Atelier n°1 Décrire sa base de données à l'aide de modèles conceptuels

Guillaume HARRY (DSI/CNRS)

Marie-Claude QUIDOZ (CEFE/CNRS)





Objectif de l'atelier

- Vous convaincre de l'intérêt de la modélisation pour l'analyse des besoins, la pérennisation, la structuration de votre base de données, ... et plus généralement pour le plan de gestion de données
- Vous avertir des pièges à éviter
- Vous présenter le logiciel Open ModelSphere
- Et en bonus, discuter autour de la création d'une bibliothèque de modèle de donnée au sein du réseau rBDD
- Vous ne serez pas devenu des experts en modélisation MAIS vous serez capable d'en comprendre les enjeux et les pièges à éviter

Plan de l'atelier

- Pourquoi modéliser ?
- Formalisme
 - Entité-Association (Merise)
 - UML (2TUP)
- Quelques éléments de réflexion
 - Entité attribut valeur
 - Normalisé mais jusqu'à où
- Bibliothèque de modèles conceptuels
- Exemple : Métadonnées (cartothèque CEFE)
- TP: Prise en main du logiciel Open ModelSphere

FIL ROUGE DE L'ATELIER

Eléments de contenu d'un DMP

- Qui seront les personnes responsables de chaque étape de la gestion ?
- Quelle sera la politique appliquée aux données : celle des agences de financement, celle de l'institution... ?
- Quels types de données seront collectés ou générés au cours du projet ?
- Comment seront organisés les données et les fichiers ?
- Comment seront décrites les données (documentation et standards de métadonnées) ?
- Comment et où seront stockées, sauvegardées et sécurisées les données ?
- Comment seront partagées les données ? Propriété intellectuelle ? Licence de réutilisation ?
- Comment seront préservées ces données à long terme ?
- Quels seront le coût et les ressources nécessaires à la gestion et au partage des données ?

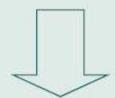
Eléments de contenu d'un DMP

- Qui seront les personnes responsables de chaque étape de la gestion ?
- Quelle sera la politique appliquée aux données : celle des agences de financement, celle de l'institution... ?
- Quels types de données seront collectés ou générés au cours du projet ?
- Comment seront organisés les données et les fichiers ?
- Comment seront décrites les données (documentation et standards de métadonnées) ?
- Comment et où seront stockées, sauvegardées et sécurisées les données ?
- Comment seront partagées les données ? Propriété intellectuelle ? Licence de réutilisation ?
- Comment seront préservées ces données à long terme ?
- Quels seront le coût et les ressources nécessaires à la gestion et au partage des données ?

Rôle des acteurs

Archiviste

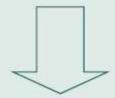




Expertise du cycle de vie + sélection des données et métadonnées associées

Informaticien

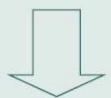




Expertise en BDD, exports et requêtes et mesures techniques à mettre en œuvre

Producteur / Utilisateur





Expertise sur le contenu informationnel des données et leur usage



Merise pour l'archivage



- Existence d'une norme ISAD/G pour décrire les archives
 - Identifie toutes les infos pertinentes à rassembler sur un document pour bien le décrire et le comprendre
 - Utilisée par les archivistes
- Dans le cas des bases de données, <u>qui est un objet plus</u> <u>complexe qu'un simple document</u>, le CINES dans son « Guide méthodologique : Pour l'archivage des bases de données » conseille de compléter la description plus contextuelle de type ISAD/G par l'utilisation de la méthode Merise (cf. §4.2).

Pour commencer: Quelques pistes d'amélioration possibles...



DOCUMENTER

- La structure de la base de données : modèles conceptuel, logique et physique, dictionnaire de données
- Son usage : types de données concernées ? traitements effectués ?



- ✓ Meilleure transmission des connaissances au sein de l'équipe
- ✓ Gain en transparence et en visibilité



- Privilégier les logiciels libres et les formats ouverts
- Assurer une sauvegarde sécurisée



Conseils pour archivage



- S'intéresser aux trois niveaux de description chacun apportant un niveau de représentation nécessaire à la compréhension des données conservées
 - Modèle conceptuel des données (MCD)
 - Modèle logique des données (MLD)
 - Modèle physique des données (MPD)
 - Ne pas oublier le dictionnaire de données
- Documenter les usages, les traitements effectués...
 - Pas de méthodologie conseillée

POURQUOI MODÉLISER?

Pourquoi modéliser ?

- Représenter les différents éléments constitutifs du système d'information
- Description des entités et leur dépendances
- La modélisation est un <u>processus important et</u> <u>indispensable</u> car elle conditionne la structure de la base de données. La structure sera déduite des différents éléments du schéma conceptuel
- Indépendant de la réalisation

Point important à prendre en compte

 Un programme est censé résoudre un problème donné

 Une base de données a pour objectif de répondre à un ensemble de problèmes qui sont en partie inconnus au moment de la création de la base

Formalisme et méthode

- Pourquoi suivre un formalisme ?
 - Utiliser un langage connu et commun
 - Deux « formalismes » très utilisés dans le domaine des bases de données
 - Entité-Association
 - UMI
- Pourquoi suivre une méthode d'analyse
 - Avoir une méthodologie!
 - Deux méthodes présentées
 - MERISE : Entité-Association
 - Two Track Unified Process (2TUP): UML

Quel formalisme choisir?

- Peu de différences si on veut modéliser la structure d'une base de données.
 - Entité pour EA classe pour UML
 - Cardinalités indiquées de façon différente
- Des différences importantes si on veut modéliser les acteurs, les interactions, ...
 - UML offre plus de possibilités

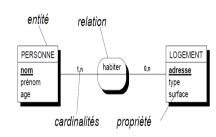
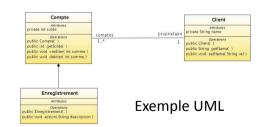


Figure 7.2 : Les concepts du formalisme entité - relation

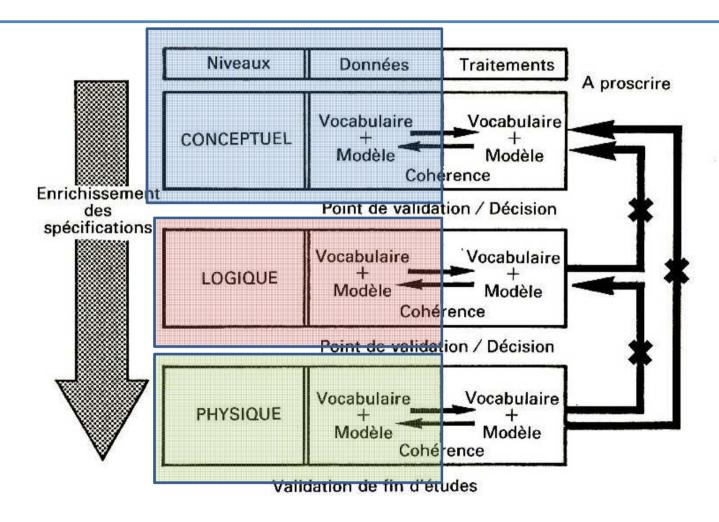


FORMALISME ENTITE-ASSOCIATION MÉTHODE MERISE

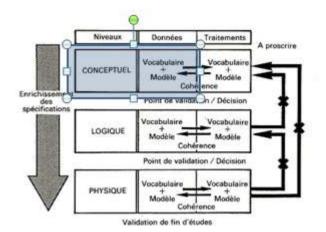
Merise

- Méthode dans le domaine des bases de données
- Résultat des travaux menés par Hubert Tardieu dans les années 1970 et qui s'inséraient dans le cadre d'une réflexion internationale, autour notamment du modèle relationnel d'Edgar Frank Codd
- Méthode française complète
 - Qui couvre de l'analyse à la réalisation
 - Qui traite séparément des données et des traitements

Merise



La méthode MERISE préconise d'analyser séparément données et traitements, à chaque niveau.



