindre une grande longueur. Surtout, il s'agit de données de longueur imprévisible, par exemple des citations, des articles de presse, des lettres, etc.

Les opérations que l'on peut faire sur les textes sont différentes des opérations sur les données alphanumériques : on peut classer par ordre alphabétique les départements français, pas des citations.

Il faut noter que, dans un logiciel tel que Microsoft Access, le type alphanumérique est appelé "type texte" tandis que le type texte prend le nom de "type Mémo".

Type logique

Les données de ce type ne peuvent prendre que deux valeurs : vrai (noté suivant le cas vrai, oui, yes, true, .t., ...) ou faux (noté faux, non, false, .f., nil ...). Les données de ce type ont pour principal intérêt d'occuper une place très réduite en mémoire (un octet ou un bit selon le cas). Les opérations que l'on peut envisager sur les valeurs logiques sont les opérations logiques classiques *et, ou, non* etc.

Type date/heure

Ce type est important, notamment pour les applications de gestion. Les dates peuvent être ordonnées, mais les opérations arithmétiques courantes n'ont pas de sens : on ne peut pas multiplier deux dates, ni les additionner mais on peut ajouter un nombre (une durée) à une date, ou soustraire une date à une autre :

- * 29 juin 1990 + 3 jours est une date, le 2 juillet 1990 ;
- * 15 février 1995 - 6 janvier 1995 est une durée (c'est-à-dire un nombre suivi d'une unité de temps), 40 jours ;

Type "objet externe"

De nombreux SGBD acceptent un type "données binaires" ou "objet externe". Dans ce cas, le SGBD se borne à mémoriser telles quelles les données qui lui sont fournies, sans posséder de fonctions lui permettant de traiter ou de mettre en forme ces données. On peut ainsi mémoriser dans un SGBD des sons ou des images numérisés, des images animées... Sous MS Access, le type de données "objet OLE" permet de stocker dans une base de données des objets aussi divers qu'un dessin, un document Word, un son etc.

2.1.3 Constitution du dictionnaire des données

Au fur et à mesure de l'étude, et pour chaque donnée identifiée :

- on lui attribue un nom, qui sera utilisé dans toute la suite de l'étude ;
- on en indique le type, le domaine de valeurs ;
- on en mentionne une description "en clair" ;
- on en étudie des propriétés telles que :
 - s'agit-il d'une donnée de base ou d'une donnée calculée?
 - s'agit-il d'une donnée stable ou non stable (pouvant évoluer au cours du temps)?

Remarque. On prendra garde à ne pas confondre *donnée numérique* (donnée sur laquelle on peut faire des calculs arithmétiques) et *donnée calculée* (donnée obtenue comme résultat d'un calcul). Ainsi, si l'on dispose d'une donnée *nom* et d'une donnée *prénom*, la donnée *identité*, formée du nom suivi du prénom, est une donnée alphanumérique *calculée*. En revanche, en l'absence de dates de naissance, l'âge est une donnée numérique non calculée.

On constitue ainsi une liste de données ; il importe, avant d'ajouter une donnée à la liste, de se poser les questions suivantes :

- Cette donnée n'est-elle pas déjà répertoriée?
- Cette donnée n'est-elle pas déjà répertoriée avec un nom différent (*synonymes*)?
- Le nom que l'on se propose d'attribuer à la donnée n'est-il pas déjà utilisé pour désigner une autre donnée (*polysèmes*)?

A la fin de ce travail, on dispose d'une liste de données sans redondances, sans synonymes ni polysèmes : le **dictionnaire des données**.

Exemple.

On veut constituer une base de données rassemblant des informations relatives à "l'historique" des études suivies par les étudiants, c'est-à-dire mémorisant les informations contenues dans des phrases telles que :

Gérard Gentil, N° étudiant 88272X, a obtenu la licence d'histoire (code HIST3) à la session de septembre 1997 (mention AB). Il avait obtenu le DEUG d'histoire (code HIST2) avec la mention passable en juin 1996.

Le dictionnaire des données est ici donné par :

Num	Nom	Type	Description	Prop.
1	Nom	alpha	nom de l'étudiant	
2	N° -Etudiant	alpha	numéro d'étudiant	
3	Cursus	alpha	filière du diplôme	
4	Niveau	alpha	nom du diplôme	
5	Code-diplôme	alpha	code du diplôme	
6	Année	num	année d'obtention	
7	Session	alpha	session d'obtention	
8	Mention	alpha	mention obtenue	

Remarque. L'étude est menée à un niveau *conceptuel*, qui suppose un certain niveau d'abstraction. Dans ce contexte, il importe de ne pas confondre les *données types* (par exemple, la donnée *Nom*, la donnée *Diplôme*) et les *occurrences* de ces données (par exemple *Gérard Gentil* ou *Jean Dupond*).

2.2 Passer des données aux entités

2.2.1 Mettre en évidence les entités

Dans la plupart des cas, les données se regroupent naturellement en différents "lots" correspondant à des objets ou concepts du domaine étudié. Dans l'exemple précédent, apparaissent ainsi deux concepts : *Etudiant* et *Diplôme*.

Dans les cas où ce regroupement n'apparaît pas de manière évidente, on peut étudier l'*univers du discours* du domaine. On y fait référence à des concepts (personnes, objets abstraits ou concrets) dans des phrases telles que celle qui a servi à introduire l'exemple précédent. De même, dans le cas "O.N.G." étudié en travaux dirigés, les phrases suivantes font partie de l'univers du discours du domaine :

L'intervenant Dupond est responsable de l'opération "Irrigation au Mali".

L'intervenant Durand est un infirmier bénévole ; il a été envoyé en mission dans le cadre de l'opération "Aide aux sinistrés en Colombie", du 1/01/99 au 25/05/99, sur le site d'Armenia.

L'opération "Vaccination" au Bangladesh se déroule sur les sites de Dacca et de Chittagong.

Dans ces phrases, apparaissent les concepts *Intervenant*, *Opération* ou *Intervention*, *Site d'intervention*.

Les représentations dans le système d'information, des concepts ou objets du domaine sont appelées **entités**. Chacune de ces entités représente un individu-type dont il existe plusieurs occurrences dans le domaine.

