

# Qualité de données en santé

7 novembre 2023

**Module 3: Qualité des données - Majeure Santé EPITA**

Dr Damien Leprovost – AP-HP / LIMICS

# Creative Commons BY-SA 4.0 License

This course material is distributed under license  
[Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)

## You are free to:

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

## Under the following terms:

 **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

 **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.



# Course outline

## Introduction aux défis de la qualité des données en santé

- I. Notions fondamentales
- II. Pourquoi c'est important

## Cadre d'évaluation de la qualité des données

- III. Dimensions de la qualité
- IV. Orchestration de l'évaluation

## Procédures et outils de contrôle de la qualité des données

- V. Contenu du framework
- VI. Outil d'évaluation de la qualité des données

## Interopérabilité et qualité des données

- VII. Interopérabilité dans les soins de santé
- VIII. Solutions existantes

# I. Notions fondamentales

De quoi on parle ?

# Les données en santé

- Aujourd'hui, les données sont au cœur des soins de santé
- Le volume des données cliniques augmente de façon exponentielle



**Passage du papier aux systèmes de santé numériques**



**Augmentation du volume et de la variété des données collectées**

# Des sources très diverses



Dossier médical électronique



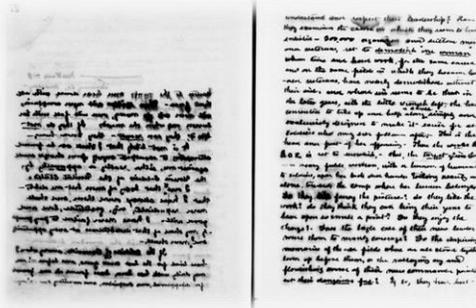
Contributions des prestataires



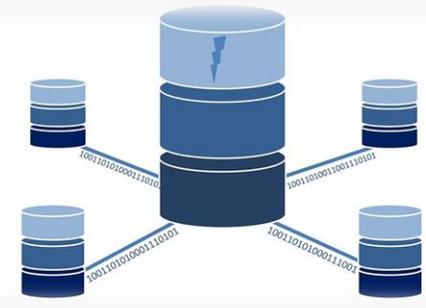
Données de l'appareil



Résultats des tests



OCR de documents



Bases de données externes

# La finalité des données de santé



- Documenter le statut du patient
- Communiquer les soins entre les prestataires
- Communiquer les plans de soins aux patients



- Identifier les événements facturables
- Maitriser les stocks d'équipements et de fournitures
- Prendre des décisions en matière de personnel



- Soutenir les projets de recherche
- Fournir des cohortes réelles

# La notion de qualité

- « Qualité » est un terme largement utilisé
- Nous sommes confrontés chaque jour à des exigences de qualité
- Mais qu'est-ce qui définit :
  - La qualité d'une voiture personnelle ?
  - La qualité de l'assistance à la clientèle ?
  - La qualité d'une vidéo ?
  - La qualité de ce cours ?



# La notion de qualité

- La qualité peut se résumer **à l'adaptation à l'objectif visé**

- 2 aspects fondamentaux :

Validation

Vérification

- 2 résultats principaux :

Confiance

Répétabilité

# Qualité des données pour les soins de santé

- La qualité des données **ne** vise **pas** à produire des données parfaites
- Données suffisamment précises, opportunes et cohérentes
  - de **prendre des décisions appropriées et fiables**



## II. Pourquoi c'est important

Risques et enjeux

# La qualité stimule la qualité

- La qualité des décisions dépend de la qualité des données sur lesquelles elles sont basées



Les décisions thérapeutiques nécessitent des informations précises et complètes pour garantir des soins de qualité



La facturation et la planification des ressources exigent une représentation fidèle de l'activité de soins



Les études scientifiques nécessitent des données fidèles pour présenter des résultats reproductibles

# Quality drives quality

- La qualité des décisions dépend de la qualité des données sur lesquelles elles sont basées



- Données incorrectes ou incomplètes concernant le patient
- Décisions susceptibles de mettre la vie en danger



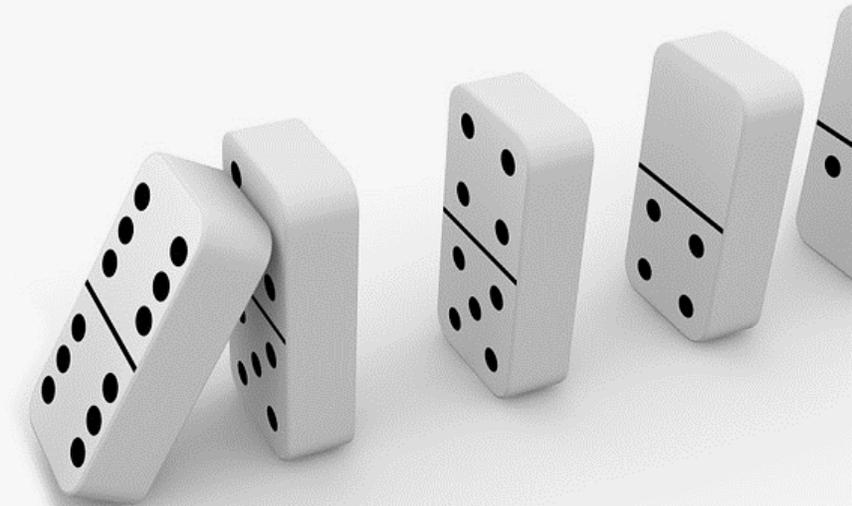
- Incidents de facturation, gaspillage d'argent
- Inventaires de fournitures incorrects, mauvaises décisions en matière d'équipement



- Hypothèses et modèles erronés
- Résultats inutilisables

# Une qualité implicite

- La plupart des utilisateurs (et des algorithmes !) ne s'interrogent pas sur la qualité des données qu'ils utilisent
- Des données impropres peuvent conduire à une mauvaise qualité de service
- Cela doit être garanti **en amont**



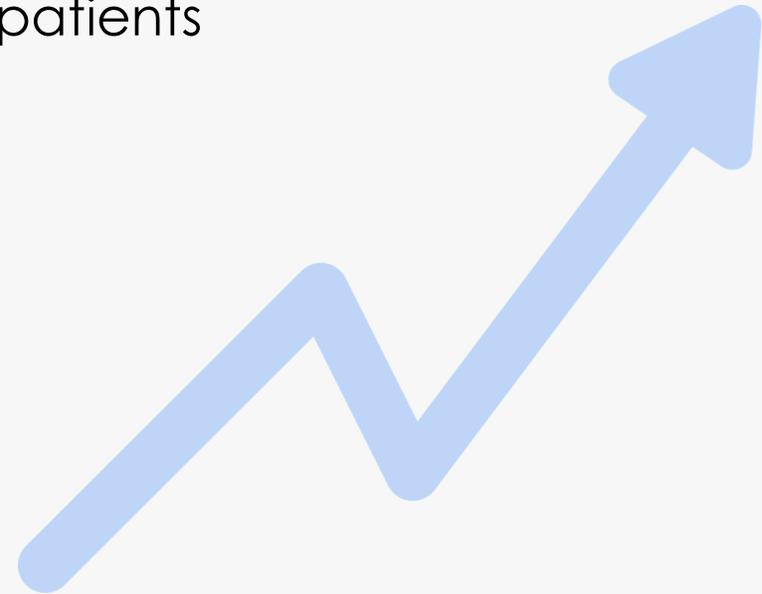
# Mais un possible cercle vertueux

- D'autre part, les utilisateurs de données peuvent avoir une plus grande confiance dans les données qu'ils obtiennent lorsqu'une approche systématique est adoptée pour évaluer la qualité de ces données.



