

Chapitre 3 – Codification et contrôle des données

Table of Contents

Dans ce chapitre, il s'agit des outils d'analyse:

1. Introduction
2. La codification
3. Définition
4. Types de codifications
5. Les contrôles
6. Conclusion

1. Introduction

- Pourquoi est-il nécessaire de codifier les informations?
 - Pour être traitées par ordinateurs, les informations ont besoin d'être structurés.
 - Cette structuration passe, obligatoirement, par l'association des codes aux différentes informations et concepts manipulés par le système d'information.
- Ces codes vont permettre de désigner chaque information de manière claire et unique.

1. Introduction

Exemple : soit le document suivant :

Bon de commande

Numéro commande :.....

Date commande :.....

Numéro client :.....

Nom client :.....

Adresse client :.....

Référence	Désignation	Quantité

Les données qu'on peut extraire de ce document sont :

- Numéro commande
- Date commande
- Numéro client
- Nom client
- Adresse client
- Référence produit
- Désignation produit
- Quantité produit

2. La codification

- ❖ Nous voyons que les désignations de ces données sont trop longues et donc très lourdes à manipuler. Donc, il faut les abréger sans perdre leurs significations.

Exemple :

- **numéro commande** → num_cde
- **date commande** → date_cde
- **numéro client** → num_cl
- **nom client** → nom_cl
- **quantité produit** → qte_prod

- ❖ on a associé un code à chaque donnée du document et cela en la désignant par un nom abrégé unique permettent de la distinguer parmi les autres données.

2. La codification

Exemple:

- Les données qu'on peut extraire de ce document précédent sont:
 - Numéro commande
 - Date commande
 - Numéro client
 - Nom client
 - Adresse client
 - Référence produit
 - Désignation produit
 - Quantité commande

Numéro commande → Num_C

Date commande → Date_C

Numéro client → Num_CI

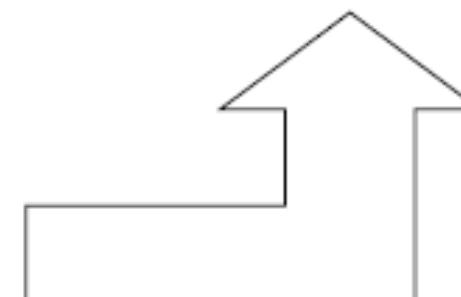
Nom client → Nom_CI

Exemple de désignation abrégé

Désignation trop longue et donc très lourdes à manipuler



mieux les abréger sans perdre leur significations.



2. La codification

Définition d'un code : est un nom abrégé ou une représentation de l'information permettant de désigner une propriété ou un concept de manière claire et unique.

Exemples

- Codes des concepts

Code d'Etudiant → Code_E

- Codes des objets

Un étudiant (bac 2008) → 2018/0049

Un autre étudiant (bac 2017) → 2017/0033

2. La codification

Définition d'une codification :

c'est l'opération qui consiste à remplacer une information, sous sa forme, naturelle par un code clair qui serait mieux adapté aux besoins de l'utilisateur de l'information.

