

Geographic Information Technology Training Alliance (GITTA) presents:

Modélisation conceptuelle des données

Responsable: Dominique Schneuwly, Regis Caloz

Table des matières

1. Modélisation conceptuelle des données	2
1.1. Concepts de modélisation	3
1.1.1. Bases de données, modèles et schémas	3
1.1.2. Modèles conceptuels de données	3
1.1.3. Processus de conception d'une BD	4
1.1.4. Terminologie: classes, objets, etc.	5
1.2. Le modèle Entité Association	6
1.2.1. Le schéma exemple	6
1.2.2. Entités et associations	6
1.2.3. Attributs, identifiants	8
1.2.4. Contraintes d'intégrité	9
1.2.5. Représentation multiple	10
1.2.6. TE faible	11
1.3. Règles de vérification d'un schéma Entité Association	12
1.3.1. La notion de dépendance	12
1.3.2. Validation des attributs d'un TE	13
1.3.3. Validation des attributs d'un TA	14
1.3.4. Validation d'un TA ternaire	16
1.3.5. Elimination des TA redondants	16
1.3.6. Transformation des attributs traduisant une association	16
1.4. Modélisation conceptuelle de données spatiales et MCD spatiaux	17
1.4.1. Concepts de modélisation spatiale	17
1.4.2. Pourquoi utiliser les MCD spatiaux et les concepts qu'ils proposent?	19
1.4.3. MCD spatiaux (MADS, Perceptory)	19
1.4.4. Exercice	20
1.5. Bibliographie	22

1. Modélisation conceptuelle des données

La modélisation spatiale consiste à identifier et à délimiter des éléments spatiaux dans l'espace géographique, à représenter ces éléments par des unités graphiques (points, lignes, polygones, etc.) et à les répartir sur différentes couches. Les données, une fois modélisées, sont stockées dans la base de données d'un SIG. Concevoir la base de données nécessite une première phase de modélisation conceptuelle qui consiste à déterminer quelles sont les structures de données pertinentes géographiques ou non (lac, route, bâtiment, personne), et les relations qui existent entre ces dernières (le bâtiment B appartient à la personne X). Nous traitons ici la modélisation conceptuelle des données : Comment élaborer un schéma de bases de données? Comment vérifier ce schéma? Les bases de données et leur implantation dans un SIG sont présentés dans le module Basic Data Management.

1.1. Concepts de modélisation

Pas de système d'information correct sans modélisation ! Pour cela il existe des concepts formels qui viennent du monde des bases de données. Dans cette unité, nous introduisons les concepts généraux ainsi que la terminologie appropriée.

Temps estimatif: 15 minutes

1.1.1. Bases de données, modèles et schémas

Une **base de données** est un ensemble de données **Cohérent** : les données sont structurées sur des définitions claires et non ambiguës pré-établies, **Intégré** : elles sont regroupées au sein d'un même ensemble, **Partagé** : elles sont utilisées par plusieurs utilisateurs et/ou types d'utilisateurs et **Défini pour les besoins d'une application**. Le cycle de vie d'une base de données (BD) comprend 4 phases: Modélisation conceptuelle de la BD, Implantation des structures et des données dans un SIG, Utilisation (requêtes d'interrogation et de mises à jour), Maintenance (correction, évolution). La première phase de modélisation est primordiale. Il n'y a pas de base de données bien structurée sans une bonne modélisation !

La phase de conception de la BD est une phase de réflexion sur la structure des données en fonction des besoins de l'application: données importantes, propriétés, contraintes, requêtes à prévoir... en accord avec les utilisateurs. Conceptuel signifie qu'on est indépendant des solutions informatiques. L'intérêt d'établir un schéma conceptuel réside dans le fait d'être accès sur une application, d'être indépendant des technologies donc portable et facilitant l'échange d'informations, établi selon un modèle formel sur des spécifications non ambiguës.

La modélisation conceptuelle des données, c'est l'activité d'élaboration du schéma conceptuel selon un modèle conceptuel. Il s'agit de poser sur le papier le schéma conceptuel qui permettra d'établir la structure des données (schéma logique) dans le modèle logique : sous forme de tables dans le modèle relationnel ou de classes dans le modèle orienté objet (voir le module BDM).

Un schéma c'est l'expression de la description de la base de données obtenue en employant un modèle de données. Un modèle conceptuel est un cadre formel pour schématiser le contenu des informations selon un formalisme établi, graphique si possible. Nous verrons dans l'unité suivante le modèle EA.

1.1.2. Modèles conceptuels de données

Un modèle conceptuel de données est un ensemble de concepts qui permettent de décrire et de manipuler des données du monde réel, et de règles d'utilisation de ces concepts. Les modèles comportent 2 parties : une partie statique qui décrit la structure des données (MCD) et les contraintes explicites sur ces données (CI), et une partie dynamique qui définit les traitements sur les données (MCT).

Les concepts de base de la modélisation sont :

- Les objets regroupés en classes et identifiés,
- Les liens entre objets avec leurs cardinalités,
- Les propriétés des objets,
- La représentation multiple des objets.

Un modèle conceptuel doit respecter les propriétés suivantes :

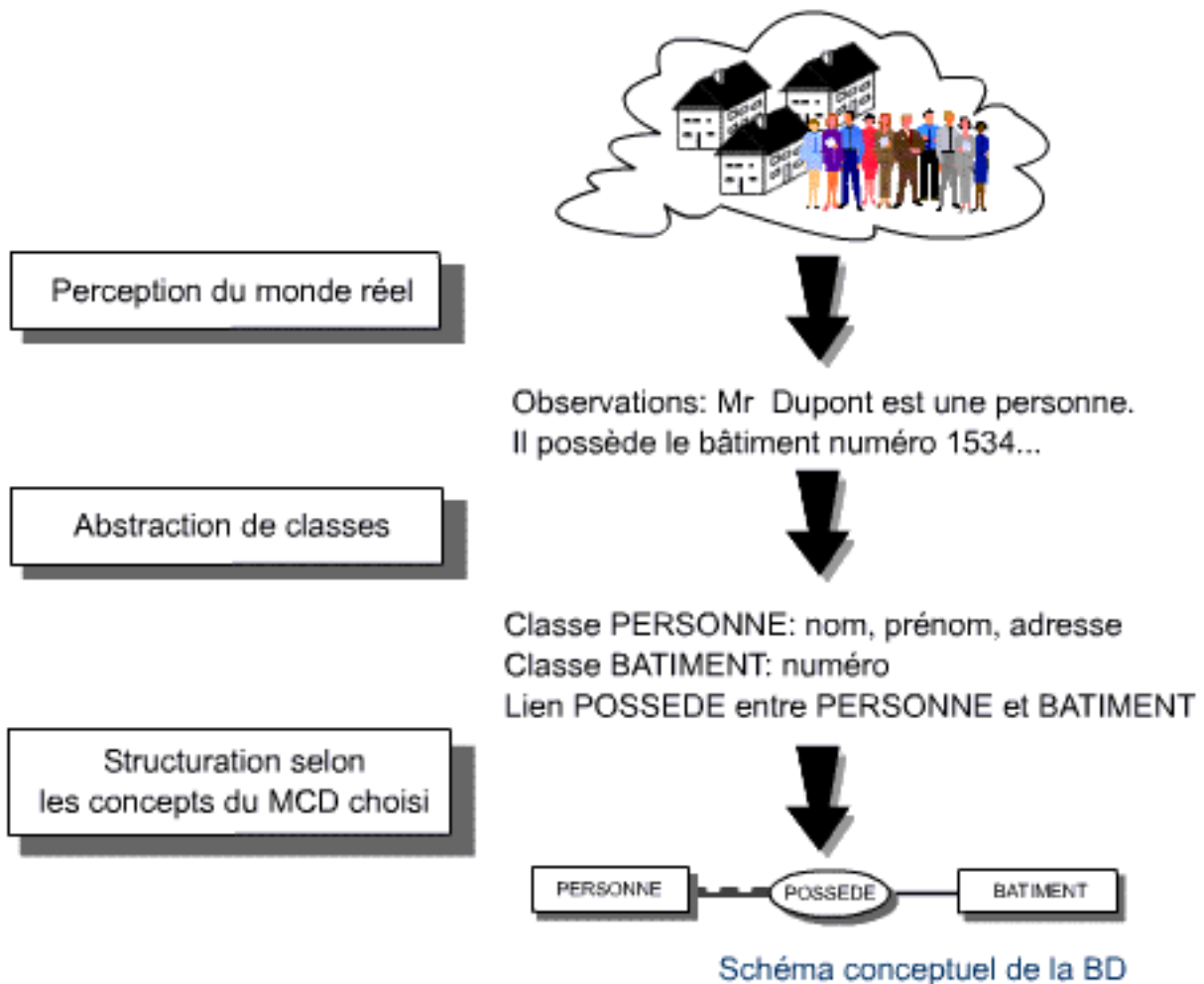
- Complétude (Description de tous phénomènes courants nécessaires à l'application)
- Fiabilité (formellement défini)

