

Polycope 1 : Modélisation des données et modélisation conceptuelle des données.

1- Modélisation des données :	2
2- Modélisation Conceptuelle des Données (MCD).	2
a- Les concepts :	6
b- La démarche de construction du MCD.	6
c- La modélisation des DF faibles, fortes, CIF et CIM.	8
d- Identifiant relatif.	8
3- Pour conclure ce polycopié...	9

1- Modélisation des données :

La conception d'un système d'information ne peut se faire sans une production de modèles : ils représentent en quelque sorte « les plans » du futur système.

Les applications informatiques comportent des données et des traitements, ces deux versants du système d'information seront représentés « graphiquement » en respectant un formalisme adapté.

La modélisation des données permet d'identifier les entités logiques et les dépendances logiques entre ces entités. C'est une représentation abstraite : les valeurs des données individuelles observées sont ignorées au profit de la structure, des relations, des noms et des formats des données pertinentes.

La méthode Merise propose 4 modèles de données : le MCD, le MOD, le MLD et le MPD.

Dans ce cours, trois seront abordés : le MCD, le MLD, et le MPD (ce dernier sera vu par le biais d'un logiciel de modélisation).

2- Modélisation Conceptuelle des Données (MCD).

➔ La situation support du cours :

Le triathlon, né dans l'archipel d'Hawaï il y a vingt ans, discipline olympique en ouverture des jeux de Sydney en septembre 2000, est un sport jeune en cours de développement. La discipline, lancée par le capitaine de vaisseau John Collins, consiste dans l'enchaînement de trois sports de base dans un ordre déterminé immuable : natation (issue du « Waikiki rough water swim » entre deux îles d'Hawaï), cyclisme sur route (issu du « Around Oahu bike race », course cycliste de l'île Oahu), course à pied (inspirée du marathon d'Honolulu).

Au niveau national, ce sport est géré par la Fédération Française de Triathlon (FFTRI), puis par des ligues régionales et enfin avec l'aide des comités départementaux auxquels participent les clubs locaux.

Le Comité Départemental de Triathlon du Val-de-Loire (CDTVL) est l'un de ces comités. Ses membres élus sont issus des bureaux des différents clubs locaux. Le CDTVL se charge notamment de l'organisation des épreuves qui se déroulent dans son département (le Loiret) et sont ouvertes aux triathlètes de la France entière.

Le CDTVL souhaite disposer d'un logiciel de gestion des triathlons qu'il organise. Les informations à prendre en compte sont décrites ci-dessous.

Les triathlètes

Chaque triathlète possède une licence délivrée par la Fédération Française de Triathlon. Le numéro de licence identifie un sportif et reste identique d'une année à l'autre.

Un triathlète n'est pas obligatoirement membre d'un club. Un sportif qui n'est pas membre d'un club s'inscrit aux compétitions de manière individuelle. Cependant, sa licence se présente comme celle d'un membre de club

Les licences se présentent ainsi :

Licence Club

Fédération Française de Triathlon	
SAISON 2013-2014	
<div>Photo du titulaire</div>	LICENCE Numéro 4526
Comité : Comité Départemental de Triathlon du Val-de-Loire	
Nom : Guimot	
Prénom : Jean	
Sexe : masculin	
Adresse : 34 rue de la Bergerie 45720 Clergie	
Date de naissance : 16 juillet 1957	
Catégorie : Vétéran	
Club : Élan Sportif de Clergie	
Date première licence : 12/10/2002	

Licence individuelle

Fédération Française de Triathlon	
SAISON 2013-2014	
<div>Photo du titulaire</div>	LICENCE Numéro 1402
Comité : Comité Départemental de Triathlon du Val-de-Loire	
Nom : Veliot	
Prénom : Alain	
Sexe : masculin	
Adresse : 3 rue des Ormes 45120 Bellat	
Date de naissance : 15 mars 1958	
Catégorie : Vétéran	
Club : Néant	
Date première licence : 12/10/2002	

Remarques :

- La date de 1ère licence est la date à laquelle le triathlète a été licencié pour la première fois.
- Si un triathlète change de statut (membre d'un club devenant individuel par exemple), son numéro de licence reste inchangé.

Tout triathlète appartient à une catégorie d'âge (benjamin, minime, cadet, junior, senior, vétéran). Chaque catégorie est identifiée par un code. Chaque catégorie est caractérisée par l'âge auquel elle débute et l'âge auquel elle se termine.

Tous les clubs ayant des triathlètes inscrits à des compétitions organisées par le CDTVl sont répertoriés. Un club est identifié par un numéro et possède un nom, une adresse et un numéro de téléphone. Il est également nécessaire de connaître le nom des entraîneurs du club : ces entraîneurs sont toujours des triathlètes membres du club qu'ils entraînent.

