

## Cours

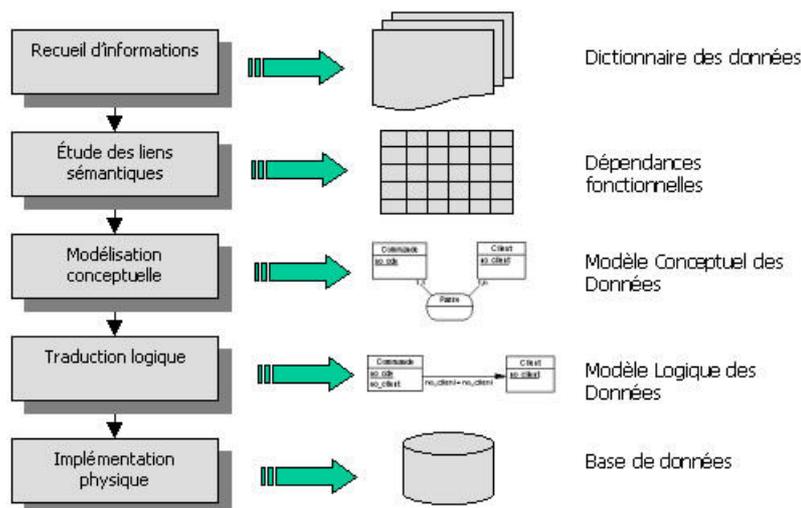
### Modélisation conceptuelle des données

#### T A B L E D E S M A T I E R E S

1	La problématique de la modélisation des données .....	1
2	Le recueil des données .....	2
2.1	Contexte .....	2
2.2	Techniques et outils .....	2
2.3	Classification des données.....	3
2.4	Typologie des données .....	3
2.5	Dictionnaire des données .....	3
3	La recherche des dépendances fonctionnelles .....	4
3.1	Notion de dépendance fonctionnelle (df) .....	4
3.2	Typologie .....	4
4	Le modèle conceptuel des données .....	5
4.1	Contexte.....	5
4.2	Concepts et formalisme .....	6
4.3	Démarche de construction du MCD.....	9

### 1 La problématique de la modélisation des données

Le but de l'analyse des données est d'obtenir un schéma de l'organisation des données **stable** et **invariant** permettant de construire une **solution physique**, concrètement, une base de données. Pour parvenir à ce résultat, on procède par étapes :



Chaque étape donne lieu à la production de documents qui aident à formaliser les spécifications du système.

- Le recueil

Cette phase est très importante mais longue et difficile. Elle consiste à réaliser un inventaire exhaustif des éléments du système d'information : règles de gestion, flux d'informations, données.

Pour l'étude des données, le recueil est consigné dans un dictionnaire des données qui sera la référence tout au long de cette étude.

- L'étude des liens sémantiques (recherche des dépendances fonctionnelles)

Cette étape n'est pas un élément formalisé de la méthode. Il s'agit d'une démarche didactique permettant de faciliter la représentation conceptuelle des données.

- La modélisation conceptuelle

Quand les dépendances fonctionnelles sont établies, on peut en déduire le schéma global des données du système d'information. Le modèle obtenu doit être minutieusement vérifié et validé car il servira de base à la phase descendante de l'étude.

- La traduction logique

Cette phase de l'étude consiste en une traduction quasi automatique du modèle conceptuel. En appliquant des règles de transformation formalisées, on obtient un nouveau modèle, appelé modèle logique des données qui pourra être traduit et implanté dans une base de données.

Toutefois, avant de procéder à la traduction de ce modèle, on cherchera à l'optimiser afin que la future base de données soit la plus efficace possible.

- L'implémentation physique

Cette étape est la phase ultime de l'étude des données. Elle consiste à passer d'un modèle exprimé dans un langage intermédiaire à une base de données opérationnelle. Il s'agit donc d'adapter le modèle aux contraintes logicielles du système de gestion de la base de données (SGBD).