- Orientation utilisateur (compréhensible, clair, lisible)
- Orthogonalité (les concepts proposés doivent être indépendants)
- Compatibilité logiciel (traduisible en SGBD existant)
- Complètement opérationnel (capacités de manipulation des données)

1.1.3. Processus de conception d'une BD

Perception du monde réel

Une BD est une représentation de la partie du monde réel qui nous intéresse. Lors de la conception d'une BD pour les besoins d'un utilisateur, après interview, l'objectif est d'élaborer le schéma conceptuel de son application. L'utilisateur a une perception du monde réel axée sur son application : chaque utilisateur a sa propre focale d'observation. Son analyse de la réalité est donc partielle (elle ne représente que les informations intéressantes pour son application), subjective (elle représente le point de vue du concepteur) et infidèle (ne représente pas la réalité telle qu'elle est, mais telle qu'elle intéresse le concepteur). Les phénomènes observés sont abstraits en classes, puis représentés et décrits dans un schéma conceptuel selon le modèle choisi. Nous pouvons donc obtenir plusieurs schémas conceptuels pour les mêmes phénomènes du monde réel observés. Cependant le schéma conceptuel obtenu doit être conforme au modèle conceptuel choisi. Le processus est décomposé en 3 phases comme dans la figure ci-dessous.



1.1.4. Terminologie: classes, objets, etc.

Une classe est l'ensemble des objets perçus comme ayant les mêmes caractéristiques. Ils auront le même type. La classe représente le modèle ou patron de ce type d'objet. La population d'une classe est l'ensemble des objets qui la composent. Un objet du monde réel est un objet de la base qui a les mêmes caractéristiques type que sa classe d'appartenance, mais qui a une identité propre (étiquette ou identifiant) avec des valeurs propres. Un objet est une instance ou occurrence d'une classe. Les termes objet, classe, type, population, occurrence, instance sont des termes génériques de modélisation. Nous verrons les termes spécifiques au modèle Entité Association dans l'unité suivante.

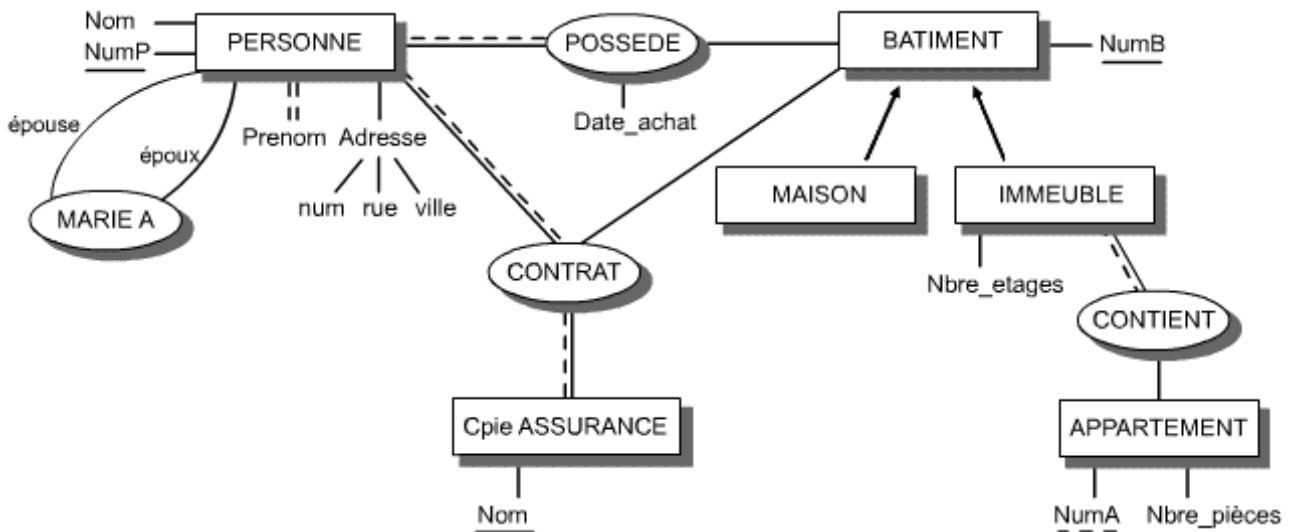
1.2. Le modèle Entité Association

Dans le modèle EA (Entité Association) ou ER (Entité Relation), les objets sont représentés par des entités, et les liens par des associations, aussi appelées relations. Nous voyons ici les concepts qui vont nous permettre d'élaborer des schémas conceptuels EA.

Temps estimatif: 30 minutes

1.2.1. Le schéma exemple

Voici un exemple de schéma conceptuel EA à partir duquel nous appréhendons les concepts de base du modèle EA.



Le schéma conceptuel EA est l'ensemble de descriptions des types d'entités TE et des types d'association TA avec leurs attributs et les liens de généralisation entre TE, et des contraintes d'intégrité (CI) associées. La base de données décrite par un schéma EA est l'ensemble des populations des TE et TA apparaissant dans le schéma conceptuel.

1.2.2. Entités et associations

Entités et TE

Une entité E est la représentation d'un objet du monde réel (concret ou abstrait) perçu par le concepteur comme ayant une existence propre, et à propos duquel on veut enregistrer des informations. Une entité existe indépendamment du fait qu'elle puisse être liée à d'autres entités de la BD. Un type d'entité (TE) est la représentation d'un ensemble d'entités perçues comme similaires et ayant les mêmes caractéristiques, et qui ont un intérêt en soi pour au moins un traitement de l'application. Toute entité appartenant à l'ensemble décrit par un TE est une occurrence du TE. L'ensemble des occurrences du TE constitue la population d'un TE.



Notre schéma conceptuel EA comprend 2 TE: PERSONNE et BATIMENT. Un exemple d'entité du TE PERSONNE est Mr Dupont. Une entité du TE BATIMENT est le bâtiment numéro 1534.

Modélisation conceptuelle des données

Le formalisme graphique proposé ici présente les TE sous forme de rectangle. On trouvera cependant dans la littérature de nombreux formalismes. Il est donc important de préciser en cas d'ambiguïté si vous en changez.

Associations et TA

Une association A est la représentation d'un lien non dirigé entre plusieurs entités (qui jouent un rôle déterminé). Un type d'association (TA) est la représentation d'un ensemble d'associations ayant la même sémantique, décrites par les mêmes caractéristiques et d'intérêt pour l'application, entre 2 ou plusieurs objets représentés par des entités.

Les TE PERSONNE et BATIMENT sont liés par un TA POSSEDE. Par exemple, le couple <Mr Dupont, 1534> est une association du TA POSSEDE et signifie que Mr Dupont possède le bâtiment 1534.

Rôles

Une association a 2 rôles de part et d'autre de l'association pour chacun des TE. Les rôles peuvent être clairement explicités sur le schéma conceptuel.

- le rôle *possède* : Mr. Dupont possède le bâtiment 1534.
- le rôle *est possédé par*: Le bâtiment 1534 est possédé par Mr Dupont.

Cardinalités

Les cardinalités des rôles permettent de contraindre les associations par les nombres minimum et maximum de participation de chaque entité du TE à l'association. Les cardinalités peuvent être notées comme sur le schéma conceptuel ci-dessus min:max ou représentées graphiquement comme indiqué ci-dessous.