# Objectifs de l'évaluation de la qualité des données

- Faciliter une planification stratégique mieux informée pour les soins de santé
- Des soins plus sûrs et de meilleure qualité pour les patients
- Des politiques mieux informées
- Contributions accrues à la recherche
- Amélioration de la santé de la population



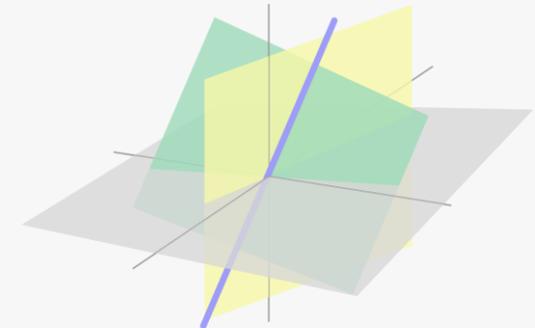
### III. Dimensions de la qualité

Principales caractéristiques de la qualité des données

# Un espace multidimensionnel

- La qualité des données est modélisée comme un concept **multidimensionnel**
- Chaque dimension de la qualité des données saisit un aspect **mesurable** particulier de la qualité des données

Les dimensions de la qualité des données représentent les points de vue, les repères ou les mesures des problèmes de qualité des données qui peuvent être compris, analysés et résolus.

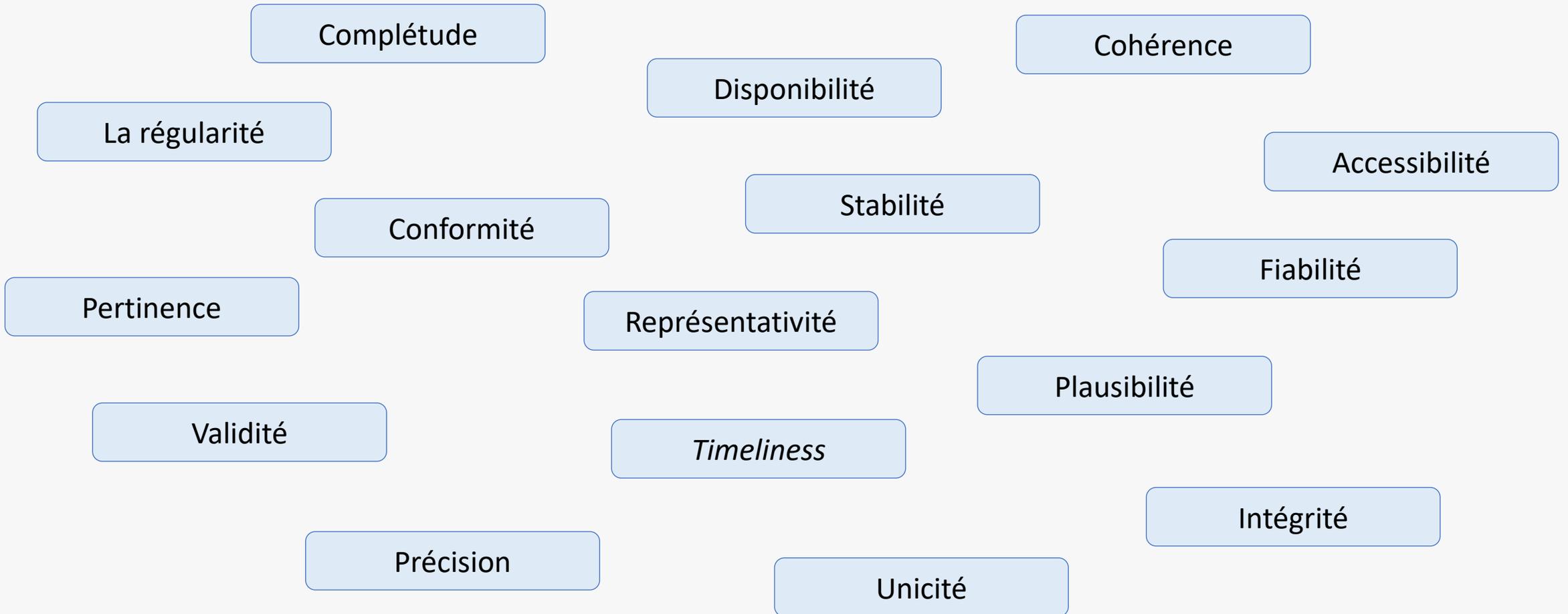


# Les dimensions ne sont pas figées

- Il s'agit de plusieurs dimensions fréquemment utilisées, mais.. :
  - pas de consensus
  - pas d'ensemble normalisé de dimensions
- Les implémentations définissent généralement de 3 à 9 dimensions
- De nombreuses dimensions proposées visent à traiter des caractéristiques similaires de la qualité des données



# Quelques dimensions utilisées dans la qualité des données



# Exemple courant : Complétude

- Définition :
  - Les données requises sont stockées dans la base de données
  - Toutes les variables associées ont des valeurs
- Exemples :
  - Le patient a une date de naissance
  - Le nombre de résultats de tests correspond au nombre de patients testés
- Objet :
  - détecter les données manquantes et/ou les valeurs manquantes

# Exemple courant : Cohérence

- Définition :
  - Les valeurs des données correspondent à leurs contraintes structurelles et syntaxiques
  - Toutes les valeurs d'une même variable correspondent au même format et aux mêmes spécifications
- Exemples :
  - L'heure d'admission du patient est enregistrée sous forme de DATETIME avec des valeurs AAAA-MM-JJ hh:mm:ss.
  - Les taux d'insuline sont stockés sous forme de FLOAT et exprimés en  $\mu\text{U/L}$ .
- Objet :
  - détecter les données qui ne seront pas utilisables

# Exemple courant : *Timeliness* (Respect des délais)

- Définition :
  - Les données sont à jour pour leur utilisation
  - Les données sont disponibles à temps
- Exemples :
  - L'unité de soins actuels du patient reflète la situation réelle du patient
  - Les résultats du test Covid-19 d'hier sont l'agrégation des résultats de toutes les unités de test de l'hôpital.
- Objet :
  - détecter l'inadéquation temporelle entre le flux d'alimentation et l'utilisation des données

# Dimensions de la qualité des données

- Chaque dimension mesure une **caractéristique indépendante** des données
- Il n'y a pas de dimensions fixes
- Les dimensions doivent être choisies judicieusement, en accord avec :
  - Les usages
  - Le contexte
  - Les ressources



# Exemple : Les dimensions du framework de Kahn

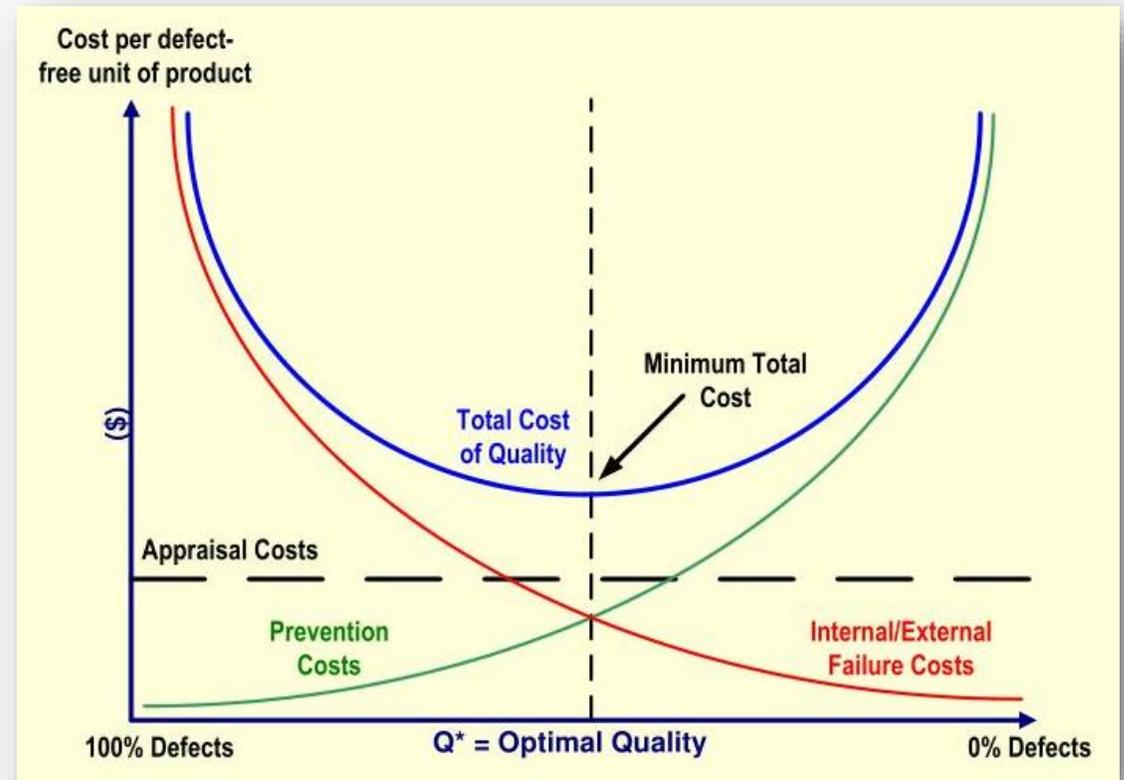
- Le framework de Kahn
  - M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051581/pdf/egems1244.pdf>
- Un cadre de qualité des données proposé en 2016 pour l'utilisation secondaire des données des dossiers médicaux électroniques
- Utilisation secondaire : utilisation des données à des fins autres que celles pour lesquelles elles ont été collectées à l'origine
  - examen opérationnel
  - l'amélioration de la qualité
  - recherche analytique
- La qualité est évaluée selon trois dimensions
  - ▶ **Conformité**
  - ▶ **Complétude**
  - ▶ **Plausibilité**

