

Conférence-Débat: Qualité de données Le rôle des standards

Montréal, 4 Avril 2023



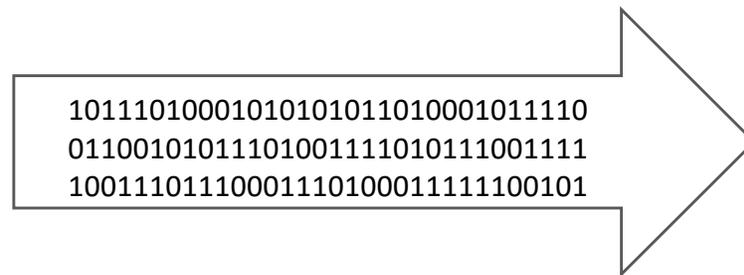
Stefan Schulz

stefan.schulz@medunigraz.at

- École de Médecine de Graz, Autriche
- Averbis GmbH, Freiburg, Allemagne

Définitions de qualité de données

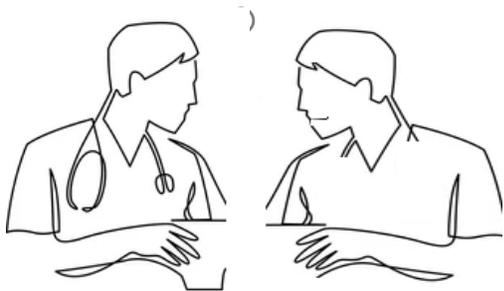
- Qualité: Ensemble des propriétés qui font que quelque chose correspond bien ou mal à sa nature, à ce qu'on en attend (Larousse)
- Qualité de données (Wang & Strong) :
 - Aptitude à l'emploi («Fitness for use»)
 - Dans un contexte déterminé, dans le cadre de la tâche à accomplir
 - Complétude, Exactitude, Rapidité, Cohérence, Accessibilité
- Qualité de la communication : qualité de la transmission et de la réception du message, en tenant compte des besoins et des attentes des parties concernées.



Exemple 2

contexte

DOB = date de naissance
 0 = masculin, 1 = féminin
 W_Ad: Poids à l'admission
 W_Tg: Poids ciblé
 W_Op: Poids à l'opération
 W_1y Poids 1 an post op



Registre Chirurgie Bariatrique

	A	B	C	D	E	F	G	H
1	Pat	ID	DOB	Sex	W_Ad	W_Tg	W_Op	W_1y
2	A.B.	342237775	21.03.1997	0	143	90	132	91
3	F.D.	294387012	04.07.1996	0	113	85	118	99
4	G.L.	917090972	03.07.2003	0	142	90	133	86
5	P.P.	644302398	01.04.1993	1	154	95	167	122
6	M.L.	877544309	26.11.1998	1	181	100	166	112

Exemple 3

contexte

Directive de codage :
« Si le traitement a été
initié et que les résultats
des tests ne sont pas
concluants, le diagnostic
suspecté doit être codé »



Diagnostique	CIM-10	Texte
Soignée pour méningite, sans confirmation de méningite	G00.9	Méningite bactérienne, sans précision



Exemple 4

Valeur DRG:
5329,20 € pour S73.3

indépendamment de
l'ajout ou non des codes
E11.90, I10, et E66.99

contexte



Diagnostiques	CIM-10	Texte
<ul style="list-style-type: none"> • Fracture comminutive de la diaphyse fémorale droite • Diabète de type 2 • Obésité • Hypertension artérielle 	S73.3	Fracture fermée de la diaphyse du fémur

Exemple 5

contexte



ECOG:

< 50% am Tag bettl., amb.Betreuung,
leichte Arbeiten möglich
invasiv ductales Mammacarcinom,
G 2 pT1c pNO(sn 0/2) DM 1,2 cm
Hochgradige duct. intraepith. NPL li,
G3 pTis(DCIS), DM 1 cm
Resektion dorsal <1mm
ER 70 % pos, PR 100 % pos
Nachresektat im gesunden
ER 95% pos, PR 20% pos, HER-2/neu
2+(FISH ausständig)
MIB1: 50% positiv

contexte



Scénarios d'utilisation primaire (1):

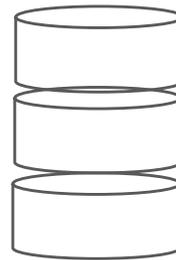
Les données sont utilisées dans le contexte pour lequel elles ont été acquises