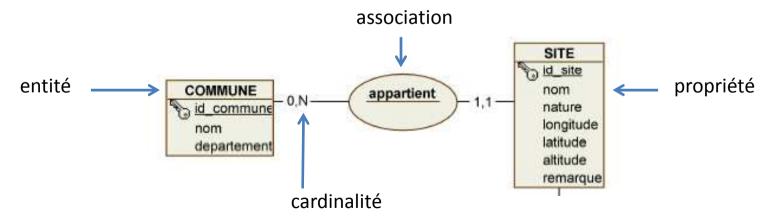
La méthode MERISE préconise d'analyser séparément données et traitements, à chaque niveau.

MODÉLISATION AU NIVEAU CONCEPTUEL

Modèle conceptuel de données (MCD)

 Représentation graphique et structurée des informations mémorisées dans un système d'information

4 notions principales :



Dictionnaire de données

- Etape très fortement conseillée
- Recense l'ensemble des données élémentaires se rapportant strictement au domaine étudié
- Revient à extraire des <u>documents</u> en notre possession et des <u>entretiens</u> l'ensemble des informations strictement indispensables à la gestion du domaine d'activité à informatiser

Formalisme du dictionnaire

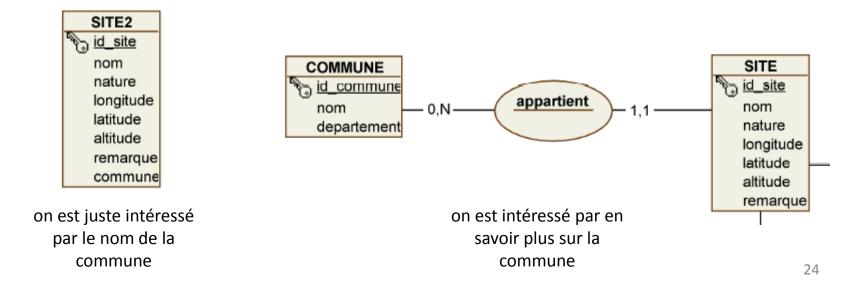
- Formalisme « minimaliste »
 - Abréviation du nom de l'attribut
 - Description précise de l'attribut et de son rôle
 - Type de donnée de l'attribut : entier, réel, texte, date...
 - Liste des contraintes sur les valeurs possibles de l'attribut
- Des ajouts intéressants
 - Nom de l'ensemble cohérent (Entité) auquel l'attribut appartient
 - Domaine des valeurs des attributs

— ...

Entité	Nom	Description	Туре	Contrainte	Domaine
PERSONNE	Nom	Nom de la personne qui a fait le relevé	Texte	Obligatoire	
OBSERVATION	Espèce	Nom de l'espèce observée	Texte	Obligatoire non nulle	Liste ouverte
SITE	Département	Nom du département dans lequel a été fait le relevé	Texte	Facultatif	Liste fermée

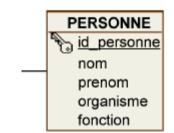
Entité

- Ensemble d'objets de même nature, concrets ou abstraits perçus d'intérêt dans le discours
- Choix du concepteur en fonction de l'intérêt que présente cette entité dans son système d'information



Choix de l'identifiant

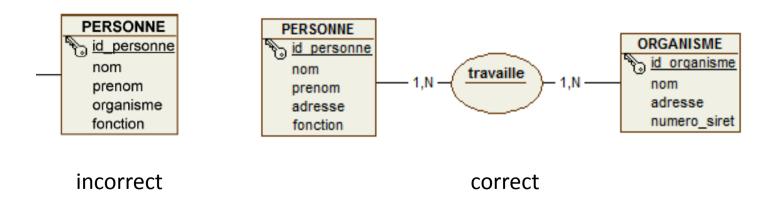
- Chaque entité doit être doté d'un identifiant
 - Propriété simple : nom
 - Propriété composée : nom + prénom
 - Propriété artificielle : id_personne
 - Possibilité d'avoir plusieurs clefs candidates



- Avantages / inconvénients
 - Une propriété artificielle est toujours <u>unique</u> vu du modèle mais elle ne garantit pas que la personne soit unique
 - Une propriété composée nécessite que chaque composant soit défini (c'est-à dire <u>non nulle</u>)
- Remarque : un identifiant doit être stable

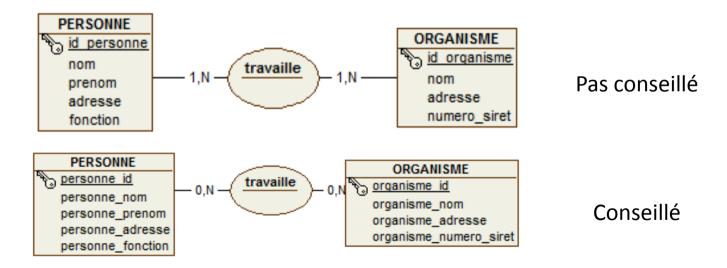
Propriété

- Une entité est décrite par une liste de propriétés
- A toute occurrence de l'entité type, il ne peut y avoir, dans la mémoire du système d'information, au plus qu'une valeur de la propriété
- Ex: une personne travaille dans deux organismes



Choix du nom de la propriété

- Chaque propriété doit figurer une seule fois sur le modèle conceptuel (conseil)
- Les mots réservés sont à proscrire



Association

- Liaison entre deux entités qui a une signification propre au système d'information
- Traduit une partie des règles de gestion qui n'ont pas été satisfaites par la simple définition des entités
- Certaines associations peuvent être porteuses de propriété



Ajout de la propriété « période » pour conserver l'historique des employeurs

Choix du nom de l'association

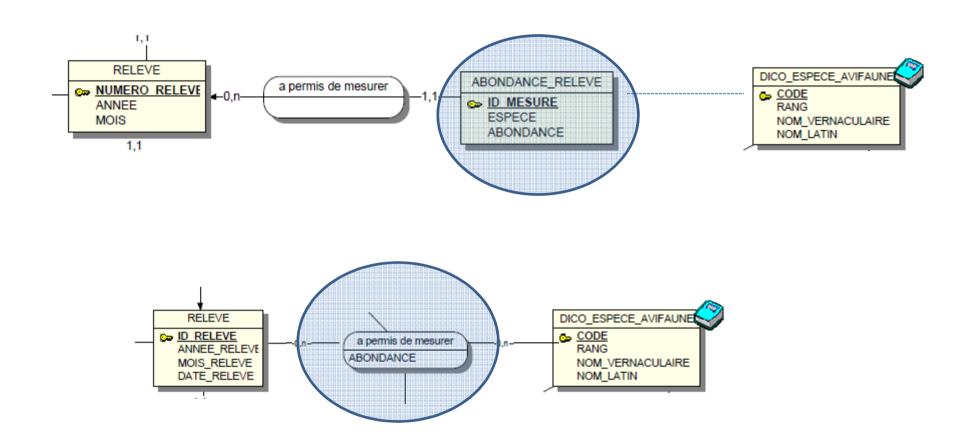
- Phase difficile qui ne doit pas apporter d'ambiguïtés
- Plein de conseils (style utiliser des noms de verbes plutôt statique, à l'infinitif, à la limite sous forme active ou passive pour orienter le sens de lecture) mais c'est plutôt au feeling



1 opération <u>est sous la responsabilité</u> d'1 personne 1 personne <u>a la responsabilité</u> d'1 opération

 Remarque : de nombreuses associations expriment une notion d'appartenance. Il est recommandé de les qualifier en évoquant l'une des entités

Association ? Entité ?