## IV. Orchestration de l'évaluation

Pourquoi un framework est nécessaire

# Rappel : La qualité est relative

- La qualité des données est une **mesure relative**
- Il dépend des ressources financières et humaines
- La qualité et la quantité des données doivent être mises en balance avec les ressources disponibles



# Pourquoi un framework est nécessaire

- La collecte de données est un processus intensif, couteux en ressources
- Les organisations doivent en tenir compte dès le départ afin de s'assurer que la quantité et la qualité des données collectées correspondent aux besoins des utilisateurs
- Cela nécessite un cadre réfléchi, pour :
  - établir une base de référence pour la qualité des données
  - identifier les domaines susceptibles d'être améliorés
  - évaluer l'impact de tout changement de pratique ou de politique
  - rendre compte sur l'amélioration de la qualité des données

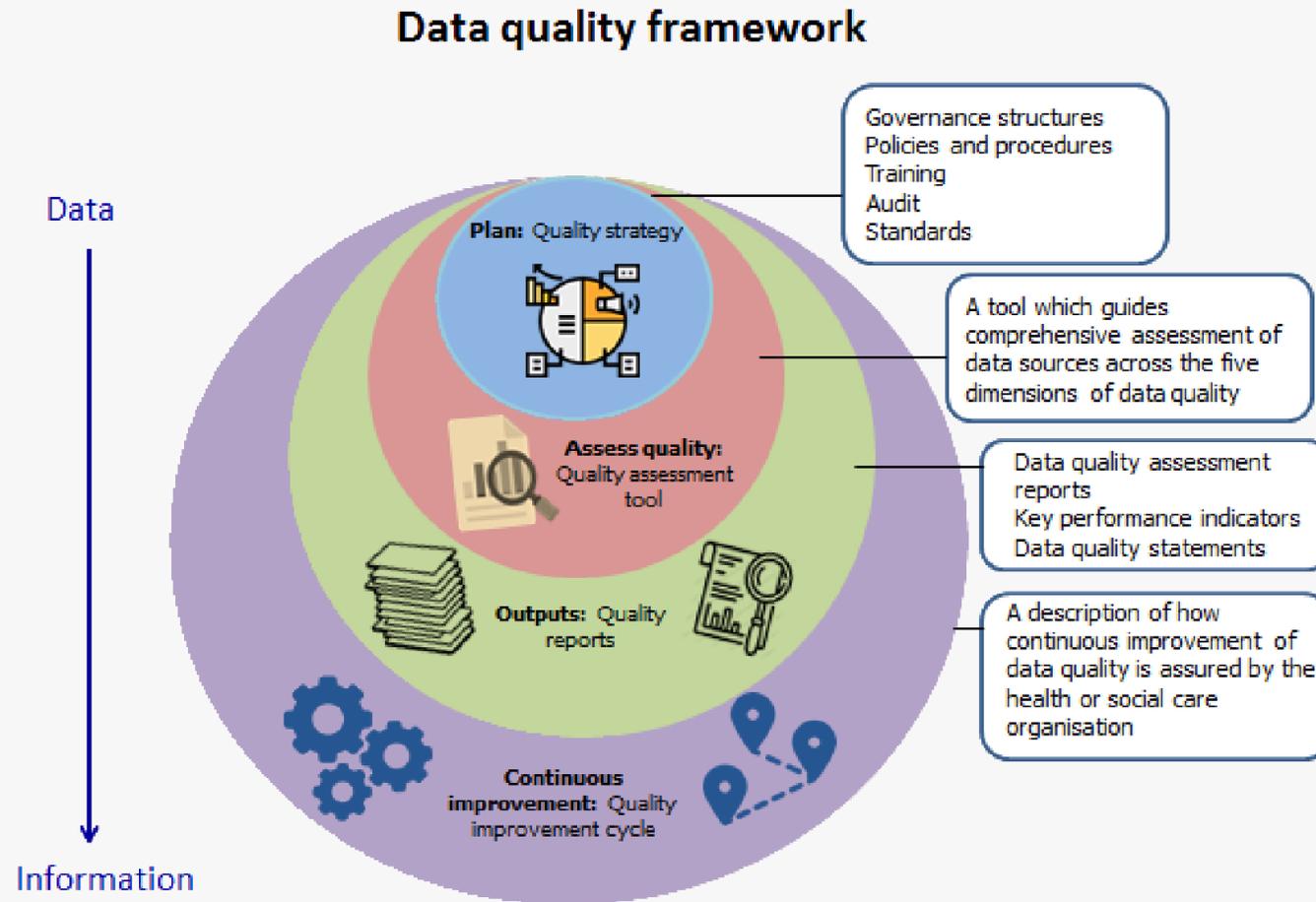


# Objectifs du framework

- Un framework de qualité des données :
  - décrit l'approche
  - fournit les outils qui peuvent être utilisés pour :
    - évaluer la qualité des données
    - documenter la qualité des données
    - améliorer la qualité des données
- Les composantes d'un framework de qualité des données sont les suivantes