- Situation idéale – contextes identiques (prestation de soins, facturation, recherche):
 - Production et interprétation de données, dans le même cadre de la tâche à accomplir
 - Les mêmes dictionnaires de données
 - Les mêmes jargons techniques
 - Les mêmes directives



Contexte clinique 1

```
101110100010101010111010001011110  
01100101011101001111010111001111  
10011101110001110100011111100101
```



```
1011101000101011101010001011110  
01101101011101001001010111010111  
10011101110110010100011000100101
```



Contexte clinique 2

Scénarios d'utilisation primaire (2):

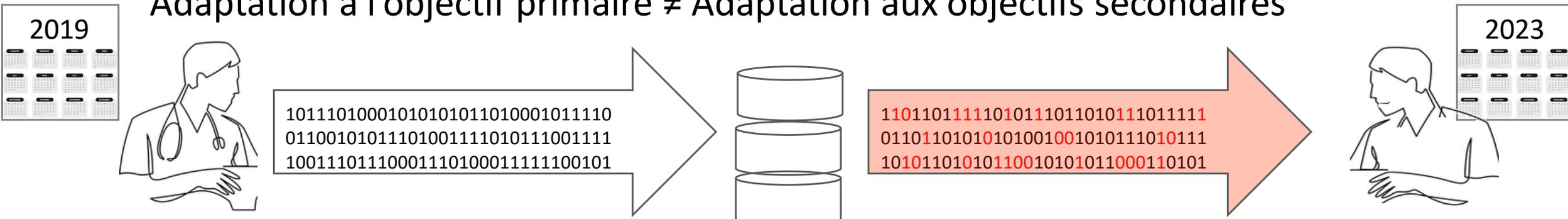
Les données sont utilisées dans le contexte pour lequel elles ont été acquises

- Efficacité de la communication
 - « Style télégraphique »: réduisant le nombre de mots / caractères au minimum
 - Abondance d'abréviations et d'acronymes
- En réalité: contextes pas identiques
 - Dialectes, régiolectes, sociolectes , idiolectes
 - Persistance de collections de données compromise (changement de langage, changement de directives)

Scénarios d'utilisation secondaire:

Les données sont utilisées dans un autre contexte pour d'autres objectifs

- Les données cliniques: trésor d'informations ():
 - recherche médicale
 - surveillance de la santé
 - gestion de la qualité des soins
 - Prise de décisions
- Problèmes
 - Confidentialité et protection des données
 - Qualité de données: incomplètes, inexactes, non standardisées:
Adaptation à l'objectif primaire \neq Adaptation aux objectifs secondaires