La notion d'entité est difficile à définir, car il s'agit d'une notion de base, sur laquelle est bâtie la modélisation. On peut cependant en donner la définition suivante.

Définition

Une **occurrence d'entité** est un exemplaire différenciable et identifiable d'une chose, d'une personne ou d'un concept abstrait pour lequel on doit gérer des informations significatives. Une **entité** est une collection homogène d'occurrences d'entités.

Exemples. Dans une entreprise, des entités *Produit*, *Client*, *Fournisseur*, ... Dans une bibliothèque universitaire, des entités *Livre*, *Etudiant*. Dans l'exemple ci-dessus, les entités *Etudiant* et *Diplôme*.

Remarque. Il ne faut pas confondre les *documents* et les *entités*. Dans certains cas, les documents sont des objets bien identifiés du monde réel (formulaires types, bons de commande, factures...). Mais ce ne sont pas pour autant des entités. Ce sont des *supports d'information* et non *des objets au sujet desquels* on mémorise de l'information.

2.2.2 Liaisons sémantiques entre les données: notion de dépendance fonctionnelle

A ce stade de l'étude, on s'efforce de rattacher chacune des données à une entité. Les données qui se rattachent à une entité constituent ses *attributs* ou *propriétés*. On s'aperçoit rapidement qu'il existe deux types de liaisons entre les données:

- Le rattachement d'une donnée à un concept (par exemple *Intitulé* se rattache naturellement à l'objet *Diplôme*);
- Une liaison entre concepts (par exemple, un *Etudiant* obtient un *Diplôme*).

Pour obtenir un modèle conceptuel des données correct, deux règles doivent impérativement être suivies:

- **Pas d'attribut multivalué.** Exemples:
 1. Un étudiant obtient plusieurs diplômes. La *liste des diplômes obtenus* n'est pas une donnée qui se rattache directement à l'entité *Etudiant*.
 2. Dans l'exemple vu en TD, une intervention se déroule sur plusieurs sites. L'entité *Intervention* ne possède donc pas d'attribut *Site*.
 3. En revanche, une intervention possède un *Objectif*, une *Date de début*, une *Durée*. Ces données élémentaires se rattachent naturellement à l'objet *Intervention* et constituent donc des attributs de l'entité de même nom.
- **Ne pas rattacher à un concept des données qui n'en dépendent pas directement.** En effet, une telle pratique conduit à introduire des *redondances*. Exemple:
 - Une intervention a un responsable et un seul. Mais, concernant ce dernier, on a répertorié tout un ensemble de données: *Nom*, *Adresse*, *Nationalité*. Faire de ces données des attributs de l'entité *Intervention* conduirait à les répéter pour chacune des interventions dont la personne est responsable. Il s'agirait de données redondantes, ce qui est la source d'éventuelles incohérences. En fait, *Intervention* et *Intervenant* sont deux entités liées par une association ... *est responsable de...*

Ces deux règles peuvent encore être exprimées de la façon suivante: **Les attributs d'une entité sont les données qui s'y rattachent par un lien de dépendance fonctionnelle directe.**

2.2.3 Identifiant d'une entité

Un exemplaire (une occurrence) d'une entité doit désigner un objet clairement individualisé du monde réel. De plus, on doit disposer dans l'univers du discours, d'une information permettant de désigner cet exemplaire sans ambiguïté. Exemples :

- Cote d'un ouvrage dans une bibliothèque ;
- Numéro d'une salle de cours ;
- Nom d'une personne, ou assemblage "Nom et Prénom" dans un carnet d'adresse.

Cette donnée est appelée **identifiant** ou **clé** de l'entité. Dans certains cas, il n'existe pas de donnée simple et commode ayant cette propriété. On peut alors créer alors un code jouant ce rôle.

Ainsi, un *Etudiant* est identifié par son *N° Etudiant*, un *Diplôme* par le *Code-diplôme*, etc.

2.3 Associations entre entités

2.3.1 Repérer les associations entre entités

Une bonne partie des données figurant dans le dictionnaire des données se rattache naturellement aux entités que l'on a créé. Mais il reste en général des données dont "on ne sait pas quoi faire". C'est le cas, par exemple, de la donnée *Mention* dans le cas *Etudiants et Diplômes*. D'autre part, certaines phrases de l'univers du discours ne sont pas représentées dans la structure des données limitée aux seules entités. Il s'agit des phrases mettant en jeu plusieurs entités :

- Tel étudiant a obtenu tel diplôme ;
- Tel intervenant est affecté sur telle intervention ;
- Tel client a commandé tel produit, ...

Ces liens sémantiques entre les données sont modélisés par des **associations** entre entités. Ces associations se repèrent facilement par des *verbes* de l'univers du discours qui expliquent le rôle d'une entité par rapport à une ou plusieurs autres.

Remarque. Une association peut faire intervenir plus de deux entités. Ainsi, dans le cas "ONG", un *Intervenant* est affecté à une *Intervention* sur un *Site* donné.

2.3.2 Attributs d'une association

Les associations peuvent également posséder des attributs. Exemples :

Un étudiant obtient un diplôme donné lors d'une session d'examen donnée, et avec une certaine mention. Les données *Session* et *Mention* sont des attributs de l'association ... *obtient*

Un intervenant est affecté à une intervention. Sa mission comporte une date de début et une date de fin de mission.

Si la modélisation est correcte, toutes les données figurant dans le dictionnaire des données devraient trouver leur place, soit comme attributs des entités, soit comme attributs des associations.

2.3.3 Remarque : entité ou association ?

Il n'est pas toujours immédiat de déterminer si un concept du monde réel doit être représenté par une entité ou par une association. Ainsi, dans un cas "bibliothèque", doit-on modéliser la situation :

- à l'aide de deux entités *Abonné*, *Ouvrage* et d'une association *emprunte* ;
- ou, à l'aide de trois entités *Abonné*, *Ouvrage*, *Emprunt* et de deux associations *effectue* et *concerne* (Un abonné effectue un emprunt concernant un ouvrage) ?

Il n'existe pas de règle absolue, mais on pourra toutefois remarquer que :

- une association-type n'a pas d'identifiant propre ; une occurrence d'une association est déterminée par les occurrences des entités-types qui participent à l'association ;
- à une occurrence de la collection des entités-types qui participent à une association, il ne peut correspondre qu' **une** occurrence de cette association ;

- enfin, entre deux solutions qui semblent équivalentes, il vaut toujours mieux choisir la plus simple.

