



Royaume du Maroc
Chef du Gouvernement

Ministère de la Transition
Numérique et de la Réforme
de l'Administration

#ADD

وكالة التنمية الرقمية

ⵜⴰⵎⴰⴳⴷⴰⵏⵜ ⵜⴰⵏⵓⵎⵓⵔⵜ ⵜⴰⵔⵉⵎⵓⵏⵉ ⵜⴰⵔⵉⵎⵓⵏⵉ

Agence de Développement du Digital

Atelier de formation sur l'Open Data au profit des organismes et institutions publics relevant de la Région Fès-Meknès

Data.gov.ma

Région Fès-Meknès

23/01/2025

Session de formation sur l'Open Data

Séance 4:

Référentiels communs et Qualité des données

14h00-15h00

Séance 4:

Référentiels de données communs et Qualité des données

Section 4.1:

Référentiels de données communs

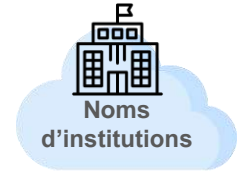


Session 4: Référentiels de données communs et Qualité de la Data

Référentiels de données communs: Pourquoi?

Le référentiel comme besoin fondamental

Certaines types de données sont fréquemment utilisées pour sécuriser la qualité et standardisation de l'information.



Données 'connues'

Données qui ne sont pas des variables, mais plutôt figées dans le temps, malgré leur évolution.

ex. Noms de rues, Codes Postaux

Données régulières

Utilisées régulièrement et couramment dans de nombreux domaines.

ex. dans le cas des rues et Codes postaux, utilisé pour adresser un courrier, pour souscrire à un service public, ...

Inputs différents

Saisis dans les SI de façons différentes (structure, orthographe,)

ex. Avenue des FAR, Av. des FAR, Avenue des Forces Armées Royales, ..

Session 4: Référentiels de données communs et Qualité de la Data

Avantage d'utiliser un référentiel commun: Exemple

Cas de la Classification Internationale des Maladies (CIM)

La Classification internationale des maladies (CIM) est un référentiel standardisé par l'OMS pour coder, analyser et comparer les maladies, troubles et causes de décès à l'échelle mondiale.

The screenshot displays the ICD-10 Version:2019 web interface. The search bar is empty, and the navigation tabs include ICD-10, Versions - Languages, and Info. The left sidebar shows a tree view of the ICD-10 categories, with 'D50-D53 Nutritional anaemias' selected. The main content area shows the following classification:

- (D50-D53)**
- Nutritional anaemias (D50-D53)**
- D50 Iron deficiency anaemia**
 - Incl.: anaemia:
 - siderotic
 - hypochromic
- D50.0 Iron deficiency anaemia secondary to blood loss (chronic)**
 - Posthaemorrhagic anaemia (chronic)
 - Excl.: acute posthaemorrhagic anaemia (D62)
 - congenital anaemia from fetal blood loss (E81.2)
- D50.1 Sideropenic dysphagia**
 - Kelly-Petersen syndrome
 - Plummer-Vinson syndrome
- D50.8 Other iron deficiency anaemias**
- D50.9 Iron deficiency anaemia, unspecified**
- D51 Vitamin B₁₂ deficiency anaemia**
 - Excl.: Vitamin B₁₂ deficiency (E53.8)
- D51.0 Vitamin B₁₂ deficiency anaemia due to intrinsic factor deficiency**
 - Anaemia:
 - Addison
 - Biermer
 - pernicious (congenital)
 - Congenital intrinsic factor deficiency
- D51.1 Vitamin B₁₂ deficiency anaemia due to selective vitamin B₁₂ malabsorption with proteinuria**
 - Imerslund(-Gräsbeck) syndrome
 - Megaloblastic hereditary anaemia
- D51.2 Transcobalamin II deficiency**
- D51.3 Other dietary vitamin B₁₂ deficiency anaemia**
 - Vegan anaemia
- D51.8 Other vitamin B₁₂ deficiency anaemias**
- D51.9 Vitamin B₁₂ deficiency anaemia, unspecified**
- D52 Folate deficiency anaemia**

Session 4: Référentiels de données communs et Qualité de la Data

Avantage d'utiliser un référentiel commun: Exemple

Cas de la Classification Internationale des Maladies (CIM)

Le domaine médical est une source infinie de Data, essentielle pour les prévisions et gestion sanitaires à l'échelle des Pays

L'Organisation Mondiale de la Santé, et tous les organes de Santé (étatiques & non étatiques) cherchent à analyser des statistiques sur les causes de décès.

Cruciale pour:

Allouer les ressources de manière efficace :

En comprenant la prévalence et l'incidence des maladies,
Mieux planifier et distribuer les ressources de santé pour répondre aux besoins de la population.

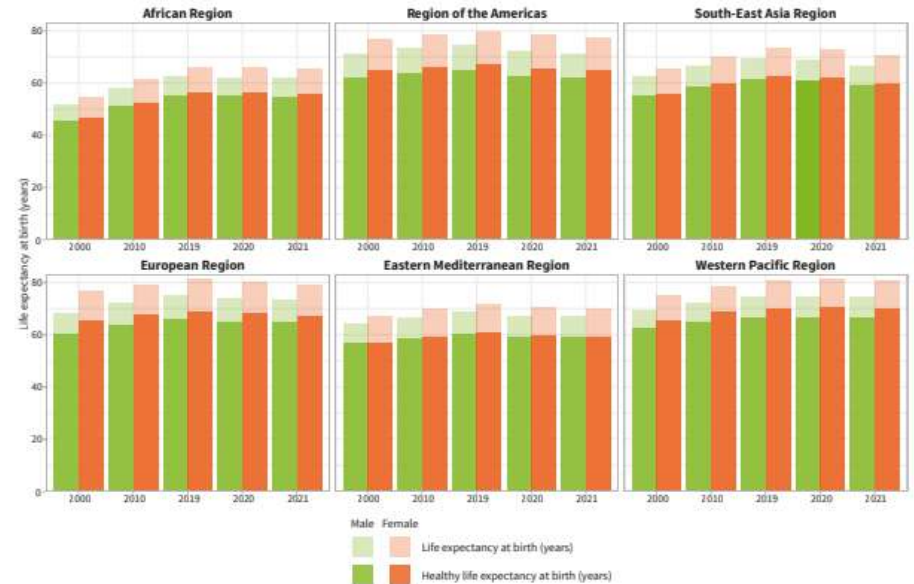
Détecter les épidémies & tendances sanitaires

De façon à identifier les priorités sanitaires

Faire des études comparatives par région ou pays

Et pouvoir corrélérer ces informations pour en tirer les bonnes conclusions

Figure 1.2 Trends in life expectancy and HALE at birth, by sex and by WHO region, 2000–2021