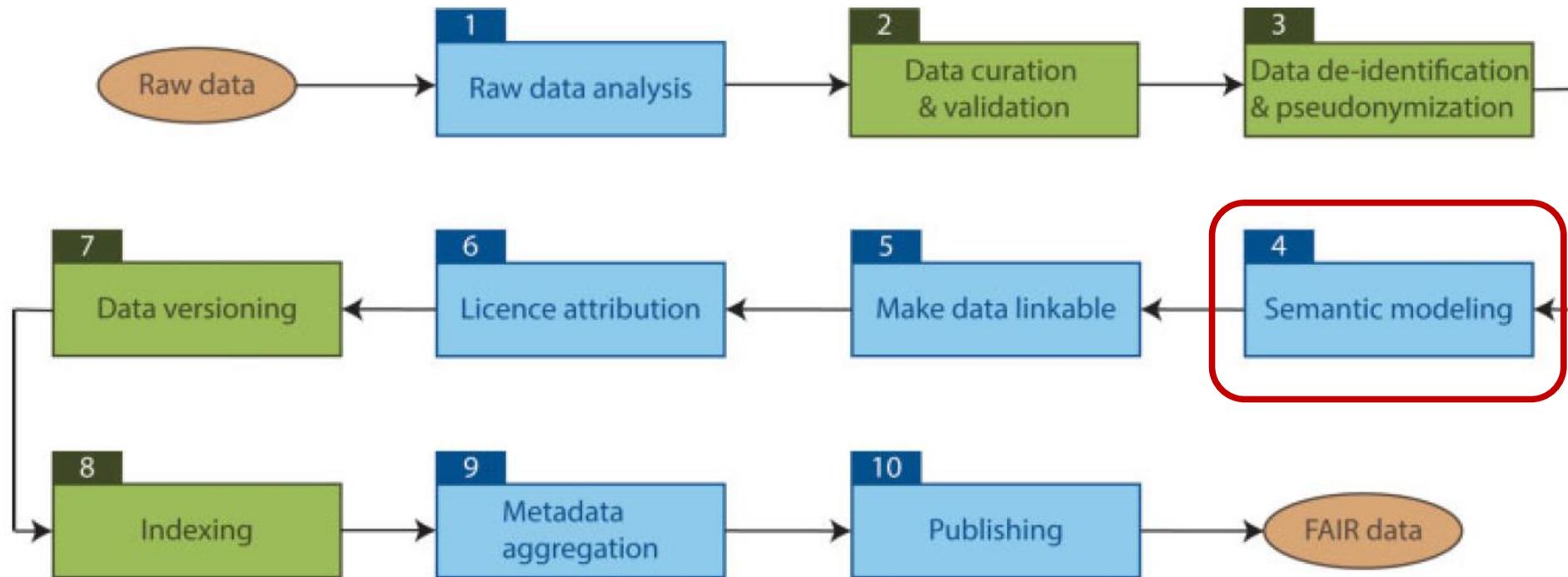


Les critères FAIR*

- Principes assurant la gestion et la réutilisation des données de recherche
 - **F** (Findable) - Trouvables : Métadonnées soutiennent leur lieu de stockage
 - **A** (Accessible) - Accessibles : protocoles de sécurité assurent l'accès, l'intégrité et la confidentialité
 - **I** (Interoperable) - Interopérables : Utilisation de terminologies et langage de représentation
 - **R** (Reusable) - Réutilisables : Les données doivent être accompagnées d'informations suffisantes pour permettre leur réutilisation et leur compréhension par d'autres chercheurs.
- Peuvent être appliquées à des données cliniques

*Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.

« FAIRification » de données cliniques



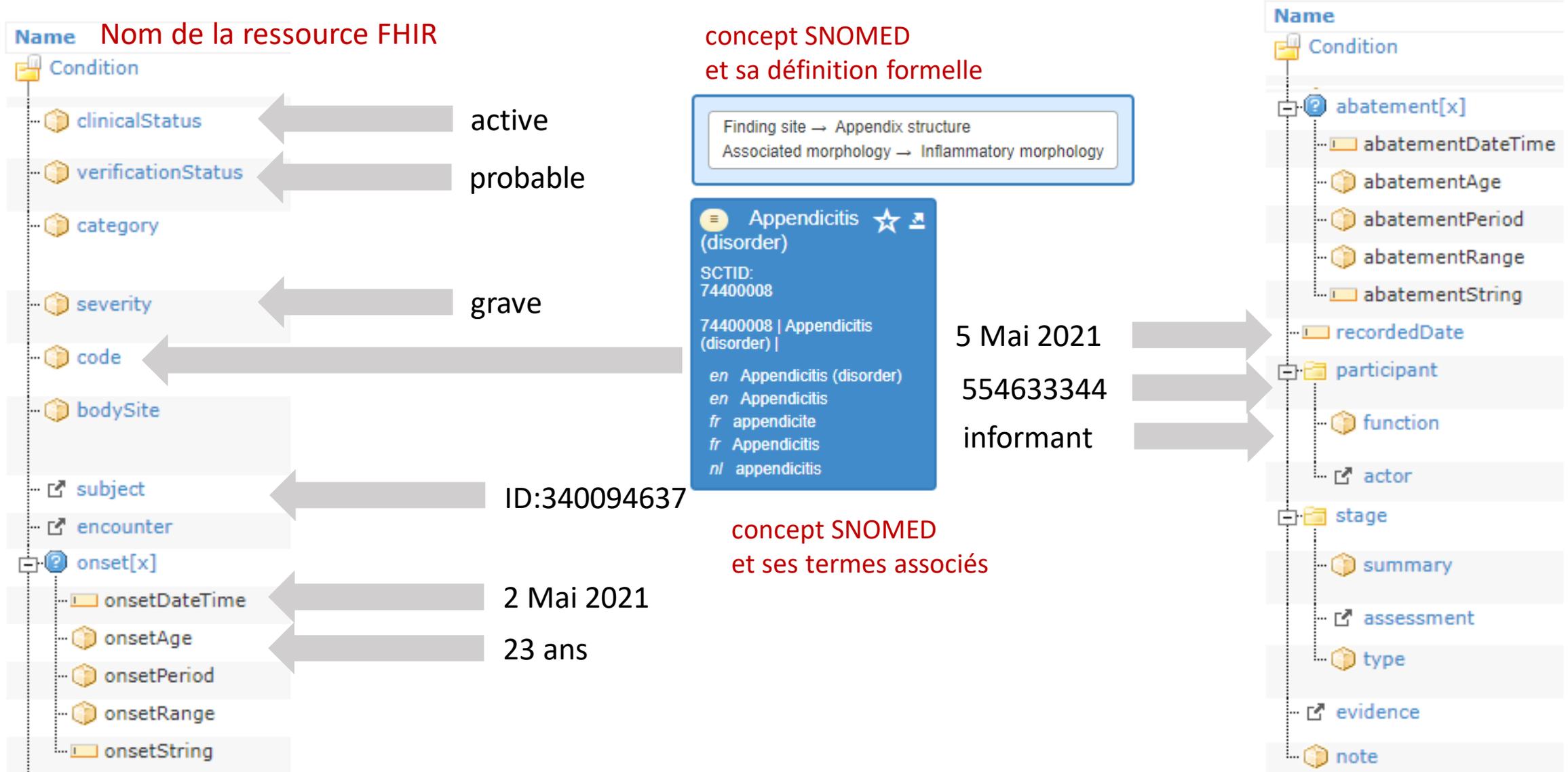
Comment définir un « modèle sémantique »