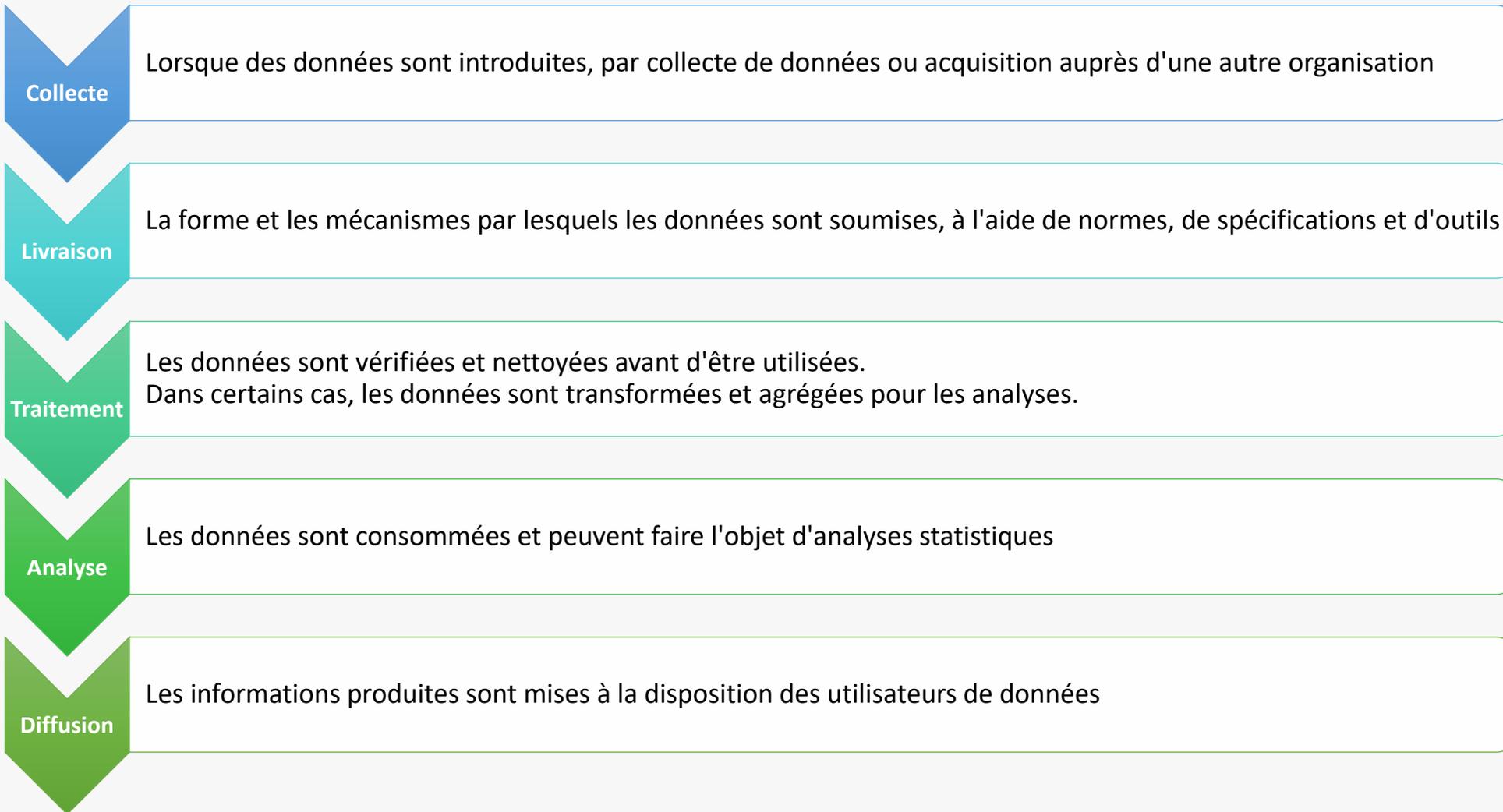


# Example: le framework HIQA



source: Health Information and Quality Authority, Ireland – <https://www.hiqa.ie>

# Cycle de vie des données



source : Health Information and Quality Authority, Irlande – <https://www.higa.ie>

# Cycle de vie des données et dimensions de la qualité

- L'applicabilité d'une dimension d'une qualité de données particulière peut varier :
  - à une étape donnée
  - d'une organisation à une autre
- Déterminer avec soin à quelle étape du cycle de vie une dimension particulière de la qualité des données doit être appliquée

		<i>Dimensions</i>			
		Complétude	Cohérence	<i>Timeliness</i>	Unicité
<i>Les étapes</i>	Capture			X	
	Soumission		X		
	Traitement	X			
	Analyse				X
	Diffusion	X	X	X	X

*Exemple de cartographie des dimensions du cycle de vie*

## V. Contenu du framework

Contrôler et assurer la qualité des données  
à l'aide de lignes directrices

# Principaux éléments du framework

## Stratégie de qualité des données

Approche formalisée de la qualité des données définissant les activités à entreprendre

## Outil d'évaluation de la qualité des données

Critères permettant l'évaluation exhaustive des sources de données via les dimensions définies

## Rapport sur la qualité des données

Rapports d'évaluation de la qualité des données, métriques et indicateurs clés de performance

## Un cycle d'amélioration de la qualité des données

Description de la manière dont l'amélioration continue de la qualité des données est gérée ; englobe les processus et méthodologies appliqués dans le cadre de l'amélioration de la qualité des données

# La stratégie de qualité des données

- Un ensemble de décisions pour gérer et améliorer en permanence la qualité des données
- Centré sur les objectifs de qualité des données de l'organisation
- Déterminer :
  - les processus de données pour améliorer
  - des solutions à mettre en œuvre
  - des personnes à engager
- Formaliser les approches permettant d'identifier, de documenter et d'assurer la mise en œuvre des pratiques en matière de qualité des données.

# Mise en œuvre de la stratégie de qualité des données

- Une stratégie de qualité des données peut être un document autonome ou être intégrée à une stratégie globale de gestion de l'information.

**Rappel :** Les décisions prises dans le cadre des soins de santé ne sont valables que dans la mesure où les informations sur lesquelles elles se fondent le sont.

Les organismes de soins de santé ont la responsabilité de consacrer le temps et les ressources nécessaires pour s'assurer que la qualité des données est contrôlée et assurée.

# Composantes de la stratégie de qualité des données

- Des informations sur les points suivants peuvent être incluses dans une stratégie de qualité des données :

## Dispositions en matière de gouvernance

La stratégie doit définir clairement les modalités de gouvernance, les rôles et les responsabilités.

La qualité des données devrait faire partie intégrante des structures de gouvernance existantes.

## Politiques et procédures

Les politiques sont des déclarations écrites décrivant les valeurs d'une organisation en ce qui concerne la qualité de ses données.

Les procédures décrivent les étapes nécessaires pour obtenir des données de bonne qualité.

Les politiques et les procédures doivent prendre en compte chaque dimension de la qualité des données.

## Programme de formation, d'éducation et de développement

Une main-d'œuvre qualifiée et compétente est essentielle pour mettre en œuvre avec succès une stratégie de qualité des données.

Ce programme devrait s'appuyer sur les résultats d'une analyse des besoins de formation.

Un recrutement et une planification des effectifs efficaces garantissent que le personnel possède les compétences et les aptitudes nécessaires.

## Audits de la qualité des données

Les audits de la qualité des données peuvent être réalisés en interne ou en externe et doivent être entrepris régulièrement.

L'audit initial peut servir de référence pour évaluer les améliorations de la qualité des données au fil du temps.

Un plan d'action pour l'amélioration de la qualité doit être élaboré sur la base des résultats de l'audit.

## Normes de qualité des données

Le respect des normes permet d'améliorer la qualité des données et de faciliter la comparaison des informations en favorisant le partage des données et l'interopérabilité.

Les normes les plus couramment utilisées sont les dictionnaires de données et les normes terminologiques.

# L'outil d'évaluation des données

- L'outil d'évaluation des données fournit un ensemble détaillé de critères permettant d'évaluer les sources de données en fonction de critères de qualité
- Les dimensions sont définies par la stratégie de qualité des données
- Les dimensions et l'outil d'évaluation doivent être **stables dans le temps**
  - Condition requise pour la comparaison et le cycle d'amélioration continue

# Rapport sur la qualité des données

- Rapport sur les résultats de l'évaluation de la qualité des données
- Les rapports sur la qualité des données peuvent inclure
  - les rapports d'évaluation interne ou externe de la qualité des données
  - l'établissement de rapports sur les indicateurs clés de performance (KPI)
  - la production de déclarations sur la qualité des données
- Afin de communiquer aux utilisateurs des données dans quelle mesure les données sont "adaptées à l'usage".
- Les rapports sur la qualité des données peuvent être établis en interne et/ou en externe.



# Reporting: Data quality assessment reports

- Describe the quality of the data according to the dimensions defined by the data quality strategy
- Should include details on data quality assessments conducted to identify the strengths and weaknesses of the organisation's data
- Should be presented using a standardised reporting structure
  - facilitate readability and comparability