Session 4: Référentiels de données communs et Qualité de la Data

Avantage d'utiliser un référentiel commun: Exemple

Cas de la Classification Internationale des Maladies (CIM)

Le référentiel s'est vite imposé comme une nécessité:

Les données proviennent principalement de certificats de décès, écrits généralement par des médecins

Le problème en l'absence de référentiel:

Plusieurs écritures possibles pour un même libellé

par exemple: "Infection pulmonaire aiguë", "Pneumonie aiguë"

Plusieurs libellés pouvant désigner la même maladie

"Diabète juvénile insulino-dépendant"

"Hyperglycémie chronique nécessitant de l'insuline"

"Insuffisance pancréatique insulino-dépendante"

"Maladie métabolique chronique (diabète sucré type 1)"

Elles désignent toutes la maladie "Diabète sucré de type 1-code E10 dans le référentiel CIM-10

Un effet de volume

Qui crée une masse d'information totalement inexploitable en l'absence de référentiel structuré

7 **55 000 entrées dans le CIM-11**

Diag	Intitulé CIM10
E10	Diabète sucré de type 1
E100	Diabète sucré de type 1, avec coma
E101	Diabète sucré de type 1, avec acidocétose
E102	Diabète sucré de type 1, avec complications rénales
E103	Diabète sucré de type 1, avec complications oculaires
E104	Diabète sucré de type 1, avec complications neurologiques
E105	Diabète sucré de type 1, avec complications vasculaires périphériques
E106	Diabète sucré de type 1, avec autres complications précisées
E107	Diabète sucré de type 1, avec complications multiples
E108	Diabète sucré de type 1, avec complications non précisées
E109	Diabète sucré de type 1, sans complication

Conséquences en l'absence de référentiel

Statistiques erronées:

Les analyses et conclusions ne reflètent pas l'état réel

Perte de valeur de la donnée

Un volume important de data qui reste inexploitable

Manque à gagner en impact

Les mesures qui doivent être prises ne le sont pas

Référentiel de l'OMS: CIM-11



1893: Le médecin français Jacques Bertillon présente sa classification

1900: Adoption à la conférence internationale, avec des révisions prévues tous les 10 ans

1948: Adoption par l'OMS (inclut, en plus des causes de décès, les maladies et traumatismes)

1948-2022: Évolutions successives et adoption de la 11e révision en 2022, intégrant les dernières avancées médicales

Session 4: Référentiels de données communs et Qualité de la Data

Avantages d'utiliser un référentiel commun

Données pouvant faire l'objet de référentiels

Ci dessous des exemples de données généralement fournies sous forme de référentiel commun dans le monde



Adresses

Noms de Rues
Codes Postaux
Villes & Communes



Identifiants d'entités

Identifiant Commun des Entreprises
Répertoire des services publics
Liste des tribunaux
Établissements hôteliers, ...



Libellés

Produits (Distribution)
Maladies & causes de décès
Motif de contravention



Economie

Nomenclature Marocaine des
Activités (NMA)
Nomenclature douanière



Transport

Noms des gares
Noms des aéroports
Noms des ports
Noms des arrêts (transport public)



Education

Filières académiques
Etablissements scolaires

Session 4: Référentiels de données communs et Qualité de la Data

Avantages d'utiliser un référentiel commun

Le double avantage du référentiel

L'utilisation de référentiels présente un double avantage: il permet de contrôler et enrichir la donnée

