



## BASES DE DONNÉES

---

# Données de production d'articles scientifiques

---

Projet (partie 1)

Maxime MEURISSE (20161278)

3<sup>e</sup> année de Bachelier Ingénieur civil  
Année académique 2018-2019

# 1 Diagramme entité-relation

Le diagramme entité-relation est fourni dans le fichier `diagramme.pdf` joint à ce document.

Les codes de mise en forme utilisés sont ceux du cours, à quelques exceptions près <sup>1</sup> :

- les attributs uniques sont soulignés ;
- les attributs uniques qui sont des clés sont soulignés et en gras ;
- les entités et relations faibles sont représentées par une bordure très épaisse.

## 1.1 Clés

### 1.1.1 Entités

Les clés des entités sont reprises dans la table 1.

Entité	Clé
Auteur	Matricule
Conférence	Nom_conf + Année_production
Article_conférence	URL
Article_scientifique	URL
Article_journal	URL
Journal	Numéro_publication + rôle « rev » de Publie_j
Revue	Nom_rev

Table 1 – Clés des entités.

Notons que les 3 entités articles ont la même clé car Article\_conférence et Article\_journal sont des enfants de Article\_scientifique (relation « is a »).

### 1.1.2 Relations

Les clés des relations sont reprises dans la table 2.

---

1. Le programme utilisé ne permettait pas de respecter scrupuleusement le code de mise en forme du cours.

Relation	Clé
Assiste	Rôle « aut » de Assiste + rôle « conf » de Assiste
Publie_a	Rôle « art » de Publie_a
Écrit	Rôle « art » de Écrit
Co-écrit	Rôle « aut » de Co-écrit + rôle « art » de Co-écrit
Contient	Rôle « art » de Contient
Publie_j	Rôle « journ » de Publie_j

Table 2 – Clés des relations.

## 1.2 Contraintes de cardinalité

Il s'agit de préciser, pour chaque ensemble d'entités participant à une relation, dans combien de tuples chaque entité peut apparaître.

### 1.2.1 « is a » des articles

La cardinalité de Article\_scientifique est (1,1) : un article est forcément soit un Article\_conférence soit un Article\_journal, et ne peut être que l'un des deux.

### 1.2.2 Assiste

La cardinalité de Auteur est (0,N) : un auteur peut n'assister à aucune conférence mais peut également assister à une ou plusieurs conférences.

La cardinalité de Conférence est (0,N) : une conférence peut être assistée par aucun auteur mais peut également être assistée par un ou plusieurs auteurs.

### 1.2.3 Écrit

La cardinalité de Auteur est (1,N) : un auteur écrit au minimum un article (sinon il ne serait pas qualifié d'auteur) et peut également en écrire plusieurs.

La cardinalité de Article\_scientifique est (1,1) : un article n'est écrit que par un et un seul premier auteur. Tout autre auteur est un auteur secondaire.

### 1.2.4 Co-écrit

La cardinalité de Auteur est (0,N) : dans cette relation, Auteur représenté un second auteur. Un auteur peut ne jamais avoir été second auteur pour un article, ou peut être plusieurs fois second auteur (pour plusieurs articles différents).

La cardinalité de Article\_scientifique est (0,N) : un article peut n'avoir été écrit par aucun second auteur (uniquement un premier auteur) ou par plusieurs seconds auteurs.

### 1.2.5 Publie\_a

La cardinalité de Conférence est  $(0,N)$  : une conférence peut ne pas publier d'article mais peut également en publier plusieurs.

La cardinalité de Article\_conférence est  $(1,1)$  : un article de conférence est forcément publié par une conférence, et n'est publié que par cette conférence.

### 1.2.6 Contient

La cardinalité de Journal est  $(1, N)$  : un journal contient au minimum un article de journal, mais peut en contenir plusieurs.

La cardinalité de Article\_journal est  $(1,1)$  : un article est forcément contenu dans un journal, et n'est contenu que dans ce journal.

### 1.2.7 Publie\_j

La cardinalité de Revue est  $(1,N)$  : une revue publie au moins un journal, mais peut en publier plusieurs.

La cardinalité de Journal est  $(1, 1)$  : un journal est forcément publié par une revue, et n'est publié que par cette revue.

