

Master 1 Humanités numériques – Bases de données

Séance 2 : Modélisation

J. Darmont – <http://eric.univ-lyon2.fr/jdarmont/>

Préambule

L'objectif de cette séance est de mettre en œuvre des bases de données, sur la base de spécifications, en suivant le processus de modélisation conceptuelle, logique puis physique décrit en cours.

Rappel : modélisation conceptuelle UML



1. Identifier les *classes*.
2. Identifier les *associations* entre classes.
3. Identifier les *attributs* de chaque classe, leur *type* et souligner l'*identifiant*.
4. Identifier les attributs de chaque association et les placer dans une *classe-association*.
5. Évaluer les *multiplicités* des associations.

Rappel : traduction logique relationnelle du modèle conceptuel

1. Chaque *classe* devient une *relation* (table). Les *attributs* de la classe deviennent *attributs* de la relation. L'*identifiant* de la classe devient la *clé primaire* de la relation.
2. Chaque *association 1-N* est prise en compte en incluant la *clé primaire* de la relation dont la multiplicité maximale est 1 comme *clé étrangère* dans l'autre relation.
3. Chaque *association M-N* est prise en compte en créant une *nouvelle relation* (nouvelle table) dont la *clé primaire* est la concaténation des clés primaires des relations participantes. Les attributs de la classe-association, s'ils existent, sont insérés dans cette nouvelle relation en plus de la clé primaire.

Notations : cléPrimaire
cléÉtrangère#

Rappel : traduction physique sous Access du modèle logique

1. La création des *tables* s'effectue via l'onglet « Créer », puis l'icône « Création de table ». Il s'agit de renseigner les noms des *champs* (attributs) et leurs *types*.
2. La *clé primaire* d'une table se définit en sélectionnant la ligne du champ adéquat, puis en cliquant sur l'icône  dans la barre d'outils (ou par un clic droit).
3. La définition des *contraintes de clés étrangères* s'effectue via l'onglet « Outils de base de données » et l'icône « Relations » . Il s'agit à chaque fois de relier un *couple clé primaire-clé étrangère* par une relation (au sens d'Access).

Cas d'étude n° 1 : Ceramom



<https://fr.pinterest.com/pin/98938523042705513/>

La base de données Ceramom¹, développée par le laboratoire Archéologie et Archéométrie² depuis la fin des années 1970, décrit des échantillons de céramiques selon diverses dimensions, la principale étant des analyses chimiques visant à dater et localiser les céramiques.

De manière simplifiée, un échantillon est caractérisé par un identifiant unique, une date de création, une datation (plus ou moins certaine), une forme, une pâte et sa décoration.

Un échantillon est physiquement stocké au laboratoire ArAr dans un lieu de stockage caractérisé par un identifiant unique, un numéro de salle, un numéro d'armoire, un numéro d'étagère et un numéro de boîte. Chaque boîte peut contenir plusieurs échantillons (ils sont étiquetés).

Un échantillon est localisé sur un site de fouille (où il a été mis au jour) caractérisé par un identifiant unique et un nom. Chaque site est situé dans une ville caractérisée par un identifiant unique et un nom. Chaque ville est située dans une région caractérisée par un identifiant unique et un nom. Enfin, chaque région est située dans un pays caractérisé par un identifiant unique et un nom.

Un échantillon de poterie est cuit selon un mode de cuisson caractérisé par un identifiant unique et une description textuelle.

Un échantillon peut faire l'objet de diverses analyses, chacune caractérisée par un identifiant unique, une date d'analyse, un type et un résultat (pour simplifier, on supposera que c'est un court compte-rendu textuel). Une analyse est réalisée par un laboratoire caractérisé par un identifiant unique et son nom.

Finalement, les échantillons peuvent faire l'objet d'études qui mènent à des publications, chacune caractérisée par un identifiant unique et l'URL (adresse web) de la référence bibliographique. De telles études publiées font souvent référence à plusieurs échantillons.