enrichissement de données

contrôle données

CODE	NOM	DCI	DOSAGE1	UNITE_DOS	FORME	PRESENTATION	PPV	PH	PRX_BR	PRIN1/AUX	
6118000230869	LIRO /EAU POUR IRRIGATION	EAU POUR PREPARATION INJECTABLE	5000	ML	SOLUTION POUR IRRIGATION	1 FIOLE 3 L	35,00		35,00	P	0%
6118000182018	ELDOXATINE 5MG/ML	OXALICATINE	500	MG	SOLUTION A DILUER POUR PERFUSION	1 BOITE 1 FLACON 40 ML	2 982,00	2 955,00	2 982,00	P	79%
61180001801645	ELDOXATINE 5MG/ML	OXALICATINE	500	MG	SOLUTION A DILUER POUR PERFUSION	1 BOITE 1 FLACON 40 ML	2 982,00	2 955,00	2 982,00	P	79%
6118000182459	VIVALAN	VILDAXATINE	160	MG	COMPRIME PELLICULE	1 BOITE 20 COMPRIME	60,70	40,30	60,70	P	0%
6118000120155	ZADRYL	CETRIZINE	1	MG	SOLUTION BUVALE	1 FLACON 60 ML	31,90	19,90	31,90	G	79%
6118000041596	MYNAZOL	FLUCONAZOLE	50	MG	GELULE	1 BOITE 7 GELULE	87,00	54,20	87,00	G	79%
6118000391974	ALFACEFAL 125 MG/5 ML	CEFAZOLOR	125	G	POUDRE POUR SUSPENSION BUVALE	1 FLACON 60 ML	45,00	29,00	45,00	G	0%
6118000041344	INDICAND 50MG	FLUCONAZOLE	50	MG	GELULE	1 BOITE 7 GELULE	35,00	21,00	35,00	G	79%
6118000100270	MEDIATOR	BENFLORXEC	150	MG	COMPRIME ENPOBE	1 BOITE 30 COMPRIME	60,00	39,70	60,00	P	0%
61180000971069	(GASTRO)UBER	LANSOPRAZOLE	30	MG	GELULE GASTRO-RESISTANTE	1 BOITE 15 GELULE GASTRO-RESISTANTE	75,00	46,70	75,00	G	79%
6118000189142	CLOFENE	DICLOFENAC	75	MG	COMPRIME ENPOBE GASTRO-RESISTANT	1 BOITE 20 COMPRIME	20,50	12,00	20,00	P	79%
6118000172496	ZELDUC	ZEPRASOLONE	30	MG	GELULE	1 BOITE 14 GELULE	319,00	211,00	319,00	P	79%
6118000170333	TRULICAN	FLUCONAZOLE	5	MG	SOLUTION POUR PERFUSION	1 FLACON 60 ML	144,30	89,80	144,10	P	79%
6118000331079	AVIT	ACIDE ACETYLSALICYLIQUE / THIAMINE (VITAMINE B1)	600/5	MG/MG	SUPPOSITOIRE	1 BOITE 10 SUPPOSITOIRE	10,50	6,80	10,50	G	0%
6118000150121	KLONOPIN	CLONAZEPAM	2	MG	COMPRIME QUADRISECALE	1 BOITE 40 COMPRIME	41,20	25,70	41,20	P	79%
6118000332182	HYDROXOCOBALAMINE	HYDROXOCOBALAMINE	5000	µg	SOLUTION INJECTABLE	1 BOITE 4 AMPOULE	59,50	39,70	59,50	P	0%
6118000140180	DISPAMOX	JAMOXOLINE	500	MG	COMPRIME DISPENSIBLE	1 BOITE 12 COMPRIME	34,60	21,60	34,60	G	79%
6118000180418	MUSARIL	TETRAZEPAM	50	MG	COMPRIME ENPOBE	1 BOITE 30 COMPRIME	51,80	34,30	51,80	G	0%
6118000100225	PRETERAX	INDAPAMIDE / PERINDOPRIL	0,625/2	MG/MG	COMPRIME SECALE	1 BOITE 30 COMPRIME	239,00		239,00	P	79%
6118000100304	CLAFORAN	CEFTAXIME	1	G	POUDRE POUR SOLUTION INJECTABLE	1 BOITE 1 FLACON	69,00	43,00	69,00	P	79%
6118000080497	RHINOPERRAL	PARACETAMOL / CHLORPHENAMINE	240/32	MG/MG	GELULE	1 BOITE 20 GELULE	14,30	9,60	14,30	P	0%
6118000391547	IRFASONE	IRFAMYCINE / PREDNISOLONE	0,25/1	G/G	POMMAGE	1 TUBE 5 G	12,00		12,00	G	0%
6118000248205	IRAFON	AMBIROL	0,25	%	SOLUTION BUVALE	1 FLACON 200 ML	39,85	25,70	39,85	G	0%
6118000161169	AMICOL	AMOXICILINE	500	MG	POUDRE POUR SUSPENSION BUVALE	1 FLACON 60 ML	39,30	24,30	39,30	P	79%
6118000240150	BENCLAMIDE	GLEBENCLAMIDE	5	MG	COMPRIME	1 BOITE 60 COMPRIME	36,20	22,60	36,20	G	79%
6118000250357	UNASVIN	AMPICILLINE / SULBACTAM	1 / 0,5	G/G	POUDRE POUR SOLUTION INJECTABLE	1 FLACON 3,2 ML	42,00	26,20	42,00	P	79%
6118000360451	LASLIX	FLUROSEMIDE	20	MG	SOLUTION INJECTABLE	1 AMPOULE 2 ML	7,40	4,60	5,60	P	79%
6118010020155	DILATOR	SALBUTAMOL	2,5	MG	SOLUTION POUR INHALATION PAR NEBULISSEUR	1 BOITE 10 REPARENT	41,20	25,70	41,20	G	79%
6118000101097	DILATOR	SALBUTAMOL	2,5	MG	SOLUTION POUR INHALATION PAR NEBULISSEUR	1 BOITE 10 REPARENT	41,20	25,70	41,20	G	79%
6118010024200	FORMOORTAL	FORMOCHORTAL	0,05	%	COLLYRE	1 FLACON 5 ML	17,80		17,80	G	0%
6118000080275	GENTAGAM 120	GENTAMICINE	120	MG	SOLUTION INJECTABLE	1 BOITE 1 AMPOULE INJECTABLE	15,90	9,80	15,90	G	79%
61180001070763	LUMILINE PROFIL 30 PEN	INSULINE HUMAINE	100	UI	SUSPENSION INJECTABLE INTERMEDIAIRE	1 CARTOUCHE 3 ML	104,00	65,00	104,00	P	79%
6118000369567	BIFRODOGYL	SPIRAMYCINE / METRONIDAZOLE	1,5 / 250	MG / MG	COMPRIME PELLICULE	1 BOITE 15 COMPRIME	102,00	63,50	79,90	P	79%
6118000341007	GLEMA	OLMESERTINE	5	MG	COMPRIME	1 BOITE 30 COMPRIME	40,00	25,00	40,00	G	79%
6118000120425	SALAZOPYRIN	SULFASALAZINE	500	MG	COMPRIME ENPOBE GASTRO-RESISTANT	1 BOITE 100 COMPRIME GASTRO-RESISTANT	141,70	89,30	141,70	P	79%
61180001070680	XANTHILUM LP	THEOPHYLLINE	300	MG	GELULE LP	1 BOITE 60 GELULE	51,50	32,20	51,50	G	79%
6118000350414	AFRLIX	CETRIZINE	1	MG	SOLUTION BUVALE	1 FLACON 60 ML	31,90	19,90	31,90	G	79%
6118000120384	GYNOMYK	BUTOCONAZOLE	100	MG	OVULE	1 BOITE 3 OVULE	62,20	41,10	62,20	P	0%
6118001950442	AVASTIN	BEVACIZUMAB	100	MG	SOLUTION A DILUER POUR PERFUSION	1 BOITE 1 FLACON	3 983,00	3 669,00	3 983,00	P	79%
61180001201963	FLIXMA 8000 LIANTHAX 60ML	FLIXONIDASE	8000	µg	SOLUTION INJECTABLE	1 BOITE 2 PERPUQE PREEMPLIE	134,00	84,00	134,00	P	79%
6118000182725	PARAPLATINE	CARBOPLATINE	50	MG	SOLUTION POUR PERFUSION	1 FLACON 5 ML	835,30	552,60	835,30	P	79%
61180001109301	TRIFAVIV	TRIFAVIV	800	MG	POUDRE POUR SOLUTION INJECTABLE	1 BOITE 1 FLACON	61,00	36,00	61,00	P	79%

Session 4: Référentiels de données communs et Qualité de la Data

Avantages d'utiliser un référentiel commun

Le référentiel comme base solide pour la Data

La publication de référentiels communs permet d'accroître la valeur fournie par la data



Une meilleure compréhension

Avec une base commune d'appellation et de description des éléments

ex. dans le cas du référentiel des diplômes, il permet d'unifier la façon dont les recrutements & évaluations sont réalisés



Une meilleure réutilisation

Avec un enrichissement de données, augmentant ainsi la valeur des données

ex. dans le cas d'un référentiel des entreprises publié, l'information ICE permet d'enrichir avec la raison sociale, adresse, activité, ...



Une meilleure interopérabilité

Rendre l'information exploitable par tous les acteurs de l'écosystème

ex. un fichier produit par une administration est directement exploitable et intégrable dans les SI, avec un effort d'intégration minimum

Session 4: Référentiels de données communs et Qualité de la Data

Exemples de référentiels disponibles

Disponibles sur le Portail National (data.gov.ma)



Codes postaux



Villes & Villages



Administrations
publiques



Etablissements
de santé



Secteurs d'Activité
économique

[Source](#)

Disponibles sur le Portail HCP (hcp.ma)



Nomenclature
Marocaine des
Activités (NMA)



Nomenclature
des diplômes



Codes
géographiques

[Source](#)

Note: l'ADD est entrain d'élaborer les référentiels des données communs prioritaires en concertation avec les acteurs concernés

Exemples de référentiels universels



Transport

Codes des aéroports
IATA

Codes standardisés à trois lettres attribués aux aéroports du monde entier



Economie

Code des devises
(ISO 4217)

Codification des devises mondiales en trois lettres (ex. : USD, EUR, MAD).