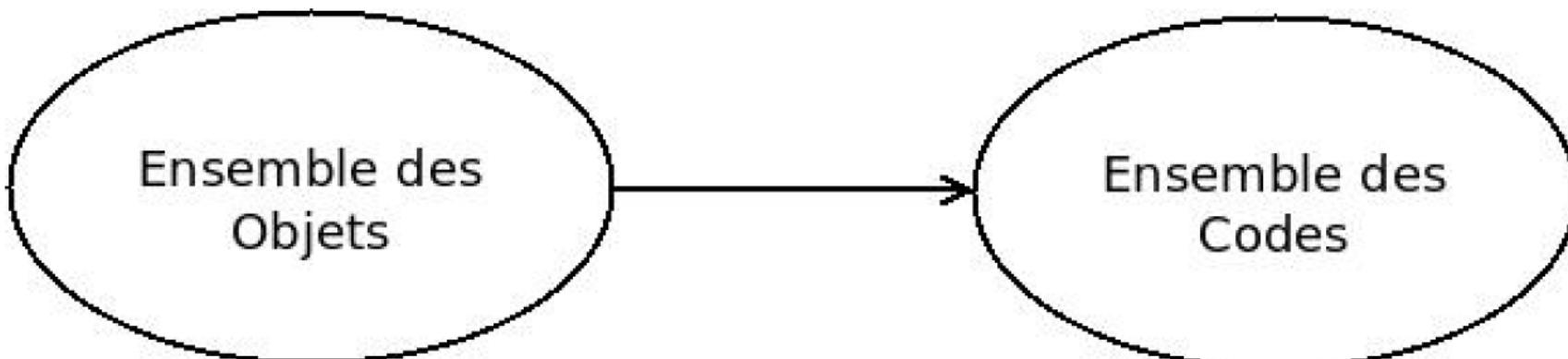


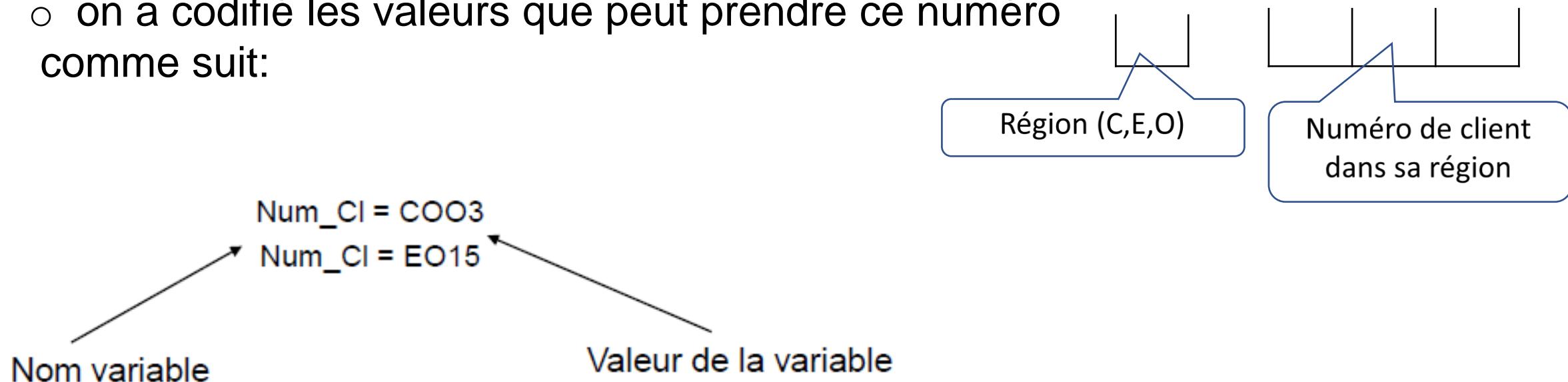
Figure 1. Codification

2. La codification

- ❖ Associer un concept issu du monde réel (système existant) à un code (issu du monde informatique) veut dire qu'il est désormais possible de le stocker, le traiter et le partager. Cela est vrai aussi pour les objets liés au concept.
 - ❖ La codification porte sur le nom de l'information à codifier, mais aussi, sur sa valeur.

Exemple :

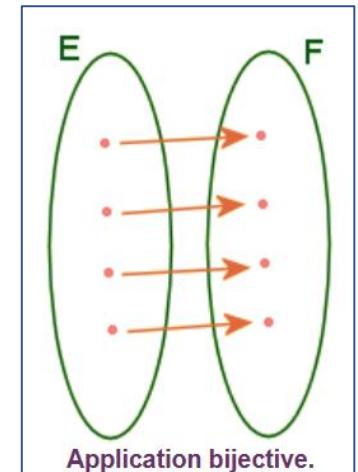
- Pour le numéro client, on a le code : num_cl
 - on a codifié les valeurs que peut prendre ce numéro comme suit:



2. La codification: 2.2 Caractéristique d'une codification

a. Non ambigu :

une codification est non ambiguë si elle associe à chaque objet un code et un seul, et à chaque code un objet et un seul (Bijection entre l'ensemble des codes et l'ensemble des objets)



b. Adaptation aux besoins des utilisateurs :

cela nécessite : **Simplicité de la fonction de codification**

Exemple:

Num Etud → numéro d'étudiant, Le cas le plus simple :

fonction de correspondance (explicite) Facilité d'interprétation des codes (pour retrouver les objets) → fonction de décodification facile à déterminer.