Cette étape peut être automatisée à l'aide d'outils spécialisés.

## 2 Le recueil des données

### 2.1 Contexte

Le but du recueil des informations est de constituer un référentiel sur lequel s'appuiera l'étude. Ce référentiel est consigné dans un document appelé dictionnaire des données. Le dictionnaire peut être géré à l'aide d'outils spécialisés.

Cette phase de recueil est effectuée en plusieurs étapes :

### 2.2 Techniques et outils

#### 2.2.1 La localisation des informations

Le premier problème à résoudre est de trouver l'information. On recherchera dans les documents, les règlements, les lois, les normes, les procédures, les bases de données, les fichiers.

#### 2.2.2 Les techniques de recueil

Pour recenser les informations, on utilise essentiellement l'étude de documents et les entretiens.

##### Les documents

L'essentiel des données peut être retrouvé sur les documents en circulation. On s'efforcera :

- De rassembler un maximum de documents,
- De s'assurer que les documents sont encore valides,
- D'utiliser des documents remplis (pas seulement des modèles) et repérer les informations réellement utilisées,
- D'attacher une attention particulière aux documents non formalisés (ils traduisent souvent l'existence d'une évolution non prise en compte des procédures ou d'une inadéquation de la procédure en cours),
- De décoder les codes et abréviations utilisés.

##### Les entretiens

- Avec la direction, il s'agit d'obtenir des informations générales : organisation des services, objectifs stratégiques, procédures et règles de gestion globales...
- Avec les utilisateurs, on s'efforcera :
  - D'obtenir une description précise et détaillée des procédures.
  - De mettre en évidence des dysfonctionnements.
  - D'obtenir des propositions, des critiques, des suggestions.
  - De rechercher des procédés non officiels.



## 2.3 Classification des données

On peut distinguer :

- Les données élémentaires : l'information se confond avec la valeur prise par la donnée. Par exemple, un nom, une date... Ces données doivent être recensées de manière exhaustive.
- Les données calculées ou déduites : elles sont obtenues par l'application d'un traitement mathématique ou logique. Ces données sont associées à des règles de calcul (règles de gestion). Il faut penser à bien identifier et conserver la règle de gestion qui permet d'arriver au résultat.
- Les données composées : certaines données sont regroupées en une même entité sémantique (par exemple une adresse). Ces informations doivent être décomposées en données élémentaires. Toutefois, s'il est montré qu'une donnée composée n'est jamais décomposée dans la chaîne de traitement de l'information, on peut envisager de la conserver telle quelle.

## 2.4 Typologie des données

### 2.4.1 Notion de domaine

Certaines données ont des valeurs prises dans le même ensemble. Par exemple, langue parlée, langue lue et langue maternelle prennent leur valeur dans un ensemble sémantique « Langues ».

On appelle domaine l'ensemble des valeurs prises par une donnée, indépendamment du contexte de son utilisation.

*Exemple :*

Donnée	Domaine	Exemple
Couleur des yeux	Couleur	Vert
Langue parlée	Langue	Français
Nom	Nom	Dupont
Prénom	Nom	Fernand
Année de naissance	Année	1973
Langue étudiée	Langue	Anglais
Bac	Bac	D

L'identification d'un domaine est une opération importante lors du recueil des données. Il s'agit de déterminer précisément l'ensemble des valeurs possibles s'il s'agit d'un domaine exhaustif ou les règles de représentation (codification, types, bornes...) dans les autres cas.

*Exemples :*

- domaine exhaustif : permis de conduire (A, A1, B, C, D, C1, E, F)
- domaine borné : notes (min 0, max 20)
- domaine typé : noms (30 caractères alphabétiques)

### 2.4.2 Types de données

Les types de données ont un sens plus restrictif que le domaine. Alors que le domaine s'applique à la sémantique d'une donnée (plus proche de la notion d'information), le type est une contrainte physique liée à la manière dont sera stockée la donnée dans le système d'information.

