Semantic harmonization of the French National healthcare database (SNDS)

Lorien Benda[1], Cécile Roseau[1], Gaëlle Rimaud[1], Régis Lassalle[2], Cécile Droz[2], Nicolas Thurin[2], Stéphanie Combes[1]

[1]Plateforme des Données de Santé (Health Data Hub), 75015, Paris, France, opensource@health-data-hub.fr
[2]Bordeaux PharmacoEpi, INSERM CIC-P 1401, Université de Bordeaux, 33000, Bordeaux, France, nicolas.thurin@u-Bordeaux.fr

Introduction

- France hosts one of world's largest claims database: the **SNDS**[1] (*National Health Data System*).
- SNDS relies on numerous specific French vocabularies : **CCAM** and **CSARR** (procedures), **NABM** (laboratory tests), **LPP** (medical devices), **CIP** and **UCD** (drugs).
- Vocabulary **standardization** is needed to **improve SNDS and script reuse**, through **common data models** (CDM).

Methods

French ontology	Level of mapping
CCAM/CSARR	80 % of the most occurrent source concepts (2019-2020, inpatient and outpatient) : mapping at the code level Others : mapping at the chapter level
CIP / UCD / NABM	Mapping at the code level
LPP	Mapping at the chapter level

1. **Translation** of source concepts by an automatic translator (Deepl).
2. **Proofreading** and **correction** of the English translation by the medical residents.
3. Mapping to the standard OMOP concepts with **USAGI** by the residents.
4. **Cross-review** of the mapping by a different resident.

French ontology	Meaning	Main target domains	Number of mapped source concepts (more 80% frequent codes / total codes)	
CCAM	Medical acts	Procedure Observation Spec Anatomic Site	686 / 8179 codes 1 387 chapters	SNOMED -CT
CSARR	Physical and speech therapy	Procedure	98 / 566 codes 94 chapters	
NABM	Biological acts	Measurement Procedure	973 codes	
LPP	Products, devices and services	On going	0 / 29 161 codes 399 chapters	
CIP / UCD	Drugs (box level / unit level)	On going	46 878 codes	RxNorm

Table 1. Summary of French vocabularies mapping

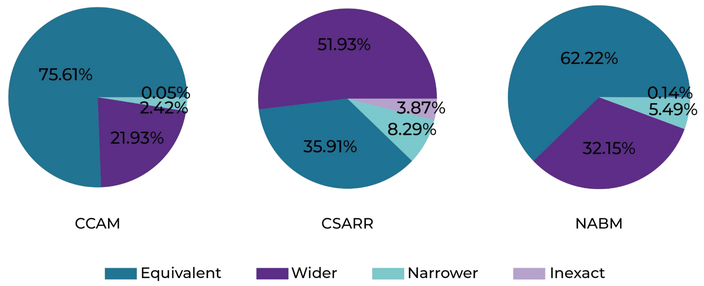

Figure 2. Level of equivalence for CCAM, CSARR and NABM codes (mapped on code level)

Results

- French ontologies map to **5 major domains** in the OMOP-CDM standard domains (Table 1, Figure 1).
- For CCAM codes, **90%** of the target codes are **procedures**.

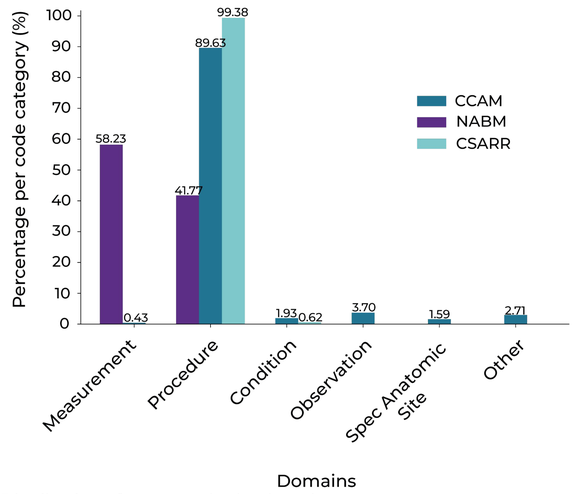

Figure 1. Distribution of target codes by domain

- Residents filled the **level of equivalence** in USAGI (Equivalent, Wider, Narrower or Inexact, Figure 2).
- Regarding CCAM codes, **22%** of the targets are **wider than the source code**, showing this ontology is particularly detailed.
- The **most frequent CCAM codes** are mapped to a **median of 3 codes,** while chapters with less detail are mapped to 1 code in median (Figure 3).

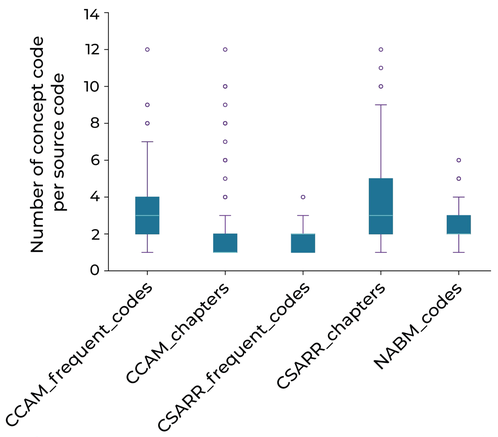

Figure 3. Number of target concepts per source code / chapter

Conclusion

- **Mapping of SNDS** ontologies belonging to **procedures**, **observation** and **measurement** to OMOP standard concepts **have been completed**.
- **Mapping** of ontologies belonging to **Drugs** (box level) and **Devices** domains are currently **ongoing**.
- Combined to **syntactic mapping**, this **semantic mapping** opens the door to the execution of **federated analysis using the OMOP CDM in the SNDS.**

References

[1] Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, Moore N. The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 2017 Aug;26(8):954-962. doi: 10.1002/pds.4233. Epub 2017 May 24. PMID: 28544284

Many thanks to the medical residents :

Alexandre Kitic, Nicolas Kitic, Raphaël Lee, Sara Tunon de Lara, François Bourquard