

## Modélisation d'une base de données

---

### Introduction :

Une base de données est un ensemble d'informations regroupées et organisées dans un objectif commun. Elle peut être gérée manuellement (fiches, dossiers, archives, microfilms, etc.) ou informatisée (tableurs, fichiers). Les évolutions continues que l'on observe dans le domaine du numérique s'appuient sur des bases informatisées amenées à gérer toujours plus de données.

C'est dans ce contexte que nous allons aborder comment, à partir du monde réel, concevoir « sur le papier » une base de données.

Ce processus se décompose en trois étapes :

- la première consiste en une analyse de l'existant ;
- à l'issue de cette étape, on établit un schéma conceptuel, résultat d'une modélisation du monde réel ;
- la dernière étape est la transformation du schéma conceptuel obtenu en une représentation exploitable par le système informatique (qui sera chargé de sa mise en œuvre). De cette étape résulte un schéma relationnel.

## 1 | L'analyse de l'existant

Penchons-nous sur la première étape relative à l'analyse de l'existant. Il faut naturellement commencer par le recueil des informations permettant de décrire le monde réel.

### a. Recueil des informations du monde réel

Il s'agit d'**identifier les différents « objets »** du monde réel concernés par le projet de base de données et les liens qui les unissent. Pour ce faire, on procède à des entretiens, des visites sur place, et à l'examen des documents et fichiers existants.



Nous utiliserons tout au long du cours l'exemple de la base de données d'un club de cuisine.



On s'intéresse à la base de données d'un club de cuisine :

- Des **adhérents** proposent des recettes.
- Ces **recettes** nécessitent des ingrédients d'un type donné, dans des quantités données, exprimées dans des unités données.

En étant attentif·ve aux sujets, verbes et compléments de chacune de ces phrases, on met facilement en évidence les objets suivants : adhérent, recette, ingrédient.

On détecte par ailleurs les liens ci-dessous.

- Entre adhérent et recette : un adhérent propose une recette ; une recette est proposée par un adhérent.
- Entre recette et ingrédient : une recette fait appel à un ingrédient ; un ingrédient intervient dans des recettes.

Si l'on veut être en mesure de garder la trace qu'un adhérent donné a proposé une recette donnée et qu'une recette donnée fait appel à un ingrédient donné, il est nécessaire d'identifier ces liens pour la conception de la future base de données.



## Caractérisation des liens

Il convient ensuite de **caractériser les liens identifiés**, c'est-à-dire se demander de quelle manière les objets sont reliés les uns aux autres. Poursuivons avec notre exemple.

- En ce qui concerne le lien entre adhérent et recette, on peut dire qu'une recette est proposée par un et un seul adhérent (il n'y a aucun intérêt à stocker la même recette, avec les mêmes étapes et les mêmes ingrédients, proposée par deux adhérents différents). Dans l'autre sens, on peut affirmer qu'un adhérent peut ou non proposer des recettes.
- Pour ce qui est du lien entre recette et ingrédient, on peut affirmer qu'une recette fait appel à un ou plusieurs ingrédients (il est difficile de ne cuisiner avec aucun ingrédient). Dans l'autre sens, on peut dire qu'un ingrédient, à un moment donné dans la base du club, peut n'être impliqué dans aucune recette ou, au contraire, intervenir dans plusieurs recettes.

### Description des données

Pour achever le recueil de l'existant, on s'attache à décrire les données relatives aux objets comme aux liens. Continuons avec notre exemple.

- Un adhérent est caractérisé par son nom, son prénom, son adresse de messagerie électronique, son adresse postale, son mot de passe, son numéro de téléphone.
- Une recette peut, par exemple, être caractérisée par son nom, son niveau de difficulté (aucune, moyenne, grande), sa catégorie (entrée, plat, dessert), sa durée, ses étapes (étape 1 : séparer les blancs des jaunes, étape 2 : monter les blancs en neige...).
- Un ingrédient est quant à lui caractérisé par son nom.
- Le lien entre recette et ingrédient qui nous indique qu'une recette peut faire appel à un ingrédient doit également préciser le type d'ingrédient (viande, légume, fruit, poisson etc.), dans quelle quantité et dans quelle unité cette quantité est exprimée.

Les données relatives aux adhérent·e·s de l'association constituent des données à caractère personnel. Tout traitement de données personnelles doit s'effectuer dans le respect des dispositions prévues par le RGPD (Règlement Général sur la Protection des Données).