Les principaux types à retenir sont :

- Alphabétique (A)
- Alphanumérique (AN)
- Numérique (on peut préciser s'il s'agit d'un entier ou d'un réel)
- Date (D)
- Logique ou booléen (L ou B)

## 2.5 Dictionnaire des données

Pendant la phase de conception, les données recueillies et spécifiées sont inscrites dans un dictionnaire. Ce dictionnaire est un outil important car il constitue la référence de toutes les études effectuées en aval.

### 2.5.1 Formalisme

Les données sont présentées dans un tableau.

Nom symbolique	Description (rôle)	Domaine ou type	Commentaires	Contraintes, règles de calcul
no_client	N° de client	N (entier)	N° séquentiel	Automatique
nom_client	Nom du client	AN 50	Nom ou raison sociale, format : tout en majuscules	Obligatoire
prix_commande	Prix unitaire HT du produit commandé	N (monétaire)	Format : 9999,99 EUR	Obligatoire
qte_commande	Quantité commandée	N (entier)		Obligatoire, > 0
montant_commande	Montant HT de la commande	N (monétaire)	Format : 9999,99 EUR	Somme (prix_commande * qte_commande)

### 2.5.2 Exploitation

L'ensemble des données recueillies constitue le dictionnaire des données brut. Toutes les données ne seront pas utilisées de la même manière. Pour la phase de modélisation des données, il convient d'épurer ce dictionnaire brut :

- des synonymes (données ayant le même sens) car ils constituent des redondances ambiguës,
- des polysèmes (mots ayant plusieurs sens) car ils peuvent provoquer des malentendus.

Les données calculées doivent être examinées avec soin. La règle est la suivante :

Si une donnée calculée peut être obtenue par l'application d'un traitement à partir de données élémentaires valides, on peut la supprimer du dictionnaire.

## 3 La recherche des dépendances fonctionnelles

### 3.1 Notion de dépendance fonctionnelle (df)

Soient A et B les ensembles de valeurs prises par deux données. Il y a dépendance fonctionnelle entre A et B lorsque, connaissant une valeur de A, quelque soit cette valeur, on détermine une et une seule valeur de B.

On symbolise la dépendance fonctionnelle par  $A \rightarrow B$  où A est appelé source de la df (on dit aussi déterminant ou partie gauche) et B la cible (on dit aussi but, déterminé ou partie droite) de la df.

Par exemple :  $no\_client \rightarrow nom\_client$

Cela signifie qu'à un numéro de client ne correspond qu'un et un seul nom. En revanche, il n'est pas impossible qu'à un nom, correspondent plusieurs clients.

### 3.2 Typologie

#### 3.2.1 Dépendance fonctionnelle forte et faible

La définition de la dépendance fonctionnelle peut être affinée :

##### Définition stricte

- la df associe à chaque valeur de A une et une seule de B : il y a **unicité** au départ
- la df est vérifiée pour toutes les valeurs de A : il y a **totalité** au départ (toutes les valeurs de A ont une image dans l'ensemble d'arrivée B).

Dans ce cas, on parle de **df forte**.

Par exemple, la dépendance fonctionnelle  $no\_commande \rightarrow no\_client$  est une df forte car il n'y a pas de commande sans client.

##### Définition large

Il y a dépendance fonctionnelle entre A et B lorsque, connaissant une valeur de A, quelque soit cette valeur, on détermine au plus une valeur de B.

Cette définition supprime la contrainte de totalité au départ. On parle de **df faible**.

Par exemple, la dépendance fonctionnelle  $no\_insee \rightarrow nom\_jeune\_fille$  est une df faible car certaines valeurs de NoInsee n'ont pas de correspondance dans l'ensemble d'arrivée.

### 3.2.2 Dépendance fonctionnelle à partie gauche composée

Une dépendance fonctionnelle peut comporter dans sa partie gauche plusieurs attributs. On parle dans ce cas de dépendance fonctionnelle à partie gauche composée. Pour connaître une valeur de l'ensemble d'arrivée C, il faut connaître un couple (ou plus) de valeurs provenant de A et de B.