Agriculture

Ontologie géopolitique

Référentiel pour les noms officiels des pays, codes et régions géographiques utilisés dans les statistiques mondiales agricoles.



Science

Nomenclature des
produits chimiques

Base de données internationale pour l'identification des substances chimiques,



Commerce

Code HS (Harmonized system)

Codification des marchandises pour les échanges internationaux, utilisée par la plupart des pays

Session 4: Référentiels de données communs et Qualité de la Data

Formats et Consistance des référentiels de données disponibles

Formats de référentiels courants

Les référentiels sont la plupart du temps dans des formats similaires aux jeux de données, avec les mêmes critères techniques



Fichiers plats (CSV, TSV)
exemple: Annuaire des agences de voyage sur data.gov.ma



Fichiers structurés (XML, JSON)



API ou Webservices

The screenshot shows the 'Annuaire des Agences de voyages' page on data.gov.ma. It features a table with columns for 'Nom de l'Agence', 'Adresse', and 'Ville'. The table contains several rows of agency information.

Nom de l'Agence	Adresse	Ville
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		
AGENCE DE VOYAGES		

The screenshot shows a data portal interface with a navigation bar and a list of files. The 'Fichiers' tab is selected, showing three main files: CSV, JSON, and XLS. Each file entry includes the format name, the date it was last updated, and the number of records. There are also buttons to download each file.

Fichiers (1)	Réutilisations et API (5)	Discussions (1)	Ressources communautaires (0)	Informations
3 FICHIERS PRINCIPAUX				
CSV	Mis à jour le 3 décembre 2024	csv	4K	CSV
JSON	Mis à jour le 2 décembre 2024	json	966	JSON
XLS	Mis à jour le 18 mars 2016	xls	2K	XLS
1 API				
API	Mis à jour il y a un peu		750	FICHER

[Source \(Portail France\)](#)



Fichiers propriétaires



Fichiers difficilement exploitables (PDF par ex.)
non recommandé

Session 4: Référentiels de données communs et Qualité de la Data

Démarche d'élaboration de référentiels

Pour quelle raison produire un référentiel de données

Un organisme peut être amené à produire ses propres référentiels, pour des besoins internes ou pour standardiser l'utilisation de ses données à l'extérieur.



Pour ses besoins internes

Pour assurer une saisie et gestion cohérente de ses propres données

Exemple:

Une commune peut créer un référentiel des établissements publics gérés, pour avoir un enregistrement des noms harmonieux à travers les différents rapports, publications, ...
ex. "Ecole Primaire Al Fath", "Ecole Al Fath"



Pour standardiser pour les tiers

Pour assurer que ses données utilisées par des acteurs tiers le soient de façon harmonieuse

Exemple:

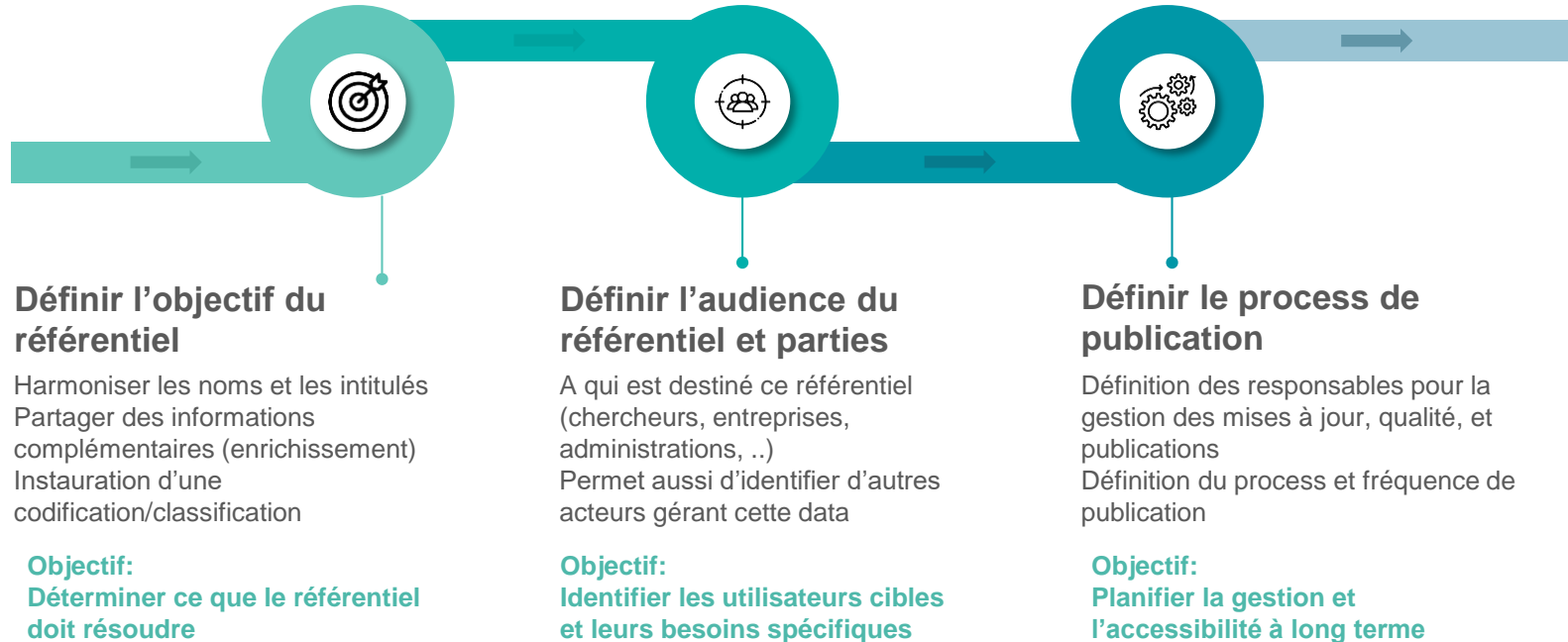
Une ville peut créer un référentiel des événements culturels avec des informations pertinentes.
Ces données peuvent être utilisées par des médias locaux, agences de tourisme, ..

Session 4: Référentiels de données communs et Qualité de la Data

Démarche d'élaboration de référentiels

Produire un référentiel de données

La démarche ci dessous est une démarche possible pour constituer et publier un référentiel de données communs, destiné à être réutilisé par des tiers



Séance 4:

Référentiels de données communs et Qualité des données

Section 4.2:

Evaluation de la Qualité des données



Session 4: Référentiels communs et Qualité de la Data

Qualité de la data

Qualité de la Data = Qualité?