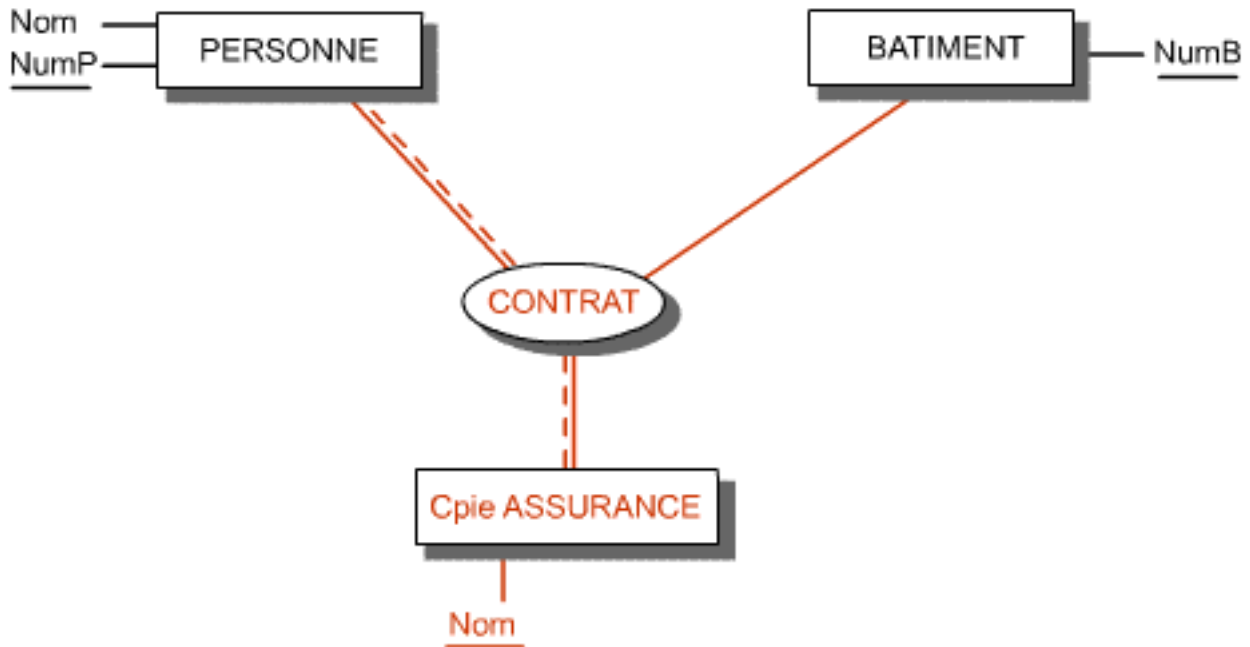
	minimum	maximum
-----	0	1
————	1	1
=====	0	n
-----	1	n
=====	m	n

Dans l'exemple, un propriétaire du TE PERSONNE possède au moins un bâtiment et peut en posséder plusieurs ; alors que chaque bâtiment du TE BATIMENT a un et un seul propriétaire. Ainsi Ducros ne peut pas posséder le bâtiment 647 puisque qu'il appartient déjà à Dupont.

PROPRIETAIRES	POSSEDE	BATIMENT
Dupont	<Dupont, 647>	647
Farmini	<Ducros, 647>	95
Ducros	<Ducros, 4621>	4194
Tartinot	<Dupont, 1534>	4621
Armigne		1534
		546

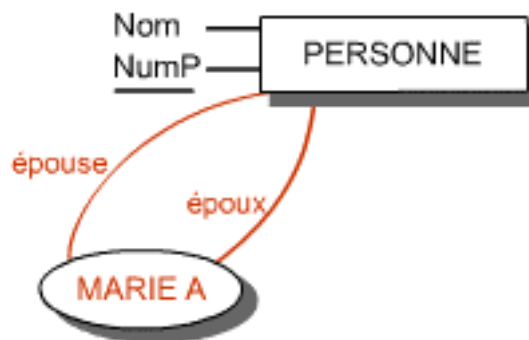
TA ternaires

Les TA sont la plupart du temps binaires, impliquant l'association de 2 entités. Ils peuvent également être ternaires en mettant en jeu l'association de 3 entités, quaternaires en permettant l'association de 4 entités, etc. Dans l'exemple ci-dessous, une occurrence du TA CONTRAT est un triplet: <personne, bâtiment, Cpie Assurance>. Un TA ternaire a 6 rôles.



TA cycliques

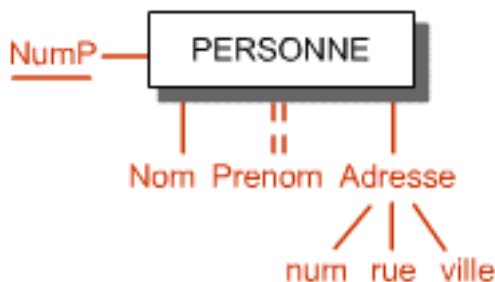
Si l'association lie 2 (ou +ieurs) entités du même type, elle est dite "cyclique". Une occurrence du TA 'marié à ' est un couple: < 1 personne/MARI, 1 personne/FEMME >. Dans ce cas, il est primordial de spécifier le rôle de chaque entité afin d'éviter toute ambiguïté.



1.2.3. Attributs, identifiants

Attributs

Est représentée par un attribut toute information intéressante qui participe à la description d'un objet ou d'un lien et qui ne fait l'objet de traitement qu'en tant que partie de cet objet ou lien. Un attribut ne dépend que de l'entité (ou de l'association, i.e. des entités liées) à laquelle il est attaché.



Les attributs peuvent être simples ou complexes, i.e. décomposés en d'autres attributs. C'est souvent le cas des dates (jour, mois, année) ou des adresses (rue, ville, code postal). Ils peuvent être monovalués: une seule valeur par occurrence (cardinalité max=1) ou multivalués: plusieurs valeurs par occurrence (cardinalité max>1). Ils peuvent aussi être obligatoires: une valeur au moins par occurrence (cardinalité min>=1) ou facultatifs (cardinalité min=0).

Ainsi l'attribut *NumP* est simple, monovalué, obligatoire. L'attribut *Adresse* est complexe, monovalué, obligatoire. L'attribut *Prénom* est simple, multivalué, facultatif.

Domaines de valeur

Le domaine définit toutes les valeurs permises pour un attribut. Par exemple on peut définir un domaine appelé Dnom comme une chaîne de caractères de longueur inférieure à 30, qu'on pourra affecter à tous les attributs qui sont des noms. Le domaine peut être statique, tel que le domaine Djour de valeurs contraintes entre 1 et 31 (noté Djour:[1:31]). Il peut aussi dépendre du contexte : contraint entre 1 et 30 si le mois est parmi les valeurs 4, 6, 9 ou 11.

Identifiants

L'identifiant d'un TE ou TA est l'ensemble minimum d'attributs tel qu'il n'existe pas 2 occurrences du TE (ou TA) ayant la même valeur pour ces attributs. Un TE, ou TA, peut avoir plusieurs identifiants possibles, ou aucun dans le cas de TE faible. Dans certains cas, on ajoute un attribut particulier tel qu'un numéro incrémental, attribut artificiel qui jouera le rôle d'identifiant. Par exemple: n°employé et nom+prénoms sont 2 identifiants possibles du TE Employé, si dans cette entreprise il n'y a jamais 2 employés ayant les mêmes nom et prénoms, ou le même numéro. L'identifiant d'un TA binaire est le couple d'identifiant des TE qui composent l'association. Cependant si le TA a une cardinalité maximum égale à 1 pour un des TE liés, alors tout identifiant de ce TE est identifiant du TA. Les identifiants des TE sont représentés graphiquement sur le schéma en les soulignant. Les identifiants des TA ne sont pas représentés sur le schéma pour ne pas le surcharger.

L'attribut *NumB* du TE BATIMENT souligné dans le schéma conceptuel est identifiant de ce TE. L'attribut *NumP* du TE PERSONNE souligné dans le schéma conceptuel est identifiant de ce TE.

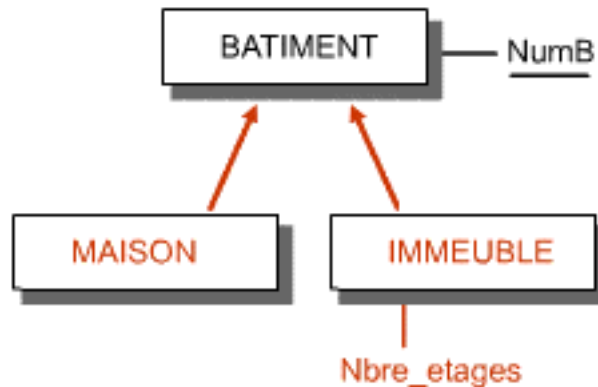
1.2.4. Contraintes d'intégrité

Les concepts d'entité, association, attribut et sous-type sont insuffisants pour décrire tout ce qui caractérise les données d'un schéma EA. Les contraintes d'intégrité CI sont des règles définissant les états, ou transitions d'état possibles de la BD. Elles permettent d'exprimer tout ce qui ne peut pas être décrit avec les concepts du modèle. Si les valeurs de la BD ne satisfont pas ces contraintes, il y a une "erreur"; on dit que la BD est incohérente. En l'état actuel des SGBD, la plupart de ces règles ne peuvent être vérifiées que par des programmes ad hoc, établis lors de la programmation des applications.

