

SCI6306 Bases de données documentaires (Automne 2023)

Christine Dufour, EBSI, UdeM

A2023 15 septembre 2023

Cours 2 : Contrôle de qualité, Modélisation sémantique
SCI6306

Christine Dufour, EBSI, UdeM

Table des matières

I - Cours 2 - Contrôle de qualité, Modélisation sémantique	3
1. + Au programme aujourd'hui.....	3
2. Contrôle de la qualité.....	3
2.1. Importance de la qualité dans les BD	5
2.2. Types d'erreur	6
2.3. Moyens d'action	6
3. Modélisation sémantique	7
3.1. Approche entités-relations (E-R)	7
3.2. Conversion d'un diagramme E-R en tables.....	9
3.3. Schéma relationnel.....	14
4. Projet de session, volet A	15
5. Ressources en lien avec le cours.....	16
Glossaire	17
Webographie	18
Crédits des ressources	19

I Cours 2 - Contrôle de qualité, Modélisation sémantique

- + Au programme aujourd'hui
- Contrôle de la qualité
 - Importance de la qualité dans les BD
 - Types d'erreur
 - Moyens d'action
- Modélisation sémantique
 - Approche entités-relations (E-R)
 - Exercice : Modélisation d'un groupe de recherche
 - Conversion d'un diagramme E-R en tables
 - Schéma relationnel
- Projet de session, volet A
- Ressources en lien avec le cours

1. + Au programme aujourd'hui

- **Contrôle de la qualité**
 - Aspects reliés au contrôle de qualité dans une base de données documentaire
- **Modélisation sémantique**
 - Approche entités-relations
 - Conversion en structure de tables relationnelles
 - Présentation du volet A du projet de session (modélisation)

2. Contrôle de la qualité

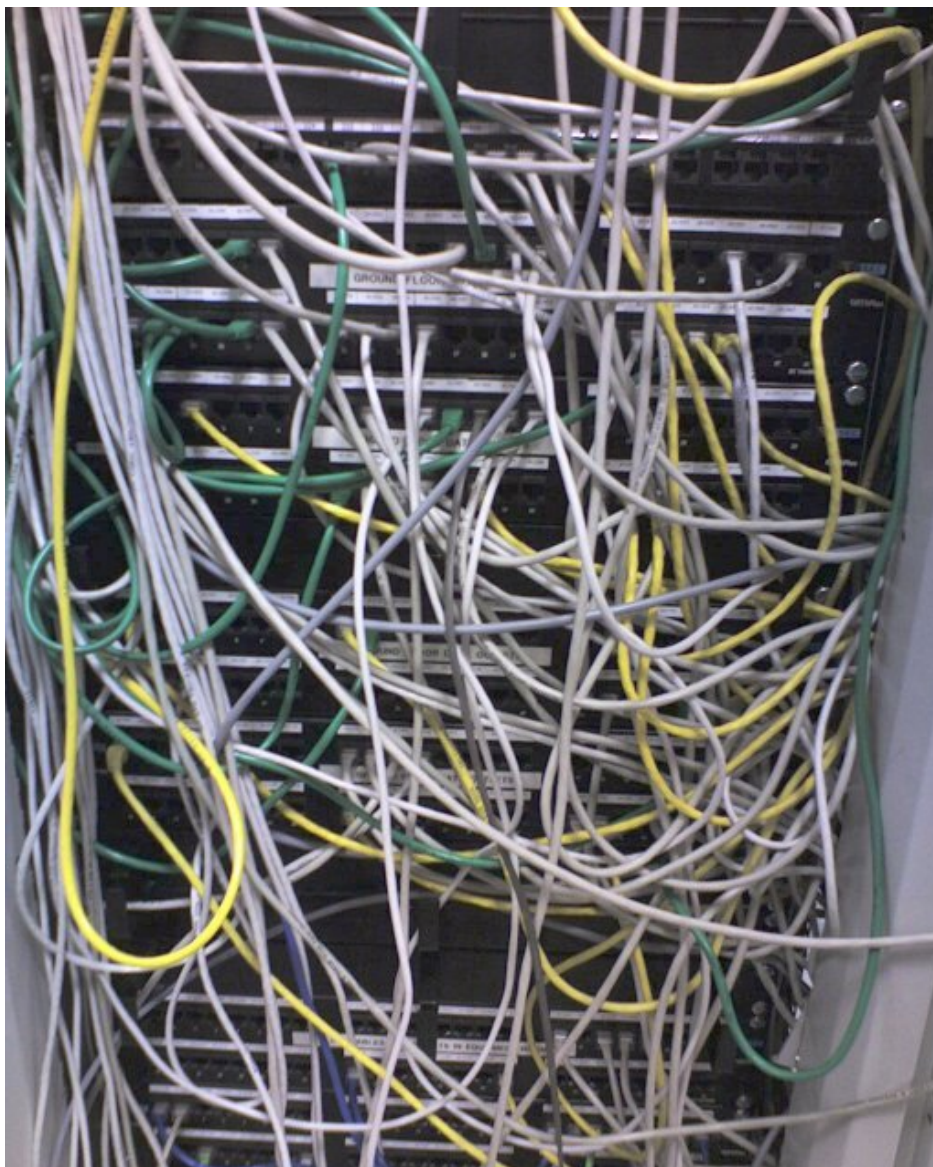
Différents aspects sont reliés au contrôle de qualité dans une base de données documentaire qu'il faut avoir en tête tout au long du processus de conception et même après! Une emphase générale est mise sur la qualité dans la société depuis les années 90 pour différentes raisons, entre autres :

- Pour des questions de crédibilité ...



HUMOR - Your Tax Dollars At Work - 001 - SOTP [https://www.flickr.com/photos/51514834@N00/7270817256]

- Pour des questions d'efficacité



"Cable ties are for wimps!" [https://www.flickr.com/photos/68423870@N00/5308965]

2.1. Importance de la qualité dans les BD

Le manque de qualité dans une base de données peut entraîner des **coûts** tant **directs** (perte de temps, d'information, coûts de correction ou remplacement) qu'**indirects** (p. ex. des dommages à la réputation et à celle des autres). Il n'est pas rare de voir dans l'actualité des exemples des problèmes liés à des BD dont la qualité n'est pas optimale.