Durant la 2e Guerre Mondiale

Les bombardiers des Alliés revenaient de mission très endommagés par des tirs ennemis.

Une équipe de scientifiques et statisticiens (Statistical Research Group) a été sollicitée pour trouver une solution et renforcer / blinder les zones qui doivent l'être.

Ils ont collecté des données sur les zones d'impact (figure)

Quelles zones devaient être blindées?

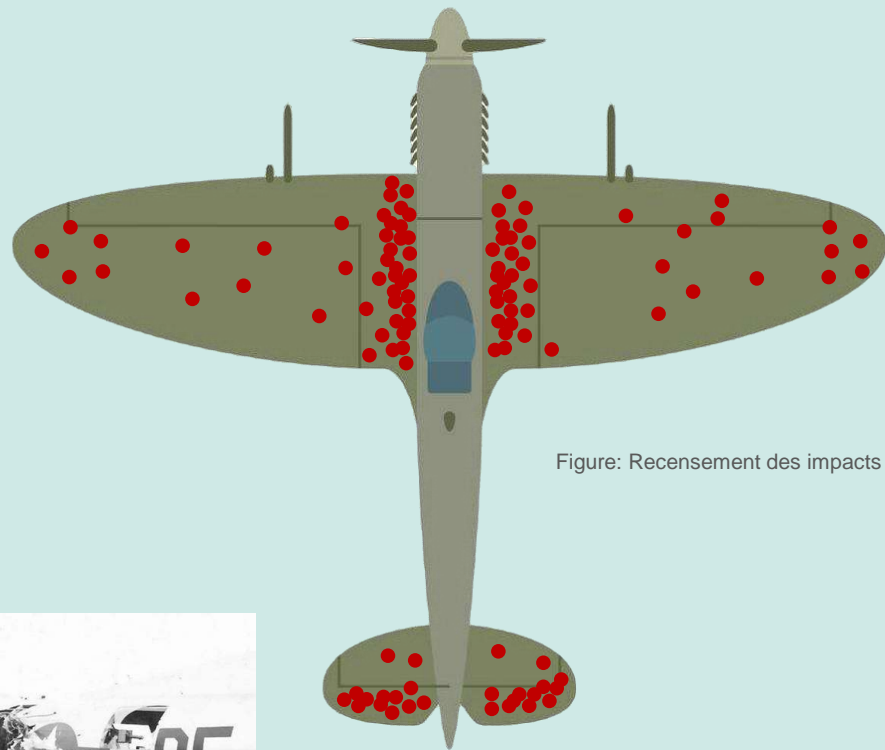


Figure: Recensement des impacts

Intuitivement les zones impactées doivent être renforcées

Session 4: Référentiels communs et Qualité de la Data

Qualité de la data



Abraham Wald

Mathématicien et membre de
l'équipe Statistical Research
Group
1902-1950

Le biais du survivant

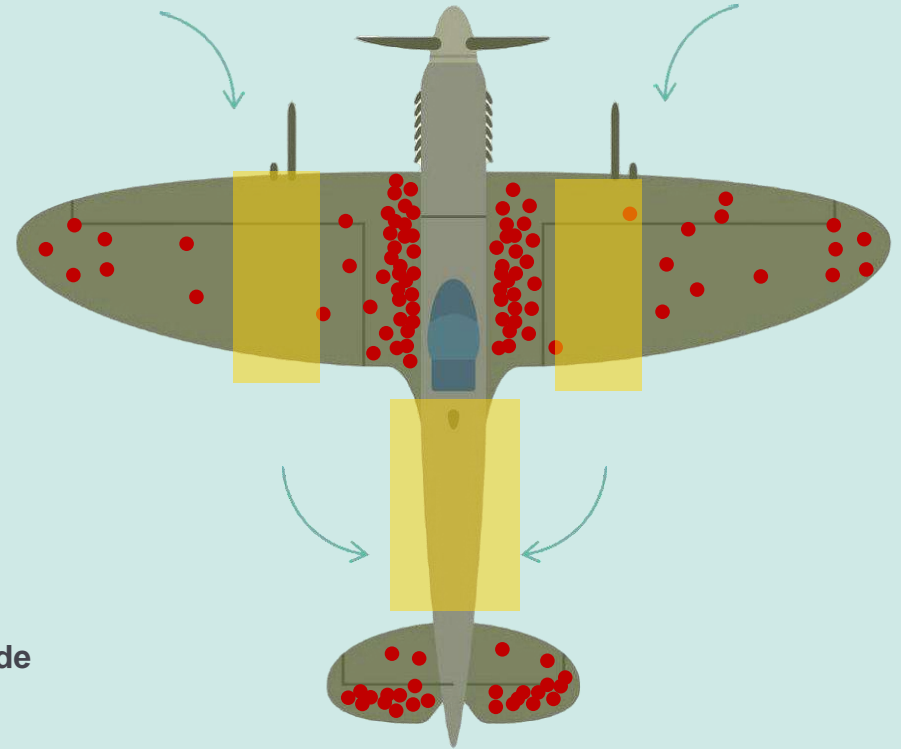
Son raisonnement ?

Les impacts de balles sur les avions « survivants » (d'où l'expression) **indiquent en réalité les points forts** :

les zones sur lesquelles un avion pouvait être touché et continuer malgré tout à voler.

En l'occurrence, celles qui précisément n'avaient **pas besoin de renforcement.**

Le blindage devait donc être posé aux **endroits sans impact.**



[Source \(Le biais du survivant - l'ADN Data\)](#)

Session 4: Référentiels communs et Qualité des données

Les différents aspects de la Qualité

Maximiser l'impact de la donnée publiée

La qualité de la data ne se limite pas à la DATA elle-même, mais également à son contexte, la compréhension de sa portée et son exploitabilité.



Clarté de la data

La donnée fournie doit être documentée avec des métadonnées claires pour comprendre l'usage et les limites des données.

Faciliter l'accès par API et mises à jour en temps réel



Interopérabilité

Le format de fichier est un des piliers de la qualité des données ouvertes pour permettre une interopérabilité, et donc une réutilisation efficace et permettant l'innovation



Fiabilité des données

Des données exactes et fiables sont la base d'un traitement de la data. Il est également essentiel que la data soit à jour et non obsolète

Session 4: Référentiels communs et Qualité des données

Frameworks et méthodes d'évaluation

Différents frameworks, pour différents contextes

Parmi les multitudes de méthodologies d'évaluation, chacune va se concentrer sur un aspect ou dimension particulière. Tout en étant des méthodologies différentes, elles visent toutes à promouvoir une donnée ouverte, selon les bons standards, et avec la bonne implication organisationnelle et technologique de l'Administration Publique

Exemples de framework



Modèle des 5 étoiles (Tim Berners-Lee)

Évalue la qualité technique des données ouvertes.