# Rapports : Indicateurs clés de performance (KPI)

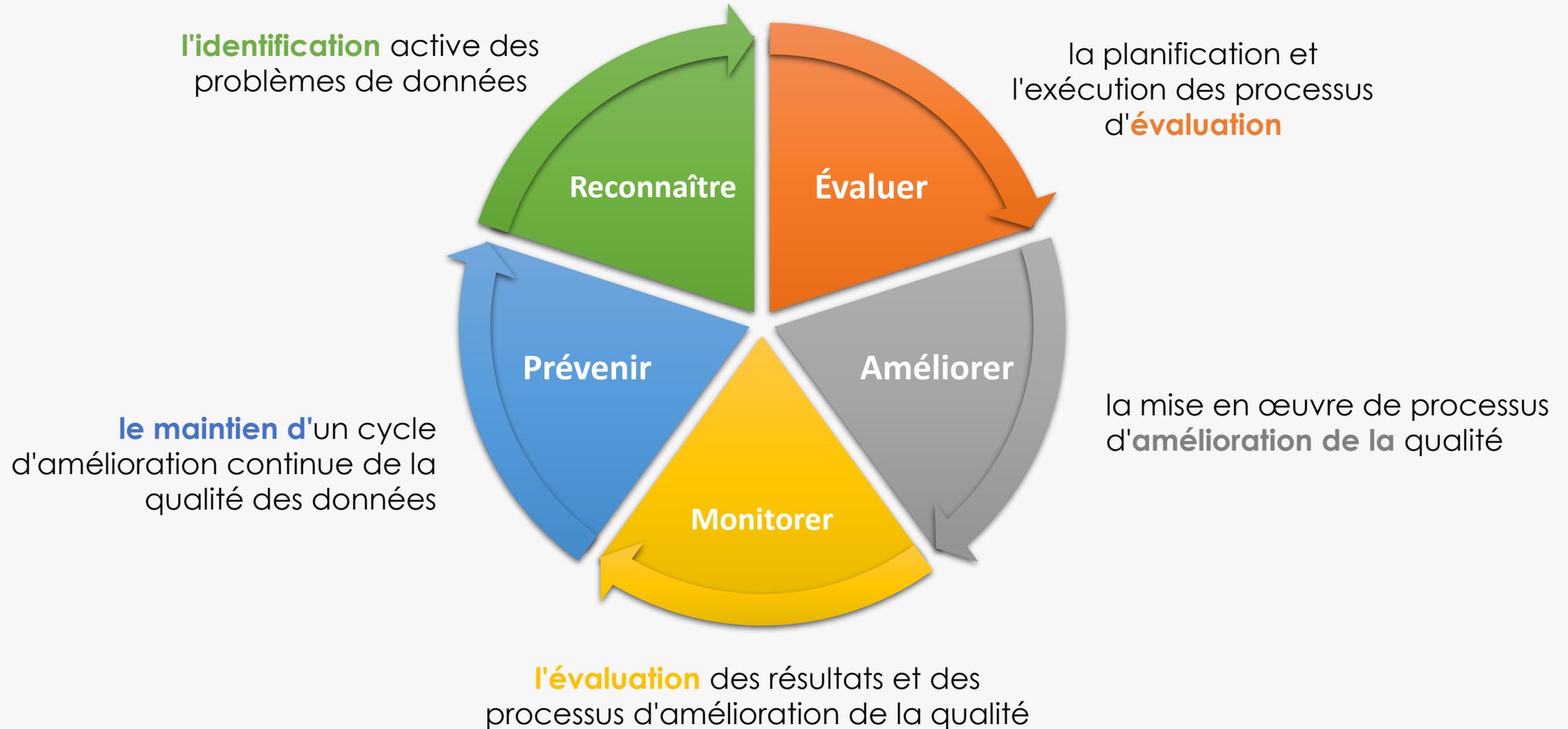
- Éléments spécifiques et mesurables
- Utilisé pour caractériser la qualité des données et rendre compte des performances
- Conçu pour évaluer les aspects clés des structures, des processus et des résultats de la qualité des données

Lorsque l'on rend compte de la qualité des données, il est important d'établir et de mesurer des indicateurs spécifiques en rapport avec chacune des dimensions de la qualité des données.

# Rapports : Déclarations sur la qualité des données

- Une déclaration préparée pour accompagner les principaux résultats publiés
- Sur la base des résultats de l'évaluation de la qualité des données effectuée
- Mettre en évidence les forces et les faiblesses des données, évaluées à l'aide des dimensions
- Le contenu varie en fonction des données publiées, mais il est susceptible d'inclure :
  - une vue d'ensemble de la collecte de données et de son objectif
  - une description de la source de données
  - un résumé des problèmes identifiés avec les données
  - un aperçu de la qualité des données pour chacune des dimensions

# Cycle d'amélioration de la qualité des données



## VI. Outil d'évaluation de la qualité des données

Identifier les métriques à monitorer

# L'outil d'évaluation de la qualité des données

- L'outil d'évaluation fournit un ensemble détaillé de critères permettant d'évaluer les sources de données en fonction des dimensions définies
- Les dimensions ne s'excluent pas mutuellement, des compromis entre les dimensions peuvent être nécessaires à certains points du cycle de vie des données

Les dimensions ne sont pas nécessairement pondérées de la même manière, car l'importance de chaque dimension varie en fonction de la source de données et du besoin.

- L'outil d'évaluation de la qualité des données est organisé en sections, chacune se concentrant sur une dimension de la qualité des données.

*Les dimensions décrites ici s'inspirent fortement du framework de qualité des données du CIHI, du framework d'assurance qualité du système statistique européen et du framework d'assurance qualité de la Division des statistiques des Nations unies.*

**Comme indiqué précédemment, les caractéristiques et les critères inclus ici ne constituent pas une liste exhaustive et peuvent être modifiés par les organisations si nécessaire pour répondre à leurs besoins et à leurs spécifications.**

# Structure de l'outil d'évaluation de la qualité des données

<b>Introduction à la dimension</b>	<p>Une explication de la dimension de la qualité des données et de l'importance de l'évaluation de la qualité des données dans le cadre de cette dimension</p>				
<b>Les caractéristiques de qualité de la dimension</b>	<p>Une description des principales caractéristiques de qualité, telles que définies par les bonnes pratiques internationales, qui fournissent un contexte aux organisations sur la manière dont la qualité des données peut être évaluée pour déterminer si elles sont "adaptées à l'objectif". Chaque caractéristique est accompagnée d'une description qui fournit des explications supplémentaires et clarifie la compréhension des caractéristiques clés.</p> <table border="1" data-bbox="677 629 1228 686"> <tr> <td>Caractéristique</td> <td>Description</td> </tr> </table>	Caractéristique	Description		
Caractéristique	Description				
<b>L'outil d'évaluation</b>	<p>Chaque caractéristique est constituée d'un ensemble de critères, présentés sous forme de questions clés, qui permettent aux organisations d'évaluer le degré de qualité de leurs données. Chaque critère est accompagné d'une section détaillée qui explique en termes simples les mesures pratiques que les organisations doivent prendre pour évaluer la qualité des données.</p> <table border="1" data-bbox="677 951 1783 1008"> <tr> <td>Caractéristique</td> <td>Critères</td> <td>Détails</td> <td>L'évaluation</td> </tr> </table>	Caractéristique	Critères	Détails	L'évaluation
Caractéristique	Critères	Détails	L'évaluation		
<b>Exemples d'évaluation de cette dimension</b>	<p>C'est une bonne pratique que de fournir des exemples de la manière dont les collectes de données nationales évaluent actuellement la qualité de leurs données sous cette dimension.</p>				

# Exemple : Précision

## Définition de la dimension

L'exactitude des données se réfère à la mesure dans laquelle les données décrivent correctement ce qu'elles ont été conçues pour mesurer.

## Introduction à l'évaluation

Les organismes de santé doivent mettre en place des processus qui leur permettent de produire et de diffuser des données reflétant fidèlement la réalité. Par exemple, des données précises sur l'incidence des maladies infectieuses sont essentielles pour planifier des interventions de santé publique telles que des programmes de vaccination.

La mesure de l'exactitude permet de quantifier la mesure dans laquelle les données reflètent correctement et régulièrement la situation qu'elles sont censées mesurer. Le degré de précision des données a des implications claires sur l'utilité et la signification des données pour l'interprétation et l'analyse ultérieure.

L'exactitude des données dépend de la mesure dans laquelle les enregistrements individuels reçus contiennent des informations complètes et valides. Ainsi, plus l'ensemble de données est complet, plus les données sont susceptibles d'être exactes.

L'exactitude de leurs données est quantifiée par le degré de différence entre l'estimation et la valeur réelle des données quantifiées. Lorsque des inexactitudes sont identifiées dans les données, elles doivent être reconnues et corrigées dans la mesure du possible.