Ainsi, dans l'exemple précédent, une modélisation sous la forme de deux entités ne permet pas de représenter l'historique des emprunts effectués par un abonné. En effet, un abonné peut très bien emprunter plusieurs fois le même ouvrage.

2.3.4 Cardinalités des associations

La notion de cardinalité d'une association est une notion délicate. C'est cependant une notion indispensable si on veut parvenir à une structure de base de données correcte.

Considérons, par exemple, l'association *... est responsable de ...* liant *Intervention* et *Intervenant*. Chaque intervention possède un responsable et un seul. On dit que la cardinalité de l'association *... est responsable de ...* par rapport à l'entité *Intervention* est **1,1**. En revanche, un intervenant est responsable de 0, 1 ou plusieurs interventions. La cardinalité de l'association *... est responsable de ...* par rapport à l'entité *Intervenant* est **0,n**.

Une telle association est dite de type *hiérarchique* ou encore *association "un à plusieurs"*. Elle est également qualifiée de *partielle* car, pour l'une des entités, la cardinalité minimale est 0.

On représente les cardinalités de l'association *... est responsable de ...* comme indiqué sur la figure 1 page 12.

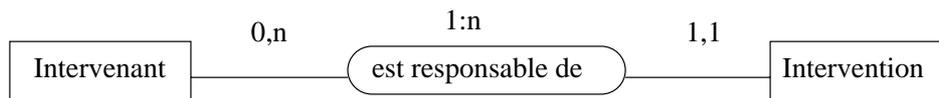


Figure 1: Cardinalités de l'association *... est responsable de ...*

Considérons l'association liant *Etudiant* et *Diplôme*. Chaque étudiant peut obtenir plusieurs diplômes, et chaque diplôme est délivré à plusieurs étudiants. Il s'agit là d'une association non hiérarchique, encore appelée *association multiple* ou *association "plusieurs à plusieurs"*. Ses cardinalités seront représentées de la manière indiquée sur la figure 2 page 12.

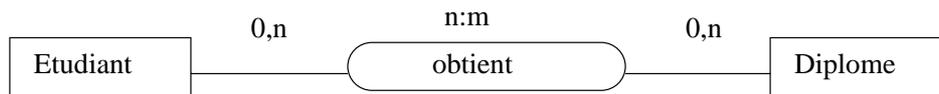


Figure 2: Cardinalités de l'association *... obtient ...*

D'une manière plus générale, les associations qui ne mettent en jeu que deux entités sont dites *binaires*. Selon les cardinalités de l'association vis-à-vis des entités participantes, on distingue :

1,1 — 1,1 ou 0,1 : Ces associations, que l'on rencontre rarement, sont dites "un à un".

1,1 — 1,n : Ces associations sont dites "un à plusieurs" ou hiérarchiques totales.

1,1 — 0,n ou 0,1 — 1,n ou 0,1 — 0,n : Ces associations sont qualifiées de hiérarchiques partielles. Elles sont également dites "un à plusieurs".

1,n ou 0,n — 1,n ou 0,n : Ces associations sont dites non hiérarchiques, multiples ou "plusieurs à plusieurs".

2.4 Validation du M.C.D.

Le M.C.D. obtenu doit refléter fidèlement le monde réel qu'il est censé modéliser. On veillera en particulier à ce que les entités correspondent bien aux objets du monde réel que l'on a voulu représenter (par exemple, une entité *livre* n'a pas le même contenu sémantique pour un bibliothécaire et pour un libraire. Pour le premier, il s'agit d'un objet physique et deux livres identiques mais physiquement distincts constituent deux occurrences différentes. Pour le second, il s'agit d'une référence de catalogue, dont il possède un certain nombre d'exemplaires (quantité en stock).

De même, on veillera à ce que les associations représentent bien un événement unique du monde réel. Ainsi, dans un cas "Bibliothèque", si l'on souhaite mémoriser un historique des prêts, il faut prévoir le cas où un abonné emprunte à plusieurs reprises le même ouvrage. L'association *... emprunte...* doit donc faire intervenir trois entités: l'ouvrage, l'abonné et la *date d'emprunt*.

D'une manière générale, un bon modèle conceptuel des données doit impérativement vérifier les règles suivantes :

- toutes les données non calculées répertoriées dans le dictionnaire des données doivent y figurer ;
- chaque donnée doit apparaître *une seule fois* dans le M.C.D.

2.5 Diagramme entités-associations

Comme synthèse de l'étude précédente, on pourra réaliser un diagramme sur lequel apparaîtront les entités (dans des rectangles), leurs attributs et les associations (dans des ovales), avec leurs attributs et leurs cardinalités.

Dans le cas "Etudiants et Diplômes", on aboutira par exemple au schéma de la figure 3 page 13.

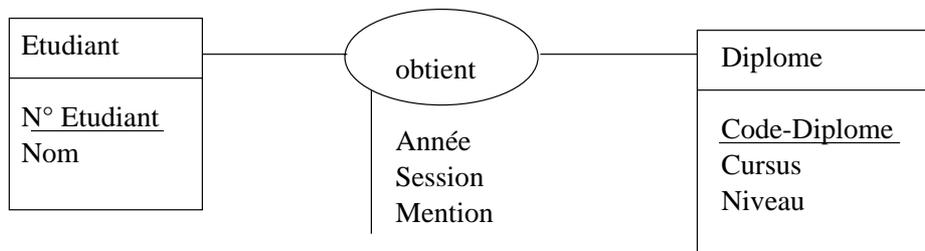


Figure 3: MCD du cas Etudiants-Etudes

3 Exemples d'études de cas

3.1 Exemple 1 : Musique classique

3.1.1 Enoncé du cas

Un de vos amis, professeur de musique indépendant, a essayé de constituer une base de données sous Microsoft Access afin de répertorier les œuvres musicales dont il dispose. Il vous consulte, pour que vous lui donniez des conseils sur la manière d'améliorer la structure de cette base de données.

Sa réalisation, largement inspirée de la métaphore de la "boîte de fiches", est constituée d'une seule table, avec les attributs indiqués dans le tableau 1 page 14.