1 étoile : Disponibilité en ligne, 5 étoiles : Interopérabilité et liens avec d'autres données.

Critère: Format & Interopérabilité des données

[Source \(5starsdata\)](#)



ODIN (Open Data Inventory)

Mesure la couverture et l'opérabilité des données ouvertes des statistiques nationales (ex. démographie, économie).

Critères : Disponibilité, accessibilité et qualité.

[Source \(ODIN: Open Data Inventory\)](#)



ODRA (Open Data Readiness Assessment)

Évalue la préparation institutionnelle d'un pays ou d'une organisation à mettre en œuvre une stratégie open data.

Critères : Cadre juridique, infrastructure, engagement des parties prenantes.

[Source \(Banque Mondiale\)](#)

Session 4: Référentiels communs et Qualité des données

Évaluer l'accessibilité de la data et son impact

La méthodologie ODIN (Open Data Inventory)

Pour compléter le modèle TBL, l'Open Data Inventory (ODIN) propose une démarche plus globale de mesure de la qualité de la donnée ouverte, en incorporant des notions d'accessibilité, actualisation, etc..

La démarche ODIN se base sur

- La présence d'un format lisible par machine
- La data en format ouvert (non propriétaire)
- La présence de méta données exhaustives
- La présence de plusieurs canaux (fichier, API, ..)
- La présence d'une licence de réutilisation proprement documentée (Conditions d'utilisation)



25e

sur 193 pays



+16

places / 2021

L'Observatoire des Données Ouvertes (Open Data Watch) a, lors de sa session de 2022, classé le Maroc à la 25ème place sur 193 en matière d'Indice des Données Ouvertes (ODIN)

Dimension	Critères	Poids (%)	Description
Couverture des données	Critères spécifiques par secteur (démographie et population, économie, environnement et infrastructure)	60	Disponibilité des données Complétude temporelle (profondeur) Granularité (niveau de détail) Fréquence de mise à jour
	Ouverture des données	10	Données disponibles en ligne, gratuites, accessibles sans barrière Disponible sous API / Temps réel
	Format	10	Données publiées dans des formats ouverts et non-propriétaires (CSV, JSON, etc.).
	Documentation	10	Métadonnées claires sur la source, la portée, les limites et l'année de publication.
	Licence ouverte	5	Licence explicite autorisant la réutilisation des données.
	Mises à jour régulières	5	Données actualisées à des intervalles réguliers et définis.

[Source \(Classement ODIN\)](#)

[Source \(ODIN: Open Data Inventory\)](#)

Session 4: Référentiels communs et Qualité des données

Évaluer l'interopérabilité

Le modèle 5 étoiles

Le modèle 5 étoiles de Tim Berners-Lee évalue la qualité des données ouvertes en promouvant leur accessibilité en ligne, leur structuration, l'utilisation de formats non-propriétaires, leur interopérabilité via des standards ouverts, et leur connexion à d'autres données pour créer un réseau d'informations liées.



Session 4: Référentiels communs et Qualité des données

Évaluer l'interopérabilité

Quelques exemples

The image shows three overlapping screenshots from the Casa Tramway website. The top-left screenshot shows the website's header with the logo 'CASV tramway CASV busway Opéré par RATP DEV' and a navigation menu with 'Accueil > Se déplacer > Lignes et Horaires'. Below the menu is a button 'Télécharger TM hiver VF (1).pdf'. The top-right screenshot shows a 'HORAIRES DE SERVICE' section for line T3, with a table of departure times: 'De Hay Al Wahda' at 22:30 and 05:30, and 'De Casa Port' at 22:30 and 06:00. The bottom screenshot shows a map of the tramway network with various lines color-coded and labeled.

Publication des données ★

Actuel : Les horaires et itinéraires des bus et tramways sont souvent disponibles en PDF sur les sites des opérateurs (comme Casa Tram ou Alsa).

Limite : Ces fichiers sont difficiles à exploiter et n'offrent pas de possibilités d'analyse automatisée.

Amélioration: Utiliser le format GTFS (General Transit Feed Specification), un standard ouvert utilisé pour structurer les données de transport public.

→ **Atteinte du 3 étoiles** ★★ ★

Autre amélioration: Interconnecter les data à d'autres jeux de données (ex. météo, évènements affectant le trafic, points d'intérêt)

→ **Atteinte du 5 étoiles** ★★ ★★ ★★

[Source \(CasaTramway\)](#)

Session 4: Référentiels communs et Qualité des données

Autres méthodes d'évaluation de la data ouverte

L'Open Data Readiness Assessment (ODRA)

L'**Open Data Readiness Assessment (ODRA)** est une méthodologie développée par la Banque mondiale pour évaluer la préparation d'un pays ou d'une organisation à l'adoption de l'Open Data (adapté pour l'échelle du Pays ou d'une Institution étatique). Elle analyse huit dimensions essentielles pour identifier les opportunités et obstacles, et proposer des recommandations adaptées.

Les 8 dimensions de l'ODRA:



Leadership

Volonté politique et engagement des dirigeants à promouvoir l'Open Data.



Politiques et réglementation

Analyse des cadres législatifs et politiques pour soutenir l'ouverture des données.



Institutions et capacités

Capacité des institutions à gérer et publier des données ouvertes.



Données

Disponibilité, qualité et accessibilité des données déjà existantes.



Demandes des utilisateurs

Identification des besoins et attentes des utilisateurs finaux des données.



Technologies

Infrastructure technique pour stocker, gérer et partager des données ouvertes.



Engagement citoyen

Participation des citoyens et implication dans les initiatives Open Data.



Écosystème

Collaboration entre les parties prenantes : gouvernement, entreprises, société civile, et autres acteurs.

[Source \(Banque Mondiale\)](#)

[Source \(Banque Mondiale-Guide\)](#)

Diagnostic : Collecte d'informations à travers des entretiens, des ateliers et l'analyse documentaire.

Recommandations : Actions concrètes pour remédier aux lacunes identifiées et maximiser les opportunités.

Plan d'action : Feuille de route adaptée aux priorités et ressources disponibles.

Session 4: Référentiels communs et Qualité des données

Autres méthodes d'évaluation de la data ouverte

ODRA: Étude réalisée par l'ADD et la Banque Mondiale

L'Agence de Développement du Digital (ADD) a réalisé, en 2020, en partenariat avec la Banque Mondiale et en concertation avec les organismes et institutions concernés, une étude portant sur le développement des données ouvertes publiques (Open Data) visant à élaborer un plan d'actions pour maximiser l'impact de ces données au Maroc, notamment dans l'amélioration des services publics, le développement de l'entrepreneuriat, et l'inclusion digitale.