À l'issue de ce travail sur l'existant, on peut d'ores et déjà constituer une première version du **dictionnaire de données** de la future base.

d. Le dictionnaire de données

### Dictionnaire de données :

Un dictionnaire de données, ou référentiel de données, recense l'ensemble des données amenées à être stockées en base. Pour chacune d'elles, on précise au minimum son code d'identification, sa description, son type et sa taille.

Voici ce à quoi pourrait ressembler le dictionnaire de données de notre club de cuisine :

Code	Description	Type	Taille
nomAdherent	Nom de l'adhérent	Alphabétique	30 caractères
prenom	Prénom de l'adhérent	Alphabétique	30 caractères
mail	Adresse de messagerie de	Alphanumérique	100 caractères

	l'adhérent		
adresse	adresse de l'adhérent	Alphanumérique	100 caractères
passe	Mot de passe de l'adhérent	Alphanumérique	30 caractères
telephone	Numéro de téléphone	Numérique	< 10000000000
categorie	Catégorie de la recette	Alphabétique	30 caractères
nomRecette	Nom de la recette	Alphabétique	100 caractères
niveau	Niveau de difficulté de la recette	Alphabétique	30 caractères
duree	Durée de la recette	Numérique	< 32767 minutes
etapes	Étapes de la recette	Alphanumérique	< 10000 caractères
nomIngredient	Nom de l'ingrédient	Alphabétique	30 caractères
quantite	Quantité de l'ingrédient dans la recette	Numérique	< 10000
unite	Unité de la quantité de l'ingrédient	Alphabétique	30 caractères

type	Type d'ingrédient	Alphabétique	30 caractères
------	-------------------	--------------	---------------

Notons au passage que nous avons précisé dans ce dictionnaire :

- un **code** qui sera adopté par la suite en lieu et place de la description de la donnée (plus longue) ;
- le **type** et la **taille** de la donnée qui conditionneront son codage futur en base.



Attention

Le code doit être dépourvu de caractères accentués et de caractères spéciaux (on écrira par exemple « duree » et pas « durée »).



À retenir

Le fait de préciser le type, la taille, et, le cas échéant, les valeurs possibles prises par une donnée définit son **domaine**. Seules les valeurs appartenant à ce domaine seront permises.

Il est temps maintenant de modéliser le monde réel.

## 2 | Le modèle entité-association

Afin de représenter l'ensemble de nos objets et leurs liens, on a recours à un formalisme particulier qui sera une aide précieuse dans cette démarche. Les formalismes les plus employés aujourd'hui à cette fin sont le **diagramme de classes** de l'approche UML et, dans une moindre mesure, le **modèle conceptuel de données** (MCD) de l'approche entité-association. Dans ce cours, nous ferons appel au second (MCD). Commençons par en étudier les notions principales.



### a. Les notions de base

## Entité :

Le terme « entité » sert à désigner un objet réel, concret ou abstrait, identifié lors de l'analyse de l'existant.

Elle est représentée par un rectangle, doté d'un titre. Elle est constituée d'attributs.

→ Dans le cas de notre club de cuisine, « recette » et « adhérent » sont des entités.

## Occurrence :

L'occurrence désigne un exemplaire (un représentant) d'une entité.

Si l'on applique le terme à notre club de cuisine :

- « mousse au chocolat » est une occurrence de l'entité « recette » ;
- « sucre » est une occurrence de l'entité « ingrédient » ;
- « Dupont Jacques » est une occurrence de l'entité « adhérent ».

## Attribut :

L'attribut désigne une donnée élémentaire qui sert à caractériser une propriété d'une entité ou d'une association.

Dans notre étude de cas, « duree » est l'un des attributs de l'entité « recette ».