Exemple : Le succès de Gangnam Style trop grand pour YouTube...

« Le problème, c'est que YouTube avait prévu dans le langage informatique un «entier» de 32 bits, soit un nombre maximum de consultations possibles de 2 147 483 647.

[...]

Les ingénieurs ont cependant anticipé le problème et revalorisé l'entier à 64 octets. Ce qui veut dire que Gangnam Style ou d'autres vidéos futures à succès pourront désormais atteindre 9 sextillions de consultations, un chiffre raisonnablement difficile à atteindre.

Source : Agence France-Presse New York. Février 2015. <https://www.lapresse.ca/arts/musique/201412/03/01-4824998-le-succes-de-gangnam-style-trop-grand-pour-youtube.php>

Exemple : Hello, I'm Mr. Null. My Name Makes Me Invisible to Computers

« For those of you unwise in the ways of programming, the problem is that “null” is one of those famously “reserved” text strings in many programming languages. Making matters worse is that software programs frequently use “null” specifically to ensure that a data field is not empty, so it's often rejected as input in a web form.

[...]

Essentially this is another spin on the Y2K problem, and what happens next will depend a lot on the quality of programming underlying the website or app that's doing the work. Most will accept “Null” without complaint. Some will loop back to the input screen and tell the user to try again, that the last name field can't be blank (But it's not blank! That's just my name!) Some will tell the user that “null” is a reserved term that can't be used. And some will just crash. The unique challenges inherent with the Null Dilemma can be a surprisingly difficult problem to solve. It turns out it's also surprisingly common, and it seems the larger the company is behind the application or the website, the more trouble it will have with my name.

Source : Null, Christopher. 11 novembre 2015. Wired. <https://www.wired.com/2015/11/null/>

Exemple : Equifax éprouve des problèmes avec le français

« S'il a été impossible d'avoir des explications de la part d'Equifax, Desjardins reconnaît de son côté qu'il y a un problème avec la reconnaissance des accents français.

[...]

M. Turcotti recommande aux gens d'éviter de mettre des accents ou des traits d'union afin de faciliter le traitement de leur dossier, et ce, même si la façon d'inscrire leur prénom sur le site diffère de la graphie sur leurs autres documents officiels et pièces d'identité.

Source : Morissette, Nathaëlle. 4 juillet 2019. La Presse. <https://www.lapresse.ca/affaires/finances-personnelle/201907/03/01-5232594-equifax-eprouve-des-problemes-avec-le-francais.php>

La qualité d'une BD est liée, d'une part, au **modèle économique du système** (clients, lectorat & de qui on parle), et d'autre part, à l'**expérience des utilisateurs et des utilisatrices**. Il ne faut pas seulement exiger la qualité, il faut aussi la créer et la mettre de l'avant :

- Comme concepteur et conceptrice de BD documentaires, cela implique par exemple l'inclusion de différents **mécanismes** pour assurer la qualité des données (masques, listes de validation, etc.) et la qualité de l' « expérience-utilisateur » (ergonomie de l'interface).
- Il faut **explicitement** aux utilisateurs et utilisatrices les moyens pris pour assurer la qualité.

2.2. Types d'erreur

Les erreurs peuvent être introduites tant **dans les données** qu'on retrouve dans la BD, que par le **système** utilisé.

Principaux types d'erreur

	Nature	Causes
Système	<ul style="list-style-type: none"> • Difficulté d'utilisation 	<ul style="list-style-type: none"> • Mauvaise ergonomie • Mauvaise modélisation/ implantation
Données	<ul style="list-style-type: none"> • Erreurs dans les données primaires : mauvaises informations, informations absentes, erreurs d'orthographe, informations périmées (par exemple liens hypertextuels) • Doublons (redondance) • Information au mauvais endroit (champ) 	<ul style="list-style-type: none"> • Erreurs typographiques (par exemple dates) • Resaisie / Reconnaissance optique de caractères (ROC) • Dictionnaire de données non suivi (raisons de la non utilisation : complexité, formation inadéquate) • Retards / indisponibilités (sources, serveurs, etc.) • Évolution des sources (Web)

Les erreurs d'orthographe peuvent être parfois très fréquentes, que ce soit des omissions, des insertions, des substitutions ou des transpositions de caractères. Le clavier QWERTY (ou AZERTY) peut être la source de plusieurs erreurs, ce dernier ayant été conçu à la base pour des raisons mécaniques et non pour optimiser l'efficacité de la saisie. Le clavier Dvorak¹ permet de réduire le nombre d'erreurs et d'être plus rapide, mais les coûts de son insertion en organisation (tant sur le plan matériel que sur le plan des ressources humaines) font qu'il n'a pas été adopté à grande échelle, aucune étude ne semblant avoir réussi à réellement prouver sa supériorité.

2.3. Moyens d'action

On peut intervenir au niveau de la qualité d'une BD à trois moments : lors de sa création, lors de la saisie des données, lors de l'utilisation de la BD :

- **Création de la BD** : Il est important de prendre le temps de *bien conceptualiser la réalité* pour laquelle la BD sera implantée. Il faut aussi exploiter l'ensemble des *mécanismes de validation* disponibles dans le SGBD (masques, listes de validation, etc.) afin de réduire les erreurs lors de la saisie de données. De plus, il faut prendre soin lors de la conception des interfaces, de développer des *interfaces ergonomiques* ainsi que d'*impliquer les utilisateurs et utilisatrices* pour s'assurer que le SGBD développé corresponde bien à leur profil.

¹ https://en.wikipedia.org/wiki/Dvorak_keyboard_layout

- **Saisie des données** : Il faut implanter des mécanismes afin de *vérifier les données qui ne sont pas contrôlées* par des mécanismes de validation comme, par exemple, une relecture humaine ou un correcteur orthographique. Il est important de sensibiliser les personnes responsables de la saisie à l'importance de leur qualité ainsi que les former à l'utilisation du dictionnaire de données.
- **Utilisation de la BD** : Il est possible d'inclure dans les interfaces de saisie ou de recherche des *rappels* quant à la forme attendue de certains contenus ou à la syntaxe du langage d'interrogation. De plus, le soin pris lors de la création de la BD afin de développer des *interfaces ergonomiques* et de bien répondre aux besoins des utilisateurs augmentera la qualité de la BD.