Cardinalité

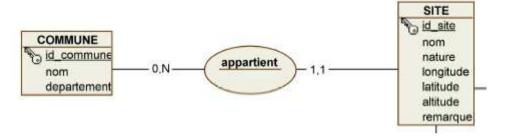
- Participation des occurrences d'une entité type aux occurrences d'une relation type
- Cardinalité minimum et cardinalité maximum
- Cardinalités les plus répandus : 0,n ; 1,n ; 0,1 ; 1,1
 - 0 exprime la participation optionnelle
 - 1 exprime la participation obligatoire
 - n exprime la multiplicité de participation



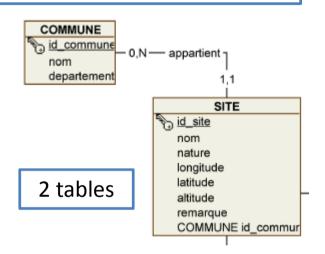
1 opération est sous la responsabilité d'1 ou plusieurs personnes

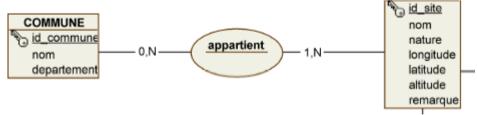
1 personne a la responsabilité d'1 ou plusieurs opérations

Conséquences des cardinalités



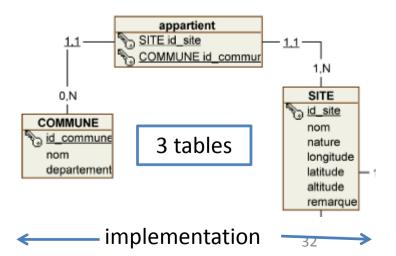
- 1 site est situé sur 1 seule commune
- 1 commune comprend 0 ou n sites





- 1 site est situé sur 1 ou + de communes
- 1 commune comprend 0 ou n sites

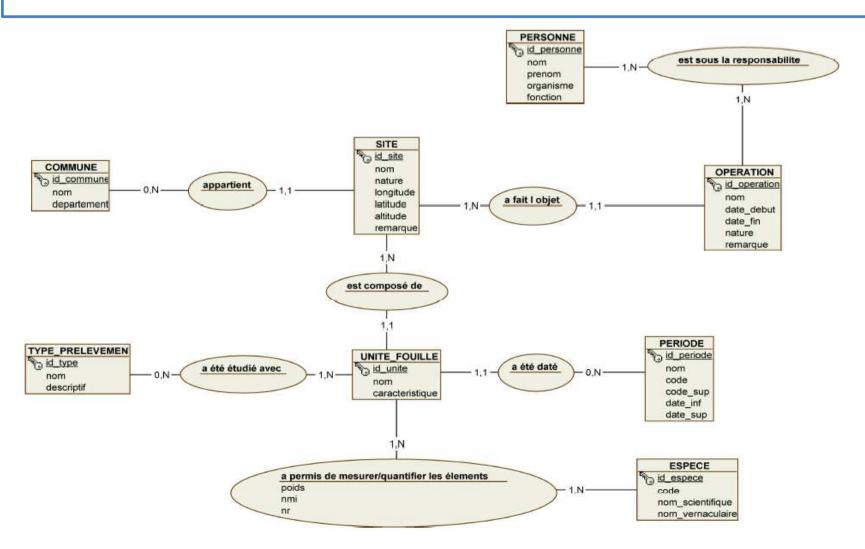




En conclusion (de cette partie)

- 4 notions mais nombreuses combinaisons possibles
- Indépendant du modèle de données (relationnel, objet, ...) et du SGBD choisis
- Pour le valider, le présenter à quelqu'un extérieur
- Elément pour assurer la pérennisation des données
- Remarque : rester modeste

Exemple de MCD



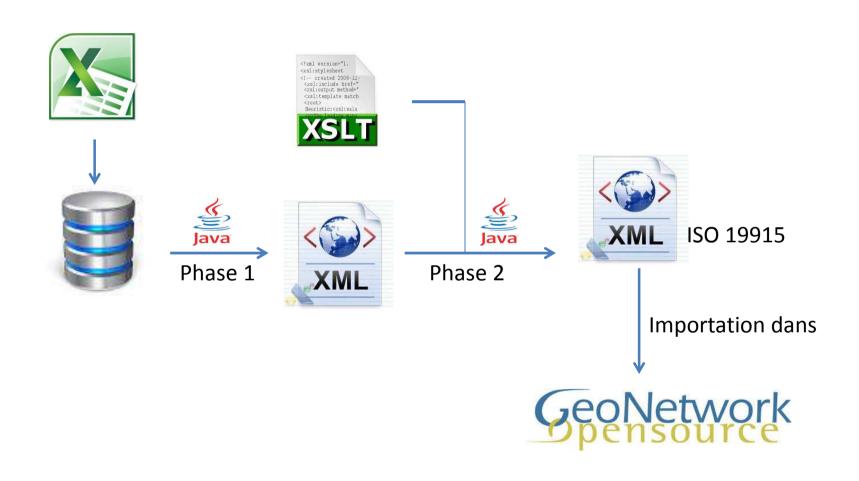
Remarque : la vie des données

- Une donnée peut avoir une pertinence qui s'étend sur une période de quelques millisecondes à la pérennité absolue
- Les SGBD se renouvellent souvent plus vite que les données qu'elles contiennent
- Pas de lien à faire entre la donnée et le SGBD qui la contient
- Travailler uniquement au niveau du modèle logique de données : c'est une fausse bonne idée