Par exemple dans notre schéma EA, on aimerait pouvoir exprimer le fait que si une personne possède un bâtiment, cette personne doit avoir contracté un contrat d'assurance pour ce bâtiment auprès d'une compagnie d'assurance. Ceci s'exprime sous forme d'une CI: soit b une entité du TE BATIMENT, soit p une entité du TE PERSONNE, s'il existe une occurrence $\langle p,b \rangle$ du TA POSSEDE, alors soit c une occurrence du TE Cpie ASSURANCE, il existe une occurrence $\langle p,b,c \rangle$ du TA CONTRAT.

1.2.5. Représentation multiple

Un TE, c'est la classe d'objets du monde réel perçus comme ayant les mêmes caractéristiques. Or un même ensemble d'objets peut être perçu d'un certain point de vue comme une seule classe, mais en même temps perçu d'un autre point de vue comme plusieurs classes, différentes malgré l'existence de caractéristiques communes. Un objet peut avoir plusieurs représentations. Ce concept de représentation multiple (aussi appelé généralisation/spécialisation) est une extension du modèle EA courant. Il est représenté par une flèche orientée du TE spécifique vers le TE générique. On l'appelle lien "est un" ou lien "is a". A toute occurrence du TE spécifique correspond une occurrence du TE générique. Inversement, à toute occurrence du TE générique correspond 0 ou 1 occurrence par TE spécifique. Il permet de raffiner les attributs ou des associations particulières selon les sous-types. Les attributs du sous-type héritent des attributs du TE génériques.

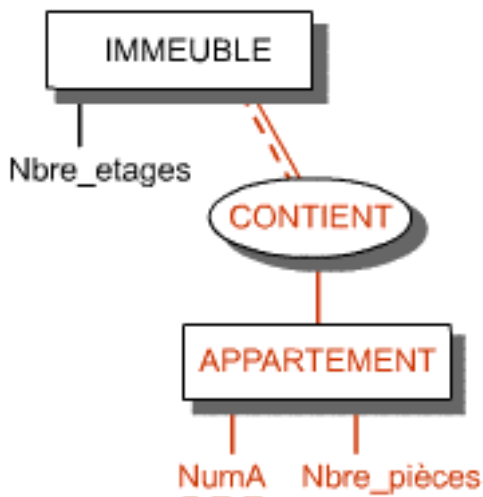


Dans notre exemple, l'ensemble des bâtiments est spécialisé selon le type : maison ou immeuble, afin de spécifier le nombre d'étages des immeubles en attributs, et les appartements qu'ils contiennent en associations. Les 2 TE MAISON et IMMEUBLE héritent de l'attribut *numB* du TE BATIMENT. De plus le sous-type n'ayant pas d'identifiant qui lui soit propre, l'identifiant des TE MAISON et IMMEUBLE est également *NumB*. Des contraintes d'intégrité sur les généralisations / spécialisations peuvent également être spécifiées pour décrire le partage de population entre sous-types d'un même sur-type:

- contrainte de couverture, pour spécifier que l'union des populations de certains TE spécifiques d'un même TE générique est égale à la population du TE générique (les bâtiments ne peuvent être que des maisons ou des immeubles).
- contrainte de disjonction, pour spécifier que les populations de certains TE spécifiques d'un même TE générique n'ont aucune occurrence en commun (un bâtiment ne peut pas être à la fois une maison et un immeuble).
- contrainte de partition, pour spécifier que la population d'un TE générique se distribue complètement et sans intersection entre certains de ses TE spécifiques : partition = couverture + disjonction (un bâtiment est soit une maison soit un immeuble).

1.2.6. TE faible

Dans notre schéma EA, il nous reste à voir un TE particulier: le TE faible APPARTEMENT. Un TE est dit faible si aucun sous-ensemble de ses attributs ne constitue un identifiant (il n'a pas d'identifiant qui lui soit interne) et si un identifiant peut être défini en intégrant un identifiant d'un autre TE qui lui est lié par un TA binaire de cardinalité (1,1), i.e. dont il dépend. Par exemple APPARTEMENT est un TE faible dépendant du TE IMMEUBLE. Du fait des cardinalités, il n'est pas possible de créer une occurrence de APPARTEMENT sans la rattacher à une occurrence existante de IMMEUBLE. On parle de dépendance d'existence. L'identifiant d'un TE faible (= celui du TA) est constitué de l'identifiant du TE dont il dépend, plus d'un (ou plusieurs) attribut du TE faible. L'identifiant de APPARTEMENT (et du TA "est dans") est: NumB + NumA.



1.3. Règles de vérification d'un schéma Entité Association

Une fois le schéma conceptuel EA établi, plusieurs types de vérification doivent être effectuées:

- vérification "syntaxique": il s'agit de vérifier que les règles du modèle entité association soient respectées (concepts du modèle + règles de vérification d'un schéma)
- par jeu d'essai: le concepteur vérifie grâce à une mini base de données que le schéma permet effectivement de stocker les informations nécessaires à l'application
- complétude par rapport aux traitements: le concepteur vérifie que le schéma contient tous les types d'information nécessaires à l'exécution des traitements prévus
- retour auprès des utilisateurs: le concepteur présente le schéma accompagné des définitions aux personnes qui utiliseront la base de données et vérifie que les informations contenues correspondent bien aux besoins
- par les règles de validation des schémas: quelques règles formelles permettent de valider le schéma : validation des attributs, élimination des TA redondants, etc.

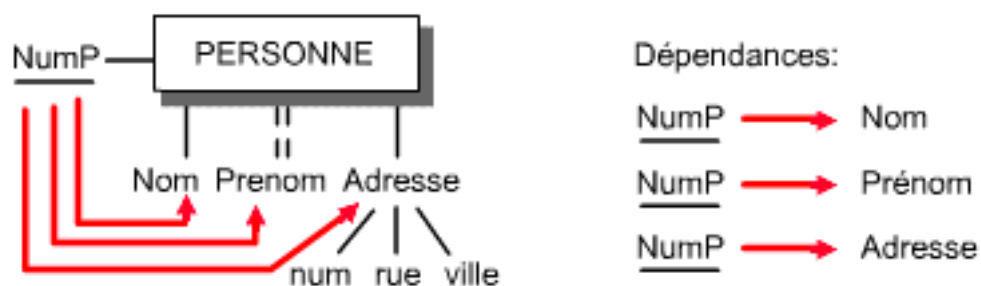
Chaque oubli, erreur, modification,, détecté lors des vérifications entraîne une mise à jour du schéma et relance les différentes phases de vérification. Cette unité présente les règles de validation de schéma après avoir introduit la notion de dépendance à partir de laquelle ces règles sont énoncées.

Temps estimatif: 45 minutes

1.3.1. La notion de dépendance

Avant de voir comment vérifier la cohérence syntaxique d'un schéma entité association, nous introduisons le concept de dépendance entre données ou entre types d'entité, qui est utile pour certaines règles de vérification. Le concept de dépendance n'est pas propre au modèle entité-association; c'est un concept générique qui est utilisé aussi bien en entité-association qu'en relationnel pour exprimer les propriétés intrinsèques des données.

Définition: étant donné un attribut, ou un ensemble d'attributs, A, d'un TE (ou TA), et B un attribut du même TE (ou TA), il y a dépendance A vers B, notée *A flèche B*, si dans la population du TE (ou TA) toutes les occurrences qui ont même valeur pour A ont toujours même valeur pour B. On dit que B dépend de A, ou que A détermine B. A est la source de dépendance, B la cible.