N°	Attribut	Exemple
1	Numéro support	12
2	Titre	Concerto N° 5, Op. 73
3	Auteur	L. Van Beethoven
4	Type-support	CD-audio
5	Partition	oui
6	Date-enreg	1965
7	Direction	Hans Swarowski
8	Instrument 1	Piano
9	Soliste 1	Friedrich Gulda
10	Instrument 2	
11	Soliste 2	
12	Intrument 3	
13	Soliste 3	
14	Instrument 4	
15	Soliste 4	
16	Orchestre	De l'opéra de Vienne
17	Œuvre associée	Sonate en mi mineur, Op. 90, Concerto N° 1
18	Tonalité	mi bémol
19	Lieu	Brest
20	Extrait	Non

Tableau 1: Base de données "Musique". Première version

Vous lui posez alors quelques questions pour mieux comprendre ses besoins. Voici quelques extraits de ce dialogue :

- A quoi servent les attributs instrument 1, soliste 1, instrument 2, soliste 2 ... Y a-t-il exactement quatre solistes pour chaque œuvre?
- *Non, il peut n'y avoir aucun soliste (pour une symphonie, par exemple), ou plus de quatre. L'attribut instrument 1 représente l'instrument joué par le soliste 1 ou, dans le cas d'un opéra, le rôle qu'il tient. Mais la numérotation de 1 à 4 est tout à fait artificielle.*
- A quoi sert l'attribut Numéro. S'agit-il du numéro de la fiche concernée?
- *Non, il s'agit du numéro du support (CD audio, cassette, mini-disque ...) sur lequel l'œuvre est enregistrée.*

- Prenons une œuvre donnée – le concerto pour piano N° 5 de Beethoven, par exemple. Quelles informations souhaitez-vous rassembler concernant cette œuvre?
- *Evidemment, je dois disposer du titre et de l'auteur, pour pouvoir le retrouver rapidement. D'autre part, pour Beethoven, il existe une classification "normalisée": cette œuvre est référencée comme l'"opus 73" de l'auteur. Enfin, j'ai également besoin de la tonalité dans laquelle l'œuvre a été écrite. Comme vous le savez, ce concerto est écrit en mi bémol.*
- Vous indiquez également que ce concerto est joué par l'orchestre de l'opéra de Vienne, sous la direction de Hans Swarowsky, avec Friedrich Gulda au piano...
- *Oui, mais je possède en fait plusieurs enregistrements ou versions de cette œuvre, joués par des orchestres et des solistes différents. Comme vous le savez, deux interprétations d'une même œuvre peuvent être très différentes...*
- Quel est le rôle de l'attribut "œuvre associée"?
- *J'indique à cet endroit la liste des œuvres présentes sur le même support.*

Après quelques questions supplémentaires, vous commencez à réfléchir et à réaliser que le cas qui vous est soumis n'est pas aussi simple qu'il en a l'air. En examinant les données, vous vous apercevez notamment que :

- Faute d'une modélisation correcte, il existe de nombreuses redondances dans les données actuelles, avec d'ailleurs de nombreuses incohérences ;
- L'attribut "œuvre associée" n'est pas une donnée élémentaire ; en fait, c'est une donnée calculée qui n'a pas lieu d'être saisie et mémorisée de manière autonome ;
- Il serait sans doute souhaitable de "normaliser" la saisie des noms d'auteurs, de chefs d'orchestre et d'artistes (par exemple en séparant nom et prénom) de façon à préserver la cohérence.

De plus, au niveau sémantique (c'est-à-dire au niveau des entités représentées dans la base de données), on peut faire les remarques suivantes :

- Il est indispensable de distinguer clairement *l'œuvre* (par exemple, le Concerto pour piano N° 5 de Beethoven) et *une version* de cette œuvre (par exemple, l'enregistrement fait par l'orchestre de l'opéra de Vienne sous la direction de Hans Swarowski) ;
- La solution actuelle concernant les interprètes d'une œuvre n'est pas satisfaisante. La limitation à quatre solistes, de même que l'ordre dans lequel ils sont cités (soliste 1 à soliste 4) sont arbitraires. L'utilisation de l'attribut "Instrument" pour mémoriser le rôle joué dans un opéra n'est pas satisfaisant non plus.
- Il importe également de veiller à définir des identifiants clairs pour les entités.

3.1.2 Dictionnaire des données

Comme résultat de votre réflexion, vous obtenez le dictionnaire des données indiqué dans le tableau 2 page 16.

Num	Nom	Type	Description
1	Code-auteur	alpha	
2	Prénom-auteur	alpha	
3	Nom-auteur	alpha	
4	Code-œuvre		
5	Titre	alpha	Titre de l'œuvre
6	Tonalité	alpha	Liste de valeurs possibles
7	Partition	logique	Oui/non : possession de la partition correspondante
8	Code-version		
9	N° -support	num	Numéro du disque ou la cassette
10	Date-enreg	num	Année d'enregistrement de la version
11	Extrait	logique	Oui/non : s'agit-il d'un extrait?
12	Type-support	alpha	Valeurs possibles : CD-audio, cassette, mini-disque
13	Lieu	alpha	Lieu où est stocké le support
14	Code-artiste		Identifiant du soliste
15	Prénom-artiste	alpha	Prénom du soliste
16	Nom-artiste	alpha	Nom du soliste
17	Spécialité	alpha	Spécialité du soliste
18	Instrument	alpha	Instrument joué par un soliste dans un enregistrement donné
19	Rôle	alpha	Rôle joué par un soliste (opéras)
20	Code-chef		Identifiant pour un chef d'orchestre
21	Prénom-chef	alpha	Prénom du chef d'orchestre
22	Nom-chef	alpha	Nom du chef d'orchestre
23	Code-orchestre		
24	Dénomination	alpha	Dénomination de l'orchestre

Tableau 2: Dictionnaire des données du cas "Musique classique"

3.1.3 Entités

Les données figurant dans le dictionnaire se répartissent entre six entités, de la façon suivante:

Auteur: Code-auteur, Prénom-Auteur, Nom-Auteur

Œuvre: Code-œuvre, Titre, Tonalité, Partition

Version: Code-version, N° -support, Date-enreg, Extrait, Type-support, Lieu

Interprète: Code-artiste, Prénom-artiste, Nom-artiste, Spécialité

Chef-orchestre: Code-chef, Prénom-chef, Nom-chef

Orchestre: Code-orchestre, Dénomination

3.1.4 Associations

Certaines données figurant dans le dictionnaire des données n'ont pas été rattachées à une entité. Ce sont en fait des attributs des associations existant entre les entités. Plus généralement, l'examen du cas étudié fait apparaître les associations suivantes:

- Un auteur *écrit* une œuvre
- Une version *concerne* une œuvre
- Un interprète *participe* à une version donnée, en jouant d'un instrument et éventuellement un rôle donné
- Un chef d'orchestre dirige un orchestre donné, dans le cadre d'une version donnée d'une œuvre.