FORMALISME UML MÉTHODE 2TUP

METADONNEE CARTOTHÈQUE CEFE

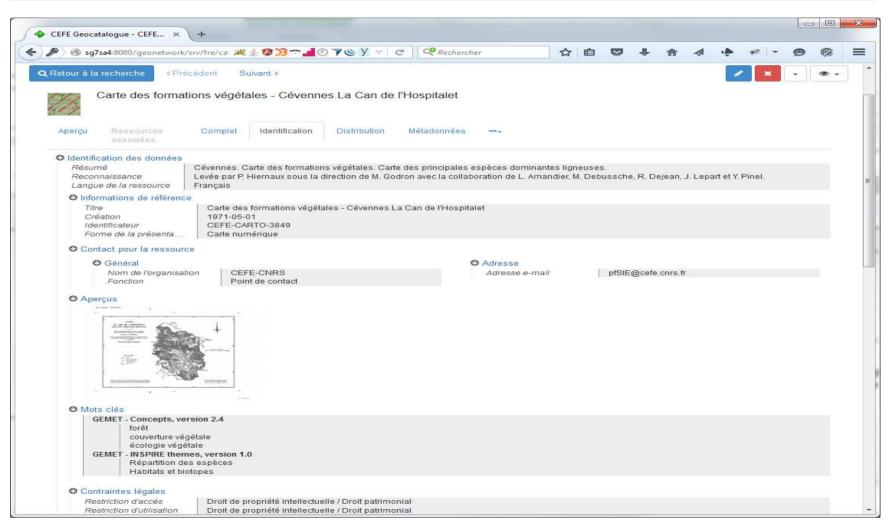
Méthodologie suivie



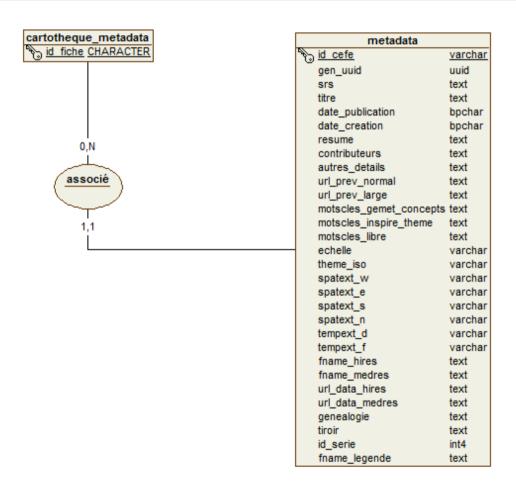
Fiche 3849 sous tableur

attribut	définition	valeur	
id_fiche	Identifiant d'origine	3849	
gen_uuid	Universally Unique Identifier de la métadonnée	7e792d5b-66c0-11e5-accf-3417eb98e42c	
srs	Référentiel de coordonnées	NTF (Paris) (EPSG:4807)	
titre	Titre	Carte des formations végétales - Cévennes.La Can de l'Hospitalet	
date_publication	Date de publication		
date_creation	Date de création	01-05-1971	
id_cefe	Identificateur de ressource unique	CEFE-CARTO-3849	
resume	Résumé	Cévennes. Carte des formations végétales. Carte des principales espèces dominantes ligneuses.	
contributeurs	Contributeurs	Levée par P. Hiernaux sous la direction de M. Godron avec la collaboration de L. Amandier, M. Debussche, R. Dejean, J. Lepart et Y. Pinel.	
autres_details	Référence bibliographique		
url_prev_normal	Donnée prévisualisation, miniature (champ obsolète)		
url_prev_large	Donnée prévisualisation, taille normale	3849_t.jpg	
motscles_gemet_concepts	Mots clefs Thesaurus GEMET	forêt, couverture végétale, écologie végétale	
motscles_inspire_theme	Most-clefs INSPIRE	Répartition des espèces, Habitats et biotopes	
motscles_libre	Mots-clefs libres	Cévennes, Centre d'Etudes Phytosociologiques et Ecologiques	
echelle	Echelle (dénominateur entier)	50000	
theme_iso	Catégorie ISO (parmi la liste)	biota	
spatext_w	Rectangle de délimitation géographique Ouest	3.40	
spatext_e	Rectangle de délimitation géographique Est	3.91	
spatext_s	Rectangle de délimitation géographique Sud	43.98	
spatext_n	Rectangle de délimitation géographique Nord	44.50	
tempext_d	Étendue temporelle : date de début (facultatif)	01-07-1970	
tempext_f	Étendue temporelle : date de fin (facultatif)	31-08-1970	
fname_hires	Donnée image haute résolution	3849_hires.tif	
fname_medres	Donnée image moyenne résolution	3849_medres.jpg	
url_data_hires	URL donnée haute résolution (champ obsolète)		
url_data_medres	URL donnée moyenne résolution (champ obsolète)		
genealogie	Généalogie (origine et processus de création de la donnée)	Les observations ont été faites sur le terrain en Juillet-Août 1970, sur photographies aériennes de la mission IFN au 1/15000 en 1967. Fond topographique IGN au 1/50000. Dessiné par le CEPE -	
tiroir	Tiroir de rangement (champ interne au CEFE)	Cévennes	
id serie	Numéro du parent	9	
fname_legende	Donnée légende, si présente dans un fichier séparé	5163_l.jpg	

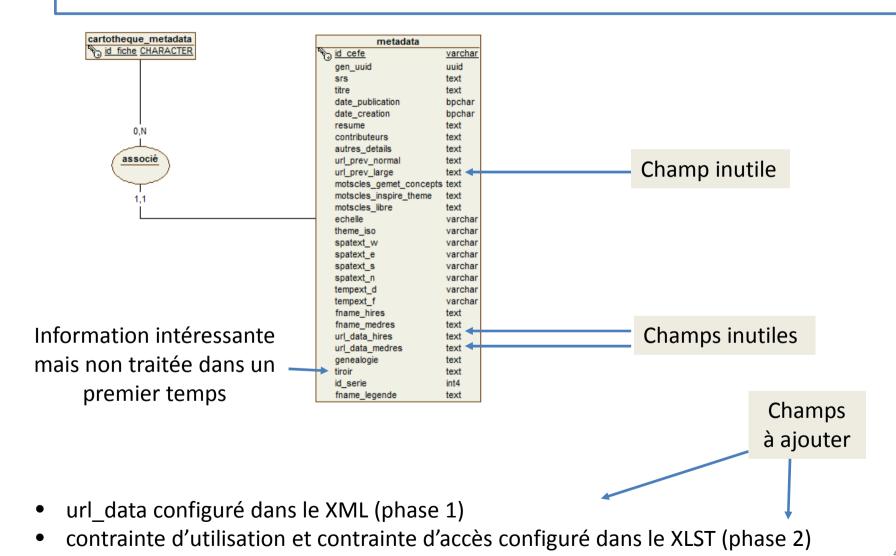
Fiche 3849 sous GeoNetwork



Modèle Conceptuel de Données



Passage en production



Il sera nécessaire de ...

 Modifier la structure de la table en ajoutant et/ou supprimant les attributs concernés

Modifier les requêtes, voir le programme

Et il faudra disposer des droits pour le faire

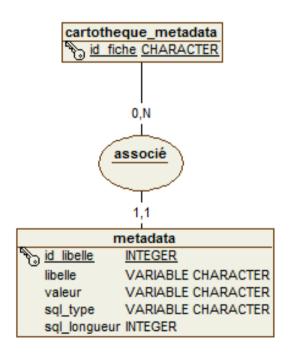
Approche « méta colonne »

- Consiste à ne pas coder les attributs des entités en dur dans le schéma conceptuel, mais à coder le couple nom de l'attribut/valeur
- Permet d'ajouter et/ou supprimer de nouveaux attributs sans modifier le schéma conceptuel
- Remarque : le type des données n'étant pas directement spécifié, il faut créer un attribut qui informe du type

Fiches 3849 & 3847

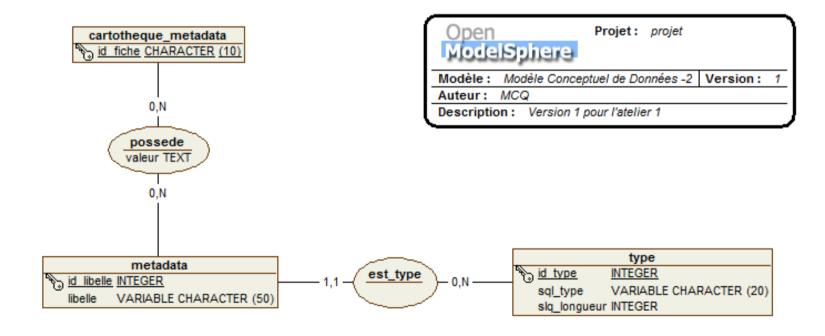
id_fiche	libelle	valeur	type	longueur
3849	gen_uuid	7e792d5b-66c0-11e5-accf-3417eb98e42c	uuid	
3849	srs	NTF (Paris) (EPSG:4807)	varchar	50
3849	titre	Carte des formations végétales - Cévennes. La Can de l'Hospitalet	varchar	200
•••	•••	•••	•••	
3847	gen_uuid	7e792d59-66c0-11e5-accd-3417eb98e42c	uuid	
3847	srs	NTF (Paris) (EPSG:4807)	varchar	50
3847	titre	Carte des formations végétales - Cévennes. Versant septentrional du Mont Lozère	varchar	200

Modèle Conceptuel de Données



- Incorrect : Ne respecte pas la 3^{iéme} forme normale
- Tous les attributs d'une entité doivent dépendre directement de son identifiant et d'aucun autre attribut. Si ce n'est pas le cas, il faut placer l'attribut pathologique dans une entité séparée, mais en association avec la première.