Objectif

Élaborer un plan d'actions pour maximiser l'impact des données ouvertes dans l'amélioration des services publics, le développement entrepreneurial et l'inclusion digitale.

L'étude s'appuie sur la méthodologie Open Data Readiness Assessment (ODRA) pour évaluer les initiatives Open Data depuis 2011 au Maroc.

Méthodologie et sensibilisation

Basée sur l'Open Data Readiness Assessment (ODRA), l'étude a évalué les forces, faiblesses et contraintes des initiatives Open Data au Maroc depuis 2011.

Environ 250 participants (administrations, établissements et entreprises publics) ont participé à des ateliers de sensibilisation sur les opportunités et principes de l'Open Data.

Objectif

Les recommandations se concentrent sur un plan d'actions réparti en 4 volets : Juridique et Réglementaire, Technique et Processus, Organisationnel et Gouvernance, Usage et Réutilisation des données.

Actions proposées : À court, moyen et long terme, pour maximiser l'impact et relever les enjeux de l'Open Data au Maroc.

[Source \(ADD\)](#)

Session 4: Référentiels communs et Qualité des données

Autres méthodes d'évaluation de la data ouverte

L'Open Data Institute (ODI)

Les Open Data Certificates de l'Open Data Institute (ODI) évaluent la qualité des données ouvertes selon quatre niveaux de certification : Bronze, Argent, Or et Platine. Chaque niveau correspond à des critères spécifiques répartis en quatre catégories : Légale, Pratique, Technique et Sociale.



Source (Open Data Institute)

Niveau	Catégorie	Critères
Bronze	Légal	Licence ouverte et réutilisable.
	Pratique	Accessibles sur le web.
	Technique	Formats lisibles par machine.
	Social	Documentation disponible.
Argent	Légal	Déclaration claire des droits.
	Légal	Conformité aux exigences de confidentialité.
	Légal	Données disponibles pendant au moins un an.
	Pratique	Liens web pour découvrir les données.
	Pratique	Données horodatées ou à jour.
Or	Technique	Formats adaptés au contenu.
	Légal	Licence en format lisible par machine.
	Pratique	Garantie d'actualisation des données.
	Pratique	Sauvegardes régulières effectuées.
	Technique	Formats ouverts et standardisés.
	Technique	URL cohérente pour le téléchargement.
Platine	Social	Coordonnées de contact.
	Légal	Provenance documentée en format lisible par machine.
	Technique	Identifiants uniques (URL) au sein des données.
	Social	Support dédié pour les utilisateurs.



Niveau Bronze:

Les données publiées sont sous licence ouverte autorisant leur réutilisation



Niveau Argent:

(en plus de Bronze): données pouvant être traitées automatiquement en plus de l'existence de la documentation nécessaire



Niveau Or:

(en plus d'Argent): conformité et utilisation des standards, avec des mises à jour périodiques



Niveau Platine:

(en plus d'Or): Documentation lisible par machine, identifiants uniques, ..

Session 4: Référentiels communs et Qualité des données

Autres méthodes d'évaluation de la data ouverte

Le self-scoring de Data.gouv.fr

Le score de qualité des métadonnées introduit par data.gouv.fr vise à aider les utilisateurs à identifier les jeux de données ayant des métadonnées complètes et fiables.

Critère	Description
Description des données	La description des données est de qualité (la description du jeu de données est suffisamment longue).
Mise à jour	La fréquence de mise à jour est renseignée et respectée.
Licence	La licence est renseignée et ouverte.
Métadonnées des ressources	Présence d'au moins une ressource avec un format ouvert déclaré.
Couverture spatiale	La couverture et la granularité spatiale est renseignée.
Couverture temporelle	La couverture temporelle des données est renseignée.



[Source \(Open Data France\)](#)

Session 4: Référentiels communs et Qualité des données

Évaluer la fiabilité de la donnée

Maximiser la sécurisation de la qualité de la data à la source

Avant d'adresser la problématique de la qualité des données, il faut que la mécanique d'input initiale de la data soit sécurisée au maximum au moment de l'input (moment où la donnée initiale est créée (soit par saisie manuelle par un agent public, ou un citoyen, ou une intégration de donnée)



Formulaires web ou SI

Si la data est saisie par un utilisateur externe ou un agent public:

- Utilisation de référentiels au maximum (adresses, villes, catégories, ...)
- Utilisation de règles de cohérence
- Utilisation de valeurs et formats obligatoires
- Structuration optimale de la donnée (répondant aux règles et standards)
- Trace des identifiants pour les besoins d'amélioration continue



Intégration de sources externes

Si la data provient de SI externes (partenaires, confrères, ..)

- Utilisation de standards communs
- Utilisation de référentiels communs
- Implémentation d'automates de contrôles (formats, structures, cohérence, ...)



IOT et équipements

Si la data provient d'automates, détecteurs ou IoT

- Contrôle et calibrage réguliers des équipements
- Mécanismes de contrôle de cohérence

Session 4: Référentiels communs et Qualité des données

Évaluer la fiabilité de la donnée

Une fois qu'on a la data (courante et historique)

Une Administration publique gère un volume important de data: opérationnelle, métier, technique, administrative, financière, etc.

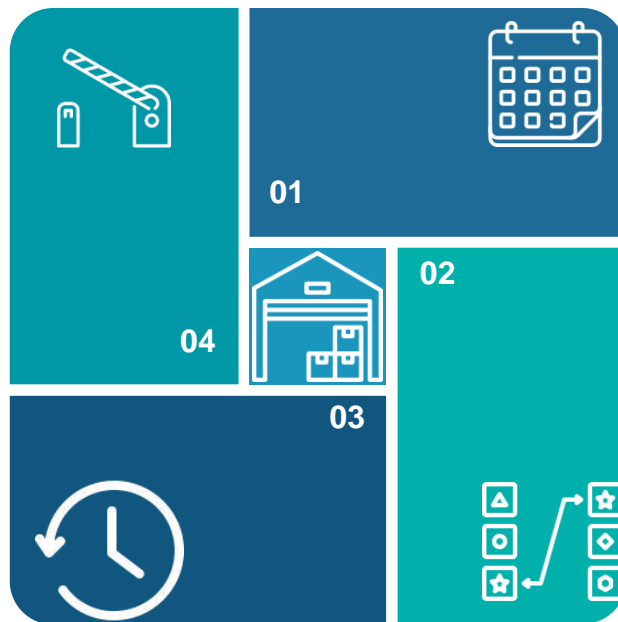
Data contrôlée à la source

Data provenant de formulaires web bien conçus, champs auto contrôlés dans les SI, ...
Taux de qualité correct



Historique et autres data

Provenant différentes années, familles métiers, etc.; dont la data n'a pas pu être contrôlée à la source
Taux de qualité inconnu



Tranches incomplètes

Tranches d'années dans l'historique ou Départements particuliers avec des zones dont la qualité est incertaine (donnée non collectée, changement organisationnel..)