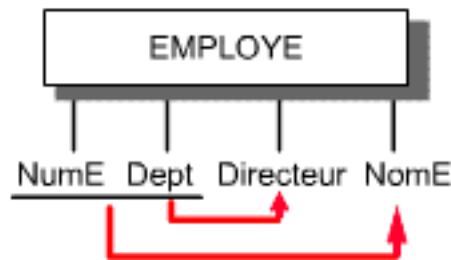
Dans l'exemple, les attributs Nom, Prénom et Adresse du TE PERSONNE dépendent de NumP. La connaissance des dépendances permet de vérifier si le schéma élaboré traduit correctement la réalité de l'application à décrire. Quelques règles permettent de corriger ou de valider le schéma.

1.3.2. Validation des attributs d'un TE

Règle 1: Dans un TE (TA) valide, tous les attributs directs (simples et complexes) dépendent de chaque identifiant entier du TE (TA). On dit aussi que l'identifiant d'un TE (ou TA) détermine tous les autres attributs du TE (TA). Sinon le TE (TA) est incorrectement défini.

Dans l'exemple, tous les attributs dépendent de l'identifiant entier NumP. Le TE PERSONNE est donc correctement défini.

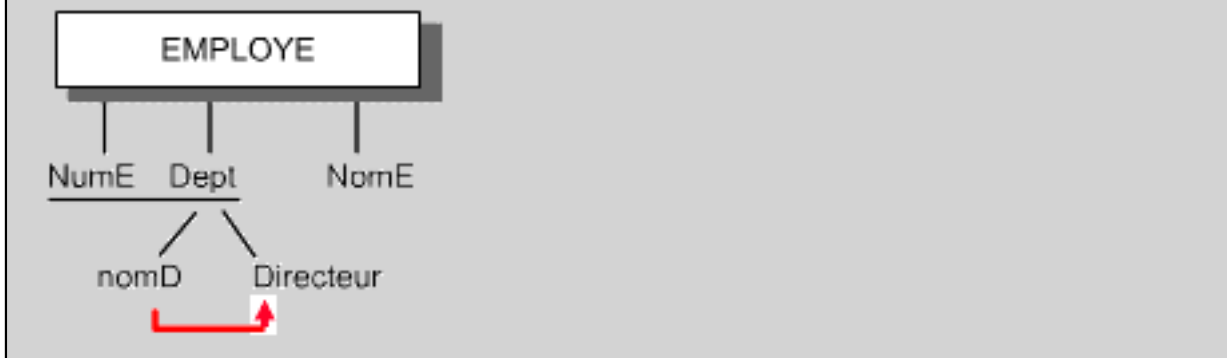
EXERCICE



Indiquez si le schéma ci-dessous est correctement défini et motivez votre affirmation. S'il ne l'est pas, proposez une solution adéquate.

SOLUTION

Non, le schéma n'est pas correct. L'identifiant de ce TE est le couple NumE et Dept (on suppose ici que les numéros sont donnés par les départements suivant une numérotation qui leur est propre, telle que deux employés différents de départements différents peuvent recevoir le même numéro). Le schéma ne respecte pas la règle 1 puisque l'attribut Directeur ne dépend que d'une partie de l'identifiant et non de l'identifiant entier (NumE, Dept). On décompose les attributs du TE comme ci-dessous. Ainsi la règle 1 est vérifiée puisque NomE dépend bien du couple identifiant et Directeur ne dépend bien que de NomD.



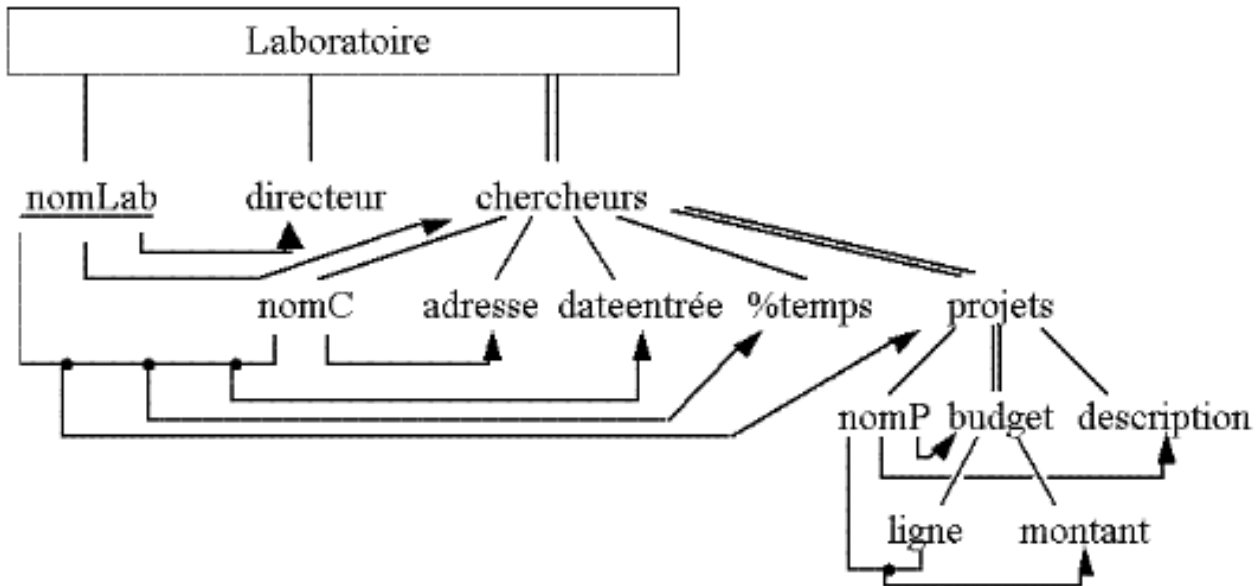
Règle 2: un attribut direct (du premier niveau) dépend de l'identifiant. Un attribut du ième niveau peut dépendre d'une combinaison d'attributs du même niveau et de niveaux supérieurs contigus.

Les attributs du TE Laboratoire ci-dessous respectent la règle 2 :

- les attributs directs, directeur et chercheurs, dépendent de l'identifiant, nomLab;
- l'attribut du 2ème niveau, adresse, dépend de nomC; ce qui signifie que l'adresse du chercheur ne dépend que du chercheur et pas du laboratoire; si le même chercheur (nomC) apparaît dans deux occurrences de Laboratoire, il y apparaîtra avec la même adresse;

Modélisation conceptuelle des données

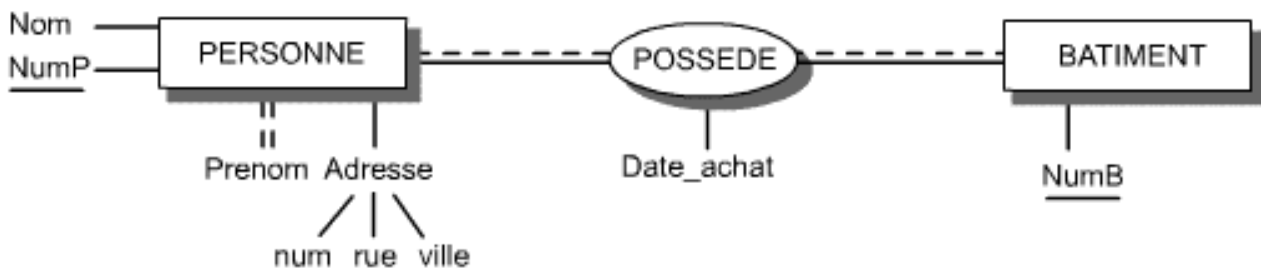
- les attributs du 2ème niveau, data-entrée, %temps et projets, dépendent de (nomC, nomLab); ce qui signifie que si un chercheur travaille dans deux laboratoires (par exemple à mi-temps), il peut y être entré à des dates différentes, travailler sur des projets différents ...
- l'attribut du 4ème niveau, montant, dépend de (nomP, ligne), ce qui signifie que le montant alloué à chaque ligne (matériels, fonctionnement, ...) dépend du projet et de la ligne.