Modèle Conceptuel de Données



Fiches 3849 & 3847 sous PostgreSQL

```
SELECT

cartotheque_metadata.id_fiche,

metadata.libelle,

possede.valeur,

type.sql_type,

type.sql_longueur

FROM

public.cartotheque_metadata,

public.metadata,

public.possede,

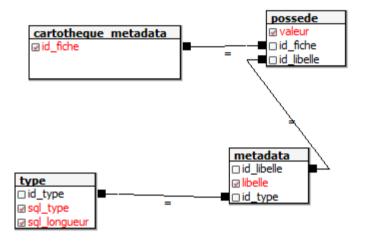
public.type

WHERE

cartotheque_metadata.id_fiche = possede.id_fiche AND

metadata.id_libelle = possede.id_libelle AND

metadata.id_type = type.id_type;
```



id_fiche characte	libelle character varying(50)	valeur text	sql_type character varying(20)	sql_longueur bigint
3849	Universally Unique Identifier de la métadonnée	7e792d5b-66c0-11e5-accf-3417eb98e42c	uuid	
3849	Référentiel de coordonnées	NTF (Paris) (EPSG: 4807)	varying character	50
3849	Titre	Carte des formations végétales - Cévennes.La Can de l'Hospitalet	varying character	200
3847	Universally Unique Identifier de la métadonnée	7e792d59-66c0-11e5-accd-3417eb98e42c	uuid	
3847	Référentiel de coordonnées	NTF (Paris) (EPSG:4807)	varying character	50
3847	Titre	Carte des formations végétales - Cévennes. Versant septentrional du Mont Lozère	varying character	200

Avantages / inconvénients

Avantages

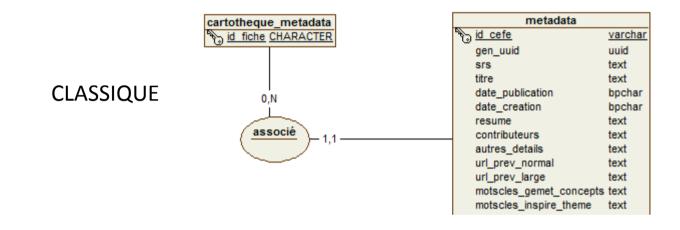
- Plus souple et plus évolutif : ajout indéfiniment d'attributs sans jamais modifier la structure même de la base
- Droit élémentaire : utilisateur

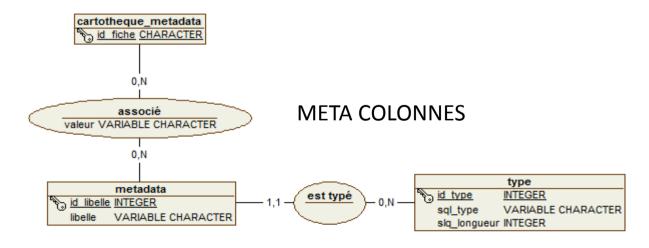
Inconvénients

- Vitesse d'exécution (nombre conséquent de jointure)
- Compréhension générale de la base par un non spécialiste
- Documentation -> Pérennisation ?

QUEL MODÉLISATION CHOISIR?

Entre classique et « méta colonnes »





Quelques éléments de réflexion

- Utilisateurs de la base de donnée
 - Chercheurs
 - Grand public
- Utilisation de la base de donnée
 - Développement logiciel sur mesure
 - Outils courants (R, QGIS, Tableur,)
- Besoin de flexibilité ?
- Pérennisation / documentation ?

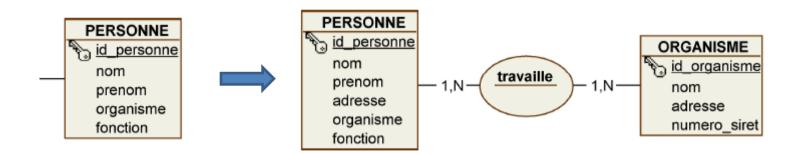
NORMALISATION / DENORMALISATION

Normalisation

- Correspond au processus d'organiser ses données afin de limiter les redondances, divisant une table en plusieurs, et en les reliant entre elles par des clefs primaires et étrangères.
- L'objectif est d'isoler les données afin que l'ajout, l'effacement ou la modification d'un champ puisse se faire sur une seule table, et se propager au reste de la base par le biais des relations.
- Pour ce faire, la normalisation introduit en tout 8 formes normales.

Première forme normale

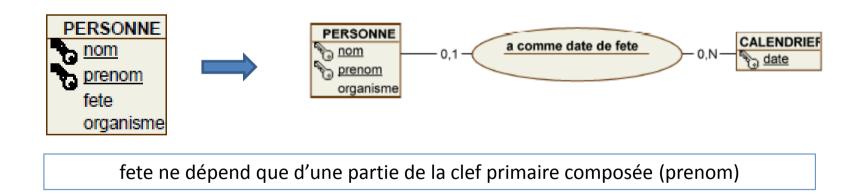
 A un instant donné dans une entité, pour un individu, un attribut ne peut prendre qu'une valeur et non pas, un ensemble ou une liste de valeurs. Si un attribut prend plusieurs valeurs, alors ces valeurs doivent faire l'objet d'une entité supplémentaire, en association avec la première.



Ex : une personne travaille dans deux organismes

Deuxième forme normale

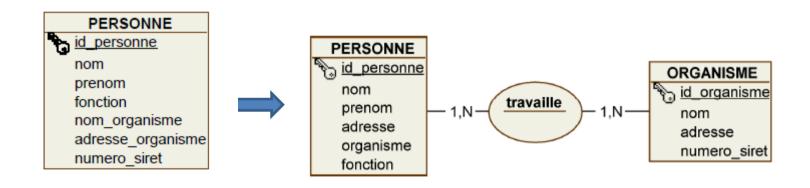
• L'identifiant peut être composé de plusieurs attributs mais les autres attributs de l'entité doivent dépendre de l'identifiant en entier (et non pas une partie de cet identifiant).



• Remarque : cette forme normale peut être oubliée si on n'utilise que des identifiants non composés et de type entier

Troisième forme normale

• Tous les attributs d'une entité doivent dépendre directement de son identifiant et d'aucun autre attribut. Si ce n'est pas le cas, il faut placer l'attribut pathologique dans une entité séparée, mais en association avec la première.



adresse_organisme et numero_siret ne dépendent pas directement de id_personne

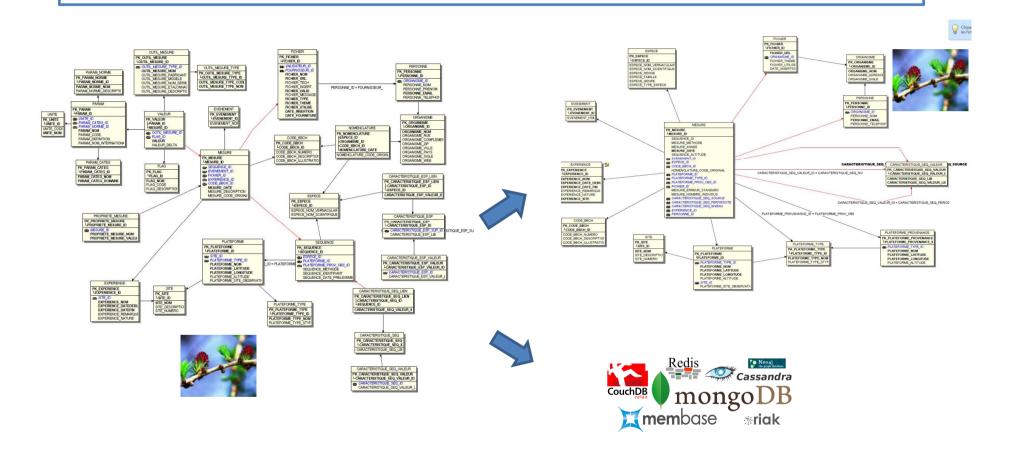
Les autres formes normales

- Les quatrièmes, cinquièmes et sixièmes formes normales évitent principalement la redondance d'information, elles sont plus précises
- En pratique la forme normale 3 est suffisante
- Les projections et les jointures sont coûteuses pour le système, ainsi une trop forte normalisation diminue fortement les performances
- La normalisation d'une base peut aussi rendre complexe les requêtes nécessaires.

Dénormalisation

- Phase implémentation
 - En fonction du temps de réponse
 - Après avoir tout tenté : indexation, partitionnement, ...
- Ajouter des attributs
 - Données calculées
 - Données redondantes
 - Clefs primaires (de taille plus réduite)
 - Clefs étrangères
- Ajouter des déclencheurs / des lignes de programme pour maintenir l'intégrité
- Remarque : la dénormalisation est profitable si vous avez beaucoup de lectures et peu d'écritures

Une meilleure solution?