# Exemple : Précision

## Caractéristiques de la dimension

Caractéristique	Description
Couverture	<p>La mesure dans laquelle les données disponibles pour les utilisateurs couvrent la population ou l'événement en question. Il est essentiel de comprendre le niveau de couverture d'une population pour pouvoir faire une déclaration sur cette population.</p> <p>La couverture peut être évaluée en définissant la population en question et en déterminant et contrôlant les taux de couverture. Toute limitation des données due à des problèmes de couverture doit être clairement expliquée.</p>
Saisie et collecte des données	<p>Les procédures mises en place pour garantir que les données sont saisies dans un format utilisable et peuvent être préparées pour la soumission.</p> <p>Cet aspect est évalué en déterminant les mesures mises en place pour garantir que les relations avec les fournisseurs de données sont positives et que des processus appropriés sont en place pour faciliter des flux de données efficaces et efficaces.</p>

# Exemple : Précision

## Caractéristiques de la dimension

Caractéristique	Description
Traitement des données	<p>La transformation des données du format dans lequel elles sont reçues en un autre format qui facilite l'analyse. Le traitement peut inclure la validation et la correction des données.</p> <p>L'évaluation consiste à vérifier si le traitement a été documenté en détail et si les données traitées peuvent être clairement différenciées des données originales reçues.</p>
Complétude et validité	<p>Le degré de présence des variables individuelles dans un ensemble de données. Les données incomplètes et non valides d'un ensemble de données ont un impact sur l'aptitude à l'emploi de cet ensemble de données.</p> <p>L'évaluation de l'exhaustivité et de la validité se fait en calculant et en contrôlant les taux de valeurs valides, invalides, manquantes et aberrantes.</p>
Révisions des données	<p>La mesure dans laquelle les données sont susceptibles d'être révisées ou corrigées à la suite de nouvelles informations ou de la correction d'erreurs dans le traitement ou l'estimation des données et le délai dans lequel les révisions sont produites.</p>

# Exemple : Précision

## Outil d'évaluation de la qualité des données

Caractéristique	Critères	Détails	Evaluation
Couverture	Les détails de la population de référence sont-ils explicitement indiqués dans tous les communiqués et la couverture de la population est-elle quantifiée ?	Indiquer si la population de référence est documentée et si la couverture a été quantifiée. Si ce n'est pas le cas, fournir des détails.  La population de référence est la population pour laquelle des informations devraient être disponibles.	Oui / Non / Partiellement
	Les problèmes de couverture majeurs susceptibles d'avoir une incidence sur l'interprétation des données sont-ils documentés et mis à la disposition des utilisateurs ?	Fournir des détails sur tout problème important lié à la couverture qui pourrait avoir une incidence sur l'interprétation des utilisateurs des données.  Examiner l'impact de tout problème de couverture identifié en termes d'impact aux différents niveaux de reporting.  Si les problèmes de couverture ne sont pas documentés, il convient d'en expliquer les raisons.	Oui / Non / Sans objet
	Des procédures ont-elles été mises en place pour identifier et traiter les doublons et les doublons potentiels dans les données ?	Décrire les processus mis en place pour identifier les doublons et calculer le taux de doublons afin de donner une indication de leur fréquence. Indiquez comment les doublons sont supprimés une fois qu'ils ont été identifiés.  Si aucune mesure n'est prise pour identifier et supprimer les doublons, expliquez pourquoi et donnez une estimation du nombre de doublons qui subsistent dans les données.	Oui / Non / Partiellement

# Précision et fiabilité

## Outil d'évaluation de la qualité des données

Caractéristique	Critères	Détails	Evaluation
Saisie et collecte des données	Les problèmes de qualité des données susceptibles d'avoir un impact significatif sur l'interprétation des données sont-ils traités et documentés à l'intention des utilisateurs ?	Fournir des détails sur les problèmes de qualité des données identifiés, y compris l'action entreprise pour résoudre le problème, le délai de résolution du problème et tout impact potentiel de ces problèmes sur l'interprétation et l'analyse des données.	Oui / Non / Sans objet
Traitement des données	Les processus de validation des données sont-ils appliqués de manière cohérente et sont-ils documentés pour les utilisateurs des données ?	Fournir des détails sur les types de contrôles de validation effectués, comme les contrôles de structure et d'intégrité, le contrôle des données manquantes, de conformité aux spécifications et de contrôle des valeurs aberrantes.	Oui / Non / Partiellement
Exhaustivité et validité	Les taux de valeurs valides, invalides, manquantes et aberrantes sont-ils documentés et communiqués à chaque publication de données ?	Fournir un profil de données pour les éléments de données qui incluent les taux de valeurs valides, invalides et manquantes.	Oui / Non / Partiellement
Révisions des données	Les révisions ou corrections apportées aux données sont-elles régulièrement analysées afin de garantir leur utilisation statistique efficace ?	Fournir des détails sur les raisons, le calendrier et la nature des révisions.  Identifier clairement les données comme préliminaires ou révisées.	Oui / Non

# L'exemple du framework de Kahn

- Framework de qualité des données proposé en 2016 pour l'utilisation secondaire des données des dossiers médicaux électroniques
- Utilisation secondaire : utilisation des données à des fins autres que celles pour lesquelles elles ont été collectées à l'origine
  - examen opérationnel
  - l'amélioration de la qualité
  - recherche analytique
- Définit trois dimensions de la qualité des données
  - Conformité
  - Complétude
  - Plausibilité
- Appliqué à deux contextes d'évaluation de la qualité des données
  - Vérification
  - Validation

M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051581/pdf/egems1244.pdf>

# L'exemple du framework de Kahn

VERIFICATION		VALIDATION	
DEFINITION	EXAMPLE	DEFINITION	EXAMPLE
<b>CONFORMANCE: DO DATA VALUES ADHERE TO SPECIFIED STANDARDS AND FORMATS?</b>			
<b>VALUE CONFORMANCE</b>			
<p>a. Data values conform to internal formatting constraints.</p> <p>b. Data values conform to allowable values or ranges.</p>	<p>a. Sex is only one ASCII character.</p> <p>b. Sex only has values "M," "F," or "U."</p>	<p>a. Data values conform to representational constraints based on external standards.</p>	<p>a. Values for primary language conform to ISO standards.</p>
<b>RELATIONAL CONFORMANCE</b>			
<p>a. Data values conform to relational constraints.</p> <p>b. Unique (key) data values are not duplicated.</p> <p>c. Changes to the data model or data model versioning.</p>	<p>a. Patient medical record number links to other tables as required.</p> <p>b. A medical record number is assigned to a single patient.</p> <p>c. Version 1 data does not include medical discharge hour.</p>	<p>a. Data values conform to relational constraints based on external standards.</p>	<p>a. Data values conform to all not-NULL requirements in a common multi-institutional data exchange format.</p>
<b>COMPUTATIONAL CONFORMANCE</b>			
<p>a. Computed values conform to computational or programming specifications.</p>	<p>a. Database- and hard-calculated Body Mass Index (BMI) values are identical.</p>	<p>a. Computed results based on published algorithms yield values that match validation values provided by external source.</p>	<p>a. Computed BMI percentiles yield identical values compared to test results and values provided by the CDC.</p>

M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016

# L'exemple du framework de Kahn

VERIFICATION		VALIDATION	
DEFINITION	EXAMPLE	DEFINITION	EXAMPLE
<b>COMPLETENESS: ARE DATA VALUES PRESENT?</b>			
<p>a. The absence of data values at a single moment in time agrees with local or common expectations.</p> <p>b. The absence of data values measured over time agrees with local or common expectations.</p>	<p>a. The encounter ID variable has missing values.</p> <p>b. Gender should not be null.</p> <p>c. Medical discharge time is missing for three consecutive days.</p>	<p>a. The absence of data values at a single moment in time agrees with trusted reference standards or external knowledge.</p> <p>b. The absence of data values measured over time agrees with trusted reference standards or external knowledge.</p>	<p>a. The current encounter ID variable is missing twice as many values as the institutionally validated database.</p> <p>b. A drop in ICD-9CM codes matches implementation of ICD-10CM</p>
<b>PLAUSIBILITY: ARE DATA VALUES BELIEVABLE?</b>			
<b>UNIQUENESS PLAUSIBILITY</b>			
<p>a. Data values that identify a single object are not duplicated.</p>	<p>a. Patients from a single institution do not have multiple medical record numbers.</p>	<p>a. Data values that identify a single object in an external source are not duplicated.</p>	<p>a. An institution's CMS facility identifier does not refer to a multiple institutions.</p>