1.3.3. Validation des attributs d'un TA

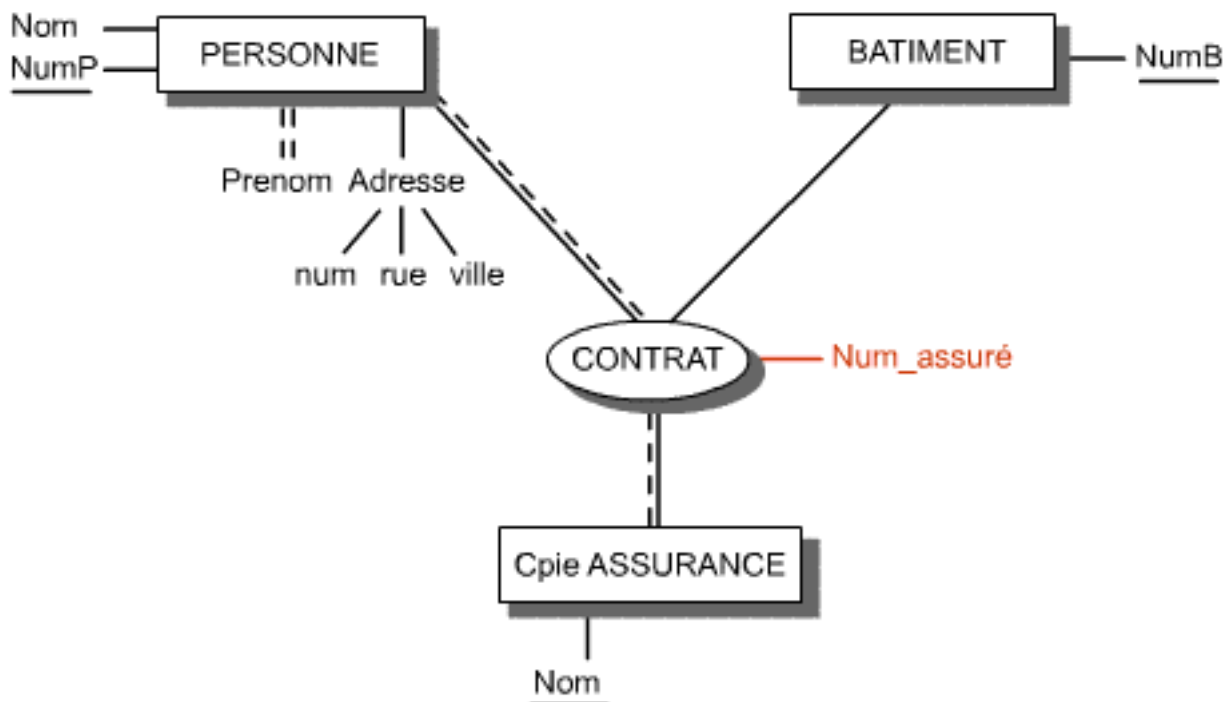
Règle 3: La règle 1 est appliquée ici aux TA. Les attributs directs du TA dépendent de l'identifiant entier, qui dans un TA est composé de tous les identifiants des TE liés au TA.

Dans cet exemple, le schéma est valide puisque la règle 3 est respectée: l'attribut Date_achat du TA POSSEDE dépend à la fois de l'identifiant NumB du TE BATIMENT et de l'identifiant NumP du TE PERSONNE.



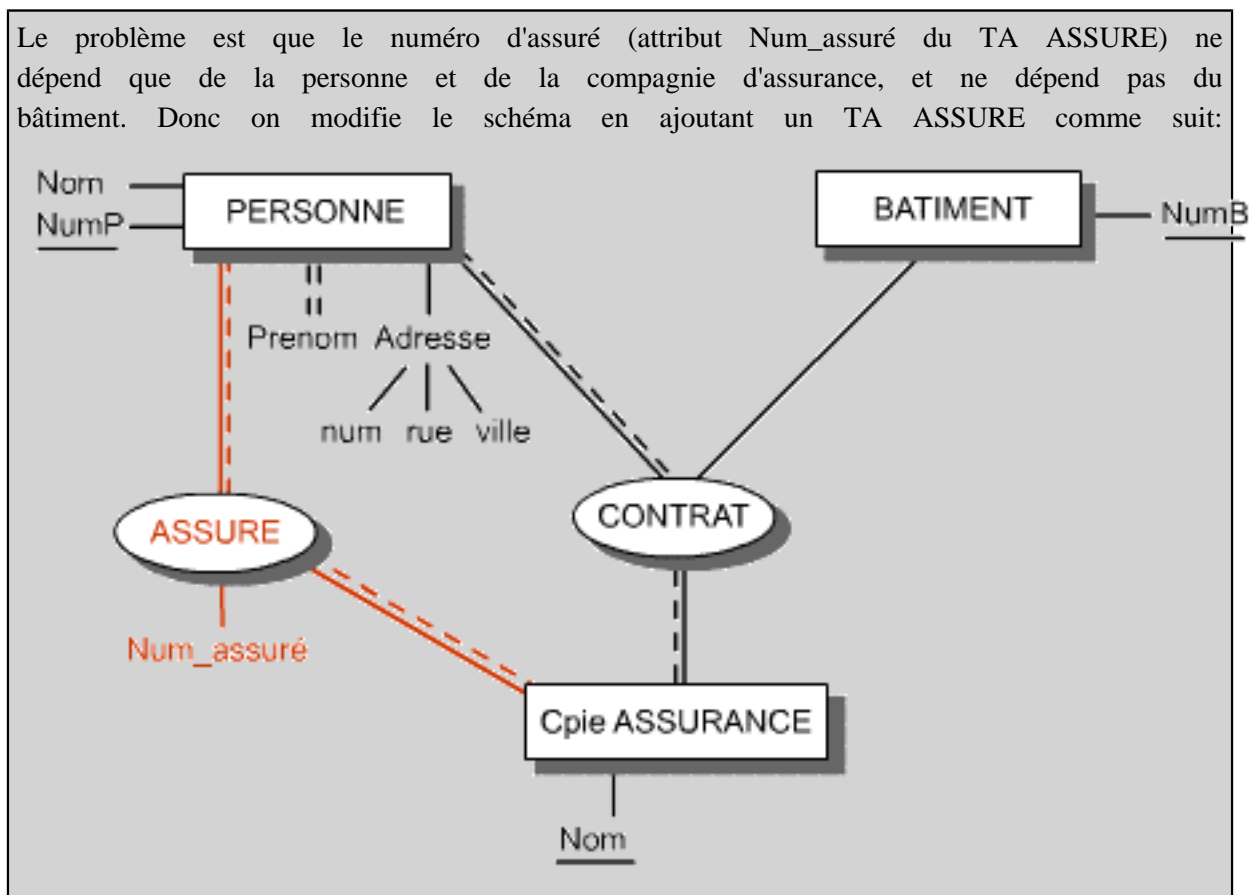
EXERCICE

Dans le schéma ci-dessous on souhaite ajouter un numéro d'assuré qu'une compagnie d'assurance assigne à chacun de ses clients. Indiquez si la solution choisie est correcte, et motivez votre affirmation. Si elle ne l'est pas, proposez une solution adéquate.



SOLUTION

Le problème est que le numéro d'assuré (attribut Num_assuré du TA ASSURE) ne dépend que de la personne et de la compagnie d'assurance, et ne dépend pas du bâtiment. Donc on modifie le schéma en ajoutant un TA ASSURE comme suit:



1.3.4. Validation d'un TA ternaire

Règle 4: Si un TA comporte l'une de ces dépendances sans les autres, il faut le décomposer.

1.3.5. Elimination des TA redondants

Règle 5: Un TA est redondant si les associations correspondantes peuvent être établies sans ambiguïté par composition des associations d'autres TA.

Par exemple dans notre schéma EA, nous aurions pu rajouter un TA entre les TE APPARTEMENT et Cpie ASSURANCE. Cependant cette information peut être retrouvée par composition des TA CONTIENT et CONTRAT : Un appartement est dans un et un seul immeuble, qui est un bâtiment assuré par une compagnie.

1.3.6. Transformation des attributs traduisant une association

Règle 6: Si l'on trouve dans un TE un attribut qui est identifiant d'un autre TE, c'est que cet attribut exprime un lien entre les TE. La règle de représentation par un TA n'a pas été respectée. Il convient donc de corriger le schéma: le lien doit être explicitement décrit comme un TA entre les deux TE et l'attribut doit être supprimé du TE.