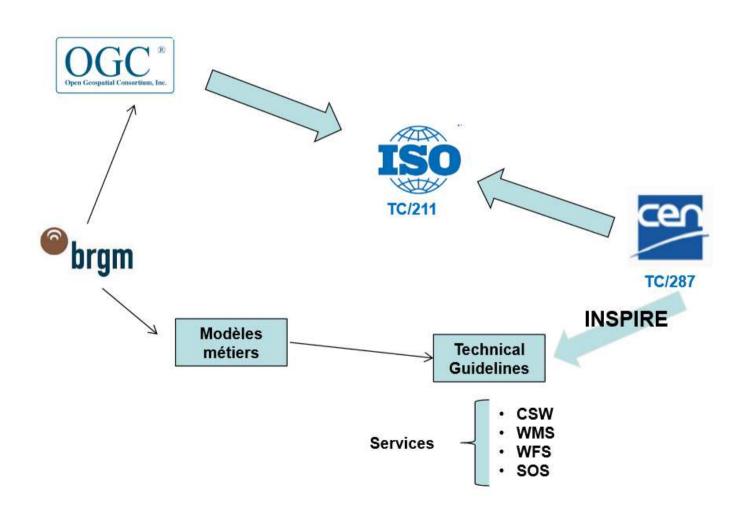


Base de données de production

Base de données de consultation

LES MODÈLES CONCEPTUELS ONT-ILS UN INTÉRÊT À ÊTRE ÉCHANGÉS ? PUBLIÉS ?

Processus autour d'INSPIRE



Data Specification par thème INSPIRE



ANNEX 3



Cadastral parcels (1/2)



Information

Legal Acts - Guidelines

Discussion Forums

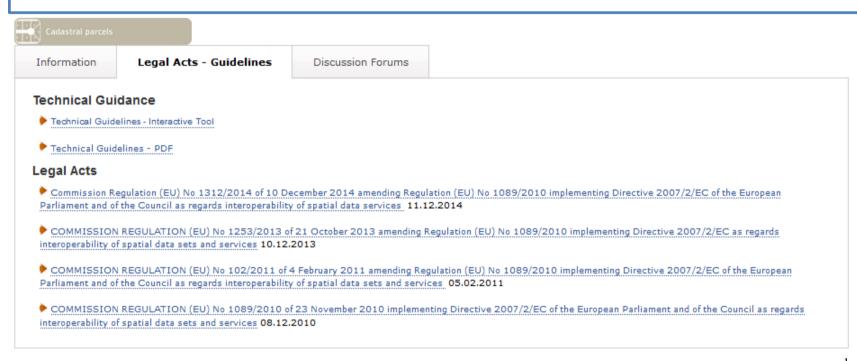
Definition

Areas defined by cadastral registers or equivalent.

Description

The INSPIRE Directive focuses on the geographical part of cadastral data. In the INSPIRE context, cadastral parcels will be mainly used as locators for geo-information in general, including environmental data. As much as possible, in the INSPIRE context, cadastral parcels should be forming a partition of national territory. Cadastral parcel should be considered as a single area of Earth surface, under homogeneous real property rights and unique ownership (adapted from UN ECE 2004 and WG-CPI, 2006). Remark: By unique ownership is meant that the ownership is held by one or several owners for the whole parcel. By homogeneous property rights is meant that rights of ownership, leases and mortgages affect the whole parcel. This does not apply to specific rights as servitudes which may only affect part of the parcel. In the definition given by the INSPIRE directive, or equivalent refers to all public agencies and institutions other than the main traditional/nominal cadastre or land registry, that register parts of the Earth's surface such as special domains, urban cadastres, public lands, which spatially complement the registrations by the main cadastre or land registry. Cadastral parcels are considered in the INSPIRE scope if they are available as vector data. Rights and owners are out of the INSPIRE scope. Buildings, land use, addresses are considered in other INSPIRE themes.

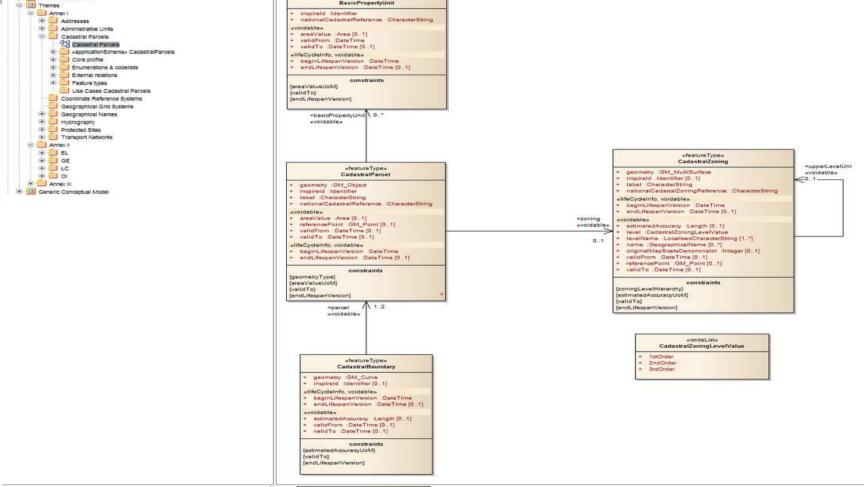
Cadastral parcels (2/2)





Cadastral Parcels (modèle UML)

INSPIRE Consolidated UML Model INSPIRE Consolidated UML Model Foundation Schemas BasicPropertyUnit Themes inspireld Identifier nationalCadastralReference CharacterString Annex I Addresses Administrative Units areaValue :Area (0...1) Cartastral Parcels



Bibliothèque de modèle conceptuel

Projet Generic Model Organism Databas

Projet Generic Model Organism Database

- Collection d'outils open source pour la gestion, la visualisation, le stockage et la diffusion des données génétiques et génomiques
- Chado le modèle relationnel de base de données
- Tripal un frontal web (basé sur Drupal) pour les bases de données
- Galaxy une plateforme web (ressemble à un ETL)

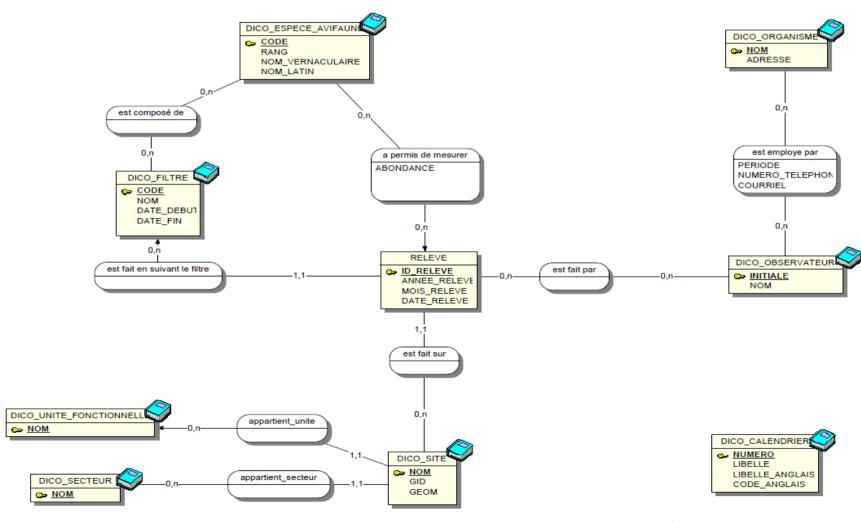
Modèle Chado

- Schéma modulaire capable de représenter un grand nombre de catégories rencontrées en biologie moléculaire : séquence, phénotypes, génotypes, ontologies, phylogénie, ...
- SGBD recommandé : PostgreSQL
- Le + sophistiqué actuellement disponible ?
- Temps d'apprentissage important.