Les triathlons

Un triathlon est une compétition ouverte à toutes les catégories d'âge.

Il existe plusieurs types de triathlon, chaque type de triathlon se distingue des autres par les distances à parcourir dans les trois sports de base (natation, cyclisme et course à pied). Chaque type de triathlon est identifié par un code et possède une désignation. À titre d'exemple, le code TROP correspond à un triathlon olympique.

Chaque triathlon possède un numéro, un nom, un type et se déroule en un lieu précis à une date donnée. Par exemple, le triathlon numéro 12, nommé « Course folle », se déroule à Orléans le 26 juin 2000 ; il s'agit d'une compétition de type TROP.

Le déroulement de la compétition

Dès que l'organisation d'un triathlon a été publiée par le CDTVL, les triathlètes peuvent s'y inscrire. À l'inscription, un numéro de dossard est attribué au triathlète pour la compétition et sa date d'inscription est mémorisée. Une inscription est identifiée par le numéro du triathlon concerné suivi du numéro de dossard attribué au triathlète. Les inscriptions sont closes 15 jours avant la date de la compétition.

Le jour du triathlon, le CDTVL recense les participants à la course. Il arrive qu'un sportif inscrit ne se présente pas, il est alors déclaré forfait.

Après le déroulement de la course, les résultats obtenus par chaque participant sont relevés :

- classement du concurrent dans la catégorie (premier, second, ...),
- temps réalisé dans les différentes épreuves (natation, course cycliste, course à pied).

Le cas des abandons en cours d'épreuve sort du cadre de cette étude et n'est donc pas à prendre en compte.

Les contrôles anti-dopage

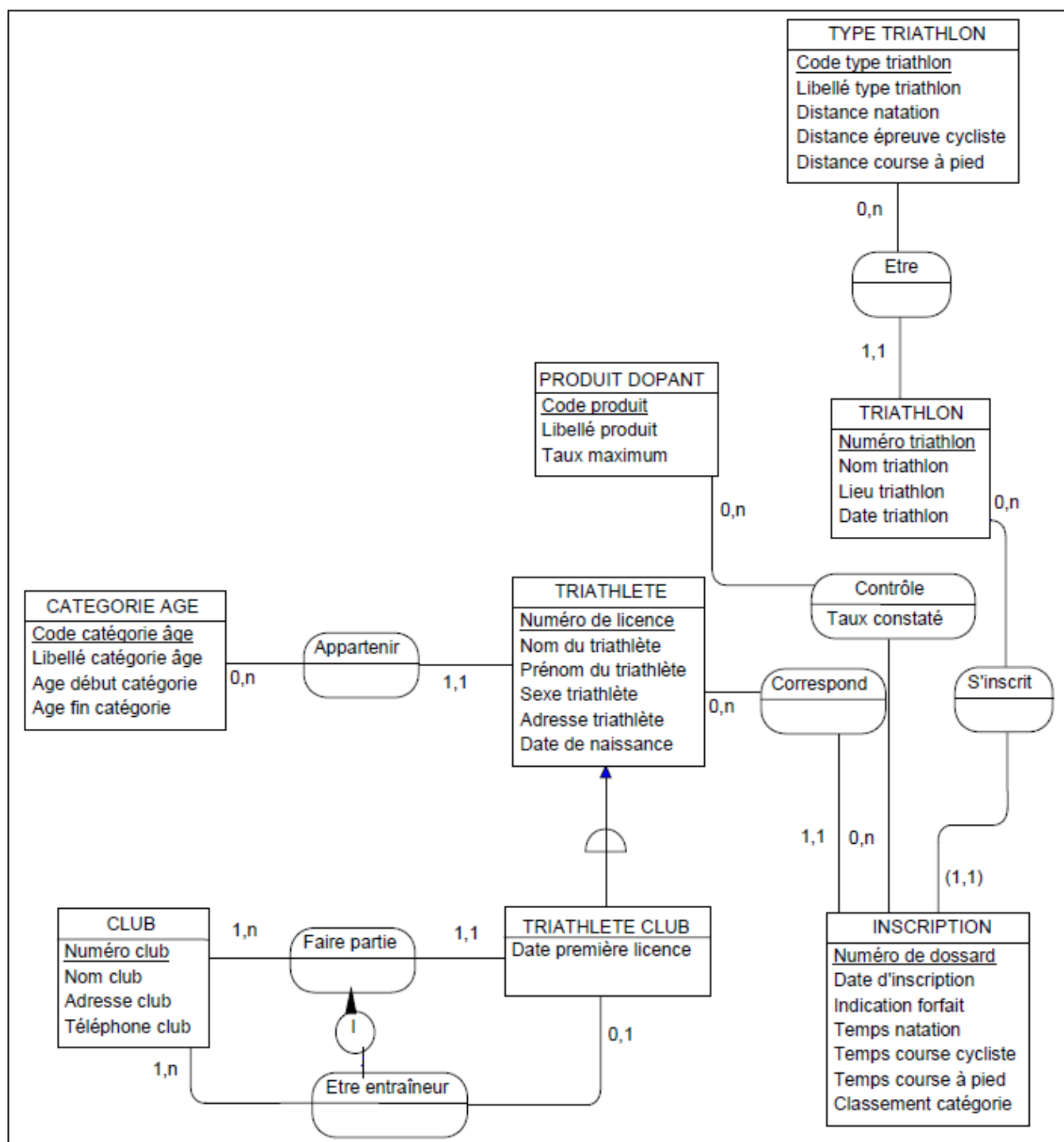
Dans le cadre de la lutte anti-dopage, la fédération nationale expérimente de nouveaux contrôles à l'issue d'une compétition de triathlon.