¹ A. Öztürk, L. Eyango, S.Y. Waksman, S. Lallich, J. Darmont. Warehousing Complex Archaeological Objects. 9th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2015), Larnaca, Cyprus, November 2015; *Lecture Notes in Artificial Intelligence*, Vol. 9405. Springer, Heidelberg, Germany, pp. 226-239.

² <http://www.arar.mom.fr>



<https://www.thefamouspeople.com/profiles/galileo-galilei-123.php>

Le projet SyMoGIH³, développé par le Laboratoire de Recherche Historique Rhône-Alpes⁴, propose un *modèle générique* des données historiques visant à stocker et à valoriser des données primaires concernant toute activité humaine (sociale, économique, intellectuelle...).

De manière simplifiée, un objet historique est caractérisé par un identifiant unique et un type, lui-même caractérisé par un identifiant unique et un libellé. Concrètement, les types d'objets historiques peuvent être des acteurs (personnes), des acteurs collectifs (entreprises, administrations...), des caractères sociaux (professions, fonctions institutionnelles, titres, diplômes, décorations...), des lieux ou des objets concrets.

Les objets historiques sont reliés à des informations caractérisées par un identifiant unique, une date estimée par l'historien, une date estimée par la source de l'information, un libellé et un type, lui-même caractérisé par un identifiant unique et un libellé. Concrètement, les types d'informations peuvent être des caractéristiques des objets historiques (patronyme, surnom, genre...), des liens de parenté, des événements (baptême, mariage, nomination, création d'entreprise...), des localisations, etc.

Une information peut être liée à plusieurs objets historiques, chacun ayant un rôle particulier et un nom d'après la source d'information. Une information peut également avoir des propriétés numériques chacune caractérisée par un identifiant unique et une valeur (par exemple, l'âge d'un acteur à sa mort ou sa fortune estimée). De plus, chaque propriété numérique est exprimée dans une unité de mesure (distance, unité monétaire, durée...) caractérisée par un identifiant unique et un libellé.

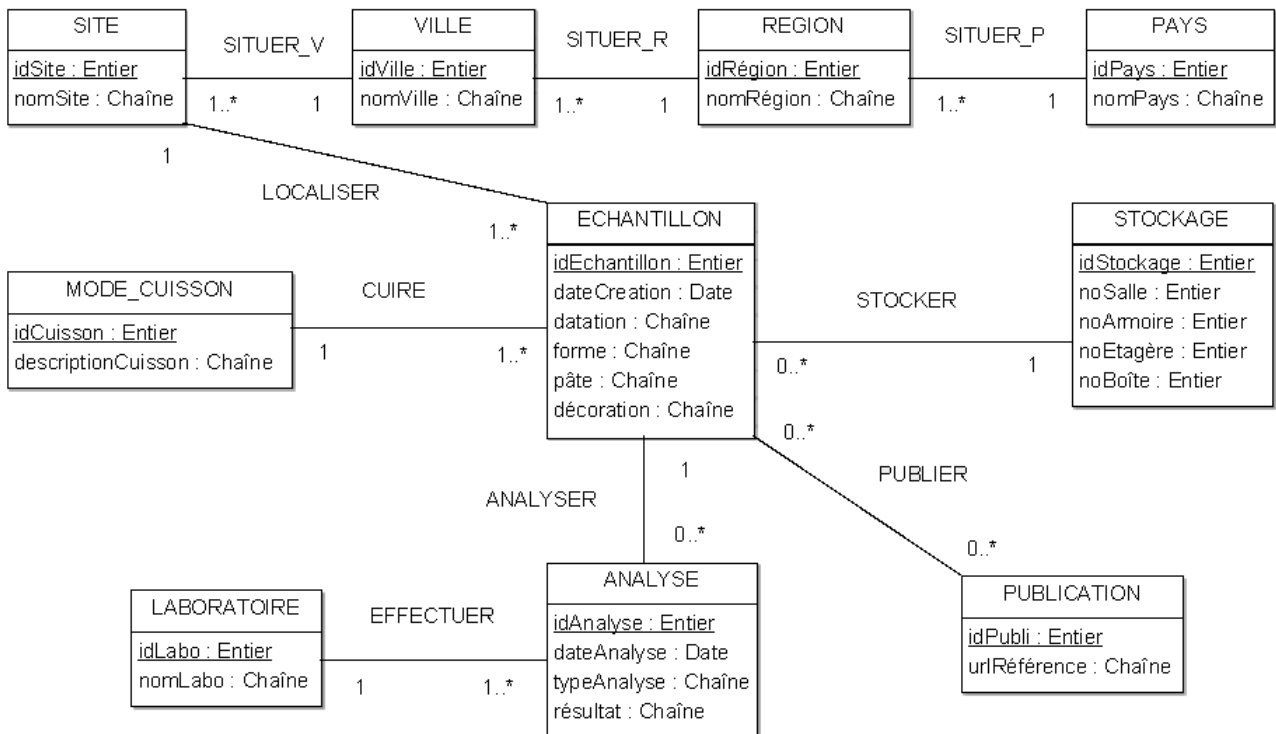
Enfin, les informations peuvent être « sourcées » par diverses sources (pardon pour la répétition), chacune caractérisée par un identifiant unique, une référence textuelle et un type, lui-même caractérisé par un identifiant unique et un libellé. Concrètement, les types de sources peuvent être des références bibliographiques, des objets matériels ou des ressources web. Pour chaque information et chaque source, une spécification textuelle est également indiquée.