3.1.5 Le M.C.D. du cas “Musique classique”

L'ensemble des résultats de l'étude précédente constitue le **modèle conceptuel des données** (M.C.D.) du cas étudié. Il peut être résumé à l'aide d'une représentation graphique telle que la figure 4 page 17.

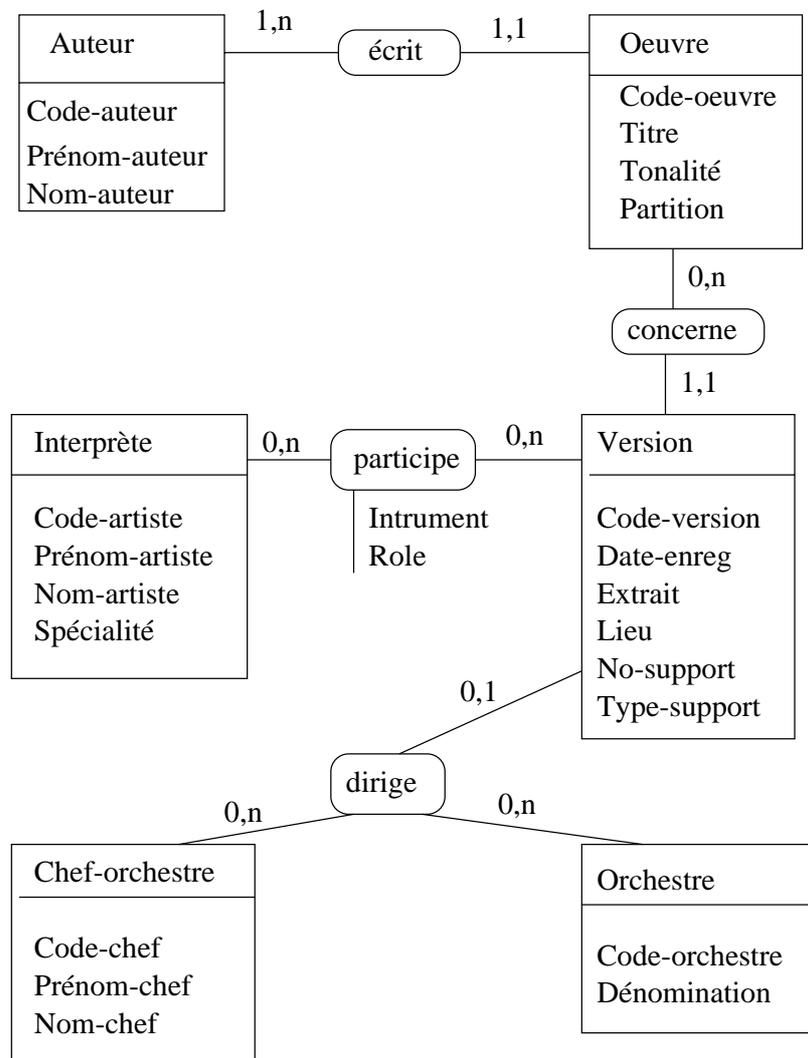


Figure 4: Le M.C.D. du cas “Musique Classique”

3.2 Exemple 2 : Un corpus de publicités

3.2.1 Enoncé

Considérons le cas décrit par l'énoncé suivant.

Source : librement adapté d'un cas décrit dans Documents pour le module RTD 1995-96, UFR de Psychologie, Université René Descartes faisant référence à M. Watiez (1992) - Approche psychosociologique du processus de socialisation alimentaire chez l'enfant français. Etude du rôle de la publicité télévisée dans la formation des représentations sur l'alimentation. Thèse de doctorat. Université Paris V.

Vous vous proposez d'analyser un corpus de 200 publicités visant la consommation de produits alimentaires plus spécialement destinés aux enfants.

Dans une première phase exploratoire, vous rassemblez de nombreuses données relatives à chacun des spots publicitaires constituant ce corpus: le produit concerné, sa marque, le fabricant et ses coordonnées, le prix moyen de vente du produit, les coordonnées du publicitaire ayant produit le spot, le réalisateur de la séquence, la durée du spot, sa date de réalisation, ...

Vous ne disposez en général que d'un spot concernant un produit donné. Mais il y a quelques exceptions. Par exemple, le corpus contient trois spots concernant les barres chocolatées "Ganymède".

A partir des premières données recueillies, vous avez défini onze catégories de produits: laitages nature, laitages fruités, fromages frais, fromages-pâtes, eaux minérales, boissons sucrées, produits pour petits dejeuners, jus de fruits, confiseries, barres chocolatées, chocolats. Chaque produit appartient évidemment à une seule catégorie.

En outre, vous avez visionné ces spots et constaté que ces publicités s'appuient sur la mise en scène de différentes valeurs: Gourmandise/Plaisir, Nature/Écologie, Forme/Santé, Aventure/Evasion, Tendresse/Affection, Prestige/Luxe, Séduction/-Érotisme, Convivialité/Partage, Tradition/Gastronomie, Innovation/Modernisme, Folie/Délire. Une publicité pour un produit peut s'appuyer sur une ou plusieurs valeurs.

Avant d'approfondir cette étude (par exemple par des moyens statistiques), vous souhaitez constituer une base de données rassemblant, de façon rationnelle, les informations recueillies. Vous souhaitez ainsi disposer d'un outil permettant d'obtenir toutes sortes de documents de travail, par exemple, la liste des spots réalisés par un publiciste donné, ou des tableaux croisant les types de produits et les valeurs mises en scène, ou encore les valeurs mises en scène et l'année de réalisation.