2. La codification: 2.2 Caractéristique d'une codification

c. La concision :

l'intérêt d'une codification est d'éviter d'avoir à manipuler des informations trop langue, pour être efficace un code doit donc comporter le moins de caractères possibles.

d. Possibilité d'extension et d'insertion :

- **Extension** : l'ensemble des objets codifiés peut s'accroître sans remettre en cause la codification choisie.
- **Insertion** : les objets nouveaux peuvent s'insérer entre les objets existants sans remettre en cause la codification choisie

2. La codification: 2.2 Caractéristique d'une codification

Exemples:

Exemple 01

Dans le cadre de réalisation d'un système de gestion pour l'université, un premier concepteur modélisait la partie concernant les étudiants. Il a proposé :

Code Etudiant → Code_E

Dans le cadre du même projet, un autre concepteur travaillait sur la partie qui concerne les employés de l'université. Il a proposé :

Code Employe → Code_E

❖ Dans ce cas, Code_E est un code ambigu.

Exemple 02

Soit les deux codes suivants proposés pour Code Etudiant :

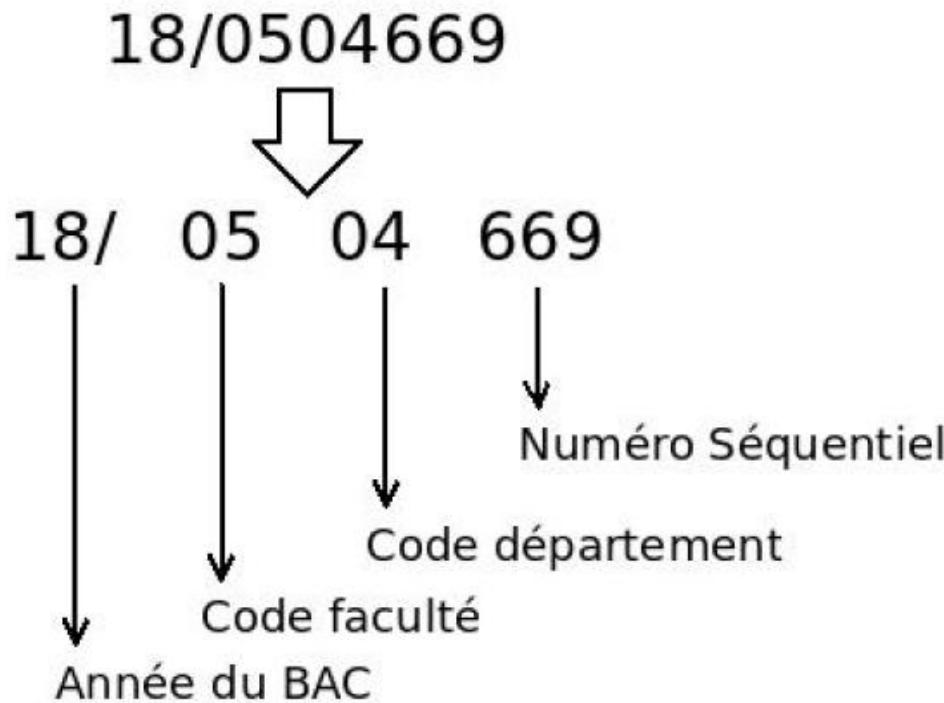


Figure 2. Première proposition pour le code d'étudiant

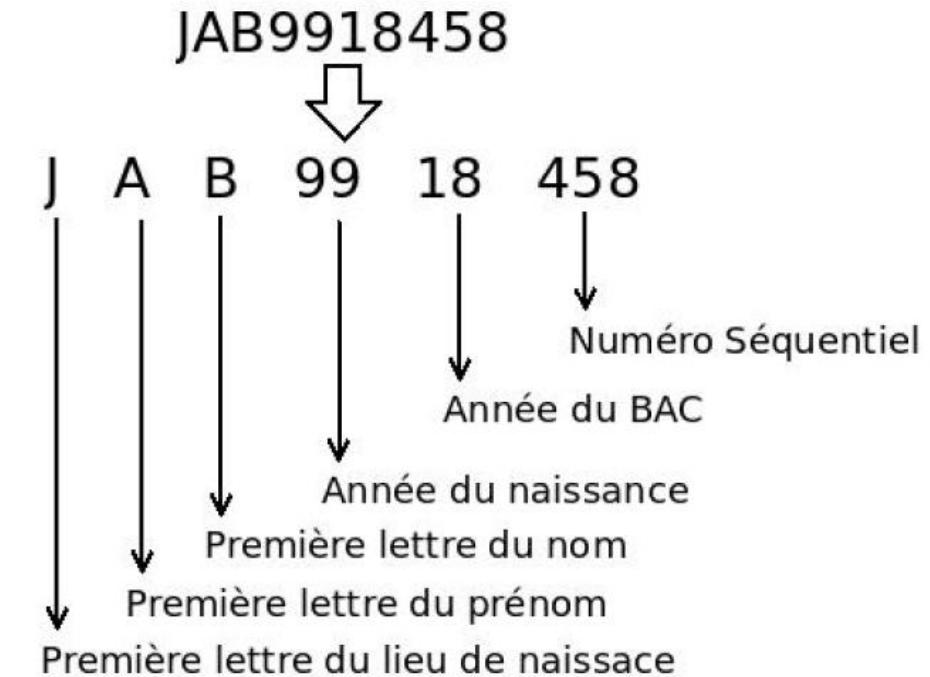


Figure 3. Deuxième proposition pour le code d'étudiant

- ❖ Malgré que les deux codes sont "corrects" vu qu'ils permettent de définir d'une manière précise et unique un étudiant. La première proposition est plus simple et plus facile à manipuler.

3. Types de codifications

a) Codification séquentielle

Associer des numéros séquentiels (consécutifs) aux informations.