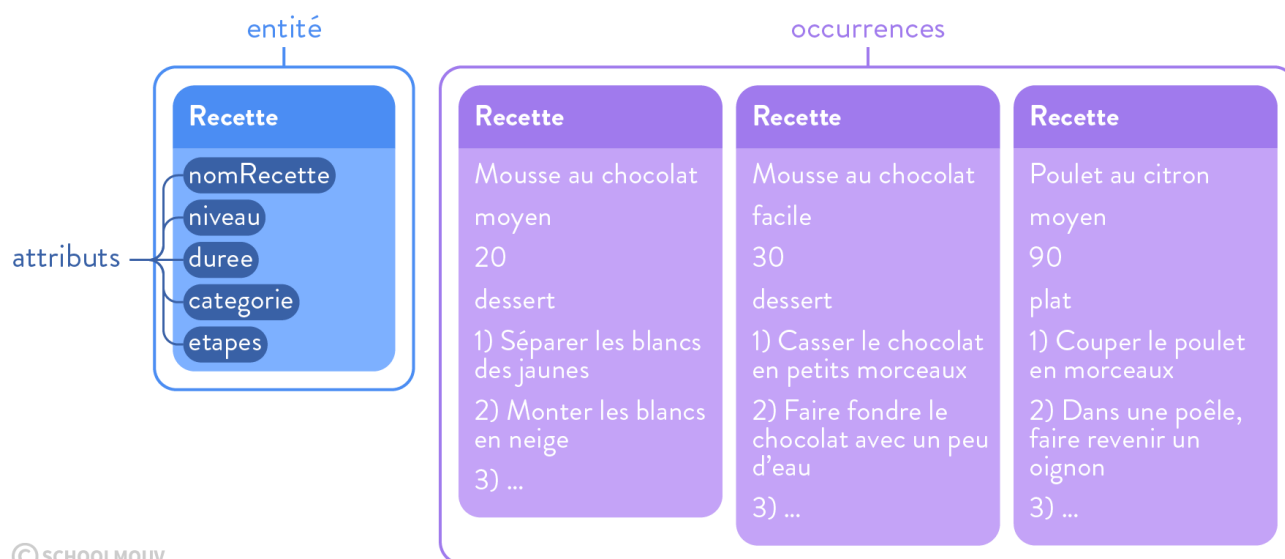
## Identifiant :

L'identifiant est un attribut, ou un ensemble d'attributs, qui permet d'identifier sans ambiguïté chaque occurrence d'une entité.



Une entité doit nécessairement être dotée d'un identifiant.

Voici, ci-dessous, la représentation de l'entité « recette » et de trois de ses occurrences.



© SCHOOLMOUV

Dans l'exemple ci-dessus, nous pouvons voir qu'il y a deux recettes de mousse au chocolat, l'une facile et plus longue, l'autre plus rapide mais plus difficile. Or, ces deux recettes portent le même nom. Ce n'est donc pas via l'attribut nom qu'il sera possible de les différencier ni, donc, de les identifier.

→ Il est par conséquent nécessaire d'ajouter un attribut à l'entité « recette » qui lui servira d'identifiant. Nommons-le « idRecette ».

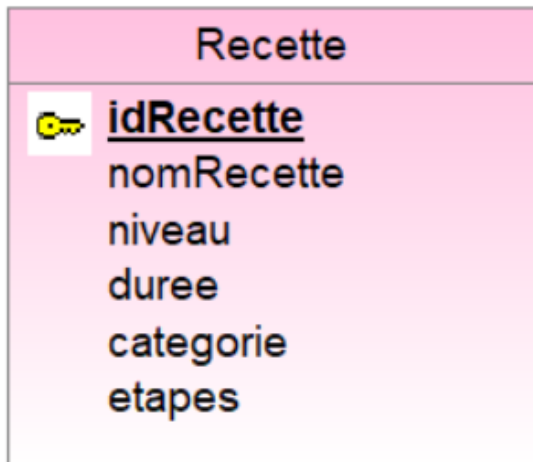
On peut noter au passage que notre dictionnaire de données s'enrichira d'une nouvelle donnée : idRecette. Cet identifiant prendra la forme d'un **numéro unique**.



Pour des raisons d'efficacité dans les traitements qui seront développés par la suite sur les données de la base, il est recommandé de privilégier



un identifiant aussi simple que possible, de préférence numérique.



Voici donc notre entité « recette » corrigée en conséquence. Notons que, selon la convention adoptée dans les MCD, on reconnaît l'identifiant en repérant l'attribut qui est souligné.

Nous avons vu que les entités peuvent faire l'objet de liaisons entre elles. Ces liaisons sont appelées « **associations** » dans l'approche entité-association. Nous avons également vu qu'elles peuvent être elles-mêmes porteuses de données, c'est-à-dire d'attributs. En effet, l'association impliquant l'entité « recette » et l'entité « ingrédient » dispose des attributs « quantité », « unité » et « type ».