Ce type de df est noté :  $(d_1, d_2) \rightarrow d_3$

$(no\_facture, code\_produit) \rightarrow qte\_facturee$

$(no\_eleve, matiere, date) \rightarrow note$

### 3.2.3 Dépendance fonctionnelle élémentaire

Une dépendance fonctionnelle est élémentaire s'il n'existe aucune donnée ou sous-ensemble de données de la partie gauche assurant une dépendance fonctionnelle vers le même but. Par définition les dépendances fonctionnelles à deux rubriques sont toujours élémentaires.

$code\_produit \rightarrow nom\_produit$  est élémentaire (deux rubriques).

$(no\_facture, code\_produit) \rightarrow qte\_facturee$  est élémentaire (ni le code produit seul, ni le numéro de facture seul permettent de déterminer la quantité).

$(no\_facture, code\_produit) \rightarrow nom\_article$  n'est pas élémentaire puisque le code du produit seul suffit à déterminer le nom du produit.

### 3.2.4 Dépendance fonctionnelle directe

Une dépendance fonctionnelle  $d_1 \rightarrow d_2$  est directe s'il n'existe aucune donnée  $d_3$  qui engendrerait une dépendance fonctionnelle transitive telle que  $d_1 \rightarrow d_2 \rightarrow d_3$

Soient les dépendances fonctionnelles :

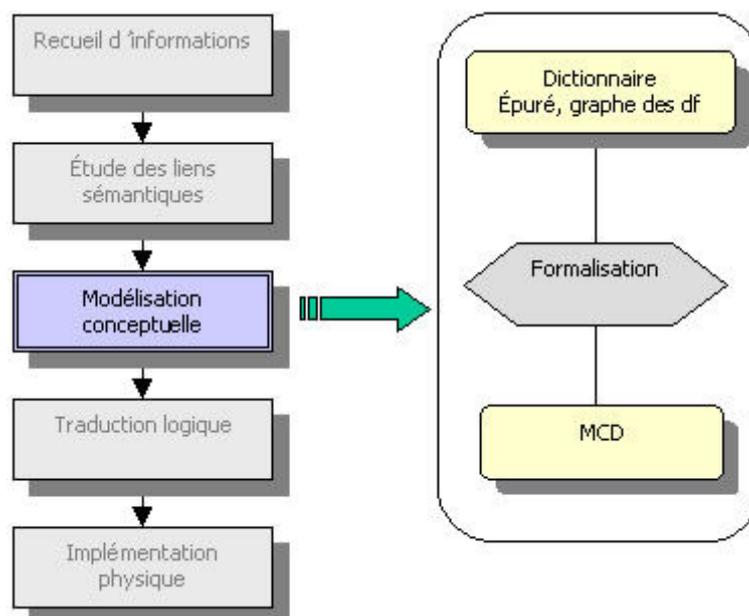
$no\_facture \rightarrow no\_representant$  ;  $no\_representant \rightarrow nom\_representant$

$no\_facture \rightarrow nom\_representant$  n'est pas une dépendance fonctionnelle directe puisque obtenue par transitivité.

## 4 Le modèle conceptuel des données

### 4.1 Contexte

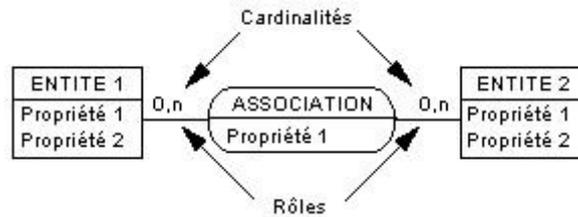
Le rôle du MCD est fondamental dans une étude d'informatisation. Il conditionne fortement la suite de l'étude. Il permet de « dessiner » la structure des données du système d'information à implanter.





## 4.2 Concepts et formalisme

Le formalisme retenu est le modèle « Entité-Association » (MEA) connu également sous le nom de « Individu - Relation ».