Avantages :

- **Non ambiguë**: chaque code (numéro) est unique,
- **Simple à mettre en œuvre**: il suffit de mettre œuvre un compteur,
- **Extension possible**: il est toujours possible d'ajouter des codes.

Inconvénients :

- **Insertion impossible** : il est impossible de créer un nouveau code entre deux codes (deux numéros) consécutifs déjà utilisés,
- **Non significative**: le code n'a aucune signification et sert uniquement à identifier d'une manière unique un objet.

3. Types de codifications

b) Codification par tranche

Associer des tranches de codes aux catégories. Chaque tranche contient des codes séquentiels.

Exemple 1:

Dans une bibliothèque, les ouvrages sont classés par catégories, comme suit: technologie, littérature, sociologie, médecine et culture générale.

La codification des ouvrages peut se faire comme suit :

De 001 à 100 : Technologie

De 101 à 200 : Littérature

De 201 à 300 : Sociologie

De 301 à 400 : Médecine

De 401 à 500 : Culture générale

3. Types de codifications

b) Codification par tranche

Exemple 2:

Les codes de réponse du protocole HTTP utilisent une codification par tranche :

1xx Réponse informationnelle	<ul style="list-style-type: none">• 100 Continue• 101 Switching Protocols• 102 Processing• 103 Early Hints
2xx Succès	<ul style="list-style-type: none">• 200 OK• 201 Created• 202 Accepted• 203 Non-Authoritative Information• 204 No Content
3xx Redirection	<ul style="list-style-type: none">• 300 Multiple Choices• 301 Moved Permanently• 302 Found• 303 See Other• 304 Not Modified