Un contrôle consiste à vérifier l'absence de produits dopants dans le sang.

Un produit dopant est identifié par un code, possède un libellé et un taux maximum autorisé par millilitre de sang.

Lors d'un triathlon, certains triathlètes, identifiés par leur numéro de dossard, sont contrôlés. On effectue donc des prélèvements sanguins et un laboratoire indépendant analyse ces prélèvements. Dès qu'elles sont connues, on enregistre les mesures établies par le laboratoire pour chacun des produits dopants recherchés.

➔ Le modèle conceptuel correspondant à cette situation :



a- *Les concepts :*

- La propriété : c'est un descripteur conforme aux choix de gestion de l'organisation. Elle est utilisée pour décrire les entités et les associations. Toute propriété est unique dans le modèle. Une occurrence de propriété est une valeur de cette propriété.
- L'identifiant : Parmi les propriétés qui décrivent une entité, l'une d'entre elles, au moins, doit permettre de caractériser chacune de ses occurrences de façon unique. La propriété qui joue ce rôle est dite « identifiant » de l'entité. Par convention de représentation, l'identifiant est la première propriété de l'entité est souligné. Le choix de l'identifiant est délicat ; on peut opter pour :
 - o une propriété « naturelle », par exemple, le nom d'un pays pour une entité « pays »,
 - o une propriété « artificielle » inventée par le concepteur pour identifier l'entité qu'il vient de concevoir (tous les numéros, références, codes, etc... en sont l'illustration).
 - o Une propriété composée en s'assurant que la règle de composition ne générera pas de doublons,
 - o Un identifiant relatif, par exemple ici, le numéro de triathlon + le n° de dossard.
- L'entité : une entité est un ensemble d'objets ou d'individus de même nature pourvus d'une existence propre et dont la définition est conforme aux choix de gestion de l'organisation. L'entité est généralement désignée par un nom commun représentatif des individus ou objets portés par l'entité. Une occurrence d'entité est un ensemble ayant une existence propre d'occurrences de ses propriétés.
- L'association : l'association n'existe qu'au travers des entités qui la composent. On appelle dimension, le nombre d'entités composant l'association. On désigne l'association **par des noms de verbes, il est souhaitable d'utiliser un verbe à l'infinitif**. Une association n'a pas d'identifiant propre (mais un identifiant implicite, nous en reparlerons).
- Les cardinalités : Le terme cardinalité, dans le formalisme entité-association, traduit la participation des occurrences d'une entité aux occurrences d'une association. Cette participation s'exprime par deux valeurs : cardinalité minimum et cardinalité maximum.

b- *La démarche de construction du MCD.*

L'analyse des données concerne un domaine d'étude déterminé qu'il convient de parfaitement bien délimiter. Puis, elle suit un processus en 3 étapes :

- La recherche des propriétés qui conduit à l'élaboration d'un dictionnaire de données,
- L'analyse des dépendances fonctionnelles entre propriétés,
- La construction du modèle conceptuel de données.

Le dictionnaire de données : sa construction passe par l'analyse de l'ensemble des documents répertoriés dans le domaine d'étude. Il s'agit d'identifier les données pertinentes et à les présenter en tableau, en respectant les règles suivantes :

- Mettre en évidence les propriétés,
- Supprimer les synonymes (exemple codetriathlon et numerotriathlon),
- Supprimer les polysèmes : une propriété polysème désigne sous une même dénomination, des informations différentes, par exemple : la propriété « nom » qui peut représenter le nom du triathlon dans un document, le nom du triathlète dans un autre. Résoudre le polysème revient à créer 2 propriétés distinctes.
- Supprimer les propriétés calculées. On ne retient comme propriétés élémentaires que celles qui interviennent dans la détermination de la propriété calculée.