1.4. Modélisation conceptuelle de données spatiales et MCD spatiaux

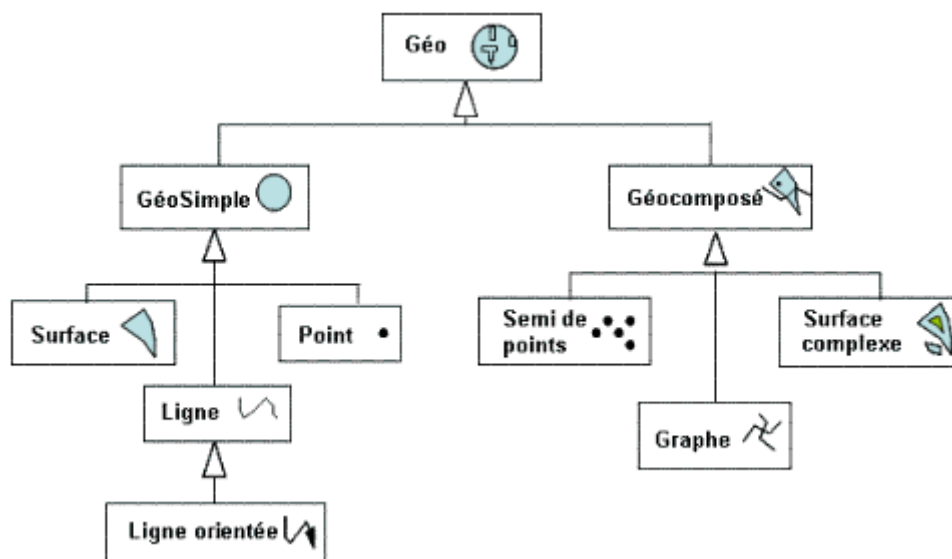
Du fait de la complexité des données spatiales, leur modélisation nécessite des concepts spécifiques. Cette unité expose les problèmes particuliers à la modélisation des données spatiales, présente les concepts dédiés: types abstraits de données (TAD) spatiaux (point, ligne surface), relations topologiques, agrégation, etc. et cite différents modèles conceptuels de données spatiales.

Temps estimatif: 60 minutes

1.4.1. Concepts de modélisation spatiale

TAD spatiaux

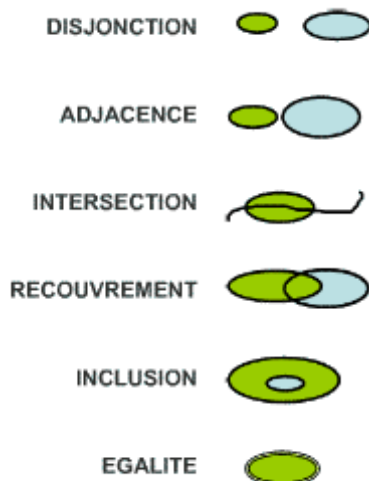
Nous avons vu que l'appréhension de l'espace conduit à choisir le mode de représentation des objets que l'on perçoit en unités d'observation ponctuelles, linéaires ou surfaciques, simples ou complexes. Les MCD issus du monde des bases de données classiques ne nous permettent pas de préciser la spatialité des objets. Depuis plusieurs années, des chercheurs développent des concepts qui étendent les modèles en permettant de prendre en compte la spatialité des objets lors de la modélisation. Notamment ils intègrent des types d'objets spécifiques appelés TAD spatiaux. Ainsi en construisant le schéma d'une application, nous pourrions modéliser l'appartenance des bâtiments à une classe surfacique par exemple ; Ou percevoir les villes comme un objet géographique complexe composé d'objets linéaires tels que les rues, de surfaciques tels que les bâtiments ou les parcs et de points tels que les stations de métro. La figure présente les différents TAD proposés par le modèle MADS (modèle ER étendu aux données spatiales).



Hiérarchie des types abstraits de données spatiales du modèle MADS

Relations topologiques

De la même manière, des liens particuliers tels que les relations topologiques entre classes d'objets spatiaux peuvent être sur le schéma. Ainsi certains introduisent des pictogrammes topologiques (voir la figure ci-dessous) à placer sur les liens. On pourra ainsi préciser par exemple la relation d'inclusion de la géométrie des lots cadastraux dans les communes, comme dans l'exemple de schéma MADS plus loin.



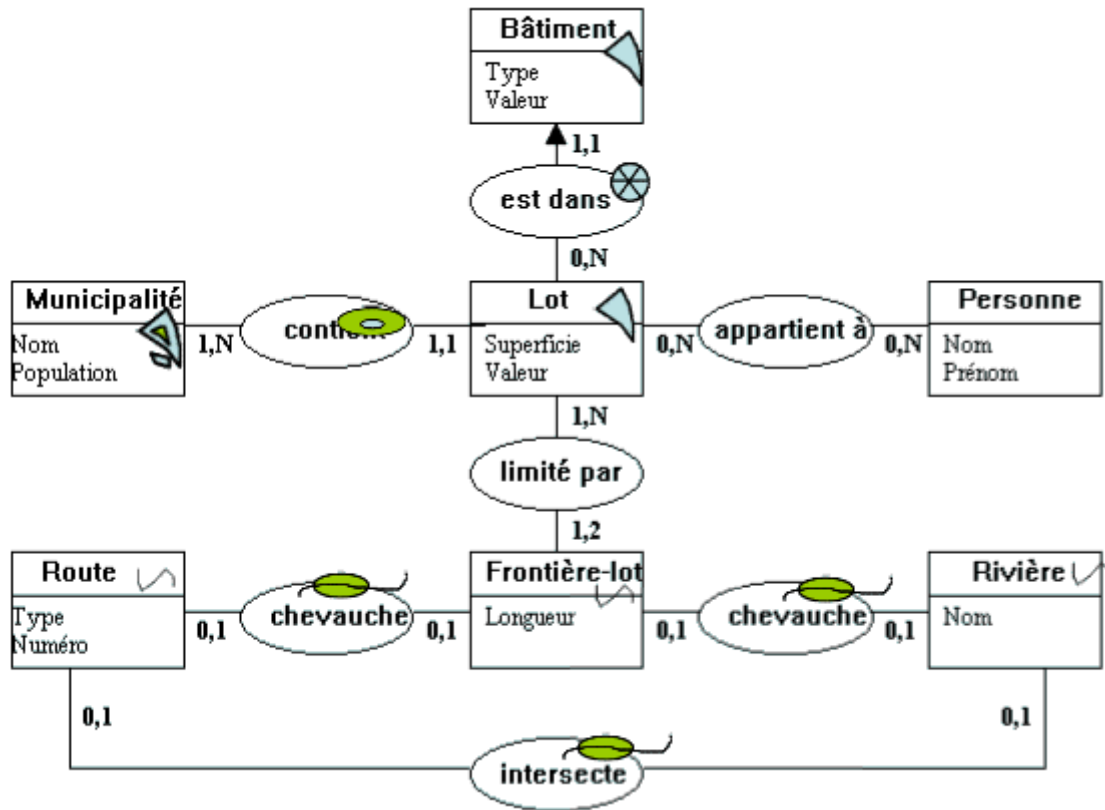
Pictogrammes des relations topologiques (modèle MADS)

Agrégation

Il existe également d'autres types de liens spatiaux particuliers: l'agrégation ou la composition, représenté par un lien fléché et un pictogramme sur le lien. Dans la figure ci-dessous, ce lien permet de préciser qu'un lotissement est composé d'un ensemble de bâtiments. En d'autres termes, la géométrie du lotissement correspond à l'agrégation des géométries des bâtiments qui le composent.

Exemple de schéma conceptuel

L'exemple proposé concerne la modélisation des municipalités régionales de comtés canadiens. Les municipalités sont représentées par des polygones complexes (zone composée de plusieurs parcelles ou comportant des trous) aux parties parfois disjointes (par exemple, la municipalité de Boucherville est composée de plusieurs îles en plus du territoire principal). Les lots cadastraux sont représentés par des polygones simples. Les rivières sont représentées par des chaînes complètes (ligne brisée comportant certains attributs topologiques). Les routes sont également représentées par des chaînes complètes. Le relief est représenté par une couche matricielle où chaque élément (ou "pixel") est une valeur d'élévation en mètres.



Exemple de schéma MADS

1.4.2. Pourquoi utiliser les MCD spatiaux et les concepts qu'ils proposent?

L'intérêt majeur réside dans le fait qu'un certain nombre de contraintes d'intégrité de la base de données sont directement liées à la définition des concepts proposés. Par exemple, la définition de la relation d'agrégation inclut la contrainte selon laquelle les géométries des composants sont inclus dans la géométrie du composés. Ainsi dans le schéma ci-dessus, ce la permet la vérification de l'inclusion des géométries lors de l'instanciation d'un lien particulier de bâtiments à un lotissement.