³ F. Beretta, P. Vernus. Le projet SyMoGIH et la modélisation de l'information : une opération scientifique au service de l'histoire. *Les Carnets du LARHRA*, LARHRA, 2012, pp. 81-107. www.symogih.org

⁴ <http://larhra.ish-lyon.cnrs.fr>

Correction cas d'étude n° 1 : Ceramom

Modèle conceptuel UML



Modèle logique relationnel

STOCKAGE (idStockage, noSalle, noArmoire, noEtagère, noBoîte)

PAYS (idPays, nomPays)

REGION (idRégion, nomRégion, idPays#)

VILLE (idVille, nomVille, idRégion#)

SITE (idSite, nomSite, idVille#)

MODE_CUISSON (idCuisson, descriptionCuisson)

ECHANTILLON (idEchantillon, dateCréation, datation, forme, pâte, décoration,
idStockage#, idSite#, idCuisson#)

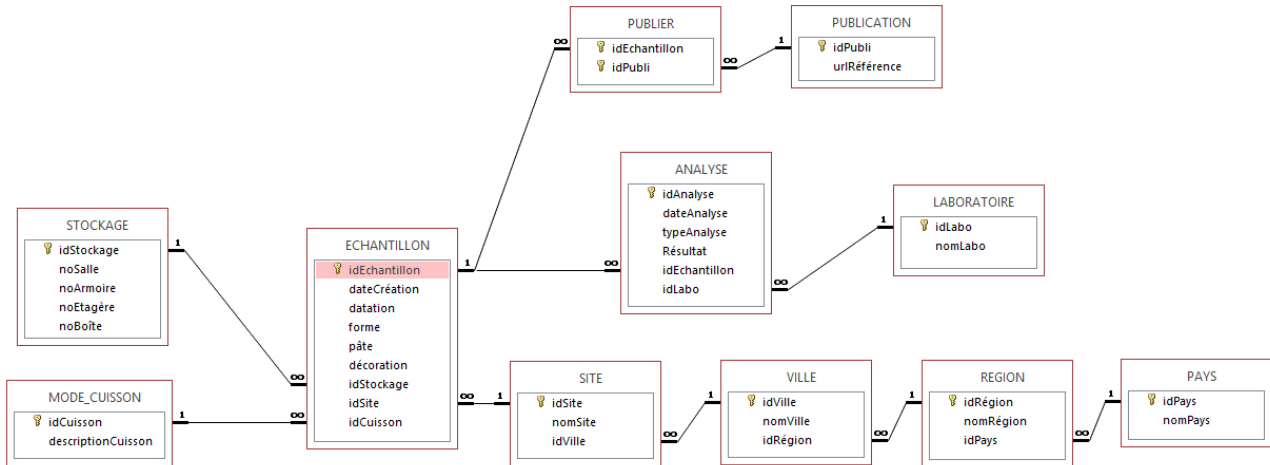
LABORATOIRE (idLabo, nomLabo)