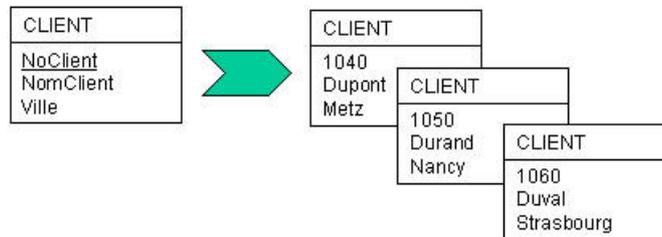


### 4.2.1 Les entités

Elles correspondent aux objets du système d'information. Par exemple, l'entité CLIENT rassemble toutes les informations communes aux clients de l'entreprise.

Les entités sont nommées.

Notion d'occurrence d'entité : soit une entité CLIENT ; dans le système d'information, cette entité sera valorisée un grand nombre de fois. Chaque client constitue alors une **occurrence** de l'entité. Une occurrence représente un « exemplaire » de l'objet.



### Propriétés

Une entité possède toujours au moins une propriété. Chacune de ces propriétés doit pouvoir être valorisée de manière unique.

Les propriétés des entités sont prélevées dans le référentiel que constitue le dictionnaire des données.

### Identifiant

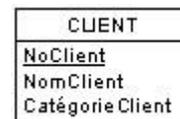
Parmi les propriétés d'une entité, il y en a une qui joue un rôle particulier :

L'identifiant d'une entité est une propriété telle qu'à chaque valeur de la propriété corresponde une et une seule occurrence de l'entité.

Par conséquent :

- on ne peut trouver deux occurrences d'une entité ayant le même identifiant
- l'identifiant se confond avec l'entité : l'entité existe si on peut valoriser l'identifiant
- si l'entité ne possède qu'une propriété, cette propriété est l'identifiant
- toutes les propriétés sont en dépendance fonctionnelle élémentaire et directe avec l'identifiant

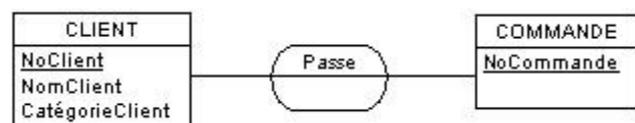
L'identifiant est représenté souligné dans le modèle.



### 4.2.2 Les associations

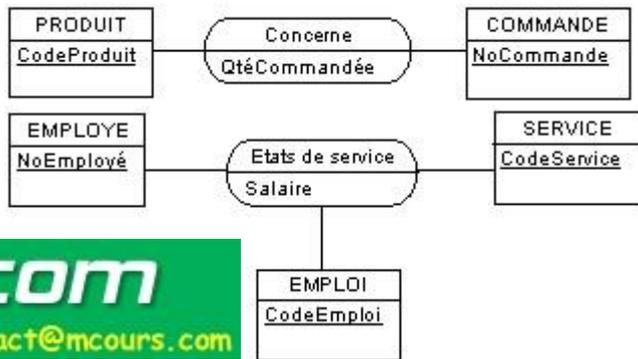
Une association est un lien sémantique entre des entités. Contrairement aux entités, les associations n'ont pas d'existence propre mais elles peuvent porter des propriétés. Les associations sont nommées, généralement à l'aide d'un verbe d'action.

Association binaire :



Association porteuse de propriété :

Association ternaire :

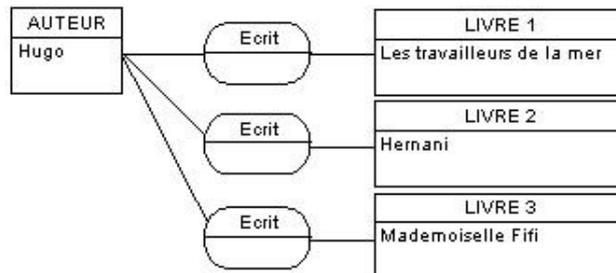


Le nombre d'entités impliquées dans l'association détermine sa **dimension** : pour les associations faisant intervenir deux entités, on parle d'association de dimension deux ou associations binaires. Pour les associations faisant intervenir plus de deux entités, on parle d'associations n-aires.