Tranches différentes

Avec des règles de gestion ou contextes différents (donnée collectée d'une autre façon, avec une autre structure, ...)

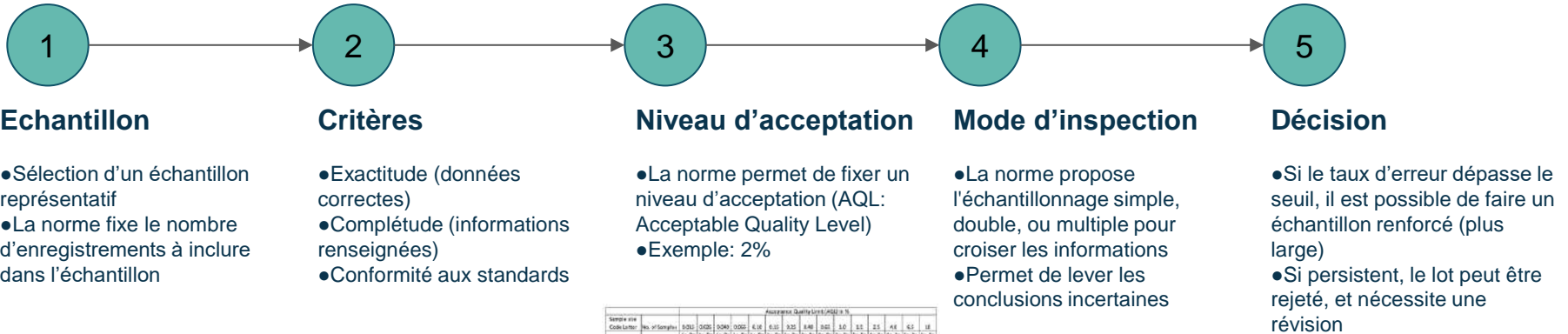


Session 4: Référentiels communs et Qualité des données

Évaluer la fiabilité de la donnée

Mesurer la qualité de la data stockée grâce à la norme ISO 2859

Malgré toutes les mesures mises en place, il est essentiel de pouvoir mesurer la qualité de la data disponible. Cela est d'autant plus vrai qu'une majorité de la data chez les institutions remonte à une profondeur historique importante dont la qualité n'est pas forcément connue.



Niveau de risque	Code Laiterie	Nbr. d'Échantillon	Acceptance Quality Level (AQL) in %															
			0.010	0.025	0.050	0.100	0.150	0.250	0.400	0.650	1.00	1.50	2.50	4.00	6.50	10.00	15.00	
A	2	32	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
B	3	65	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
C	4	125	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
D	5	250	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
E	6	500	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
F	7	1000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
G	8	2000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
H	9	4000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
I	10	8000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
J	11	16000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
K	12	32000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
L	13	65000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
M	14	130000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
N	15	260000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
O	16	520000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
P	17	1050000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
Q	18	2100000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
R	19	4200000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		
S	20	8500000	100	125	160	200	250	315	398	500	630	790	1000	1250	1600	2000		

Source ISO2859

Session 4: Référentiels communs et Qualité des données

Fiabiliser les données

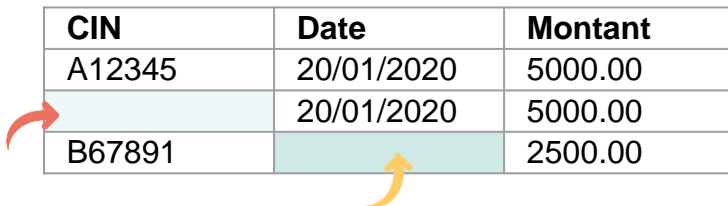
Bien définir ce qu'est une donnée correcte

La fiabilisation du patrimoine Data d'une institution est une opération titanesque. Afin d'avoir des résultats concrets et atteignables, il est essentiel d'adopter une approche stratégique, visant cibler les données que l'on souhaite fiabiliser, ainsi qu'une méthode industrialisée pour ce faire.

Complétude (présence)

- Une donnée obligatoire manquante n'a pas le même impact qu'une donnée facultative
- En tenir compte dans la notation

CIN	Date	Montant
A12345	20/01/2020	5000.00
	20/01/2020	5000.00
B67891		2500.00



Poids de la qualité

- Toutes les erreurs ne se valent pas
- En tenir compte dans la notation (pondération par exemple ou score)

Date: 20/01/2020 ↔ Date: 01/01/2020

CIN: A12345 ↔ CIN: A123

Session 4: Référentiels communs et Qualité des données

Fiabiliser les données

Adopter une approche stratégique pour la fiabilisation

La fiabilisation du patrimoine Data d'une institution est une opération titanesque. Afin d'avoir des résultats concrets et atteignables, il est essentiel d'adopter une approche stratégique, visant cibler les données que l'on souhaite fiabiliser, ainsi qu'une méthode industrialisée pour ce faire.



Prioriser par valeur

- Identifier les familles de données qui génèrent le plus de valeur (généralement les données métier)
- Prioriser les opérations de contrôle et fiabilisation sur les populations de données sélectionnées

Cas d'exemple:

Les Banques ont souvent recours à cette approche sur différentes lignes métier: Garanties & Crédits, Comptes, Recouvrement & Juridique, ...



Prioriser par besoins

- Prioriser selon les besoins émanant de l'intérieur ou extérieur
- Cela peut concerner des types de processus ou des tranches d'années, ou des régions

Cas d'exemple:

Il est possible de mettre en place une cellule de contrôle qui fiabilise les données avant publication, en continu
exemple: Fiabilisation de l'inventaire physique (très courant)



Prioriser par règles

- Utiliser des règles de gestion pour détecter les populations de données les plus susceptibles d'avoir des erreurs
- Extraire et contrôler les données n'ayant pas respecté les règles de gestion

Cas d'exemple:

En 2013, l'OMPIC a développé une solution de fiabilisation industrielle et paramétrable des données, visant à fiabiliser sur plusieurs années les données du Registre de Commerce.

Les campagnes ont été proposées selon différents critères (CA, Région, Secteur) ou règles métier (Incohérence de données financières, incohérences métier, ...)



Fiabilisation industrielle

- Il peut être nécessaire de développer des **outils** de fiabilisation et de traitement industriel du contrôle (suivi, traçabilité, automatisation)
- Également possible de sous traiter le traitement de la data