

Dictionnaire des données et graphe des dépendances fonctionnelles

1	Le relevé d'informations	1
2	Le dictionnaire des données (DD).....	2
3	Le graphe de dépendances fonctionnelles (GDF).....	5

1 Le relevé d'informations

1.1 En théorie c'est ça ...

Le relevé d'information est la chose la plus difficile de l'analyse merise ou UML. Cependant, les sujets sont souvent bien faits, à défaut d'être complètement réels, et le relevé d'information est facilité par le discours et la progression des exercices.

Concepts, données et valeurs

Avant d'aborder le MCD, on va pouvoir relever trois types d'informations qui sont des concepts, des données ou des valeurs.

Les concepts sont des choses, des objets que l'on va gérer. Ils deviendront probablement des entités du schéma entité association vu plus loin.

Les données sont des informations élémentaires sur les concepts.

Données et concepts sont souvent masqués par des valeurs de ceux-ci ou exprimés par des euphémismes ou des synonymes.

Donnée

Une donnée est une information élémentaire que l'on ne peut pas décomposer.

Exemple oui : description, prix, quantité, numéro d'identification, nom, prénom, ville, code postal, etc. ...

Exemple non : fournisseur, article, ingénieur

Dans le discours du sujet, une donnée est souvent rattachée à un concept.

Exemple oui : description de l'article, nom de l'employé, numéro du client, numéro de commande, etc. ...

Exemple non : description (tout court, sans dire de quoi)

Concept

Un concept est une information complexe qui peut être décomposée en données.

Exemple oui : article, commande, client, employé, etc. ...

Exemple non : description, organisation gérée, tapis Z7 cendré.

Pour détecter un concept, on va relever les noms de choses, d'objets, de personnes ou de types. Ce seront souvent des concepts.

Ce seront sûrement des concepts si des données y sont rattachées.

Valeur

Certains concepts ou données sont masqués par des valeurs exposant différentes occurrences de ceux-ci.

Exemple oui : Strasbourg, Nantes, Rennes, Lyon et Mulhouse sont des occurrences du concept 'ville'.

Ces mêmes valeurs sont, de plus, des occurrences de la donnée 'nom ville'.

Exemple non : type article, ville.

Dans tous les cas, il ne faut relever que les données et concepts pertinents pour l'étude, gérés par le système décrit, et ne pas se laisser perturber par des informations qui constituent le bruit de l'étude.

A noter que les informations sur l'entreprise elle-même ne sont pas à relever, seules les données avec lesquelles on travaille sont à noter.

Exemple descriptif : la société InfoTech travaille avec des fournisseurs à Strasbourg, Lyon et Nancy.

- InfoTech n'est pas une donnée, c'est la description de l'organisation ;
- Fournisseur est un concept (dont on ne sait rien d'autre)
- Strasbourg, Lyon, ... sont des occurrences de la donnée 'nom ville' du concept 'ville'.
- À indique qu'il y a un lien, une association, une relation entre les concepts fournisseur et ville

1.2 ... mais en pratique ...

Certains concepts sont ambigus. Ainsi dans certains cas, des concepts seront relégués au rang de donnée.

Exemple descriptif : la société InfoTech travaille avec des fournisseurs à Strasbourg, Lyon et Nancy. Les noms des villes ne font pas l'objet d'une gestion particulière.

- InfoTech n'est pas une donnée, c'est la description de l'organisation ;
- Fournisseur est un concept (dont on ne sait rien d'autre)
- Strasbourg, Lyon, ... sont des occurrences de la donnée 'nom ville' du concept 'ville'.
- À indique qu'il y a un lien entre 'Fournisseurs' et 'Ville'. Mais comme la ville n'a pas besoin d'être de façon précise, codifiée (ajout au sujet) le nom de la ville sera noté seul et le concept risque de disparaître.

De toutes façon, si on note les données dans un dictionnaire, les concepts ne seront pas notés en dehors du brouillon, à titre d'indicateur.

2 Le dictionnaire des données (DD)

2.1 Définitions

Le dictionnaire des données contient toutes les données nécessaires, relevées dans l'analyse de l'organisation (dans l'analyse du sujet). Il est organisé sous forme d'un tableau de cinq colonnes.

Exemple descriptif :

N° Donnée	Code Donnée	Description Donnée	Type et longueur	Observations
1	numFou	Numéro du fournisseur	Numérique 5	Identifiant de Fournisseur
2	nomFou	Raison sociale	Alphanumérique 30	
3	dateLiv	Date de livraison	Date	
4	nomVille	Nom de la ville	Alphanumérique 30	
...				

Les données relevées ici concernent les données concrètes et utiles à l'organisation, souvent repérées par des noms communs ou propres dans le descriptif de l'organisation.

Les données concernant l'organisation elle-même ne sont pas à relever.