3.2.2 Dictionnaire des données

Dans le cas étudié, une lecture attentive du texte et un effort d'imagination concernant les documents sur lesquels les notes ont été prises pourraient nous conduire au dictionnaire des données donné dans le tableau 3 page 19.

3.2.3 Entités

Les données figurant dans le dictionnaire se répartissent entre six entités, de la façon suivante:

Produit: Nom-P, Prix-P

Num	Nom	Type	Description	Prop.
1	nom-P	alpha	nom ou marque du produit	
2	nom-F	alpha	raison sociale du fabricant	
3	Adr-F	alpha	coordonnées du fabricant	
4	Prix-P	num	prix moyen du produit	
5	Nom-Pub	alpha	raison sociale de l'agence de publicité	
6	Adr-Pub	alpha	coordonnées de l'agence	
7	Réalisateur-Spot	alpha	nom du réalisateur du spot	
8	Durée-spot	num	durée du spot en secondes	
9	Date-spot	num	année de réalisation du spot	
10	Catégorie	alpha	catégorie du produit	
11	Valeur	date	valeur mise en scène	

Tableau 3: Dictionnaire des données du cas “Spots publicitaires”

Fabricant: Nom-F, Adr-F

Publiciste: Nom-Pub, Adr-Pub

Spot: Date-Spot, Durée-Spot, Réalisateur-Spot

Valeur: Valeur

Catégorie: Catégorie

3.2.4 Indentifiants des entités

Les différentes occurrences des entités *Produit*, *Fabricant*, *Publiciste* sont clairement identifiées par les données *Nom-P*, *Nom-F*, *Nom-Pub*. En revanche, les données recensées jusqu'à présent ne permettent pas d'identifier clairement les spots publicitaires. Par exemple, il peut y avoir plusieurs spots de 120 secondes réalisés par Dupond en 1995. Dans l'univers du discours, diverses périphrases permettent de lever l'équivoque. En traitement automatique, il est indispensable de s'affranchir de cette ambiguïté en attribuant à l'entité *Spot* un attribut supplémentaire, *N°-Spot* qui prendra des valeurs systématiquement différentes pour chacune des occurrences de l'entité *Spot* et permettra de les identifier.

Par commodité, nous ajouterons également des identifiants *Code-P*, *Code-F*, *Code-Pub* aux trois entités *Produit*, *Fabricant*, *Publiciste*.

Ces attributs, de même que les attributs *Valeur* et *Catégorie* sont les *identifiants* des six entités que nous avons mises en évidence.

3.2.5 Associations

L'examen du cas étudié fait apparaître les associations suivantes:

- Un spot *met en scène* des valeurs
- Un spot *est produit par* un publiciste
- Un spot *concerne* un produit
- Un produit *est fabriqué par* un fabricant
- Un produit *appartient à* une catégorie.

3.2.6 Le M.C.D. du cas “Spots Publicitaires”

Le **modèle conceptuel des données** (M.C.D.) du cas étudié peut être résumé à l'aide de la figure 5 page 20.

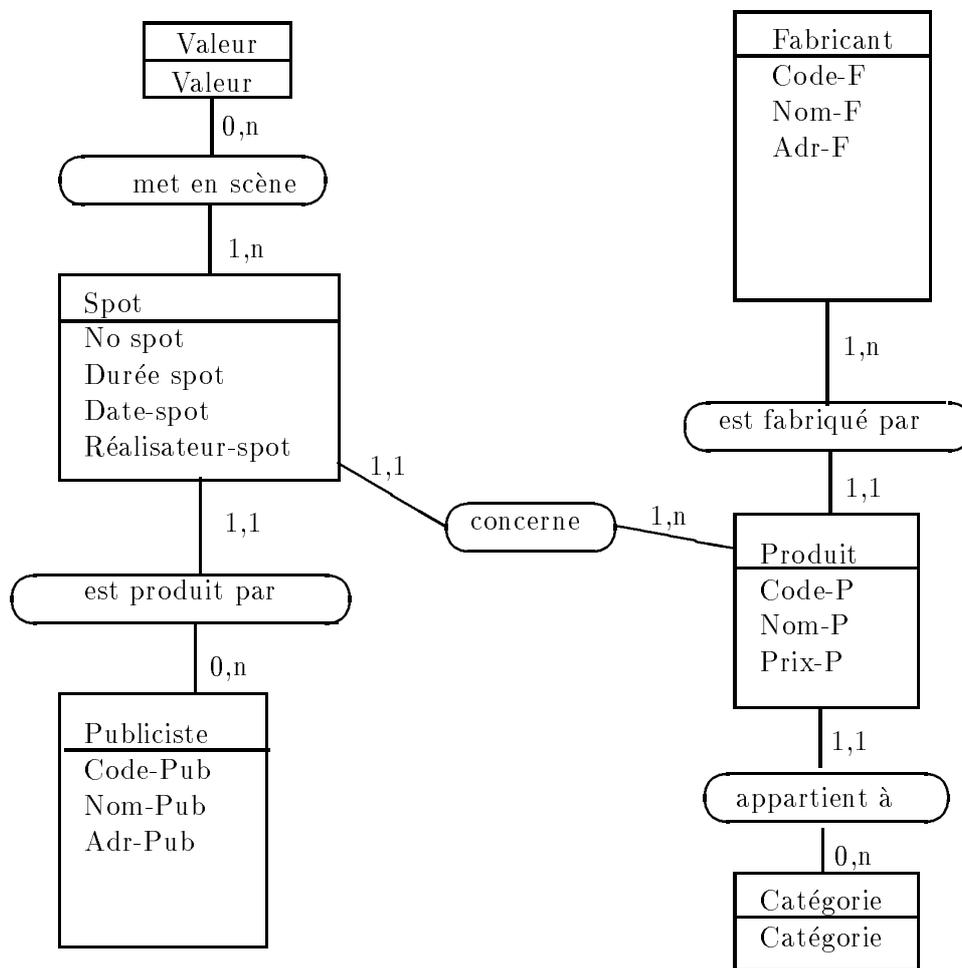


Figure 5: Le M.C.D. du cas “Spots Publicitaires”

4 Modèle logique des données

4.1 Position du problème

Dans les chapitres précédents, nous avons adopté deux points de vue :

- D'une part, nous avons développé certains aspects du modèle relationnel, avec ses tables et ses opérations (projection, sélection, jointure). Ce modèle est le cadre théorique sur lequel sont construits les logiciels de gestion de bases de données relationnels tels que Microsoft Access.
- D'autre part, nous avons développé le modèle entités-associations, qui est une méthode visant à modéliser la réalité en vue de constituer une base de données.