ANALYSE (idAnalyse, dateAnalyse, typeAnalyse, résultat, idEchantillon#, idLabo#)

PUBLICATION (idPubli, urlRéférence)

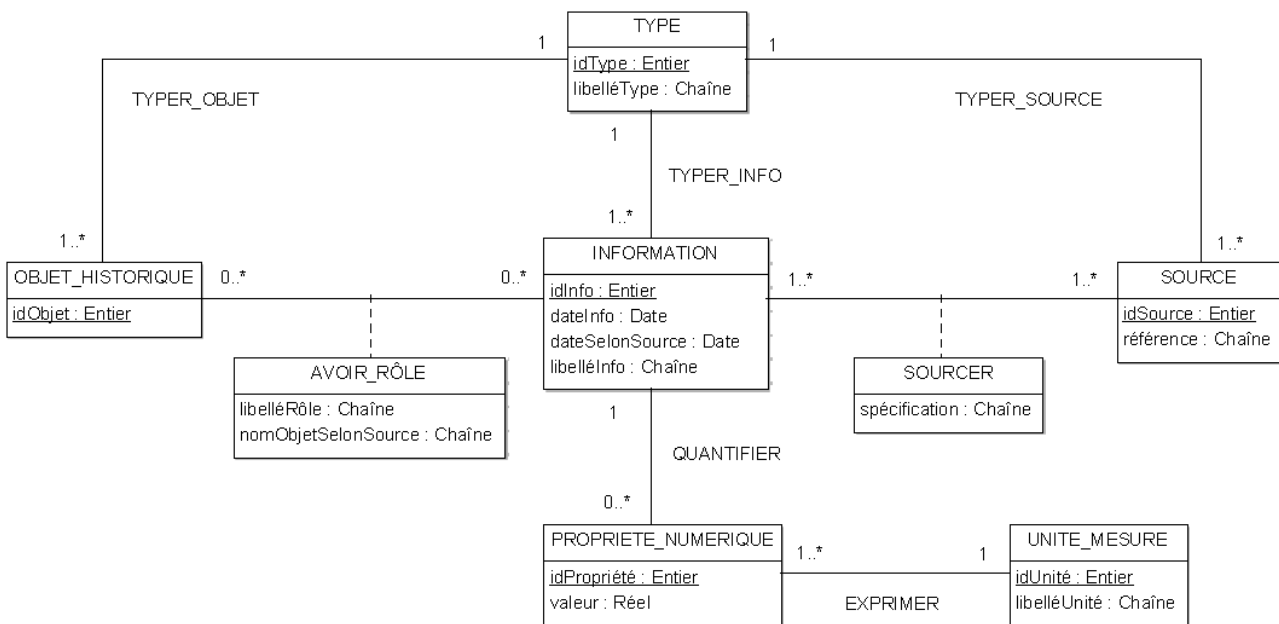
PUBLIER (idEchantillon#, idPubli#)

Modèle physique Access



Correction cas d'étude n° 2 : SyMoGIH

Modèle conceptuel UML



Modèle logique relationnel

TYPE (idType, libelléType)

OBJET_HISTORIQUE (idObjet, idType#)

INFORMATION (idInfo, dateInfo, dateSelonSource, libelléInfo, idType#)

AVOIR_RÔLE (idObjet#, idInfo#, libelléRôle, nomObjetSelonSource)

UNITE_MESURE (idUnité, libelléUnité)

PROPRIETE_NUMERIQUE (idPropriété, valeur, idInfo#, idUnité#)

SOURCE (idSource, référence, idType#)

SOURCER (idInfo#, idSource#, spécification)

Modèle physique Access