3. Types de codifications: b) Codification par tranche

Exemple 2:

Les codes de réponse du protocole HTTP utilisent une codification par tranche :

4xx Erreur (par le client)	<ul style="list-style-type: none">• 400 Bad Request• 401 Unauthorized• 402 Payment Required• 403 Forbidden• 404 Not Found
5xx Erreur (par le serveur)	<ul style="list-style-type: none">• 500 Internal Server Error• 501 Not Implemented• 502 Bad Gateway• 503 Service Unavailable• 504 Gateway Timeout

3. Types de codifications

b) Codification par tranche

Avantages :

- **Non ambiguë:** chaque code (numéro) est unique,
- **Insertion et Extension possible:** il est toujours possible d'insérer de nouveaux codes dans une catégories même si les catégories suivantes contiennent déjà des codes,.
- **Possibilité de contrôle:** il est possible de mettre en œuvre qui contrôle est-ce que la valeur saisie est valide ou pas.

3. Types de codifications

b) Codification par tranche

Inconvénients:

- Possibilité de saturation d'une tranche: l'insertion ne devient plus possible,
- Un seul regroupement est possible.

3. Types de codifications:

c) Codification articulée

Consiste à attribuer des codes découpés en zones, chaque zone est appelée descripteur, chaque descripteur a une signification particulière relative à l'objet codifié.

Exemple :

L'exemple type est le code attribué à l'immatriculation d'un véhicule :