- Supprimer les propriétés génériques : une propriété générique structure un ensemble de propriétés élémentaires. La propriété « adresse » structure les propriétés élémentaires « Rue », « Ville », « Code Postal ». On retient comme propriété les composées de la propriété générique.

L'analyse des dépendances fonctionnelles entre propriétés :

- Il y a dépendance fonctionnelle entre deux propriétés a et b, si, connaissant une valeur de a, on obtient **une valeur de b et une seule**. On note $a \rightarrow b$, a est un identifiant.

Exemples : TRIANumLicence \rightarrow TRIANom
 TRIAPrenom

INSCDossard \rightarrow TRIANumLicence

INSCDossard + DOPCode \rightarrow TxConstaté

- Les dépendances fonctionnelles doivent être directes :

Exemple : Si INSCDossard \rightarrow TRIANumLicence et TRIANumLicence \rightarrow TRIANom, alors :

INSCDossard \rightarrow TRIANom : mais cette DF n'est pas directe.

- Les dépendances fonctionnelles doivent être élémentaires :

INSCDossard + DOPCode \rightarrow TxConstaté est une DF élémentaire,

INSCDossard + DOPCode \rightarrow DOPTxMaximum n'est pas élémentaire.

- L'analyse des dépendances fonctionnelles peut être synthétisée dans un tableau, appelé « matrice des DF », outil précieux qui évite bien des erreurs lors du tracé du MCD. La matrice correspondant au modèle conceptuel ci-dessus est la suivante :

		1	4	11	17	22
1	DOPCode	1				
2	DOPLibelle	1				
3	DOPTauxMaximum	1				
4	INSCDossard		1			
5	INSCDate		1			
6	INSCIndicationForfait		1			
7	INSC TpsNatation		1			
8	INSC TpsCourseCycliste		1			
9	INSC TpsCoursePied		1			
10	INSC ClasstCategorie		1			
11	TRIANumLicence		1	1		
12	TRIANom			1		
13	TRIA Prenom			1		
14	TRIASexe			1		
15	TRIAAdresse			1		
16	TRIA DateNaissance			1		
17	TRINumero		(1)		1	
18	TRINom				1	
19	TRILieu				1	
20	TRIDate				1	
21	TxConstaté					1
22	1+4					1
23						

Et le passage de la matrice au MCD est aisé :

- A chaque identifiant non concaténé va correspondre une entité,
- Une DF entre deux identifiants va donner une association avec cardinalité maximum à 1 sur la patte côté source de DF.
- Un identifiant concaténé va donner une association avec des cardinalités maximum à N sur chacune des pattes.

c- *La modélisation des DF faibles, fortes, CIF et CIM.*

- Une DF faible est une DF qui n'est pas toujours réalisée ; elle est modélisée par des cardinalités 0,1 sur la patte côté source de DF. Elle ne peut pas être porteuse de données.
- Une DF forte est une DF qui est toujours réalisée ; elle est modélisée par des cardinalités 1,1 sur la patte côté source de DF. Elle ne peut pas être porteuse de données.
- Une CIF (contrainte d'intégrité fonctionnelle) est une DF forte stable dans le temps, elle est modélisée par des cardinalités 1,1 sur la patte côté source de DF.
- Une CIM (contrainte d'intégrité multiple) est modélisée par une association avec des cardinalités maximum à N sur chacune des pattes. Cette association peut être porteuse ou non de données.

d- *Identifiant relatif.*

Les cardinalités de l'association « correspondre » côté « inscription » sont notées (1,1) pour signaler la présence d'un identifiant relatif : une inscription est identifiée par le numéro de triathlon + le numéro de dossard.

Le numéro de dossard est un numéro séquentiel (1, 2, 3...) et il ne pouvait, seul, identifier l'entité « inscription », car il y aurait eu des doublons. L'identification relative permet de contourner le problème :

garder le numéro d'ordre que constitue le dossard comme identifiant de l'entité « inscription », mais associé au numéro de triathlon afin que l'identifiant soit unique.

L'entité « Inscription » est dite « Entité faible », car elle dépend totalement de l'entité « triathlon » qui est dite « entité forte ».

3- Pour conclure ce polycopié...

Travail à faire :

- Compléter la matrice de dépendances fonctionnelles du cas « Triathlon ».
- Retrouver les règles de gestion illustrant ces dépendances fonctionnelles.
- Fiche d'exercice : cas « CRAB »