2.2 Construction

Une fois toutes les données relevées, il faut contrôler qu'il n'y ai pas de :

- **imprécisions** : les descriptions sont suffisamment précises (dans une commande, quantité ne suffit pas, préciser quantité commandée)
- **polysèmes** : deux données portant le même nom ne signifiant pas la même chose. Dans ce cas, il faut modifier le nom de la donnée en précisant son rôle (nom ? préciser nom fournisseur et nom ville),
- **synonymes** : deux données différentes peuvent signifier la même chose. Dans ce cas, choisir une donnée et supprimer les synonymes (code article et numéro d'identification de l'article, référence article, ...),

Pour régler la plus part des problèmes, le premier travail est de préciser les descriptions trouvées dans le sujet.

Pour cela, on précisera pour chaque donnée la nature, le rôle et le concept auquel elle appartient.

Exemples oui :

- nom de l'employé : nature=nom=texte, rôle=nom, concept=employé ;
- nom du fournisseur : nature=nom=texte, rôle=nom, concept=fournisseur;
- quantité d'article commandé : nature=quantité, rôle=quantité, concepts=article, commande
- date de commande : nature=date, rôle=date de la commande, concept=commande

Exemple non :

- description : nature=texte, rôle=description, concept inconnu
- fournisseur : nature et rôle inconnus, concept=fournisseur

Les données calculées (totaux, prixTTC si on dispose du prix HT et du taux de TVA, ...) ne sont pas retenues mais elles peuvent indiquer la présence de données non évoquées qui ont permis leur calcul (un prix TTC révèle la présence d'un prix HT et d'un taux de taxe, le prixTTC ne sera pas une donnée, les prix HT et taux le seront).

Elles seront indiquées dans le dictionnaire mais n'auront ni code, ni numéro de donnée.

Les concepts ne sont jamais mentionnés dans le dictionnaire des données.

Le travail suivant est d'éliminer les homonymes et polysèmes puis les synonymes.

Ensuite, et ensuite seulement, on attribuera un code à chaque donnée.

Ce code reprendra les nature, rôle et concepts de la description afin d'être plus clair.

2.3 Convention de notation des codes des données et concepts

Une bonne solution est d'abrégier le concept puis d'y associer le rôle et la nature. Le sens importe peu, l'essentiel est de rester cohérent et utiliser le même système tout au long de l'étude.

Le code des données une fois défini, il ne bougera plus et sera implémenté physiquement dans la base de données. Soyez donc très attentif aux codes que vous donnez.

Une convention d'écriture récente utilise l'association (la concaténation) de mots dont chaque mot commencera par une majuscule.

Exemple oui : nomFournisseur, artDesc, descArt, etc. ...

Exemple non : nomfournisseur, artdesc, descart : peu lisibles.

Dans cette convention, un concept, une entité, une relation commencera par une majuscule :

Exemple : Fournisseur, Article, LigneCommandeAtricle, ...

Une donnée commencera par une minuscule :

Exemple : nomFou, artDesc, ...

Cette convention est issue de la programmation objet en langage java et C

La convention précédente utilisait des caractères soulignés ('_', *angl.* : *underscore*) comme suit :

Exemple : nom_fournisseur, desc_art

Cette notation est aléatoire car il est facile de faire une erreur "d'orthographe" au cours de l'étude.

Exemple : qté_article_commandé, qté_articleCommandé

2.4 Les types principaux de données

Les types de données sont principalement :

- Numérique : composé exclusivement de chiffres (valeurs, numéro d'ordre, de série, compteurs, ...), longueur x chiffres dont y décimales ;
- Alphanumérique : composé de caractères alphabétiques ou numériques, les espaces et quelques signes de ponctuation sont tolérés (. , ! \$ * / + = - _ @ & # % ...), mais il convient de les éviter autant que possible ;
- Date : nombre représentant une date (pas de longueur à spécifier) ;
- Booléen : champ pouvant prendre les valeurs vrai ou faux (ou oui ou non, pas de longueur à spécifier).

Ils permettent de préciser la nature de la donnée. Cependant, la détermination précise du type peut être difficile au début de l'étude. D'où la nécessité de préciser la nature de la donnée dans sa description.

3 Le graphe de dépendances fonctionnelles (GDF)

3.1 Dépendance fonctionnelle, définition

Une dépendance fonctionnelle est une interrelation, un lien, une association, une relation entre deux données ou deux groupes de données.

On distingue une source et une cible.

La définition de dépendance fonctionnelle est la suivante :

Pour une valeur source, on peut déterminer une et une seule valeur cible

J'explique :

Si, connaissant une valeur d'une donnée a, on peut déterminer une et une seule valeur d'une donnée b, alors on peut dire que b dépend fonctionnellement de a.

(en maths on pourrait écrire : $b = f(a)$; $\text{nomFou} = f(\text{numFou})$)

Exemple : la valeur F0262 du n° de fournisseur détermine que le nom du fournisseur est plasti'Form (à Arras).

Le numéro du fournisseur détermine le nom du fournisseur (un et un seul).