### Définition

#### Association :

Une association correspond au lien qui existe entre deux entités. Elle peut être composée ou non d'attributs.

On la représente dans un MCD par un segment reliant les deux entités concernées. Une ellipse y est tracée en son milieu : elle contient un terme pour la désigner (généralement un verbe), ainsi que les éventuels attributs qui la caractérisent.

À chaque extrémité de ce segment, on précise les **cardinalités** de l'association.



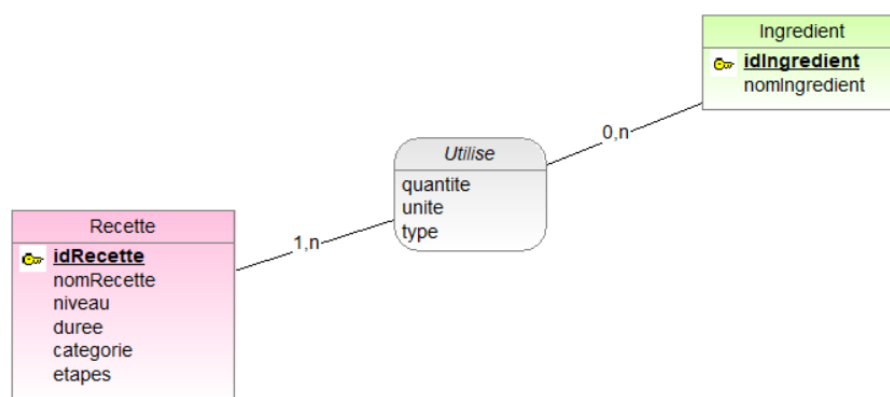
### Définition

## Cardinalités :

Les cardinalités d'une association entre deux entités précisent ses caractéristiques.

Dans la pratique, on détermine deux nombres qui correspondent à la quantité minimale et maximale d'occurrences relatives aux entités impliquées dans l'association, dans un sens et dans l'autre.

Dans notre exemple, une recette fait appel à **1** ou plusieurs ingrédients. Et un ingrédient intervient dans **0** ou **+** de recettes. Cela se traduit par les cardinalités formalisées comme suit dans notre MCD.



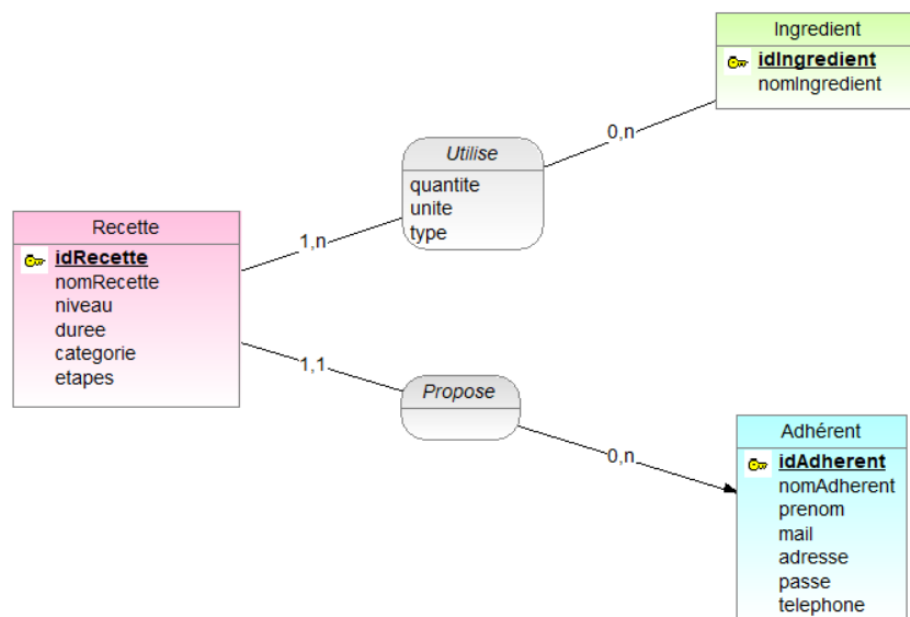
## Codification des cardinalités :

- 0, 1 veut dire **aucun** ou 1 ;
- 1, 1 veut dire **1** et 1 seul ;
- 0, n (ou 0, N) veut dire **0** ou + ;
- 1, n (ou 1, N) veut dire **1** ou + ;
- a, b au minimum a, et au plus b.