### 1.2.8 Remarques

Avec les cardinalités actuelles, il n'est pas possible d'encoder certaines entités dans la base de données. Par exemple, pour encoder un journal, il faut encoder une revue, mais pour encoder une revue, il faut encoder un journal (on tourne en rond).

En pratique, ce problème peut être résolu en mettant certaines cardinalités minimum à 0, notamment :

- Auteur (dans la relation Écrit) :  $(0,N)$  ;
- Revue (dans la relation Publie\_j) :  $(0,N)$  ou Journal (dans la relation Contient) :  $(0,N)$ .

## 1.3 Contraintes d'intégrité additionnelles

Quelques contraintes d'intégrités supplémentaires sont à souligner :

- Conférence : Numéro doit être positif, Nom\_rue et Ville doivent exister
- Article\_scientifique : URL doit être au format «  $(\text{http(s) ://})(\text{www.})\langle \text{nom de domaine} \rangle$  », DOI doit être un numéro
- Article\_journal : Page\_début doit être plus petit ou égal à Page\_fin
- Journal : Numéro\_publication doit être un numéro
- Revue : Facteur\_impact doit être un nombre positif

On considère également que tous les attributs représentant une date doivent être inférieur à la date du jour. En effet, on suppose que rien n'est encodé à l'avance dans la base de données; tout ce qui est encodé a déjà eu lieu ou est publié.

## 1.4 Relations et entités faibles

Dans ce modèle, Journal est une entité faible (avec la relation faible Publie\_j). En effet, le numéro de publication identifie le journal de manière unique *au sein de la revue qui le publie*. Pour être identifié complètement de manière unique, il faut donc connaître le numéro de publication du journal et le nom de la revue qui le publie.

# 2 Modèle relationnel

## 2.1 Conversion vers le modèle relationnel

On commence par traduire le diagramme entité-relation en un modèle relationnel.

### 2.1.1 Ensemble d'entités « normaux »

- Auteur(Matricule, Nom\_aut, Prénom\_aut, Date\_doctorat)
- Conférence(Nom\_conf, Année\_production, Numéro, Nom\_rue, Ville)
- Revue(Nom\_rev, Facteur\_impact)
- Sujets\_recherche(#URL, Ordre, sujet\_recherche)

### 2.1.2 Ensemble d'entités faibles

- Journal(Numéro\_publication, #Nom\_rev)

### 2.1.3 Ensemble d'entités liés par une relation « is a »

- Article\_scientifique(URL, DOI, Titre, Date\_publication, #Matricule)
- Article\_conférence(URL, Façon\_présenté, #Nom\_conf, #Année\_production)
- Article\_journal(URL, Page\_début, Page\_fin, #Numéro\_publication, #Nom\_rev)

### 2.1.4 Relations

- Assiste(#Matricule, #Nom\_conf, #Année\_production, Tarif\_appliqué)
- Co-écrit(#Matricule, #URL)

Les relations Publie\_a, Écrit, Contient et Publie\_j ne sont pas conservées dans le modèle. En effet, celles-ci sont à chaque fois liées à une entité dont la cardinalité maximum est 1.

## 2.2 Dépendances fonctionnelles et à valeurs multiples

### 2.2.1 Dépendances fonctionnelles

La notion de clé est un cas particulier de dépendance fonctionnelle : un ensemble d'attributs  $X$  sera une clé d'une relation  $R$  par rapport à un ensemble fonctionnel  $F$  si  $X$  est un ensemble minimum tel que  $F \models X \rightarrow R$ .

On a donc, pour chacune des relations définies précédemment,

- Auteur : Matricule  $\rightarrow$  R
- Conférence : Nom\_conf, Année\_production  $\rightarrow$  R
- Revue : Nom\_revue  $\rightarrow$  R
- Sujets\_recherche : URL, Ordre  $\rightarrow$  R
- Article\_scientifique : URL  $\rightarrow$  R
- Article\_conférence : URL  $\rightarrow$  R
- Article\_journal : URL  $\rightarrow$  R
- Assiste : Matricule, Nom\_conf, Année\_production  $\rightarrow$  R

Pour les relations Journal et Co-écrit, l'ensemble des attributs-clés correspond à l'ensemble de tous les attributs. Il y a donc peu d'intérêt à dire que ces attributs forment une dépendance fonctionnelle.