Ces deux modèles ne sont pas identiques. Le premier définit des opérations d'un point de vue algébrique abstrait, alors que le second propose une méthode respectant la sémantique des données. Mais, une association au sens d'un MCD n'est pas une jointure ; dans un MCD, une donnée n'est l'attribut que d'une seule entité (ou association) alors que la notion même de jointure entre tables suppose que la même donnée soit présente comme attribut dans les deux tables. La notion de cardinalité d'une association n'a pas d'équivalent dans le modèle relationnel.

Il importe donc, à ce stade de l'étude, de rapprocher les deux points de vue et d'énoncer des règles permettant, à partir d'un MCD, de déterminer les tables nécessaires pour construire un modèle relationnel correct.

Le **modèle logique des données (MLD)** relationnel est la définition (au niveau structure) des tables d'une base de données relationnelle représentant fidèlement le MCD.

En fait, le passage d'un MCD de type entités-associations à un modèle logique des données conforme au modèle relationnel ne pose pas de véritable problème, car il se fait de façon "presque" automatique. Il existe d'ailleurs des logiciels définissant mécaniquement la structure des tables à partir d'une description du MCD.

Rappelons d'une part le but poursuivi dans la mise en œuvre d'une base de données : *permettre l'accès aux données tout en minimisant la redondance dans le stockage de ces mêmes données* et les concepts de base du modèle relationnel : *des tables définies par leurs attributs, et des opérations fondamentales entre ces tables : sélection, projection et surtout jointure.*

Notation. Les structures des tables sont généralement notées de la manière suivante : on indique le nom de la table, suivi par l'indication entre parenthèses des attributs constituant la table ; on souligne l'attribut ou la combinaison d'attributs qui constitue la clé de la table. Par exemple :

Spot(No spot, Durée spot, Date-spot, Réalisateur-spot)

4.2 Traduction des entités

A chaque entité du MCD, on fait correspondre une table dans le MLD. Cette table comprend comme attributs au moins tous les attributs de l'entité. Nous verrons que d'autres attributs sont souvent nécessaires pour représenter les associations. Sa clé est l'attribut clé de l'entité.

Ainsi, dans le cas "Spots publicitaires" étudié, on définira les tables suivantes :

Spot(No spot, ...)
 Publiciste(Code-Pub, ...)
 Valeur(Valeur, ...)
 Fabricant(Code-F, ...)
 Produit(Code-P, ...)
 Catégorie(Catégorie, ...)

4.3 Traduction des associations

Comme nous allons le voir ci-dessous, on peut toujours représenter une association par une table. Mais il existe parfois des solutions plus économiques.

4.3.1 Traduction d'une association par une table

Une association du MCD peut être représentée par une table possédant les attributs suivants:

- les clés des entités participant à l'association;
- les attributs éventuels propres à l'association.

La clé de cette table est alors la juxtaposition des clés des entités participantes¹.

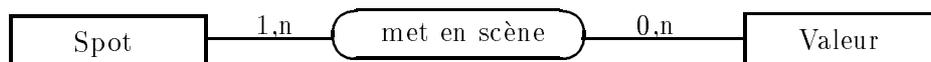


Figure 6: Association *Met en scène*

Par exemple, l'association représentée sur la figure 6 pourra être représentée par la table:

Met-en-scène(No spot, Valeur)

L'accès aux données se fait alors à l'aide de deux jointures, entre les tables *Spot* et *Met-en-scène* d'une part, entre les tables *Met-en-scène* et *Valeur* d'autre part.

Dans le cas d'une association "plusieurs à plusieurs" (telle que celle de l'exemple précédent) ou d'une association mettant en jeu plus de deux entités, on voit difficilement comment représenter la structure d'une manière plus simple. En revanche, d'autres solutions sont envisageables dans le cas d'associations hiérarchiques.

4.3.2 Traduction des associations hiérarchiques

Règle : Une association binaire de type $(0, n) - (1, 1)$ ou $(1, n) - (1, 1)$ peut être représentée en ajoutant à la table correspondant au côté $(1, 1)$ de l'association, la clé de l'autre table. Cet attribut supplémentaire est appelé **clé étrangère**.



Figure 7: Association *concerne*

Ainsi, dans le cas étudié, un spot concerne un produit et un seul (cf. figure 7). L'association sera représentée à l'aide d'un attribut supplémentaire dans la table des spots: *Code-P*, c'est-à-dire le code du produit que le spot met en scène. Pour accéder aux informations relatives au produit faisant l'objet d'un spot donné, il faudra faire une jointure entre les tables *Spot* et *Produit*, sur le champ commun *Code-P*.

Dans le cas des associations binaires de type $(0, n) - (0, 1)$ ou $(1, n) - (0, 1)$, les deux solutions peuvent être envisagées. Elles pourront être traitées de la même façon qu'une association de type $(0, n) - (1, 1)$, la clé étrangère restant vide lorsque le n-uplet considéré n'est lié à aucun n-uplet de

¹Rappel: dans un M.C.D. correct, une occurrence d'une association est parfaitement déterminée par les occurrences des entités qui y participent.

l'autre table. Elles pourront aussi être traitées comme des associations “plusieurs à plusieurs”. Le critère permettant de choisir entre ces deux solutions est la proportion d'occurrences de l'entité du côté (0,1) de la relation qui sont effectivement associées à une occurrence de l'autre entité.

Exemple. Une bibliothèque rassemble 100000 ouvrages dont en moyenne 2000 sont empruntés à un moment donné. Est-il raisonnable de représenter l'association ... *emprunte*... à ajoutant une clé étrangère à la table *Ouvrage*?