Et en interne dans nos laboratoires

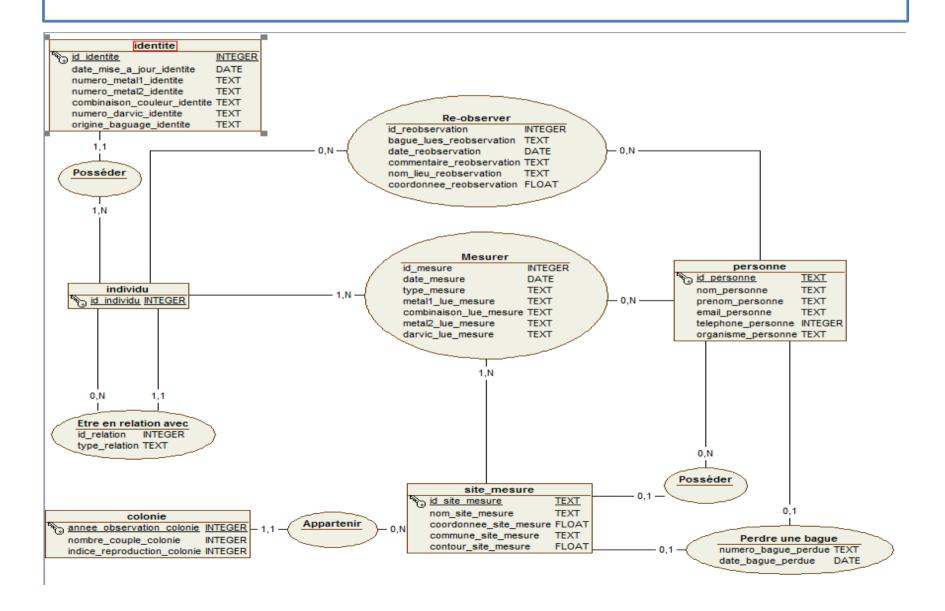
- Des tentatives de modèle « générique »
 - Relevé d'espèces
 - Suivi de population

Relevé d'espèces



Base de données AVIMED (CEFE/CNRS)

Suivi de population



BIBLIOGRAPHIE / WEBOGRAPHIE

Quelques références (1/5)

- Règles de transformation du MCD en MLD
 http://steveostine.free.fr/pdf/Rgles%20de%20transformation%20du%20MCD%20au%20MLD.pdf
- Initiation à la conception de base de données relationnelles avec MERISE
 http://ineumann.developpez.com/tutoriels/merise/initiation-merise/
- Conception d'une base de données http://cyril-gruau.developpez.com/uml/tutoriel/ConceptionBD/
- UML 2 pour les bases de données avec 20 exercices corrigés, Christian Soutou, Eyrolles, 1^{iere} édition
 - http://www.essai.rnu.tn/Ebook/Informatique/uml2pourlesbasesdedonnees.pdf
- UML 2 pour les bases de données, Modélisation Normalisation Génération SQL – Outils, Christian Soutou avec la collaboration de Frédéric Brouard, Eyrolles, 2^{ième} édition
- UML en action De l'analyse des besoins à la conception, Pascal Roques, Franck Vallée, Eyrolles, 4^{ième} édition

Quelques références (2/5)

- Bases de données relationnelles et normalisation : de la première à la sixième forme normale http://fsmrel.developpez.com/basesrelationnelles/normalisation/
- Spécifications de données pour INSPIRE http://inspire.ec.eu/opa.eu/index.cfm/pageid/2

Quelques références (3/5)

• Chado - Getting started

http://gmod.org/wiki/Chado

La technique des méta données (colonnes)

http://sqlpro.developpez.com/cours/modelisation/metadonnees/

Entity-attribute-value model

https://en.wikipedia.org/wiki/Entity%E2%80%93attribute%E2%80%93value model

Quelques références (4/5)

• Guide méthodologique. Pour l'archivage des bases de données

https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/ca323bcc-804a-43e3-822f-5d18fa304dff/GM_archivage_BDD.pdf

- ISAD(G): Norme générale et internationale de description archivistique http://www.ica.org/download.php?id=1688
- Journée de sensibilisation à la sécurisation et à la pérennisation des données (Novembre 2014)

http://rbdd.cnrs.fr/spip.php?article172

Quelques références (5/5)

Logiciel OpenModelSphere

http://www.modelsphere.org/fr/open modelsphere.html

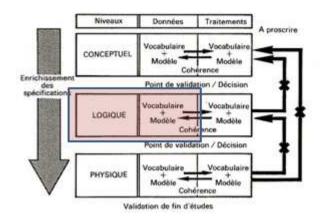
- Open ModelSphere: a free computer-aided software engineering tool
 http://www.marco-savard.com/OpenModelSphere/
- Forum du club des développeurs

http://www.developpez.net/forums/f429/general-developpement/alm/outils/autres/

Open ModelSphere – La communauté

https://www.linkedin.com/grp/home?gid=2503967

FIN



La méthode MERISE préconise d'analyser séparément données et traitements, à chaque niveau.

ANNEXE 1 : MODÉLISATION AU NIVEAU LOGIQUE

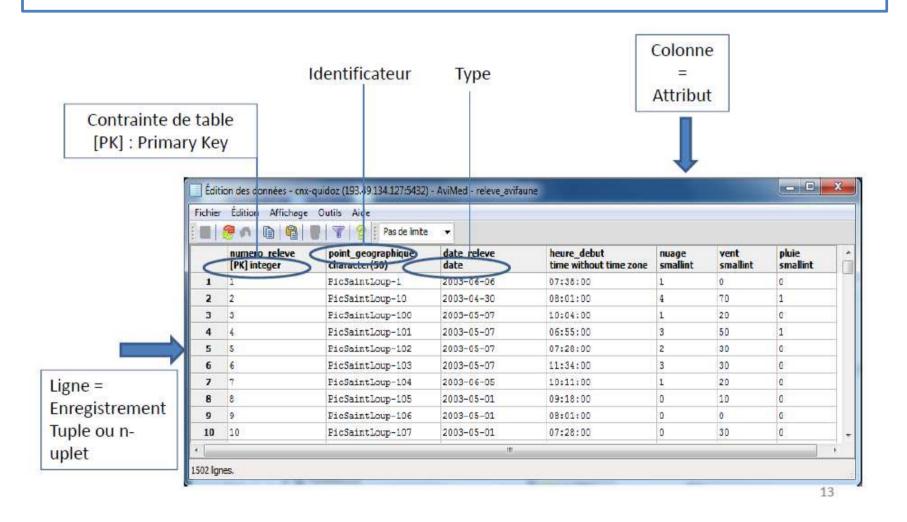
Terme : bases de données

- Ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'exploitation (ajout, mise à jour, recherche de données)
- Organisée selon un modèle de description de données
 - Hiérarchique, réseau, relationnel, objet, relationnel objet
- Relationnel
 - En 1970, Edgar F. Codd, chercheur chez IBM, propose dans une thèse mathématique d'utiliser les informations présentes dans les enregistrements pour assurer les <u>liens</u> entre les informations et de regrouper les enregistrements dans des <u>tables</u>
 - Modèle le plus courant
 - Présentation simple des données sous forme de tables
 - Puissance et simplicité du langage de requête

Modèle logique de données (MLD)

- Représentation (graphique ou non) de la structure de la base de données selon le modèle de description des données retenu (relationnel dans notre cas)
- Traduction du MCD en MLD (ou MRD ou MLR) en appliquant des règles de transformation
- Remarque : MLD est indépendant du SGBD retenu

Concepts du modèle relationnel (1)

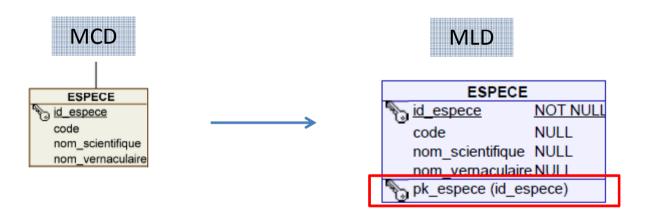


Concepts du modèle relationnel (2)

- Clef primaire sert à identifier une ligne de manière unique
- Clef étrangère permet
 - de gérer les relations entre plusieurs tables
 - de garantir la cohérence des données (intégrité référentielle)

Règle 1: Transformation des entités

Toute entité est transformée en table. Les propriétés de l'entité deviennent les attributs de la table. L'identifiant de l'entité devient la clé primaire de la table.