On obtient donc la fermeture de l'ensemble des dépendances de chacune des relations. Il est à noter que ces fermetures contiennent en plus les dépendances triviales et composées.

**Remarque** On peut également noter qu'on a la dépendance Article\_{scientifique, conférence, journal} : DOI  $\rightarrow$  R. En effet, comme mentionné dans l'énoncé, le DOI est un numéro unique associé à un article. On ne peut donc pas rencontrer plusieurs tuples qui ont des DOI identiques pour un article, mais des URL différentes.

### 2.2.2 Dépendances à valeurs multiples

On sait qu'une dépendance  $X \rightarrow A_1 A_2 \dots A_n$  est équivalente à l'ensemble des dépendances  $X \rightarrow A_1, X \rightarrow A_2, \dots, X \rightarrow A_n$ .

De plus, par la propriétés de reproduction, toutes les dépendances fonctionnelles sont également des dépendances à valeurs multiples.

On a donc,

- $D_{\text{Auteur}} = \{\text{Matricule} \twoheadrightarrow \text{Nom\_aut}, \text{Matricule} \twoheadrightarrow \text{Prénom\_aut}, \text{Matricule} \twoheadrightarrow \text{Date\_doctorat}\}$
- $D_{\text{Conférence}} = \{\text{Nom\_conf}, \text{Année\_production} \twoheadrightarrow \text{Numéro}, \text{Nom\_conf}, \text{Année\_production} \twoheadrightarrow \text{Nom\_rue}, \text{Nom\_conf}, \text{Année\_production} \twoheadrightarrow \text{Ville}\}$

- $D_{\text{Revue}} = \{\text{Nom\_revue} \rightarrow\rightarrow \text{Facteur\_impact}\}$
- $D_{\text{Sujets\_recherche}} = \{\text{URL}, \text{Ordre} \rightarrow\rightarrow \text{sujet\_recherche}\}$
- $D_{\text{Article\_scientifique}} = \{\text{URL} \rightarrow\rightarrow \text{DOI}, \text{URL} \rightarrow\rightarrow \text{Titre}, \text{URL} \rightarrow\rightarrow \text{Date\_publication}, \text{URL} \rightarrow\rightarrow \text{Matricule}\}$
- $D_{\text{Article\_conférence}} = D_{\text{Article\_scientifique}} + \{\text{URL} \rightarrow\rightarrow \text{Façon\_présenté}, \text{URL} \rightarrow\rightarrow \text{Nom\_conf}, \text{URL} \rightarrow\rightarrow \text{Année\_production}\}$
- $D_{\text{Article\_journal}} = D_{\text{Article\_scientifique}} + \{\text{URL} \rightarrow\rightarrow \text{Page\_début}, \text{URL} \rightarrow\rightarrow \text{Page\_fin}, \text{URL} \rightarrow\rightarrow \text{Numéro\_publication}, \text{URL} \rightarrow\rightarrow \text{Nom\_revue}\}$
- $D_{\text{Assiste}} = \{\text{Matricule}, \text{Nom\_conf}, \text{Année\_production} \rightarrow\rightarrow \text{Tarif\_appliqué}\}$

Pour les relations Journal et Co-écrit, la remarque est similaire à celle faite pour les dépendances fonctionnelles.

Ces dépendances à valeurs multiples ont été déduites des propriétés des dépendances à valeurs multiples. Elles correspondent donc à des dépendances dérivées.

Ces ensembles contiennent également les dépendances triviales.

## 2.3 La 4<sup>e</sup> forme normale

Une possibilité pour avoir un schéma de relation en 4FN est que pour toute dépendance  $X \rightarrow\rightarrow Y$ ,  $X$  est une super-clé (ou une clé, puisqu'une clé est par définition une super-clé).

Dans le modèle construit précédemment, tous les membres de gauche de l'ensemble des dépendances à valeurs multiples de chaque relation sont des clés. En effet, les dépendances à valeurs multiples ont été déduites des dépendances fonctionnelles, celles-ci ayant été trouvées sur base des clés des relations.

Le modèle présenté est donc en 4FN (et donc par conséquent en BCNF).