- Les associations n-aires sont généralement difficiles à qualifier par un verbe d'action, on utilise alors un substantif représentatif de l'objet de gestion représenté.
- Les associations n-aires de dimension supérieure à quatre sont suspectes et assez rares. Il faut vérifier qu'il ne s'agit pas d'entités « cachées ». Il s'agit souvent d'objets abstraits faisant intervenir la notion de temps (historiques en particulier).

### Occurrences d'une association

Comme pour les entités, on ne représente sur le MCD que l'association de manière générique. Le nombre d'occurrences de l'association n'est pas directement représenté mais sera borné à l'aide des cardinalités.



### Identification d'une association

Une association est identifiée par la « concaténation » des identifiants des entités qu'elle relie. Cette identification est implicite et n'est pas représentée.

### Rôles d'une association

Chaque « patte » d'une association vers une entité représente le rôle joué par l'association. Ces rôles peuvent être nommés, cela peut aider à clarifier le modèle dans certaines situations complexes.

### Cardinalités

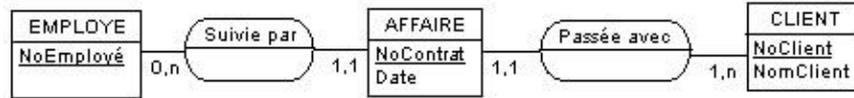
- Définition

La cardinalité représente le nombre d'occurrences minimum et maximum d'une association par rapport à une entité.

- La cardinalité minimale représente le nombre de fois « au minimum » où une occurrence de l'association participe aux occurrences de l'entité. Cette cardinalité est choisie parmi 0 ou 1.
- La cardinalité maximale représente le nombre de fois « au maximum » où une occurrence de l'association participe aux occurrences de l'entité. Cette cardinalité est choisie parmi 1 ou n où *n* indique une cardinalité maximale supérieure à 1 mais non quantifiée. Si la valeur de *n* est connue, on peut la mentionner.

- Formalisme

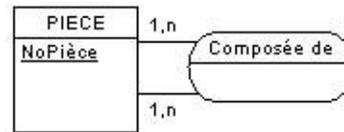
Les cardinalités sont mentionnées par couple du côté de l'entité à considérer. La cardinalité minimale est représentée en premier, la maximale en second.



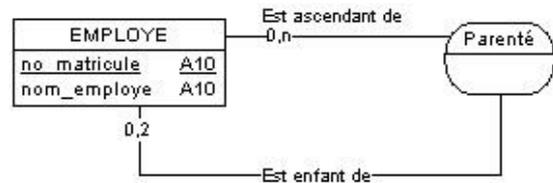
- un employé peut (0) suivre plusieurs (n) affaires
- une affaire est suivie par un (et un seul) employé
- une affaire concerne un (et un seul) client
- un client passe une (sinon ce n'est pas un client) ou plusieurs (n) affaires

**Associations réflexives**

Ce sont des associations qui relient deux fois la même entité.



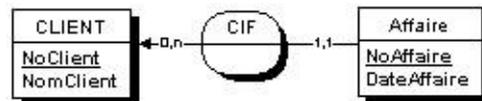
Dans les associations réflexives, la notion de rôle prend tout son sens.



**4.2.3 Association binaire fonctionnelle et contrainte d'intégrité fonctionnelle**

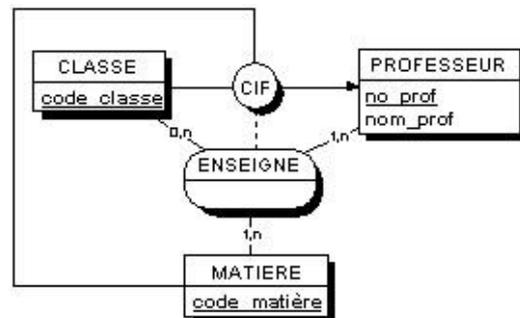
Quand on détermine une association présentant une cardinalité maximale à 1, on est en présence d'une association binaire fonctionnelle appelée parfois « contrainte d'intégrité fonctionnelle » (CIF). Une CIF traduit la présence d'une dépendance fonctionnelle entre les identifiants des entités participant à l'association.