- Description de la signification des entités et des relations avec précision et sans ambiguïté
- Un bon modèle sémantique devrait représenter une vue consensuelle de la domaine
- Utilisation des ressources existants issus de processus de normalisation
- Deux structures sémantiques complémentaires
 - Ontologies (modèles de signification): inventaires des classes d'entités d'intérêt pour la domaine avec ses définitions formelles et informelles. Leur codes sont attribués aux données de patients. Exemple: SNOMED CT
 - Modèles d'information (modèles d'utilisation) : décrivent la manière dont les données sont utilisées et échangées. Exemple: HL7 FHIR

SNOMED CT – LOINC – FHIR

- SNOMED CT – Ontologie de la santé
 - Énorme système de codage avec 350 000 concepts
 - Riche en axiomes logiques (OWL-EL) que facilitent le raisonnement automatique
 - Standard international
 - Adopté dans plus de 40 pays
- LOINC – Catalogue de tests et mesures, chacune compose par les axes
 - Composant, Propriété, Chronologie, Système de mesure, Echelle, Méthode
- HL7 FHIR (« Fast Healthcare Interoperability Resources »)
 - Standard international pour l'échange de données électroniques de santé
 - Offre des modèles « ressources » pour structurer plusieurs scénarios d'acquisition de données. Exemples:
 - « condition » - Maladies, symptômes, certitude diagnostique, état clinique, référence temporelle, provenance des données
 - « observation » - tests de laboratoire, mesures physiques, avec valeurs quantitatifs ou qualitatifs

Exemple: SNOMED CT code intégré dans une instance de ressource FHIR



Améliorer la qualité de données – quoi faire?

- En théorie: saisir les données structurées, en utilisant les standards, dérivées des codifications spéciales, comme CIM pour la facturation.
 - Pas très réaliste: goulot d'étranglement de la saisie de données – les professionnels de la santé préfèrent le texte libre
 - Réalisable uniquement dans des cas particuliers où il y a le temps et la motivation
 - Production parallèle de récits difficilement évitable
- Perspective:
 - hier: l'humain s'adaptait à l'ordinateur
 - demain: l'ordinateur s'adaptera à l'humain
- Technologies:
 - Traitement automatique de langues (TAL)
 - IA conversationnelle
 - Ressources: vocabulaires, corpus annotés



Diagnosen

Code	Text	Seite	Status
Keine Diagnosen gefunden			

Labor- & Vitalparameter

Name	Wert	Einheit
Keine Labor- & Vitalparameter gefunden		

Medikamente

Name	Stärke	Form	morgens	mittags	abends	zur Nacht
Keine Medikamente gefunden						



Speech to Structure



Eingabe

Annotation

Questions?