Couramment dans un SIG, les relations topologiques d'intersection, d'inclusion, d'overlay, etc. sont générées *on the fly* lors de requêtes. Il peut s'avérer intéressant de stocker le résultat dans une table lorsque les requêtes sont fréquentes de manière à éviter à l'utilisateur des temps de calculs qui peuvent parfois être longs et coûteux. De plus cela permet de mettre clairement en évidence l'importance de ce type d'information dans un schéma.

1.4.3. MCD spatiaux (MADS, Perceptory)

Il existe différents MCD spatiaux. Nous citerons les 2 principaux actuels: Perceptory et MADS. Ils sont basés sur des MCD provenant du monde des bases de données, respectivement UML et ER, et ont été étendus aux concepts spatiaux.

Perceptory

Perceptory, développé par la chaire de SIRS de l'université de Laval au Canada, est une extension dans le méta-modèle du formalisme UML, dont les éléments de base sont : le paquetage, les classes, les attributs, les opérations, les associations, la généralisation. Le modèle intègre le concept de TAD spatiaux par la notion de pictogramme appelés PVL (Plug-in for Visual Language). Reportez vous à l'article ou à l'URL suivants pour de plus amples informations sur le modèle Perceptory :

"Bédard, Yvan 2000" (Bédard Yvan 2000)

<http://sirs.scg.ulaval.ca/perceptory/pvl.asp#PVLspatial>

MADS

Le modèle MADS (Modélisation d'Applications à Données Spatiales) est développé à l'EPFL au laboratoire de bases de données. Il est basé sur le modèle ER étendu à la spatialité. Il propose les concepts suivants: TAD spatiaux. Pour en savoir plus, reprenez vous à l'article

"MADS: A Conceptual Model for Spatio-Temporal Applications" (Anonyme)

1.4.4. Exercice

ENONCE DU PROBLEME

Le service du cadastre de la commune de Durdon gère son réseau d'eau potable, en particulier les conduites et les réservoirs qui alimentent les parcelles et les bâtiments. Vous êtes chargé d'établir une base de données qui permettra le stockage de toutes les informations sur ce réseau.

Les bâtiments ont un numéro qui les identifie et une adresse. On souhaite connaître le ou les propriétaires de ces bâtiments dont on stocke le numéro, le nom, la liste des prénoms et l'adresse. On veut aussi conserver la date d'achat d'un bâtiment par son propriétaire. Les bâtiments sont situés sur des parcelles identifiées par leur numéro.

Ces bâtiments peuvent avoir des compteurs d'eau. Ces compteurs ont un diamètre et un nombre de chiffres, un numéro identifiant, une date de pose et une marque.

Les bâtiments peuvent être raccordés à des segments de conduite en eau potable, qui possèdent un numéro, un secteur, une date de pose, un type de matériau, l'état (bon, à réparer, à changer) et un diamètre. Notons que les segments sont numérotés par secteur. Il peut exister un même numéro de segment pour deux secteurs différents (p.ex. on peut trouver un segment numéro 441 dans les secteurs 16 et 28).

On distinguera les bâtiments selon les catégories suivantes : maison individuelle, immeuble, bâtiment industriel). Les maisons individuelles ont la particularité de n'être connectés qu'à un seul segment de conduite, ou pas du tout. Les immeubles et les bâtiments industriels sont connectés à un ou plusieurs segments.

Ces segments d'eau potable peuvent être desservis par un réservoir dont on connaît le numéro, le volume, l'altitude et la date de construction. Ils peuvent également être liés à des bornes hydrantes, qui ont un numéro, un taux de pression et une date de pose.

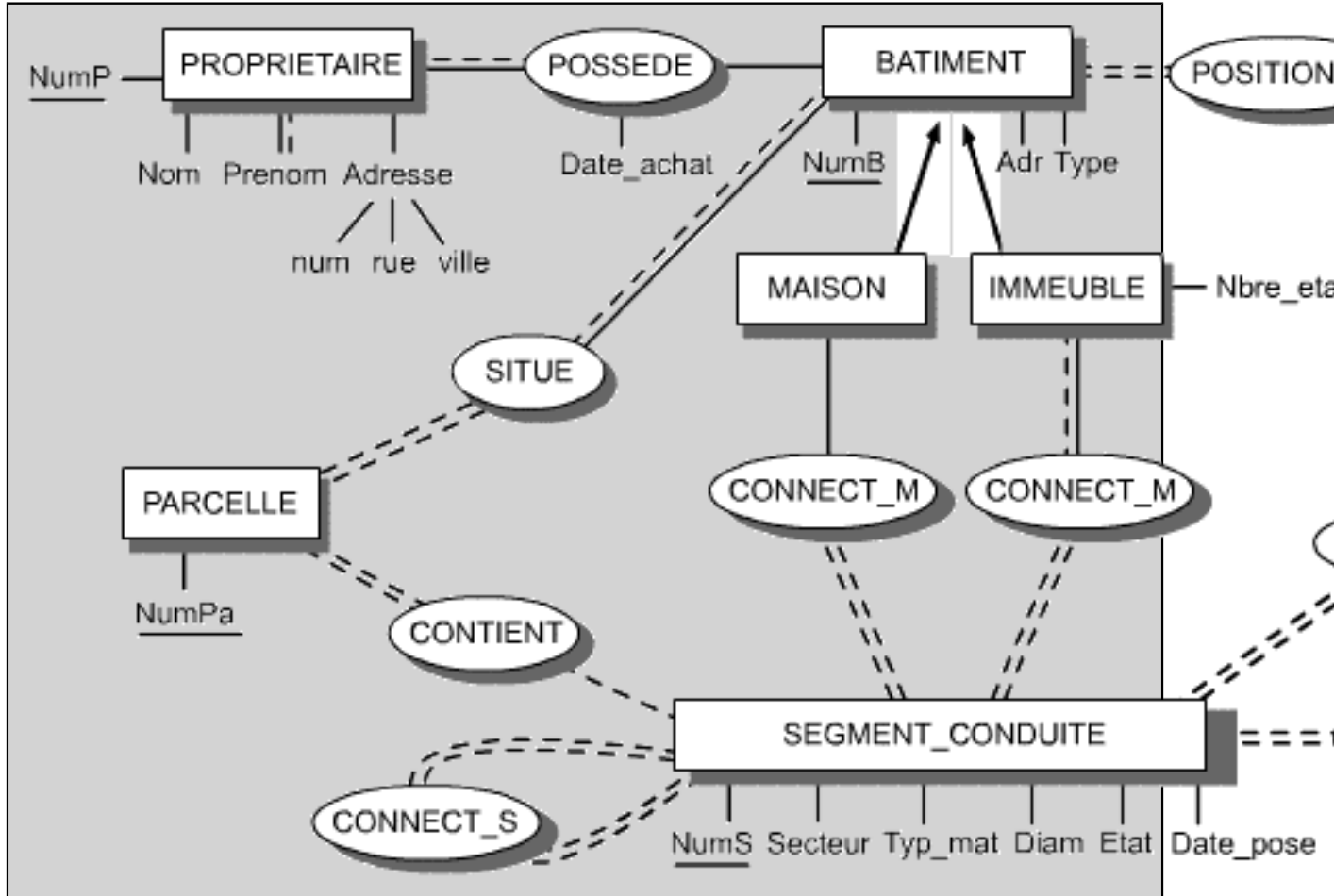
Pour chaque segment d'eau potable, il est indispensable de connaître la pression d'eau. Pour cela on devra être capable de trouver à travers le réseau pour retrouver les segments connectés à un réservoir.

Dans une première phase, le canton n'ayant pas encore d'outil SIRS, une base de données non spatiale sera développée sur un SGBD de type Oracle. Pour cette phase, on demande que chaque entité puisse être identifiée de façon unique. Etablissez un schéma EA pour cette base de données. Il faut décrire le schéma sous forme graphique comme présenté dans le cours. Donnez en quelques mots les définitions en français des TE et

Modélisation conceptuelle des données

des TA de votre schéma (uniquement ceux dont la dénomination paraît ambiguë). Indiquez les cardinalités et les domaines de valeur des attributs (lorsque ceux-ci sont particuliers). Donnez un exemple de contrainte d'intégrité.

SOLUTION



1.5. Bibliographie

- **Anonyme.** *MADS: A Conceptual Model for Spatio-Temporal Applications.*
- **Bédard Yvan,** 2000.