Une BD étant en constante **évolution**, il est utile de prévoir des moyens pour recueillir les **commentaires** des utilisateurs et utilisatrices, que ce soit par un lien vers une adresse courriel sur le site Web, par un formulaire Web de rétroaction ou par l'utilisation de plateformes collaboratives externes (ex. Facebook) ou internes (c'est-à-dire intégrées au site Web comme, par exemple, la documentation php¹).

3. Modélisation sémantique

La modélisation d'une base de données relationnelles comporte les trois étapes habituelles pour toute modélisation :

- La **conceptualisation** de la réalité à représenter
- La **représentation** de cette conceptualisation
- La **validation** de la modélisation proposée

Les **deux premières étapes** (conceptualisation et représentation) peuvent être effectuées conjointement lors de la **modélisation sémantique** de la base de données. La **validation** peut passer quand à elle par un **retour vers la ou les personnes à la source du besoin** pour une base de données ainsi que par l'exercice de produire des **exemples de contenu** pour chacune des tables prévues dans votre modélisation.

La **modélisation sémantique** d'une base de données relationnelle permet d'ajouter à la compréhension d'une base de données et d'ainsi pouvoir répondre plus intelligemment aux interactions de l'utilisateur ou de l'utilisatrice. Cette modélisation est utile au processus de *conception systématique* des bases de données. Elle se fait habituellement en *dehors du SGBD*.

L'objectif de cette modélisation est de représenter une certaine réalité selon le modèle relationnel pour pouvoir construire, par la suite, une base de données relationnelle. Plusieurs approches peuvent être utilisées pour modéliser une BD, dont l'**approche entités-relations (E-R)** qui est parmi les plus connues et utilisées.

3.1. Approche entités-relations (E-R)

Cette approche est fondée sur le modèle E-R (entités-relations) défini par Chen (1976) et raffiné par la suite. On y retrouve représentés plusieurs *objets sémantiques* :

- **Entités** : objets que l'on peut et veut distinguer (par exemple, des étudiant.e.s, des cours, etc.).
- **Relations** : connections entre des entités (par exemple, des inscriptions qui relient les étudiant.e.s aux cours). Une relation possède une **cardinalité** qui représente la manière dont les enregistrements des deux tables connectées sont liés.
- **Attributs** : propriétés décrivant une entité ou une relation (par exemple, le nom d'un.e étudiant.e qui décrit l'étudiant.e, la note obtenue à un cours qui caractérise l'inscription d'un.e étudiant.e à un cours) que l'on veut documenter dans la BD. Certains attributs permettent d'identifier de manière unique les enregistrements dans une table comme, p. ex., NO_COURS dans la table COURS. On parle en ce cas d'une *clé primaire**.

Un des résultats de la modélisation entités-relations est un **diagramme Entités-Relations** qui encapsule *visuellement* les différents objets sémantiques identifiés. Classiquement, les **entités** sont représentées par des *rectangles*, les **relations** par des *losanges* et les **attributs** par des *ovales*. Le diagramme Entités-Relations

¹ <https://www.php.net/manual/fr/preface.php>

qui suit représente la BD INSCRIP que nous utiliserons comme exemple tout au long de la session. Il s'agit d'une BD pour gérer les cours, les professeur.e.s ainsi que les étudiant.e.s pour un établissement d'enseignement (voir *Documentation complémentaire - Dictionnaire de données d'INSCRIP* pour la description détaillée de cette base de données).

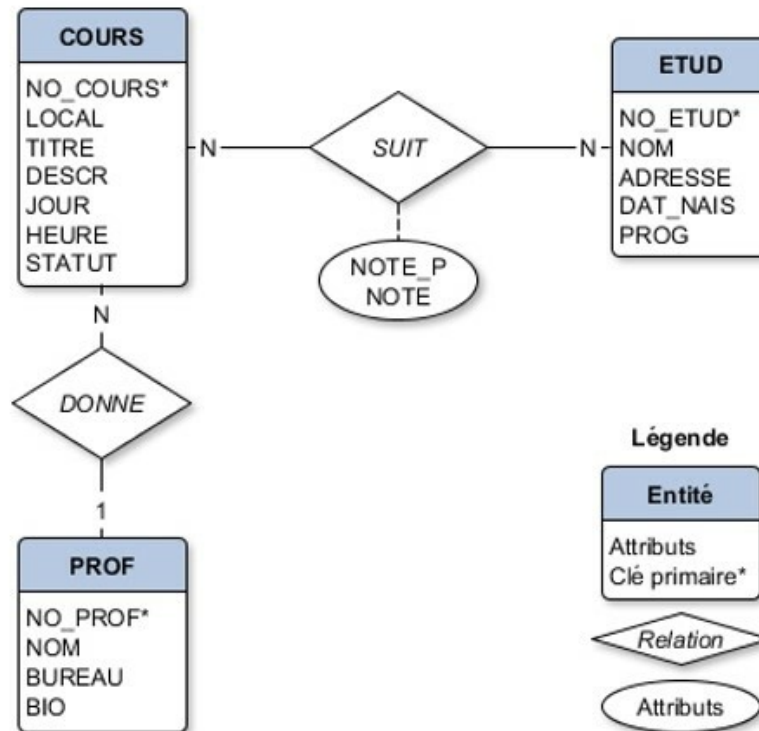


Diagramme Entités-Relations de la base de données INSCRIP

On retrouve dans cette base de données **3 entités**, soit les *cours*, les *professeur.e.s* et les *étudiant.e.s*. **Deux relations** connectent ces entités, soit une relation *entre les cours et les étudiant.e.s* qui représente l'*inscription* d'un.e étudiant.e à un cours, et une relation *entre les cours et les professeur.e.s* pour représenter le lien d'*enseignement*. Ces deux relations illustrent les **deux principaux types de cardinalité** que l'on retrouve dans une base de données relationnelle :