Le nom du fournisseur dépend du numéro du fournisseur (déterminé par).

Attention, une dépendance fonctionnelle n'est pas réversible.

Exemple : plusieurs fournisseurs pourraient avoir le même nom (homonymes).

Une dépendance fonctionnelle est symbolisée par une flèche : $\text{numFou} \rightarrow \text{nomFou}$.

En théorie, la cible doit toujours exister.

En pratique, pour un numéro de commande, il n'y a pas forcément de numéro de livraison mais lorsqu'il y en a un, il est toujours unique. On peut noter ce type de dépendances par une flèche en pointillé.

3.2 La matrice des DF

A l'aide du dictionnaire des données, il est possible d'établir la matrice des DF directes.

Cette matrice est un tableau faisant apparaître verticalement et horizontalement toutes les données.

Elles seront source horizontalement et cible verticalement.

Je ne me sers pas de la matrice des DF mais il faut savoir qu'elle existe.

		1	2	3
1	numFou			
2	nomFou	1		
3	dateLiv			

3.3 La liste des DF

Après avoir constitué le dictionnaire des données, pour construire le graphe des DF, il faut repérer toutes les DF entre les données du dictionnaire.

Exemples oui :

source (ou origine) \rightarrow cible (ou but, destination, ...)

numFou \rightarrow nomFou, nomVille

codeArticle \rightarrow descriptionArticle

livNum, artNum \rightarrow qtéArtLiv

Exemples non :

nomFou \rightarrow nomVille

- homonymes possibles
cdeNum -> qtéArtCde
- le n° de commande ne suffit pas à déterminer la quantité d'article, il peut y avoir plusieurs articles dans la commande.

Pour améliorer la liste des DF, il faut contrôler et éliminer les transitivités.

Si $a \rightarrow b$, c et $b \rightarrow c$ alors $a \rightarrow c$ et $b \rightarrow c$. on simplifie la première DF.

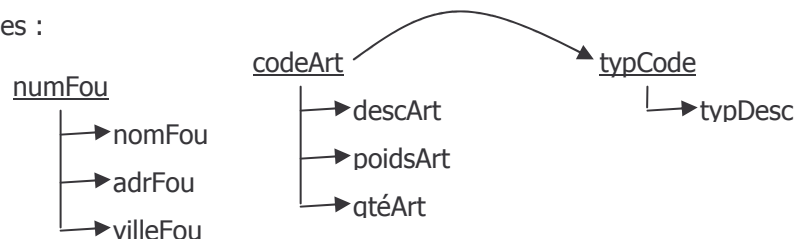
Exemple : cdeNum \rightarrow cdeDate, cliNom et cliNum \rightarrow cliNom

Ça veut dire : cliNum \rightarrow cliNom et cdeNum \rightarrow cdeNom

3.4 Construire le graphe des DF

Ce graphe des DF est une représentation graphique des dépendances fonctionnelles entre les données.

Exemples :



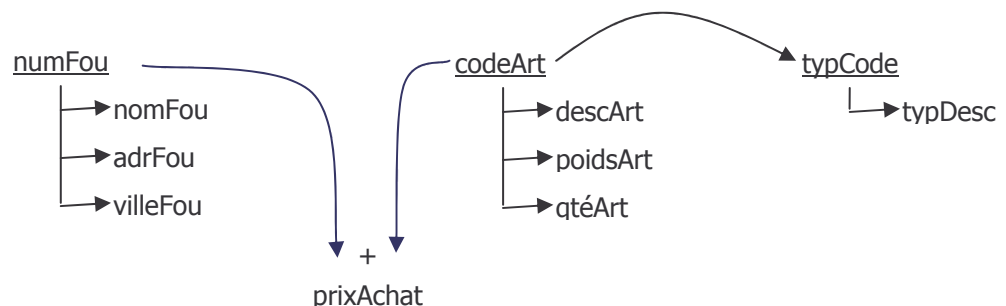
Dans certains cas, une donnée peut dépendre de plusieurs attributs.

Exemple : un article est acheté chez différents fournisseurs avec des prix différents. Un fournisseur peut vendre plusieurs articles.



La notation du GDF est souple et on peut aussi le décrire comme suit :

Exemple :



(noter le signe + qui signifie que les données numFou et codeArt doivent être conjointes)

Grâce au graphe des DF il est alors possible de construire un MCD sûr à 80%, les autres 20% sont constitués de réflexion sur l'existence d'associations, non hiérarchiques et non porteuses de propriété entre les concepts, et de questions telles que :

Cette donnée est-elle codifiable, doit-elle être codifiée,

Y a-t-il toujours une valeur pour cette donnée, est-elle unique ?

Cette association ternaire est-elle juste ?

Les associations ayant la même collection d'entité ont-elles le même ensemble d'occurrences ?

A suivre, avec le MCD (doc cgo122) puis le passage du GDF au MCD (doc cgo123).