0	2	3	3
---	---	---	---

N° séquentiel

1

Catégorie de véhicule

0	2
---	---

Année de 1^{ère} mise en circulation

1	0
---	---

Numéro de wilaya

3. Types de codifications: c) Codification articulée

Exemple

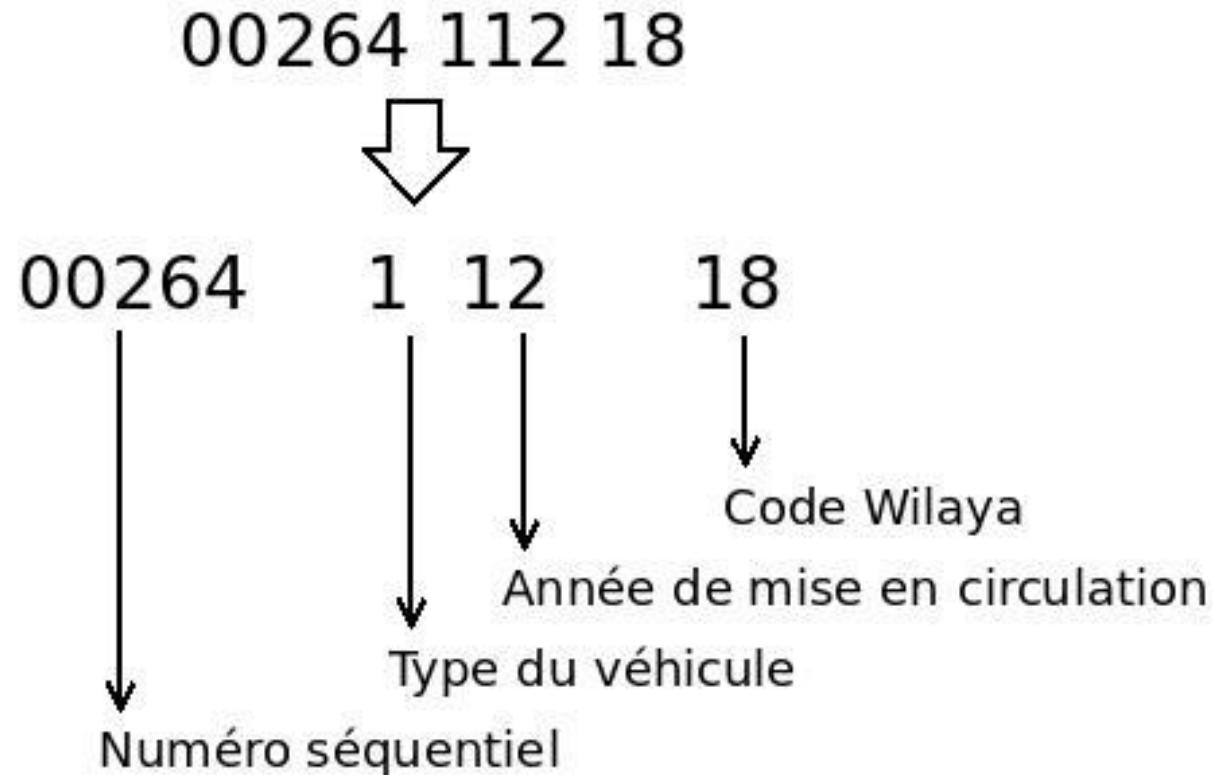


Figure 4. Matricule d'un véhicule

3. Types de codifications: c) Codification articulée

Avantages :

- **Non ambiguë** : chaque code (numéro) est unique,
- **Insertion et extension possibles**: il est toujours possible d'insérer de nouveaux codes dans une catégorie même si les catégories suivantes contiennent déjà des codes,
- **Possibilité de regrouper les objets selon un critère donné**: chaque descripteur offre une possibilité de regroupement,
- **Codification très utilisée**
- **Possibilité de regrouper les objets selon un critère donné** : chaque descripteur offre une possibilité de regroupement,
- **Possibilité de contrôle**: il est possible de mettre en œuvre qui contrôle est-ce que la valeur saisie est valide ou pas.

3. Types de codifications:

c) Codification articulée

Inconvénients :

- **Codes trop longs, donc lourds à manipuler:** vu qu'il est composé de plusieurs 'codes',
- **Possibilité de saturation d'un descripteur:** l'insertion ne devient plus possible,
- **Instabilité :** si un descripteur change, c'est tout le code qui va être changé

4. Types de codifications:

d) Codification par niveau

C'est un cas particulier de la codification articulée. Les descripteurs sont des niveaux.

Exemple

Le code postal peut être vu comme une codification par niveau parce que la ville (ou la région) est une partie de (appartient à) la Daïra qui fait, à son tour, partie de la Wilaya.

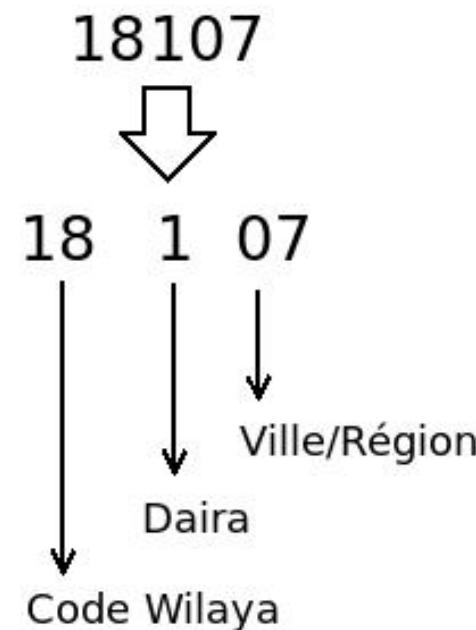
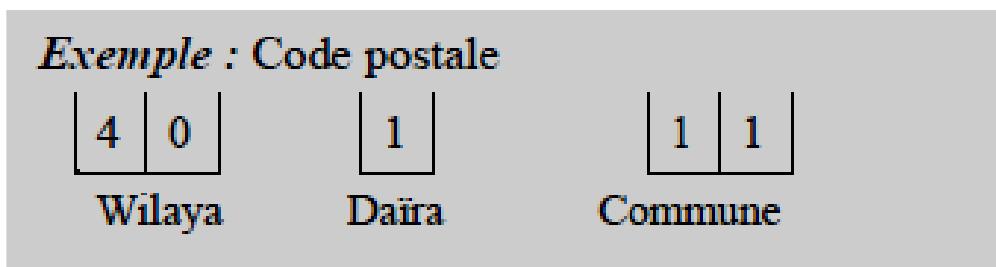


Figure 5. Code Postal

3. Types de codifications:

d) Codification par niveau

Avantages :

- Mêmes avantage que pour la codification articulée
- Facilité de recherche .