4.3.3 Les tables du cas étudié

La mise en œuvre des règles énoncées ci-dessus conduit, pour le cas étudié, aux tables suivantes:

- Spot(No spot, Durée spot, Date-spot, Réalisateur-spot, *Code-P*, *Code-Pub*)
- Publiciste(Code-Pub, Nom-Pub, Adr-Pub)
- Fabricant(Code-F, Nom-F, Adr-F)
- Produit(Code-P, Nom-P, Prix-P, *Code-F*, *Catégorie*)
- Catégorie(Catégorie)
- Valeur(Valeur)
- Met-en-scène(No spot, Valeur)

Remarque. A ce stade de l'étude, on peut se poser la question de l'utilité des tables *Valeur* et *Catégorie*. Une justification simple peut être la suivante: il s'agit d'avoir des données de référence dans lesquelles les valeurs saisies dans les tables *Produit* et *Met-en-scène* sont obligatoirement prises.

4.4 Exercices

4.4.1 Exercice 1

Trouver la liste des jointures à faire pour obtenir la liste des publicistes qui ont travaillé avec un fabricant donné.

4.4.2 Exercice 2

On considère le “MCD” représenté sur la figure 8.

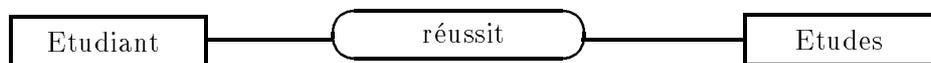


Figure 8: MCD Etudiant-Etudes

Données: nom, prénom, n° étudiant, niveau, cursus, mention, code, année, session.

Exemples de phrases de l'“univers du discours”: Gérard Gentil, n° étudiant 88272X, a obtenu la licence d'histoire (code HIST3) à la session de septembre 1992, avec la mention AB. Il avait obtenu le DEUG d'histoire (HIST2) en juin 1990, avec la mention Passable.

Classer les données ci-dessus comme attributs des entités ou de l'association. Précisez l'identifiant choisi pour chaque entité.

Traduire le modèle entités-associations obtenu dans le modèle relationnel.

4.4.3 Exercice 3

Une cinémathèque souhaite mettre à disposition du public une base de données relative à un certain nombre de films. Pour un film donné, on veut pouvoir connaître son réalisateur et son ou ses acteurs principaux. On veut aussi pouvoir trouver les films réalisés par un réalisateur donné, les films dans lesquels a joué tel acteur, etc. On constate que:

- chaque film possède un unique réalisateur;
- chaque réalisateur a fait plusieurs films (en moyenne, 20 films);
- chaque film a au moins un acteur principal; certains films en ont plusieurs;
- chaque acteur a joué dans un ou plusieurs films (en moyenne, 25 films).

On veut pouvoir consulter les données suivantes: le nom et une photo de chaque réalisateur, le nom (nom légal ou "vrai nom"), le pseudonyme (ou "nom de scène") éventuel et une photo de chaque acteur, le titre et l'année de sortie de chaque film.

1) a) Quelles sont les entités qui apparaissent dans ce système d'information? Quels sont leurs attributs? Définir une clé pour chacune des entités (en créant au besoin des attributs numéro ou code).

b) Préciser les associations entre ces entités et leurs cardinalités.

2) On réalise la base de données envisagée à l'aide d'un SGBD relationnel. Préciser les tables à définir, leurs attributs et la clé primaire de chacune d'elles.

3) La base de données concerne 500 films, 20 réalisateurs et 40 acteurs principaux. Par ailleurs, on estime que chaque photo nécessite environ 30Ko d'espace de stockage.

a) Evaluer l'espace mémoire nécessaire au stockage de la table des acteurs.

b) Pour 10% des films, on dispose d'une bande annonce stockée sous forme d'un fichier contenant une image animée. Une bande annonce nécessite en moyenne 10 Mo d'espace de stockage. On souhaite attacher ces bandes annonces à la table des films. Quel est alors l'espace mémoire nécessaire au stockage de cette table (y compris les bandes annonces)?

Rappels: 1 Ko \simeq 1000 octets; 1 Mo \simeq 1000000 octets; 1 caractère \sim 1 octet.

Table des matières

1	Généralités sur les bases de données	1
1.1	Introduction	1
1.1.1	Fichiers manuels et bases de données	1
1.1.2	Exemples	2
1.1.3	Remarques	2
1.2	Définitions	2
1.2.1	Base de données	2
1.2.2	Système de gestion de bases de données	3
1.3	Mise en œuvre d'un SGBD	3
1.4	Concepts de base du modèle relationnel	4
1.4.1	Les tables	4
1.4.2	Les opérations de base sur les tables	4
1.4.3	Pourquoi développer des modèles conceptuels des données?	5
2	Modèle conceptuel des données	6
2.1	Constitution du dictionnaire des données	6
2.1.1	Etude des rubriques de documents:passer des rubriques aux données	6
2.1.2	Les types de données couramment acceptés par un SGBD	6
2.1.3	Constitution du dictionnaire des données	8
2.2	Passer des données aux entités	9
2.2.1	Mettre en évidence les entités	9
2.2.2	Liaisons sémantiques entre les données:notion de dépendance fonctionnelle	9
2.2.3	Identifiant d'une entité	10
2.3	Associations entre entités	11
2.3.1	Repérer les associations entre entités	11
2.3.2	Attributs d'une association	11
2.3.3	Remarque :entité ou association?	11
2.3.4	Cardinalités des associations	12
2.4	Validation du M.C.D.	13
2.5	Diagramme entités-associations	13
3	Exemples d'études de cas	14
3.1	Exemple 1 :Musique classique	14
3.1.1	Enoncé du cas	14
3.1.2	Dictionnaire des données	15
3.1.3	Entités	16
3.1.4	Associations	16
3.1.5	Le M.C.D. du cas "Musique classique"	17
3.2	Exemple 2 :Un corpus de publicités	18
3.2.1	Enoncé	18
3.2.2	Dictionnaire des données	18
3.2.3	Entités	18
3.2.4	Identifiants des entités	19
3.2.5	Associations	19
3.2.6	Le M.C.D. du cas "Spots Publicitaires"	20

4	Modèle logique des données	21
4.1	Position du problème	21
4.2	Traduction des entités	21
4.3	Traduction des associations	22
4.3.1	Traduction d'une association par une table	22
4.3.2	Traduction des associations hiérarchiques	22
4.3.3	Les tables du cas étudié	23
4.4	Exercices	23
4.4.1	Exercice 1	23
4.4.2	Exercice 2	23
4.4.3	Exercice 3	24