- *Relation 1-N* : La relation DONNE est une relation 1-N parce que, dans le contexte représenté, un.e professeur.e peut être associé.e à *plusieurs* cours (d'où le *N* du côté de l'entité COURS), tandis qu'un cours n'est associé qu'à *un.e seul.e* professeur.e comme il n'y a pas de cours donné simultanément par plus d'un.e professeur.e (d'où le *1* du côté de l'entité PROF).
- *Relation N-N* : On retrouve une relation N-N dans INSCRIP, soit la relation SUIT. En effet, un.e étudiant.e peut suivre *plusieurs* cours - d'où le *N* du côté de l'entité SUIT, et un cours peut être suivi par *plusieurs* étudiant.e.s - d'où le *N* du côté de l'entité ETUD.

L'identification de la cardinalité des relations est importante, comme la cardinalité aura un **impact direct** sur la manière dont la conceptualisation sémantique sera convertie en structure de tables et de champs. Cet impact sera illustré dans une section subséquente.

Selon le contexte, il arrive que l'on puisse hésiter pour un certain élément entre le définir **comme une entité ou comme un attribut**. Par exemple, dans INSCRIP, on retrouve un attribut LOCAL comme caractéristique de l'entité COURS. Pourquoi ne pas en avoir fait une entité à part entière ? Il y a deux questions à se poser :

- Est-ce qu'on s'intéresse directement aux locaux comme tels ou bien indirectement comme une caractéristique des cours?
- Et si aucun cours ne se donne dans un local, veut-on conserver dans la base de données des informations sur ce local?

La réponse peut varier en fonction de l'objectif de la BD :

- Si la base de données sert à gérer des cours, alors le local nous intéresse comme caractéristique d'un cours uniquement; il s'agit donc d'un attribut.
- Si la base de données sert à gérer l'attribution des locaux, alors le local nous intéresse même si aucun cours ne s'y donne. On voudrait ainsi pouvoir conserver des informations autres que le numéro du local, p. ex. le nombre de places ou le type d'équipements. Il s'agit en ce cas d'une entité.

*Pour plus d'information sur la distinction entité et attribut, voir Comment distinguer une entité d'un attribut / Marcoux^{MARCOUX, 2007} *.*

Comme la modélisation sémantique d'une BD relationnelle se fait souvent à l'extérieur du SGBD, il est légitime de se demander **avec quoi dessiner** le diagramme entités-relations. Deux éléments sont à retenir pour faire un choix :

- Il faut choisir une application capable de facilement faire des **formes**, du **texte** et des **lignes**. Il faut pouvoir respecter la *symbolique prédéfinie*. Ce peut être un logiciel de traitement de texte (Word, Writer), de présentique (PowerPoint, Keynote), de diagramme et synoptique (yEd, Visio), de dessin (FireWork), etc.
- Il faut aussi prendre en compte **vos préférences**. Utilisez un outil accessible et que vous connaissez bien. Il faut aussi vous rappeler qu'il ne s'agit pas d'une œuvre d'art. Si vous voulez en explorer un, yEd¹ de la compagnie yWorks (gratuit, Windows/Linux/Mac) est installé sur les postes au laboratoire et fait bien le travail (les diagrammes E-R dans le matériel de cours ont été produits avec yEd). Vous retrouverez sur le site du cours un mode d'emploi abrégé² pour cet outil.

a) Exercice : Modélisation d'un groupe de recherche

Soit le contexte d'un **groupe de recherche universitaire** ayant les caractéristiques suivantes :

- Le groupe représente une communauté de *chercheur.se.s* (professeur.e.s, post-doctorant.e.s, étudiant.e.s, etc.) se regroupant autour d'une thématique afin de produire de nouvelles connaissances par le biais de différents *projets de recherche*.
- Le groupe est associé à des projets de recherche, *financés ou non*, regroupent en tout ou en partie les *membres* du groupe de recherche.
- Certains projets peuvent être le sujet du *mémoire* ou de la *thèse* d'un.e des étudiant.e.s membre du groupe de recherche qui est encadré en ce cas par un.e des professeur.e.s du groupe de recherche.
- Parmi les « livrables » du groupe de recherche on retrouve la diffusion des *résultats* (articles, etc.).

Exercice : Identifiez, en groupe de 2-3 personnes, les différents éléments permettant de représenter ce contexte (entités, relations, principaux attributs). Nous ferons un retour ensemble par la suite.

3.2. Conversion d'un diagramme E-R en tables

Le diagramme E-R représente la **modélisation sémantique** d'un contexte et **non directement sa structure en tables** (bien qu'il n'en soit pas si loin que ça!). L'étape suivant l'élaboration du diagramme E-R est ainsi de le convertir en structure de tables et de champs afin de pouvoir ensuite l'implanter dans un SGBD. On peut découper le processus de conversion d'un diagramme E-R en tables en **6 grandes étapes** :