Inconvénients :

- Mêmes inconvénients que pour la codification articulée

3. Types de codifications:

e) Codification mnémonique

Consiste à abréger la désignation d'un objet à l'aide d'un ensemble réduit de caractères qui doit être évocateur de l'objet codifié. Il existe 2 sortes:

- Des codes dits «consonants», obtenus par suppression des voyelles, exemple:
FACTURE → FCTR,
- Des codes dits « abréviatifs » : obtenus par abréviation de la désignation de l'objet, exemple : FACTURE → FACT.

Consonants:

Total → ttl

Abréviatifs :

Total → tot

3. Types de codifications:

e) Codification mnémonique

Le code est une abréviation de la désignation littéraire d'un objet.

Exemple : le numéro de compte est désigné par NUMCOMPT.

Avantages	Inconvénients
✓ Très pratique car près du sens de l'objet.	✓ Ne se prête pas toujours au décodage automatique.

Choix d'une codification

Pour choisir le type de codification, il faut savoir :

- a) Etudier l'utilisation future des codes à définir.
- b) Etudier le nombre d'objets à codifier ainsi que la croissance de ce nombre.
- c) Etudier la répartition statistique des objets à définir.
- d) Rechercher s'il existe déjà des codifications pour les objets étudiés.
- e) Définir une codification avec les personnes qui auront à l'utiliser et par souci de standardisation tenir compte des codes déjà utilisés.
- f) Tester très sérieusement la codification définie et l'utiliser avant de l'adapter définitivement.

Les contrôles

Les contrôles détectent les informations erronées qui risquent de devenir sources de problèmes pour le système. Les lieux de contrôles sont les postes de saisies, les centres de calcul informatique et les postes de diffusion des résultats.

Comment contrôler les informations ?

Contrôler une information permet de vérifier sa justesse et sa conformité par rapport à la réalité de l'organisation.

Le contrôle comporte la détection des erreurs et la correction des erreurs détectées.

Les types de contrôle:

Il existe deux types de contrôle de l'information : les contrôles directs et les contrôles indirects.

1) Contrôles directs :

Il s'agit des contrôles qui s'effectuent sur l'information elle-même sans tenir compte des autres informations existant dans le système.

Les principaux types de contrôles directs sont :

- a) Le contrôle de présence ou de non présence.
- b) Le contrôle de type.
- c) Le contrôle de cadrage.

1) Contrôles directs :

a) Contrôle de présence ou de non présence

- Consiste à vérifier l'existence ou non d'une information sur le support où elle devrait se trouver.
- Le support peut être un document ou un fichier.
- En effet, le rajout d'une nouvelle occurrence d'information nécessite d'abord de vérifier que celle-ci n'existe pas déjà sur le support (non présence), cela nous évitera d'avoir des doublons de l'information.
- De même, avant toute opération de manipulation d'une occurrence d'information, nous devons d'abord vérifier son existence (présence) sans quoi la manipulation serait impossible.

1) Contrôles directs :

a) Contrôle de présence ou de non présence

Exemples :

- Avant d'insérer les notes d'examens d'un étudiant, on doit d'abord vérifier que ce dernier existe (présence).
- L'insertion d'un nouvel étudiant nécessite de vérifier que ce dernier n'existe pas dans le fichier des étudiants (non présence).

1) Contrôles directs :

b) Contrôle de type :

Consiste à vérifier que le type de l'information est conforme à ce qu'il devrait être.

Exemple :

- N° de téléphone = 02145T566 ==> Erreur : « T » n'est pas un chiffre et le numéro de téléphone est numérique.

1) Contrôles directs :

c) Contrôle de cadrage :

Désigne la position d'une information dans une zone de saisie ou de remplissage.

Exemples :

- Les types alphanumériques sont alignés à gauche tandis que les types numériques sont alignés à droite.