On notera au passage que, dans notre exemple, nous avons fait apparaître un nouvel attribut : idIngrédient. Nous avons en effet décidé d'affecter un attribut qui sera numérique et unique pour identifier sans ambiguïté un ingrédient, plutôt que d'utiliser son nom.

## b. Mise en pratique

Il s'agit maintenant de représenter dans le MCD l'ensemble des entités de la base de données, accompagnées de leurs attributs (dont leur identifiant). On formalisera également les associations impliquant les entités, accompagnées de leurs cardinalités et de leurs éventuels attributs. Dans notre cas, nous avons identifié les entités suivantes : adhérent, recette, ingredient. Deux associations ont été recensées : l'une entre adhérent et recette, l'autre entre et recette et ingredient. Il en résulte le MCD suivant :



Maintenant que nous avons modélisé les données du monde réel selon un MCD, il va nous falloir le traduire en modèle relationnel. C'est un modèle qui permet d'organiser les données dans une base informatisée.

## 3 | Le modèle relationnel

Nous verrons plus loin, dans le cours consacré aux SGBD, qu'il existe plusieurs modèles de données possibles pour traduire un MCD : hiérarchique, réseau, relationnel ou relationnel-objet.



Nous adoptons ici le modèle relationnel qui est le plus répandu de nos jours et qui a maintes fois fait ses preuves.

## a. Principes du modèle relationnel



### Définition

#### Modèle de données :

Un modèle de données définit la manière dont l'information est structurée dans la base de données. Il est une représentation d'informations du monde réel, exploitable par l'informatique.

Le **modèle relationnel** s'appuie sur la notion de « relation », d'où son nom. On représente une relation par une **table**. Les lignes de cette table sont nommées « tuples », et ses colonnes sont nommées « champs ».

Diagram illustrating a relation (table) with columns (champs) and rows (tuples).

idRecette	nomRecette	niveau	duree	categorie	etapes
1	Mousse au chocolat	moyen	20	dessert	1) Séparer les blancs des jaunes. 2) Monter les blancs en neige. 3) ...
2	Mousse au chocolat	facile	30	dessert	1) Casser le chocolat en petits morceaux. 2) Faire fondre le chocolat avec un peu d'eau. 3) ...
3	Poulet au citron	moyen	90	plat	1) Couper le poulet en morceaux. 2) Dans une poêle, faire revenir un oignon haché. 3) ...
4	Soupe à l'oignon	facile	45	entrée	1) Éplucher et couper les pommes de terre en morceaux. 2) Faire de même pour les oignons. 3) ...

© SCHOOLMOUV

Cette relation peut s'écrire sous la forme  $\rightarrow \text{Recette}(\text{idRecette}, \text{nomRecette}, \text{niveau}, \text{duree}, \text{categorie}, \text{etapes})$ .

→ C'est ce qu'on appelle le schéma relationnel de la relation recette.

On doit pouvoir accéder à un tuple donné de la relation par le contenu d'un ou plusieurs de ses champs identifiant de manière unique ce tuple. Le ou les champs en question constituent une **clé**.

Une relation peut comprendre plusieurs clés possibles. On les nomme « clés candidates ». On nommera plus précisément « clé primaire » celle d'entre elles qui est choisie. Les clés primaires de l'ensemble des tuples sont stockées dans un index (sorte de répertoire) qui indique l'emplacement où trouver chacun des tuples qui constituent la relation.



La clé primaire doit contenir le minimum de champs possible.

Maintenant que nous savons comment est structuré le modèle relationnel, étudions comment il est conçu à partir du modèle conceptuel qui, rappelons-le, est une vue tout à fait abstraite de la future base de données.