Processus pour convertir un diagramme E-R en structure de tables

¹ <https://www.yworks.com/products/yed>

² https://cours.ebsi.umontreal.ca/sci6306/co/site_guide_yed.html

Étapes	Précisions
1. Transformation des entités en tables de données	<ul style="list-style-type: none"> • Entité = table • Attribut de l'entité = champ de la table • Clé primaire = soit un ou des attributs (champs) existant pouvant naturellement servir de clé primaire, soit un champ créé spécifiquement comme clé primaire
2. Traitement des relations	<ul style="list-style-type: none"> • Relation 1-N : ajout, dans la table de l'entité du côté N, de la clé primaire de l'entité du côté 1. Ce champ ajouté est ce que l'on appelle une <i>clé externe</i> * qui lie la table de l'entité du côté N à la table de l'entité du côté 1. • Relation N-N : création d'une table du même nom que la relation qui comprend deux clés externes (une vers chaque table liée). Chacune des clés externes correspond à la clé primaire de la table reliée. La combinaison de ces deux clés externes servira de clé primaire à cette nouvelle table. • Attribut d'une relation : pour une <i>relation 1-N</i>, les attributs de cette relation deviendront des champs dans la table du côté N. Pour une <i>relation N-N</i>, les attributs de la relation seront des champs dans la table ajoutée pour représenter la relation N-N.
3. Réévaluation des clés primaires en fonction de l'ensemble des champs créés	<ul style="list-style-type: none"> • Modification des clés primaires en conséquence, si besoin • Voir à ce sujet le texte <i>Clé primaire multichamp / Marcoux^{MARCOUX, Yves, 2007}*</i>
4. Définition des types pour chaque champ	<ul style="list-style-type: none"> • En fonction des contenus attendus, choix du type de données le plus adapté
5. Détermination des champs à valeur NULL (champs facultatifs) ou permettant la chaîne vide	
6. Règles d'écriture particulières (peuvent être générées par exemple via un masque)	<ul style="list-style-type: none"> • Précision des règles d'écriture pour les différents champs

Exemple : Exemple de conversion pour la base de données INSCRIP

Soit le diagramme E-R de la base de données INSCRIP

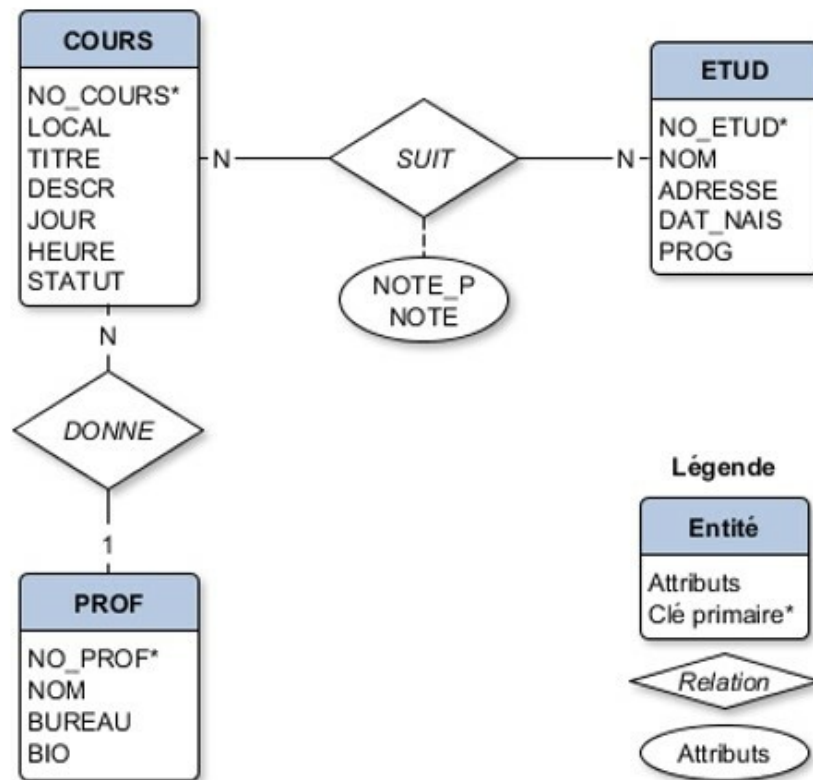
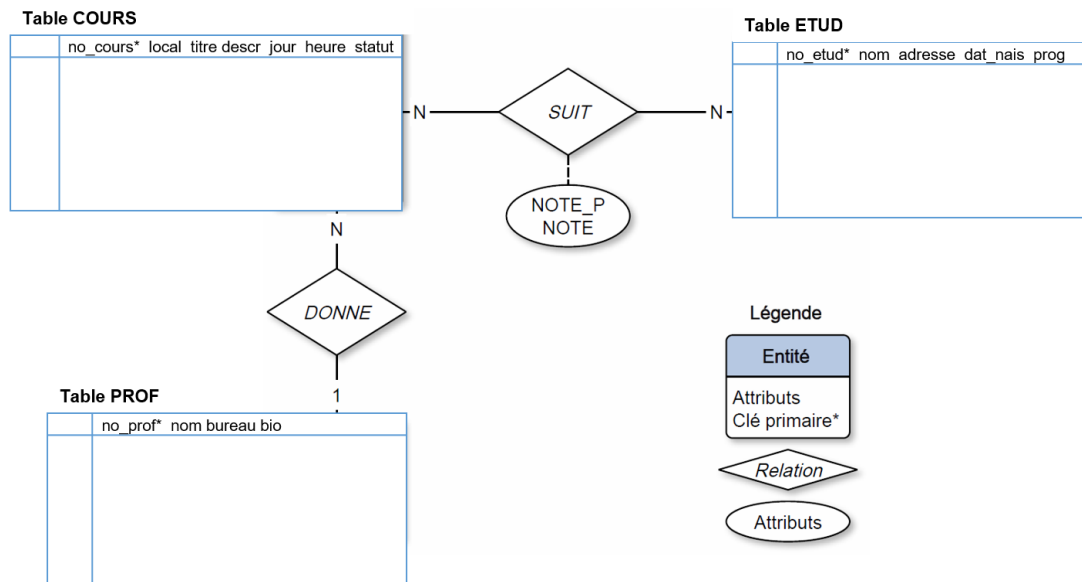