	A
1	Alphanumérique
2	25
3	

Figure 6. Exemple du cadrage sous Excel

2) Contrôles indirects

Il s'agit des contrôles qui vérifient la conformité d'une information par rapport à l'ensemble des informations se trouvant dans le système. Ceci se fait par comparaison d'informations entre elles.

Les types de contrôles indirects sont :

- a. Le contrôle de cohérence interne.
- b. Le contrôle de cohérence externe.
- c. Le contrôle de vraisemblance.

2) Contrôles indirects:

a) Le contrôle de cohérence interne :

Ce type de contrôle convient à la codification articulée. Son principe consiste à vérifier une partie d'une information par rapport à d'autres parties de la même information.

Exemples :

- Dans une information de type « date », il faut vérifier la cohérence du numéro du jour par rapport à celle du mois : la date du 31/05/2007 est fausse car le mois de mai compte 30 jours seulement.

2) Contrôles indirects:

a) Le contrôle de cohérence interne :

Exemples :

- Soit la codification suivante relative à des véhicules roulants :
[Code Véhicule][Nombre de roues].
 - L'occurrence [CA][02] signifie «Camion à 2 roues », or ceci est faux car il n'existe pas de camion ayant seulement 2 roues.
- Si la Wilaya 18 compte 6 Daïras, alors le code postale "18721" ne peut pas être correct.

2) Contrôles indirects: b) Le contrôle de cohérence externe :

Il s'agit de vérifier la conformité d'une information par rapport à d'autres informations externes.

Exemples :

Soit la date de naissance = 25/10/2005 d'un étudiant.

L'immatriculation d'un étudiant de l'USTHB est structurée comme suit :

[Durée de formation][Année du BAC][N° séquentiel].

Soit l'occurrence de matricule de cet étudiant nouvellement inscrit: [5][2007][3569].

Nous avons donc l'information sur l'année d'obtention du BAC = 2007.

Le calcul de l'âge de l'étudiant permet de déduire que ce dernier a 2 ans, or il est impossible d'avoir le BAC à cet âge ==> **on détecte alors une erreur sur la date de naissance.**

2) Contrôles indirects:

b) Le contrôle de cohérence externe :

Exemple

La date d'obtention du diplôme doit être supérieure (après) la date d'inscription.

2) Contrôles indirects:

d) Contrôle de vraisemblance

S'assurer que la valeur de l'information est semblable a une valeur correcte (par comparaison a des règles, des valeurs précédentes, ou des valeurs issues du secteur professionnel).

Exemples :

- Le matricule étudiant **[6][2007][8974]** porte une information invraisemblante car il n'existe pas de formation sur 6 années à l'USTHB.
- L'occurrence de code postal **[50 210]** est invraisemblante car il y a 48 wilaya seulement en Algérie, d'où la valeur 50 n'est pas correcte.
- La date doit être valide (30/02/2020, par exemple, a une forme correcte mais ce n'est pas une date valide).

2) Contrôles indirects:

Remarque :

D'autres types de contrôles existent comme l'usage de fonctions mathématiques. Ces dernières sont appliquées sur les occurrences de code et permettent de vérifier certaines propriétés comme la longueur d'un code par exemple.

Ordre d'exécution des contrôles

Les différentes contrôles sont exécutés dans l'ordre suivant :

1. Présence,
2. Type,
3. Cadrage,
4. Vraisemblance,
5. Cohérence interne,
6. Cohérence externe.

Conclusion

- ❖ Dans ce chapitre, nous avons vu la notion de codification. Cette notion permet de ramener les notions et concepts manipulés par l'organisation vers le monde informatique en proposant une représentation adéquate.
- ❖ Cette représentation permettra l'utilisation de l'outil informatique de stocker, traiter et communiquer ces concepts.
- ❖ La codification nécessite des actions du contrôle pour une utilisation correcte et pertinente de cette représentation.
- ❖ Ainsi, nous avons évoqué les différents contrôles possibles sur une information.