Stefan Schulz

stefan.schulz@medunigraz.at

- École de Médecine de Graz, Autriche
- Averbis GmbH, Freiburg, Allemagne

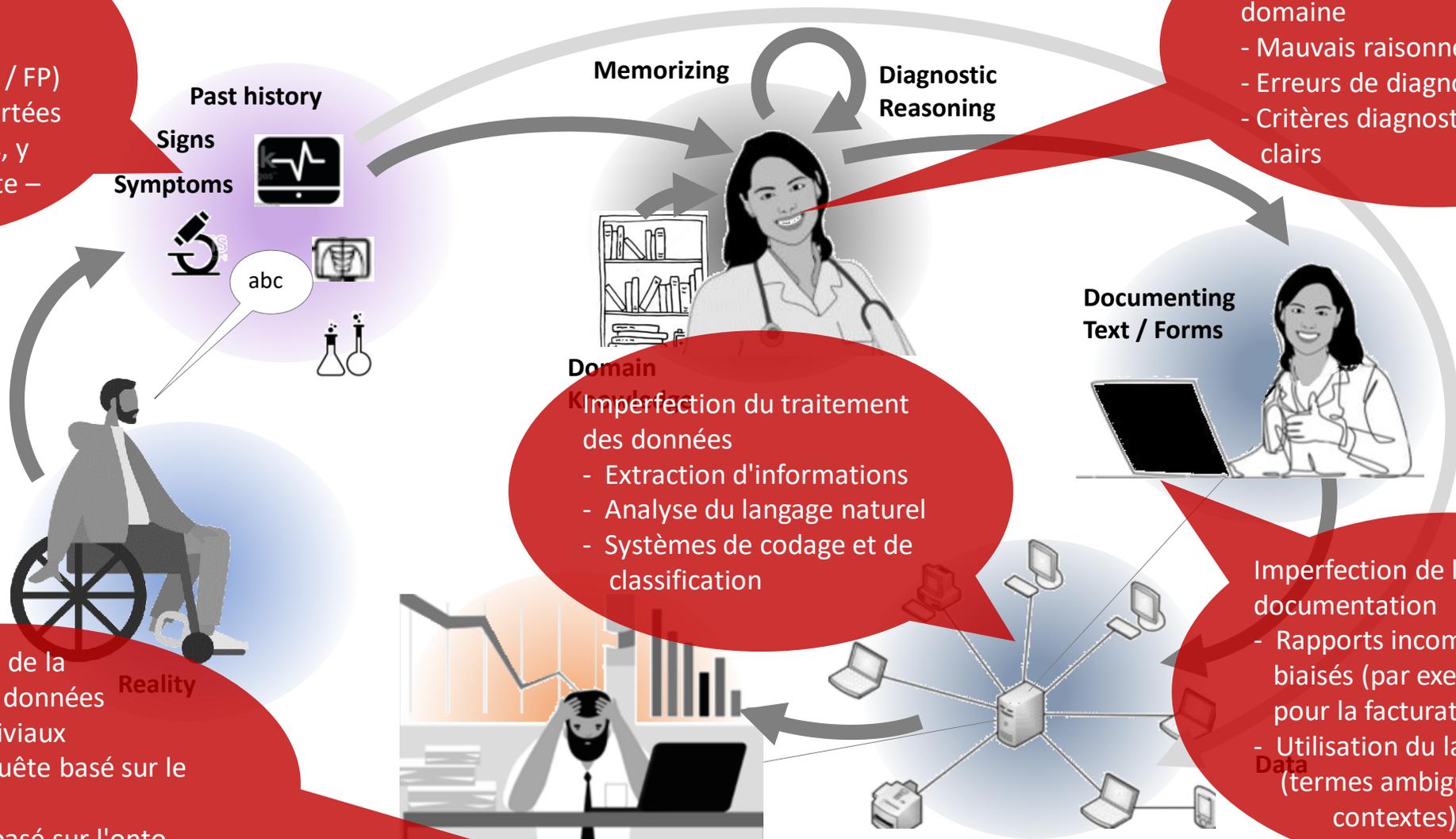
La réalité des données cliniques

Inexactitude et incomplétude

- Résultats de laboratoire (FN / FP)
- Données rapportées par les patients, y compris du texte – Histoire

Imperfection du diagnostic

- Manque de connaissance du domaine
- Mauvais raisonnement
- Erreurs de diagnostic
- Critères diagnostiques peu clairs



Imperfection du traitement des données

- Extraction d'informations
- Analyse du langage naturel
- Systèmes de codage et de classification

Imperfection de la récupération des données

- Des outils conviviaux
- Langage de requête basé sur le langage naturel
- Raisonnement basé sur l'ontologie (par exemple via des taxonomies)

Imperfection de la documentation

- Rapports incomplets, erronés, biaisés (par exemple, données pour la facturation)
- Utilisation du langage naturel (termes ambigus, flous, contextes)