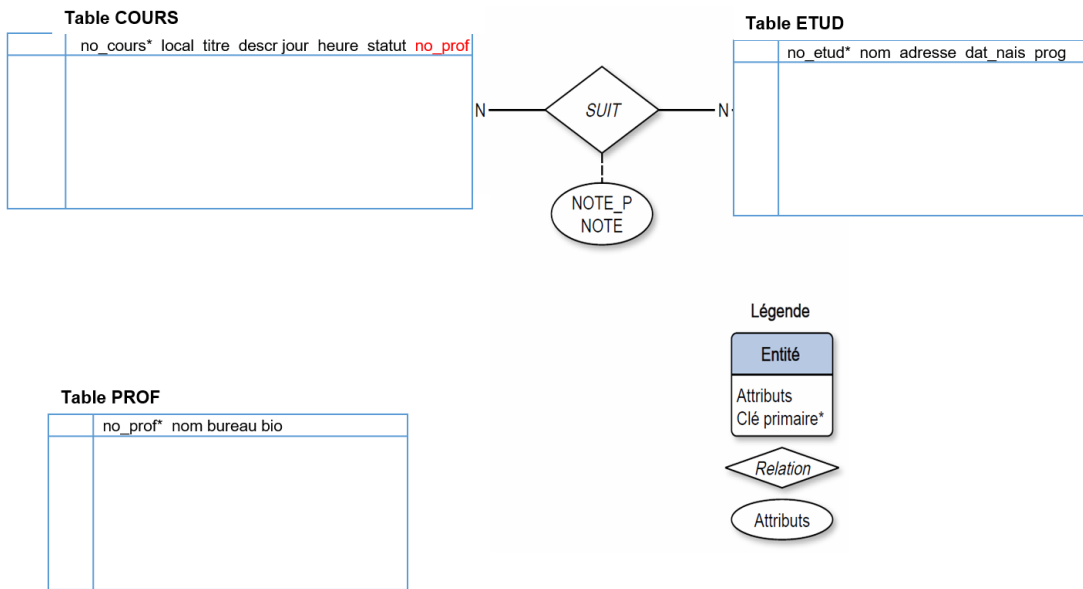
Diagramme Entités-Relations de la base de données INSCRIP

La **première étape** consiste à **transformer** les **entités** en tables et les **attributs** en **champs**. Par exemple, l'entité COURS devient la table COURS et ses différents attributs (p. ex. LOCAL) deviennent des champs pour cette table :



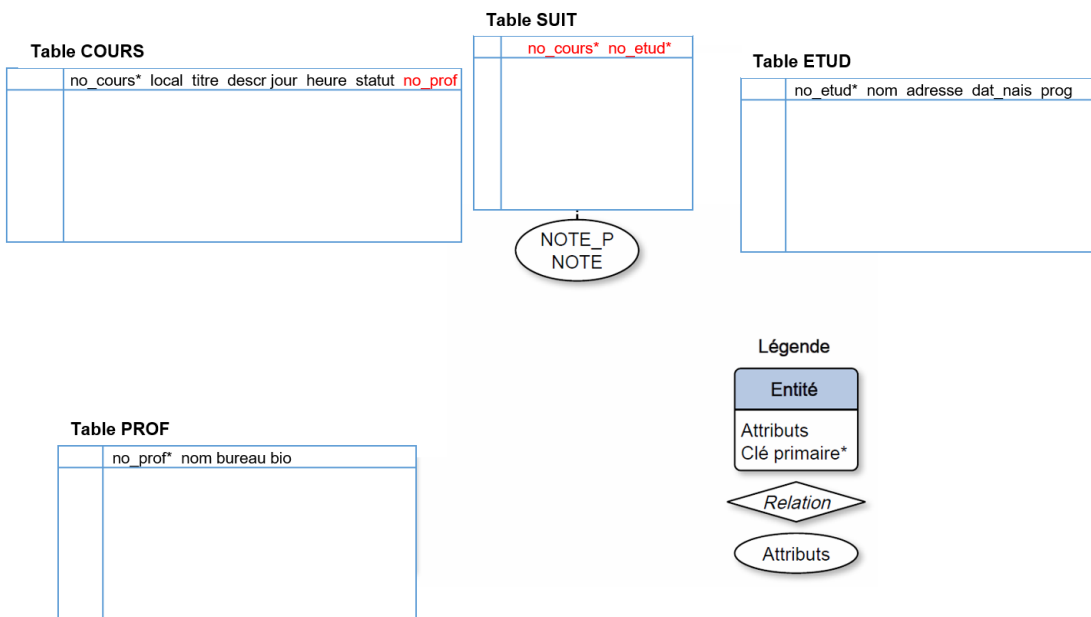
Étape 1 : Transformation des entités en table

La **deuxième étape** vise à **transformer** les **relations 1-N** en **clés externes**. La seule relation 1-N dans le diagramme E-R d'INSCRIP est la relation DONNE liant les entités COURS et PROF. La transformation ajoutera dans la table du côté N (soit COURS), le ou les champs servant de clé primaire à la table du côté 1 (soit PROF). On ajoute ainsi le champ NO_PROF dans la table COURS :



Étape 2 : Transformation des relations 1-N

La deuxième étape traitera aussi les **relations N-N** qui seront transformées en **tables**. Pour INSCRIP, la seule relation N-N est la relation SUII. Cette dernière deviendra la table SUII et sa clé primaire sera la combinaison des clés primaires des deux tables qu'elle relie, soit le champ NO_COURS pour la table COURS et le champ NO_ETUD pour la table ETUD :



Étape 2 (suite) : Transformation des relations N-N

Pour terminer la deuxième étape, il faut transformer les **attributs des relations N-N** en **champs** dans la table créée pour cette relation. Dans INSCRIP, la relation SUIV comporte deux attributs, soit NOTE et NOTE_P. Ces deux attributs deviendront ainsi des champs dans la table SUIV :

no_cours*	local	titre	descr	jour	heure	statut	no_prof

no_etud*	nom	adresse	dat	nais	prog

no_prof*	nom	bureau	bio

no_cours*	no_etud*	note	p	note

Étape 2 : Transformation des attributs de relation en champs

Maintenant que l'ensemble des tables ont été créées, il faut prendre le temps à la **troisième étape** de réévaluer les champs identifiés lors de la conception du diagramme E-R comme clés primaires. Sont-ils toujours pertinents ou y a-t-il d'autres champs qui pourraient agir comme clé primaire de manière plus efficace? Pour INSCRIP, les choix initiaux faits sont les plus pertinents. Il n'est donc pas nécessaire de changer.