M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016

# L'exemple du framework de Kahn

VERIFICATION		VALIDATION	
DEFINITION	EXAMPLE	DEFINITION	EXAMPLE
<b>ATEMPORAL PLAUSIBILITY</b>			
<p>a. Data values and distributions agree with an internal measurement or local knowledge.</p> <p>b. Data values and distributions for independent measurements of the same fact are in agreement.</p> <p>c. Logical constraints between values agree with local or common knowledge (includes "expected" missingness).</p> <p>d. Values of repeated measurement of the same fact show expected variability.</p>	<p>a. Height and weight values are positive.</p> <p>a. Counts of unique patients by diagnoses are as expected</p> <p>a. Distribution of encounters per patient or medications per encounter distributions are as expected</p> <p>b. Serum glucose measurement is similar to finger stick glucose measurement.</p> <p>b. Oral and axillary temperatures are similar.</p> <p>c. Sex values agree with sex-specific contexts (pregnancy, prostate cancer).</p> <p>d. Height values are similar when taken by two separate nurses within the same facility using the same equipment.</p>	<p>a. Data values and distributions (including subgroup distributions) agree with trusted reference standards or external knowledge.</p> <p>b. Similar values for identical measurements are obtained from two independent databases representing the same observations with equal credibility.</p> <p>c. Two dependent databases (e.g., database 1 abstracted from database 2) yield similar values for identical variables.</p>	<p>a. HbA1c values from hospital and national reference lab are statistically similar under the same conditions.</p> <p>a. Distribution of patients with cardiovascular disease diagnoses are similar to CDC rates for the same age and sex groups</p> <p>a. Readmission rates by age groups for Medicare patients agree with CMS values</p> <p>b. Diabetes ICD-9CM and CPT codes are similar between two independent claims databases serving similar populations.</p> <p>c. Recorded date of birth is consistent between EHR data and registry data for the same patient.</p>
<b>TEMPORAL PLAUSIBILITY</b>			
<p>a. Observed or derived values conform to expected temporal properties.</p> <p>b. Sequences of values that represent state transitions conform to expected properties.</p> <p>c. Measures of data value density against a time-oriented denominator are expected <i>based on internal knowledge</i>.</p>	<p>a. Admission date occurs before discharge date.</p> <p>b. Date of an initial immunization precedes date of a booster immunization.</p> <p>c. Similar counts of patient observations between extraction-transformation-load cycles.</p> <p>c. Counts of emergency room visits by month shows expected spike during flu season.</p> <p>c. Medications per patient-day are as expected</p>	<p>a. Observed or derived values have similar temporal properties across one or more external comparators or gold standards.</p> <p>b. Sequences of values that represent state transitions are similar to external comparators or gold standards.</p> <p>c. Measures of data value density against a time-oriented denominator are expected <i>based on external knowledge</i>.</p>	<p>a. Length of stay by outpatient procedure types conforms to Medicare data for similar populations.</p> <p>b. Immunization sequences match the CDC recommendations.</p> <p>c. Counts of emergency room visits by month shows spike during flu season that are similar to local health department reports.</p> <p>c. Medications per patient-day matches claims data.</p>

M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016

# L'exemple du DQD d'OHDSI



- OHDSI définit un modèle de données commun (CDM) pour l'analyse systématique de bases de données disparates d'observation médicales
  - Le CDM OHDSI OMOP : <https://www.ohdsi.org/data-standardization/the-common-data-model/>
- L'utilisation d'un CDM permet de **partager la mise en œuvre** des règles et des outils d'évaluation.
- Deux contributions notables :
  - Les règles ACHILLES (Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems)
  - Le tableau de bord de la qualité des données

# L'exemple du DQD d'OHDSI



- Basé sur le cadre de Kahn
- Evaluation de métriques de la qualité des données
- Évalue les métriques par rapport à un seuil prédéfini
- Restitue les évaluations dans une interface visuelle

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	95%

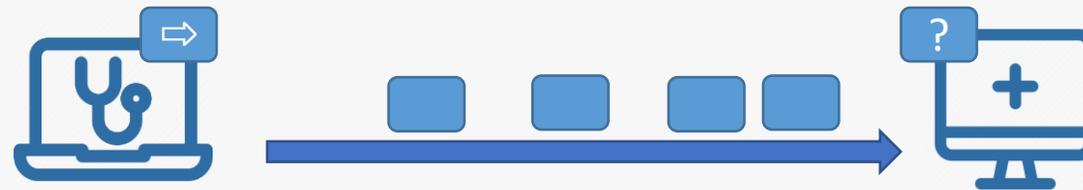
Exemple en ligne: <https://data.ohdsi.org/DataQualityDashboard/>

# VII. L'interopérabilité dans les soins de santé

Défis et risques

# Interopérabilité dans les systèmes de données de santé

- La capacité de différents systèmes d'information et applications logicielles à :
  - communiquer
  - échanger des données
  - utiliser les informations échangées



- Les schémas et les normes d'échange de données doivent permettre de partager les données indépendamment de l'application ou de son fournisseur
- L'interopérabilité est la capacité des systèmes d'information sur la santé à fonctionner ensemble à l'intérieur et au-delà des frontières de l'organisation

# Interopérabilité structurelle

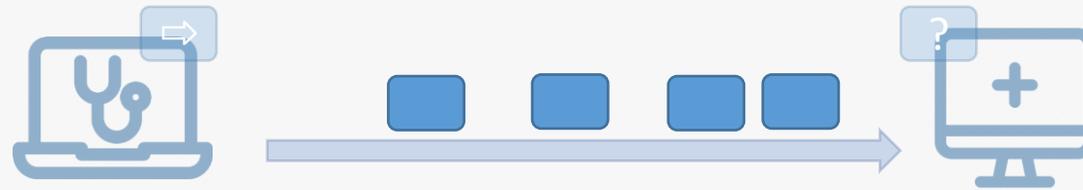
- Capacité de deux systèmes (ou plus) à émettre et à recevoir des données
- Effectuer une tâche donnée sans intervention supplémentaire de l'opérateur



- Nécessité d'un accord sur les protocoles, les API, la sécurité
- Indépendant du domaine d'utilisation

# Interopérabilité sémantique

- Capacité des systèmes informatiques à échanger et à interpréter les informations
- L'expéditeur et le destinataire ont tous deux des données qui signifient exactement la même chose



- Exige un accord sur les codes et les usages
- Spécifique au domaine et aux concepts

# Interopérabilité organisationnelle

- Capacité de transmettre et d'utiliser des données pertinentes dans le cadre des soins aux patients
- Utiliser les nouvelles informations d'une manière différente de ce qu'ils auraient fait sans elles, ce qui leur permet d'obtenir des résultats différents



- Le niveau humain d'interopérabilité
- Nécessite des changements dans le flux de travail et dans la manière dont les praticiens et les systèmes cliniques fonctionnent à un niveau de détail très fin

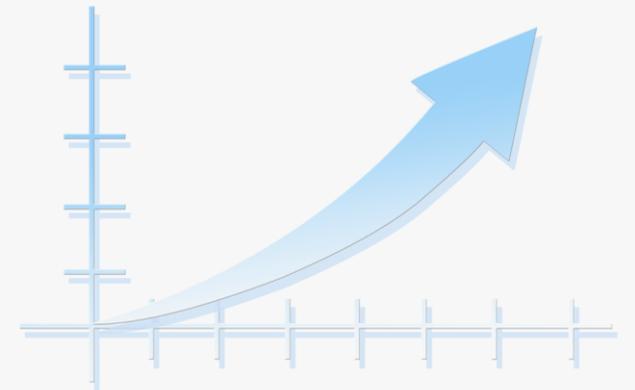
# Interopérabilité dans les soins de santé

- L'interopérabilité clinique peut nécessiter un **effort important de réorganisation** des flux de travail cliniques
- Plus nous comprenons les différents aspects de l'interopérabilité, moins nous risquons de sous-estimer le travail nécessaire pour rendre les systèmes de santé interopérables
- Les interopérabilités technique, sémantique, processuelle et clinique sont interdépendantes.
  - toutes sont nécessaires pour obtenir des avantages significatifs pour l'entreprise

En bref, l'interopérabilité vise à éviter un nombre considérable de doublons, de gaspillages et d'erreurs.