Certains auteurs ou outils de génie logiciel représentent les CIF avec un formalisme différent mais cette notation est désormais déconseillée :



L'association nommée est remplacée par une « pastille » CIF et la « patte » portant la cardinalité maximale à n est remplacée par une flèche. Cela signifie qu'il existe une dépendance fonctionnelle entre NoAffaire et NoClient.

Une CIF peut mettre en relation plusieurs entités. Elle traduit alors la présence d'une dépendance fonctionnelle à partie gauche composée. Cette notation est désormais remplacée par la contrainte d'unicité (Merise/2) dans laquelle le terme « CIF » est remplacé par un « 1 ».



La CIF ou contrainte d'unicité indique que la connaissance de l'occurrence d'une entité est déterminée par la connaissance d'un couple d'occurrences des entités participant à l'association : la connaissance d'un Professeur est déterminée par la connaissance du couple Classe, Matière.

### 4.3 Démarche de construction du MCD

Le MCD est élaboré à l'aide du dictionnaire des données épuré et éventuellement du graphe des dépendances fonctionnelles.

#### 4.3.1 Repérage des entités

Les entités correspondent aux objets du SI. On peut les repérer intuitivement ou s'aider du graphe des dépendances fonctionnelles.

Repérer une entité revient à repérer une propriété identifiante (une source de dépendances fonctionnelles).

- L'entité doit être nommée. On choisira des identificateurs concis et parlants, en rapport avec le domaine de gestion étudié.
- La propriété identifiante devient l'identifiant souligné de l'entité.
- Les données but de dépendances fonctionnelles deviennent des propriétés de l'entité.

#### 4.3.2 Repérage des associations binaires fonctionnelles (CIF)

- **CIF binaires** : ces associations matérialisent la présence d'une dépendance fonctionnelle entre identifiants.
- **CIF n-aires** : ces associations représentent des dépendances fonctionnelles à partie gauche composée n'impliquant que des identifiants.

#### 4.3.3 Repérage des associations porteuses de propriétés

Elles correspondent aux dépendances fonctionnelles à partie gauche composée dont le but n'est pas une source de dépendance fonctionnelle. Cette donnée devient une propriété de l'association porteuse.

#### 4.3.4 Repérage des associations « creuses »

Ces associations non porteuses de propriétés binaires ou n-aires sont plus difficiles à repérer car elles ne font pas intervenir de notion de dépendance fonctionnelle. Elles sont l'expression de règles de gestion faisant intervenir des objets du SI.

#### 4.3.5 Placer les cardinalités

Le placement des cardinalités permet de vérifier la validité des dépendances fonctionnelles.

#### 4.3.6 Vérifier le modèle

En principe, si les dépendances fonctionnelles ont été correctement spécifiées et si les règles de construction ont été respectées, le MCD est correct.

Si l'étude des dépendances fonctionnelles n'a pas été utilisée, il faut vérifier le modèle :

- Toutes les propriétés doivent être élémentaires (non décomposables)
- Chaque entité possède un et un seul identifiant, rôle joué par une seule propriété.
- Les propriétés d'une entité doivent être en dépendance fonctionnelle avec l'identifiant.
- Les propriétés d'une association doivent être en dépendance fonctionnelle élémentaire avec l'identifiant implicite de l'association.
- Une propriété ne peut figurer qu'une fois dans le modèle (pas de redondance).
- Une propriété d'une entité ne peut être en dépendance fonctionnelle avec une autre propriété non identifiante de cette entité (pas de transitivité).