C'est à la **quatrième étape** qu'il est temps de s'interroger quant aux meilleurs **types de données** pour les différents champs que l'on retrouve dans les tables. Les quatre principaux types sont (1) *caractères* (= information textuelle), (2) *numérique* (= chiffres), *date/heure* (= date et/ou heure), *logique* (=Oui ou Non / Vrai ou Faux). P. ex., quel type de données s'attend-on à retrouver pour le titre d'un cours? Un champ de type caractère semble un choix assez évident! Mais de quelle longueur? Il faut en effet prendre soin, entre autres pour les champs de type caractère, d'en identifier la longueur maximale.

Le cœur de la **cinquième étape** est l'identification des champs qui acceptent la valeur NULL, c'est-à-dire qui sont des champs facultatifs. Par exemple, devrait-on toujours avoir les informations nécessaires pour remplir le champ BIO pour un.e professeur.e? La décision, dans certains cas, est facile lorsque l'information pour un champ n'existe pas toujours. Imaginons, par exemple, que nous ayons prévu un champ pour le numéro de cellulaire dans la table ETUD. Bien que cela soit plus rare, il se peut qu'un.e étudiant.e n'ait pas de téléphone cellulaire. Le champ doit en ce cas être facultatif. Dans d'autres cas, la décision peut être moins directe car demandant une réflexion sur la charge de travail impliquée par la saisie de certains champs. Par exemple, le champ BIO pour un.e professeur.e est un champ qui peut demander un certain travail de création et de saisie de l'information. Même si, dans l'idéal, on pourrait souhaiter qu'il soit toujours rempli, il peut être plus raisonnable de le laisser facultatif.

Finalement, en **dernière étape** vient la précision des **règles d'écriture** pour chacun des champs avec l'identification, s'il y a lieu, de moyens de s'assurer de respecter ces règles d'écriture (p. ex. par des masques de saisie). Certains champs seront à saisie relativement libre, comme la biographie des professeur.e.s. D'autres auront une forme bien précise comme le numéro identifiant un cours.

*Pour des explications complémentaires sur la conversion d'un diagramme entités-relations en structure de tables relationnelles, voir le texte Conversion d'un diagramme E/R en structure de tables relationnelles / Marcoux^{MARCOUX, Yves, 2007} *.*

3.3. Schéma relationnel

Le résultat de la conversion d'un diagramme E-R en structure de tables permet de développer un élément de documentation fort important pour une BD relationnelle, soit son **schéma relationnel**. Il s'agit de son *dictionnaire de données* dans lequel sera précisé les éléments suivants :

- **Contexte** général de la base de données
- Structure générale de la base de données, incluant son **diagramme entités-relations**
- Structure des **tables**, incluant la description des champs (à quoi ils correspondent, quel est leur format, les valeurs possibles, etc.)
- **Contraintes additionnelles** sur les contenus s'il y a lieu (c'est-à-dire des restrictions sur certains contenus qui ne peuvent se définir par les caractéristiques d'un champ (par exemple des restrictions d'un champ par rapport à la valeur d'un autre champ))
- Exemples de **contenus valides** (pour illustrer les règles de saisie)

Vous retrouverez à titre d'exemple dans la *Documentation complémentaire du cours* le *schéma relationnel de la BD INSCRIP*. Des informations complémentaires sont données ci-dessous concernant les différentes sections attendues.

Contexte général de la base de données

Cette section présente le contexte global de la base de données. Il faut y être assez précis pour bien comprendre la **finalité de la base de données** et les **éléments contextuels** ayant une influence sur cette dernière.

Structure générale de la base de données

Cette section du document présente la schématisation de la **modélisation sémantique** effectuée pour représenter le contexte de la base de données. En sus du **diagramme entités-relations** on y retrouve une **brève description de chacune des tables**. Il est à noter qu'il s'agit ici de la modélisation entités-relations et non de la transformation en tables et en champs de cette dernière. On ne doit ainsi pas y retrouver les clés externes, comme ces dernières proviennent de la transformation des relations.

Structure des tables

Cette section donne toutes les informations nécessaires pour définir les tables dans un SGBD relationnel (type de données, statut obligatoire ou non, clés primaire et/ou externe, index plein texte) ainsi qu'être en mesure de bien comprendre la nature des contenus attendus pour chacun des champs (description des champs et précision, s'il y a lieu, des règles d'écriture). Pour chacune des tables prévues qui proviennent de la transformation de votre diagramme entités-relations, un **premier tableau résume les caractéristiques de ses champs** :

- *Nom du champ* : Des noms significatifs assez courts sont préférables, exempts de caractères accentués et d'espace.
- *Type* : Il s'agit ici des types génériques de données, les types spécifiques étant définis au moment de l'implantation dans le SGBD. Quatre types sont possibles : *Caractère* (pour les chaînes de caractères alphanumériques), *Numérique* (pour des données comportent uniquement des nombres), *Date/Heure* (pour tous les formats de date et d'heure allant d'une année uniquement jusqu'à la version longue comportant l'heure et la date complète), *Logique* (pour un champ acceptant seulement deux valeurs comme Vrai ou Faux).
- *Taille* : Cette caractéristique s'applique pour les champs de type *Caractère* et *Numérique*. Pour un champ de type *Caractère*, il s'agit du nombre maximal de caractères attendus pour une entrée. Pour un champ de type *Numérique*, il s'agit du nombre de chiffres de la valeur maximale attendue. Par exemple, si la valeur maximale attendue est 99, la taille sera de 2. Pour une valeur maximale de 99 999, la taille sera de 5.
- *Obligatoire* : Il s'agit ici de préciser si un champ doit nécessairement comporter une valeur ou s'il peut être laissé vide.

À la suite de ce tableau, il faut indiquer le ou les champs qui serviront de **clé primaire** pour la table. Si la clé primaire est la combinaison de plusieurs champs, elle est indiquée sous la forme (*champ1, champ2*). On indique, s'il y a lieu, la ou les **clés externes** sous la forme *CHAMP vers TABLE*. Finalement, on précise le ou les **index plein texte (s'il y a lieu)**. Il est en effet utile lors de la modélisation d'identifier les éventuels champs comportant des données textuelles assez denses pour lesquels une recherche plein texte serait utile (en ce cas, un index plein texte est nécessaire).