## b. Passage du modèle conceptuel au modèle relationnel

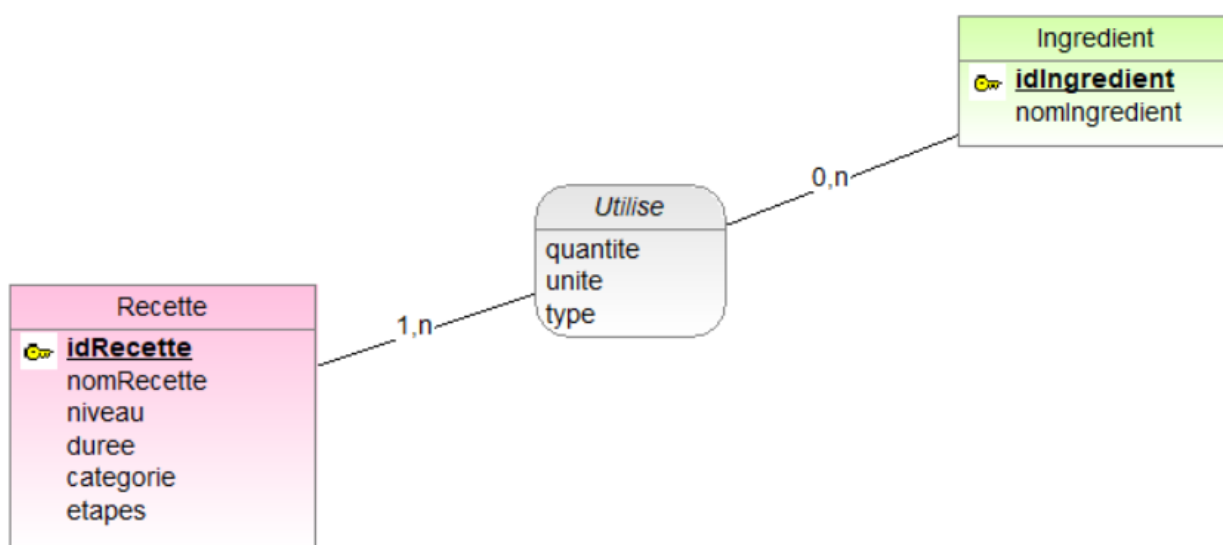
Nous disposons à ce stade d'un MCD, constitué d'entités, d'attributs et d'associations.

Deux règles principales doivent être observées pour passer au modèle relationnel.

- 1 Une entité du MCD devient une relation du modèle relationnel. Ses attributs deviennent les **champs** de cette relation ainsi créée. L'identifiant de l'entité devient la **clé** de la relation dont elle est à l'origine.

- 2 Une association du MCD devient une **relation**. Les champs qui la composent sont les **identifiants** des entités impliquées et, le cas échéant, les **attributs** qui caractérisent l'association. La clé de cette relation ainsi créée est composée des deux identifiants évoqués à l'instant.

Appliquons ces règles à une partie du MCD relatif à la base de données du club de cuisine. En observant les entités Recette et Ingredient et l'association qui les implique, on obtient les schémas relationnels suivants :



Recette(idRecette, nomRecette, niveau, duree, categorie, etapes) avec

idRecette : clé primaire

Ingredient(idIngredient, nomIngredient) avec

idIngredient : clé primaire

Utilise(idRecette, idIngredient, quantite, unite, type) avec

idRecette, idIngredient : clé primaire

idRecette : clé étrangère en référence à idRecette dans recette

idIngredient : clé étrangère en référence à idIngredient dans ingredient

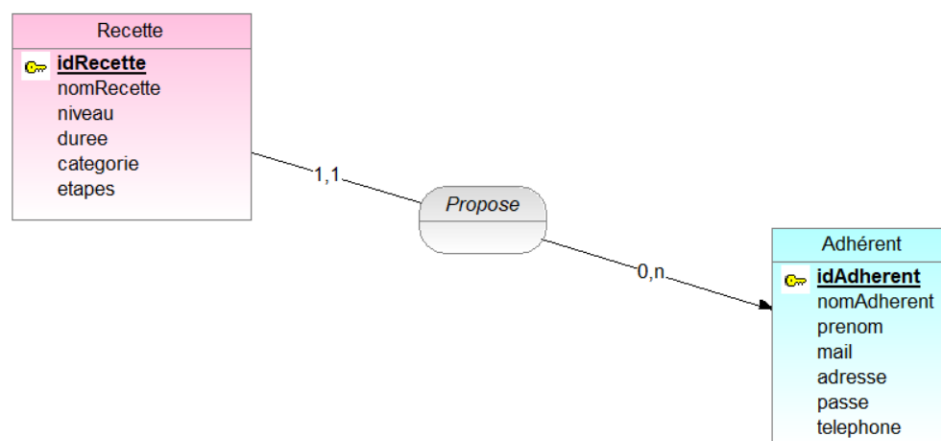
La notion de **clé étrangère** apparaît dans le schéma relationnel de Utilise : cela signifie que ce sont des champs qui prennent forcément leurs valeurs parmi celles prises par la clé qui s'y rapporte. idRecette dans Utilise ne peut prendre ses valeurs que parmi celles prises par la clé idRecette dans Recette. Autrement dit, on ne peut pas trouver, dans la relation Utilise, un

tuple qui se rapporterait à une recette inexistante dans la relation recette ou un ingrédient inexistant dans la relation ingredient. C'est ce qui contribue à maintenir la cohérence des données au sein de la base.