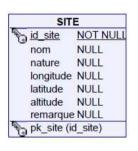


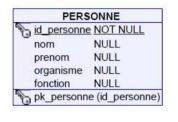
La clef primaire s'appelle pk_espece Elle fait référence à l'attribut « id_espece » de la table « ESPECE »

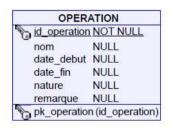
Première étape de la transformation

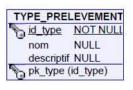
• 8 tables qui correspondent aux 8 entités



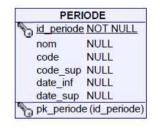


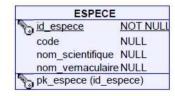






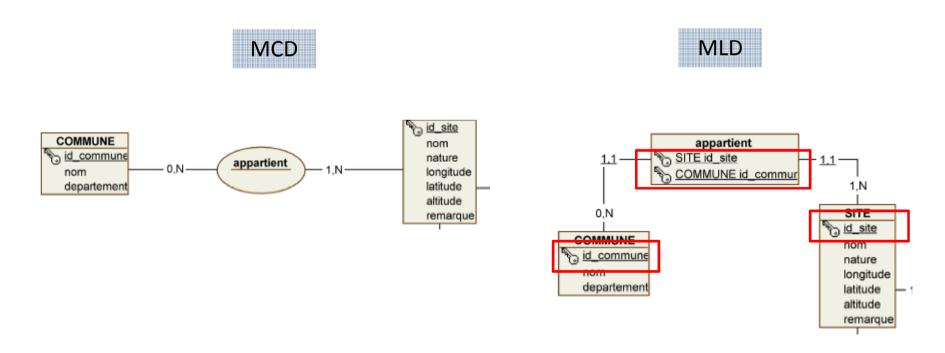






Règle 2: Transformation des relations binaires du type (x,n) - (x,n) avec x = 0 ou 1

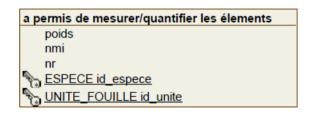
On crée une table supplémentaire ayant comme clé primaire une clé composée des clés primaires des 2 tables. Lorsque la relation contient elle-même des propriétés, celles-ci deviennent attributs de la table supplémentaire. Une propriété de la relation qui est soulignée devra appartenir à la clé primaire composée de la table supplémentaire.



Deuxième étape de la transformation

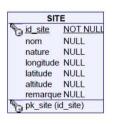
 3 tables en plus qui correspondent aux 3 relations binaires du type (x,n) – (x,n)



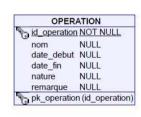






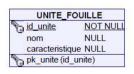


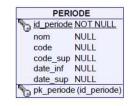






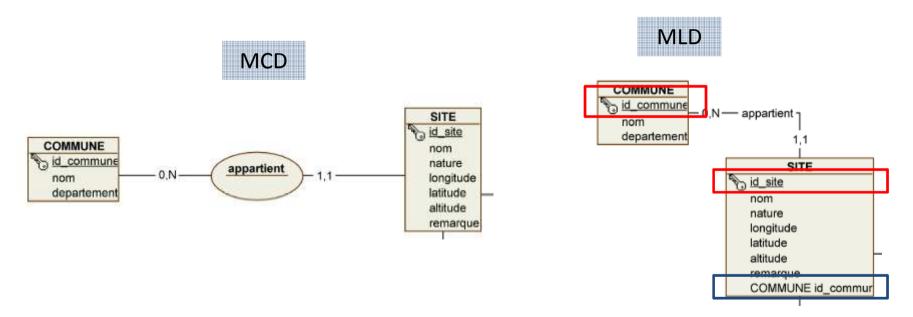






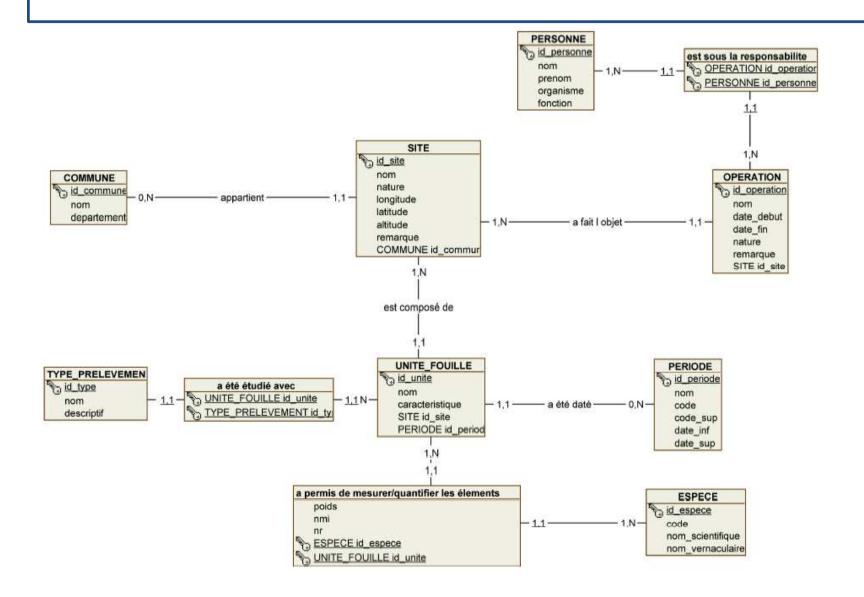
Règle 3: Transformation des relations binaires du type (x,n) - (x,1) avec x = 0 ou 1

Afin de représenter la relation, on duplique la clé primaire de la table basée sur l'entité à cardinalité (x,n) dans la table basée sur l'entité à cardinalité (x,1). Cet attribut est appelé clé étrangère. Les deux tables sont liées par une flèche nommée selon la relation, qui pointe de la table à clé étrangère vers la table qui contient la clé primaire correspondante.



L'attribut « id_commune » qui est clef primaire de la table COMMUNE devient clef étrangère dans la table SITE

Troisième étape de la transformation



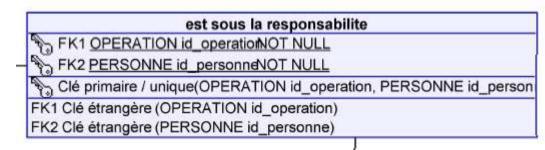
Remarque importante

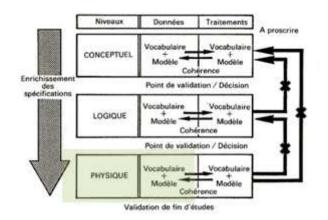
- ✓ Présentation simplifiée des règles de transformation
- ✓ Il faut ajouter la notion de clef étrangère dans la règle 2 (relations binaires du type (x,n) (x,n)) pour représenter la relation entre les deux tables.

⚠ On crée une table supplémentaire ayant comme clé primaire une clé composée des clés primaires des 2 tables. Lorsque la relation contient elle-même des propriétés, celles-ci deviennent attributs de la table supplémentaire. Une propriété de la relation qui est soulignée devra appartenir à la clé primaire composée de la table supplémentaire.



Régle2





La méthode MERISE préconise d'analyser séparément données et traitements, à chaque niveau.

ANNEXE 2 : MODÉLISATION AU NIVEAU PHYSIQUE

Modèle physique de données (MPD)

- Génération du script SQL adapté aux spécificités du SGBD retenu
 - Commandes
 - Types de données
- Cette phase est intéressante uniquement si :
 - Vous utilisez un logiciel de modélisation
 - Toutes les propriétés des entités ont été définies dans la réalisation du modèle conceptuel de données

ANNEXE 3 : METADONNEE CARTOTHÈQUE CEFE

Modèle Conceptuel de Données V2