À la fin du schéma relationnel, un tableau présente la **description précise des contenus attendus** pour les différents champs en indiquant, lorsque pertinent, les règles d'écriture. La précision est importante comme ces informations guideront les personnes en charge de la saisie. La description des champs est donc un facteur important dans la qualité des données saisies.

Contraintes additionnelles

On retrouve dans cette section toutes les **contraintes par rapport aux contenus** de la base de données qui **n'ont pu être contrôlées par le biais de la définition des champs**. Ces contraintes additionnelles pourront, par exemple, être mises en place par programmation.

Exemple de contenus valides

Cette dernière section présente des **exemples de contenus valides pour tous les champs de toutes les tables**. Il n'est pas nécessaire de mettre beaucoup d'informations; l'important est de donner des **exemples des principales règles d'écriture à suivre**. Il est toutefois important de bien couvrir les différentes règles d'écriture comme cette section servira d'exemple à la ou les personnes en charge de la saisie.

4. Projet de session, volet A

L'objectif du premier volet du projet de session est de **modéliser un cas concret** en utilisant l'approche entités-relations, cas concret décrit ci-dessous. Le **livrable** pour le volet A sera le **schéma relationnel** produit à partir du gabarit fourni. Vous trouverez le protocole, le gabarit à utiliser ainsi qu'un forum dédié sur StudiUM :

- **Protocole du volet A** : format Web pour lecture à l'écran¹ | format PDF pour impression²
- **Gabarit pour le schéma relationnel**³
- **Forum pour demander des précisions sur le contexte**⁴

Méthode : Contexte du projet de session

Le contexte du projet de session est celui d'un **projet pilote** pour concevoir une **plateforme de visualisation des projets de recherche étudiants de l'EBSI** qui sera ajoutée au site Web de l'école. Ce projet est découpé suivant les trois étapes classiques pour le développement d'une BD Web, soit :

- **Volet A** : *modélisation sémantique* de la réalité étudiée
- **Volet B** : implantation de la modélisation dans *MySQL* et saisie de données
- **Volet C** : développement d'une *interface Web* pour pouvoir saisir certaines informations ainsi que produire certains types de documents

Le prototype de l'équipe qui obtiendra la meilleure note pour le Volet C sera retenu comme base pour finaliser la plateforme qui sera intégrée au site de l'EBSI. Le nom des membres de l'équipe sera indiqué sur le site de l'EBSI (à moins que les membres de l'équipe ne le veuillent pas).

¹ https://studium.umontreal.ca/pluginfile.php/8505770/mod_resource/content/9/index.html [accès restreint]

² https://studium.umontreal.ca/pluginfile.php/8505771/mod_resource/content/8/sci6306_a23_protocole_voletA.pdf [accès restreint]

³ https://studium.umontreal.ca/pluginfile.php/8505772/mod_resource/content/8/sci6306_schema_relationnel_modele.docx [accès restreint]

⁴ <https://studium.umontreal.ca/mod/forum/view.php?id=4798968> [accès restreint]

5. Ressources en lien avec le cours

Matériel de cours

- *Notes de cours [cf. sci6306_cours2_notes.pdf]*

Lectures obligatoires

- *Conversion d'un diagramme E/R en structure de tables relationnelles / Marcoux^{MARCOUX, Yves, 2007}**
- *Comment distinguer une entité d'un attribut / Marcoux^{MARCOUX, 2007}**
- *Clé primaire multi-champ / Marcoux^{MARCOUX, Yves, 2007}**
- *Bases de données relationnelles / Dufour^{DUFOUR, Christine, 2016}**

Glossaire

Clé externe (ou clé étrangère)

Une **clé externe (ou clé étrangère)** est un champ dans une table qui pointe vers la clé primaire d'une autre table de données. Elle sert donc à lier des tables.

Imaginons par exemple une BD où on retrouve une table contenant des informations sur des personnes ainsi qu'une table contenant des informations sur différentes villes. On pourrait vouloir lier les deux tables afin d'associer à une personne des informations complémentaires au nom de sa ville de résidence actuelle (par exemple, le nombre d'habitants) sans avoir à consigner ces informations complémentaires dans la table PERSONNE (ce qui évitera bien de la redondance dans cette table, plusieurs personnes pouvant résider dans la même ville !). On pourrait retrouver dans la table PERSONNE un champ RESIDENCE_ACTUELLE qui pointerait vers le champ VILLE de la table GEOGRAPHIE.

Clé primaire

La **clé primaire** d'une table de données est un champ ou la combinaison de plusieurs champs qui permettent d'identifier uniquement chaque enregistrement de la table de données. Par exemple, si une table de données comporte des informations sur des personnes, une clé primaire potentielle pourrait être le numéro d'assurance sociale.

Webographie

DUFOUR, Christine. 2016. *S4.3 Bases de données relationnelles*. In *MOOC Architecture de l'information, Séquence 4 – Web et bases de données*. <https://archinfo00.hypotheses.org/235>

MARCOUX, Yves. 2007. *Comment distinguer une entité d'un attribut*. <https://marcoux.ebsi.umontreal.ca/enseign/6306/entite-vs-attribut.htm>

MARCOUX, Yves. 2007. *Clé primaire multi-champ*. <https://marcoux.ebsi.umontreal.ca/enseign/6306/cle-primaire-multichamp.htm>

MARCOUX, Yves. 2007. *Conversion d'un diagramme E/R en structure de tables relationnelles*. <https://marcoux.ebsi.umontreal.ca/enseign/6306/conversion-ER-relat.htm>

Crédits des ressources

HUMOR - Your Tax Dollars At Work - 001 - SOTP

[<https://www.flickr.com/photos/51514834@N00/7270817256>] p. 4

<http://creativecommons.org/licenses/publicdomain/4.0/fr/>, [monstersforsale](#)

[<https://www.flickr.com/photos/51514834@N00/>]

"Cable ties are for wimps!" [<https://www.flickr.com/photos/68423870@N00/5308965>] p. 4

<http://creativecommons.org/licenses/by/2.0/fr/>, [mrtom](#) [<https://www.flickr.com/photos/68423870@N00/>]