# Un besoin croissant d'interopérabilité

- Besoin d'accès facile au dossier du patient dans un contexte de multiples prestataires de soin
- Explosion de la volumétrie et de la diversité des dossiers médicaux
- Soutien renforcé à la gestion des maladies chroniques
- Réduction des erreurs médicales
- Limitation de l'augmentation des coûts de santé



# Obstacles à l'interopérabilité dans les soins de santé



**Complexité du domaine**



**Problèmes de normalisation**



**Des données massives**



**Systèmes existants**



**Manque de compétences et de ressources**



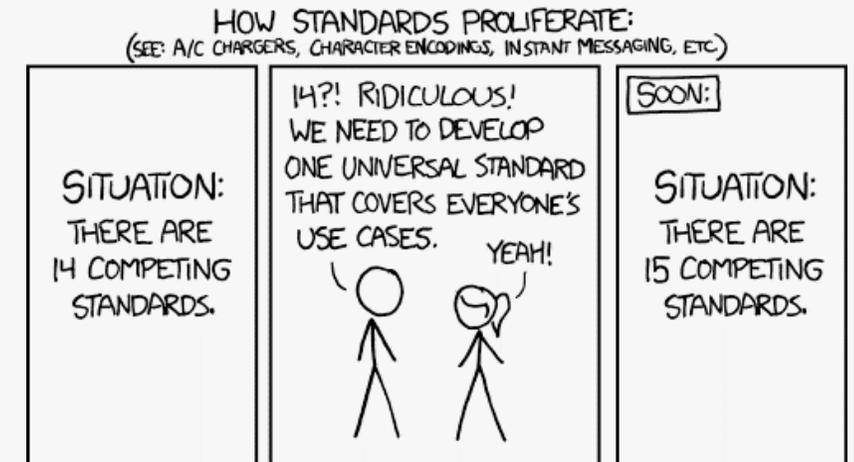
**Ecosystèmes et marchés cliniques**

## VIII. Solutions existantes

Normes et terminologies existantes

# Initiatives en matière d'interopérabilité dans les systèmes de santé

- Plus de 40 organismes de normalisation (SDO) différents
  - Définir des normes
  - Publier des guides et des profils de mise en œuvre
- Grande variété de types de données
  - Vocabulaire / Normes terminologiques
    - SNOMED-CT, LOINC, ICD
  - Normes de contenu
    - HL7 CDA, OHDSI OMOP CDM
  - Normes de transport
    - FHIR, DICOM
  - Normes de confidentialité et de sécurité
    - Règles HIPAA, "GDPR"
  - Normes d'identification
    - EMPI, MRN, ObjectID



xkcd – <https://xkcd.com/927/> – CC NY-CA 2.5

## Focus sur : HL7 FHIR

- Norme pour l'échange électronique d'informations sur les soins de santé
- Les ressources FHIR décrivent les formats et les éléments de données de santé échangeables.
- Standardisation des API
- Facilite les échanges interopérables avec les normes existantes



- La première version normative est **FHIR Release 4**, publiée en octobre 2019.
- Les Connectathons accueillent des ateliers de développement et de test, ainsi que des réunions de groupes de travail

## Focus sur : OHDSI OMOP CDM

- Norme pour le stockage des bases de données d'observation des dossiers médicaux électroniques
- Décrit les données dans un format commun (modèle) et une représentation commune (terminologies, vocabulaires, schémas de codage)
- Conçu pour des analyses systématiques basées sur un CDM
- Facilite les échanges interopérables selon les normes existantes



- La version actuelle du CDM est **OMOP CDM v5.4**, publiée en septembre 2021
- Le symposium annuel de l'OHDSI accueille des panels, des ateliers et des tutoriels

# Vocabulaires et terminologies

- Quelques vocabulaires utiles à connaître :

Nom	Distribution	Domaine	Type	Codes (environ)
CIM-10	Ouvert (OMS)	Conditions, observations	Classification	12 000
SNOMED-CT	Licence requise	Conditions, procédures, indications, médicaments, dispositifs, ...	Terminologie	311 000
LOINC	Gratuit	Observations, mesures, documents, ...	Terminologie	95 000
MeDRA	Licence requise	Conditions, procédures, indications, ...	Classification	107 000
RxNorm	Ouvert (US NLM)	Médicaments (US)	Terminologie	250 000
MeSH	Ouvert (US NLM)	Indexation des articles de journaux	Vocabulaire	41 000
CCAM	Ouvert (FR HAS)	Procédures	Classification	7 600
ATC	Ouvert (OMS)	Médicaments	Classification	6 300

# La série des normes ISO 8000

- Normes ISO relatives à la qualité des données
- Découlent et spécialisent la série des normes ISO 9000 relatives au management de la qualité



- Traitent tous les aspects de la qualité de données :
  - Vocabulaire de la qualité des données
  - Concepts et mesures
  - Profilage et évaluation des données
  - Processus de référence et maturité des processus

# Données FAIR

- Les humains s'appuient toujours plus sur l'informatique pour traiter les données
  - En raison de l'augmentation du volume, de la complexité et de la vitesse de création des données
- Les principes FAIR mettent l'accent sur la capacité d'action des machines
  - la capacité des systèmes informatiques à trouver, accéder, interopérer et réutiliser des données sans intervention humaine ou avec une intervention humaine minimale



## Findable

Les données et les métadonnées doivent être facilement repérables par les humains et les machines. Pour cela, il faut leur attribuer un identifiant unique, les décrire avec des métadonnées riches et les indexer correctement.



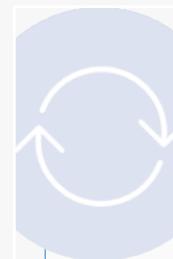
## Accessible

Les données ne doivent pas seulement être trouvables, elles doivent aussi être accessibles ; cela peut aller de l'accès immédiat aux détails sur la manière d'obtenir l'authentification ou l'autorisation si nécessaire.



## Interoperable

Les données doivent être interopérables et utiliser un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances. Si les données sont cloisonnées, certains éléments peuvent être inaccessibles ou longs à obtenir.



## Reusable

Garantir la réutilisation des données signifie que ceux qui ont besoin d'y accéder ultérieurement peuvent le faire facilement et efficacement, en ayant la certitude qu'elles sont exactes, quels que soient la durée et le format de leur stockage.

# Course URL and license

Permalink : <https://www.damien-leprovost.fr/enseignements/DataQualityHealthcare-FR.2023.pdf>

This course material is distributed under license  
**Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)**

**You are free to:**

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

**Under the following terms:**

 **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

 **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.



# Références bibliographiques

- F. Dravis, Data Quality Strategy: A Step-by-Step Approach, ICIQ, 2004
  - <http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202004/Papers/DQStrategy.pdf>
- M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016
  - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051581/pdf/egems1244.pdf>
- European Statistical System, Quality Assurance Framework of the European Statistical System, 2015
  - [http://ec.europa.eu/eurostat/documents/64157/4392716/qaf\\_2012-en.pdf](http://ec.europa.eu/eurostat/documents/64157/4392716/qaf_2012-en.pdf)
- UN Statistics Division. Guidelines for the template for a generic national quality assurance framework, 2012
  - <http://unstats.un.org/unsd/dnss/docs-nqaf/GUIDELINES%208%20Feb%202012.pdf>
- Canadian Institute for Health Information, CIHI's Information Quality Plan, 2017
  - <https://www.cihi.ca/sites/default/files/document/corporate-information-quality-plan-2016-en.pdf>
- Health Information and Quality Authority (Ireland), Guidance on a data quality framework, Oct. 2018
  - <https://www.hiqa.ie/reports-and-publications/health-information/guidance-data-quality-framework-health-and-social-care>
- Healthcare Information and Management Systems Society, Definition of Interoperability, 2013
  - <https://www.himss.org/sites/hde/files/d7/FileDownloads/HIMSS%20Interoperability%20Definition%20FINAL.pdf>