Un cas particulier doit néanmoins être traité différemment : il s'agit de celui d'une association qui comporte une cardinalité 1, 1. Dans cette situation, on ne crée pas une relation pour une telle association. Tous les champs de l'association sont simplement intégrés à la relation issue de l'entité associée à la cardinalité 1, 1.

Pour illustrer ce cas, prenons l'exemple suivant : une recette est proposée par 1 et 1 seul adhérent, et un adhérent peut proposer aucune ou plusieurs recettes. Le MCD qui en résulte est le suivant :



Les schémas relationnels qui en découlent sont alors :

Adherent(idAdherent, nomAdherent, prenom, mail, adresse, passe, telephone) avec

idAdherent : clé primaire

Recette(idRecette, nomRecette, niveau, duree, categorie, etapes, idAdherent) avec

idRecette : clé primaire

idAdherent : clé étrangère

L'association modélisée dans le MCD impliquant les entités recette et adhérent n'a pas donné lieu à la création de la relation Propose(idRecette, idAdherent). Les champs qui la constituent ont été intégrés à la relation Recette.

On remarque que parmi ces champs, idRecette est déjà présent dans Recette.

Voici une illustration des relations obtenues sous forme de tableau, qui ne sont pas exhaustifs. Vous pourrez constater que les clés étrangères (mises en évidence dans les colonnes colorées) ne prennent que des valeurs parmi celles prises par les clés primaires auxquelles elles se réfèrent.

### Adherent

idAdherent	nomAdherent	prenom	mail	adresse	passerelle
1	Dupont	Jacques	dup@orange.fr	3 rue blanche 23000 Gueret	
2	Dampierre	Gérard	gege@free.fr	7 rue Dumont 29000 Brest	
3	Forestier	Denise	dd@gmail.com	9 rue aux loups 12000 Rodez	
4	Vielle	Rose	vr@sfr.fr	1 av. des prés 10000 Troyes	



## Recette

idRecette	nomRecette	niveau	duree	categorie	etapes	idAdhe
1	Mousse au chocolat	moyen	20	dessert	1) Séparer les blancs des jaunes. 2) Monter les blancs en neige. 3) ...	2
2	Mousse au chocolat	facile	30	dessert	1) Casser le chocolat en petits morceaux. 2) Faire fondre le chocolat avec un peu d'eau. 3) ...	3
3	Poulet au citron	moyen	90	plat	1) Couper le poulet en morceaux. 2) Dans une poêle, faire	2

					revenir un oignon haché. 3) ...	
4	Soupe à l'oignon	facile	45	entrée	1) Éplucher et couper les pommes de terre en morceaux.  2) Faire de même pour les oignons.  3) ...	1

## Ingredient

idIngredient	nomIngredient
1	oeuf
2	chocolat
3	sucre
4	beurre
5	oignon
6	poulet

7	pommes de terre
---	-----------------

## Utilise

idRecette	idIngredient	quantite	unite	type
1	1	6	oeuf	viande
1	2	200	gramme	epicerie
1	3	3	cuillère à soupe	epicerie
2	2	1	tablette	epicerie
2	1	6	oeuf	viande
3	6	1	kg	viande
3	5	2	gros oignon	legume
4	7	350	g	legume
4	5	5	oignon moyen	legume
4	4	2	cuillère à café	laitage

Cela nous donne un aperçu de la manière dont les données de notre base y sont concrètement organisées.

### Conclusion :

Nous savons maintenant modéliser le monde réel selon un MCD et transcrire un MCD en un modèle relationnel, également dénommé « modèle logique ». Ce modèle logique permet la création concrète de la structure de la base donnée. Cependant, il est important, avant la création de cette structure, de s'assurer que le modèle relationnel ne présente pas d'anomalies. Nous allons ainsi découvrir dans le cours qui suit le principe de la normalisation et des contraintes d'intégrité.

