Analyzing microbiology *In Vitro* Diagnostics data using the OMOP CDM and SNOMED CT encoding

M. Le Gall, J-F. Gorse, X. Gansel | bioMérieux

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Background

This work was undertaken as part of the European VALUE-Dx project [1], aiming to combat antimicrobial resistance (AMR) and improve patient outcomes.

OHDSI and its OMOP Common Data Model [2], is well-known to support decentralized data aggregation and analytics, while preserving data privacy (see fig. 1).

Analyzing specifically laboratory microbiological data implies capturing data at a lower level than hospital or regional EHR to gain a more detailed level of information

- laboratory information system
- middleware or even *in vitro* diagnostics (IVD) devices

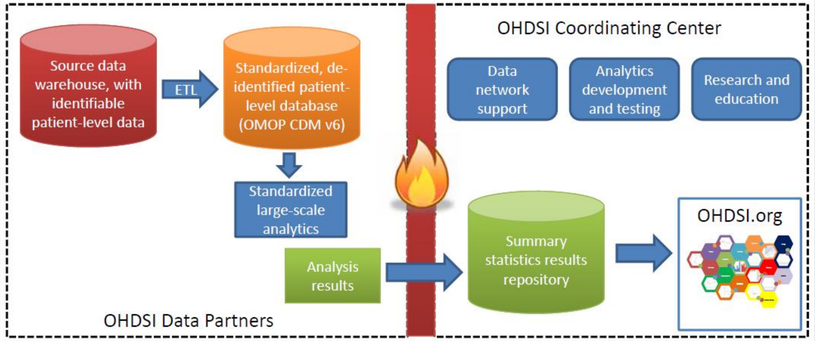

Figure 1- How OHDSI works (from G. Hripcsak - MedInfo conference, 2019)

OMOP CDM makes extensive use of standardized vocabularies. Notably LOINC® and SNOMED CT® are used to encode the OMOP «Standardized Clinical Data Tables» (see fig. 2) SPECIMEN, MEASUREMENT and OBSERVATION.

We reused our previous work [3] showing that *in vitro* diagnostics (IVD) systems tests and test results are very well described using LOINC® and SNOMED CT®, up to supporting SNOMED CT® mediated analytic. Indeed, we were able to map 91% (1,349/1,482) of our taxa (VITEK® 2 and VITEK® MS IVD systems), 65% (13/20) of our ordinal test results, 89% (320/361) of our drugs and between 64% (7/11) and 98% (39/40) of our specimen breakdown to SNOMED CT®.

Topic

In the context of VALUE-Dx, our use cases supporting usage of OMOP are

- Identify candidate laboratories for clinical studies based on antimicrobial resistance tests and test results
- Support microbiology laboratory results demography observations per laboratory or region
  - Describe tests implemented
  - Describe test results and observations (including high level results interpretation such as Multi-Drug Resistant phenotype)

To our knowledge, no observational study addresses microbiology laboratory data.

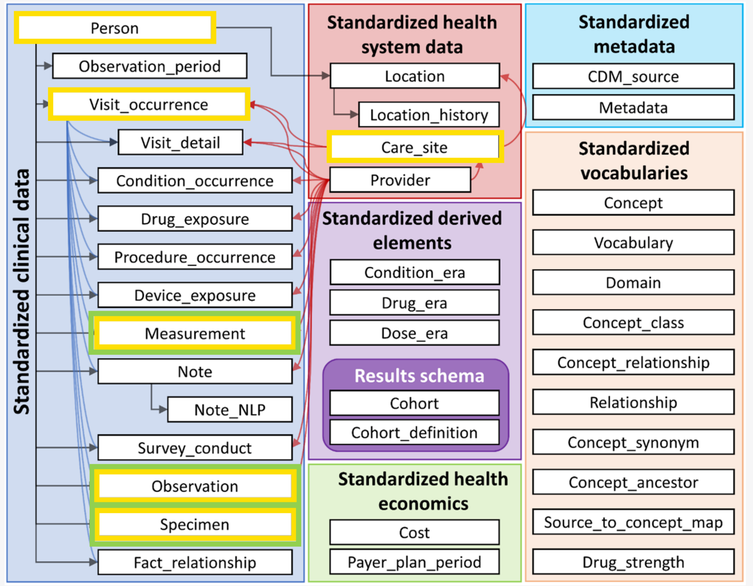

Figure 2- OMOP CDM v6 tables in Yellow are candidate to host laboratory data. In Green table targeted to store LOINC® and SNOMED® CT encoded element

Objectives

Our work aims at

(i) Demonstrate the capabilities and limitations of OMOP CDM to represents LOINC® and SNOMED CT® encoded microbiology IVD laboratory data

(ii) Envisage options to solve those limitations aiming at preparing future analysis

Analyzing microbiology *In Vitro* Diagnostics data using the OMOP CDM and SNOMED CT encoding

M. Le Gall, J-F. Gorse, X. Gansel | bioMérieux

METHODS

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Problem statement

Aligning data logical view from a laboratory perspective with the OMOP CDM v6 is not straightforward as shown in fig. 3.

The concepts of "isolate", hierarchy of tests and of test results are absent from OMOP CDM v6. Note that Some active discussions exist on the OMOP forum

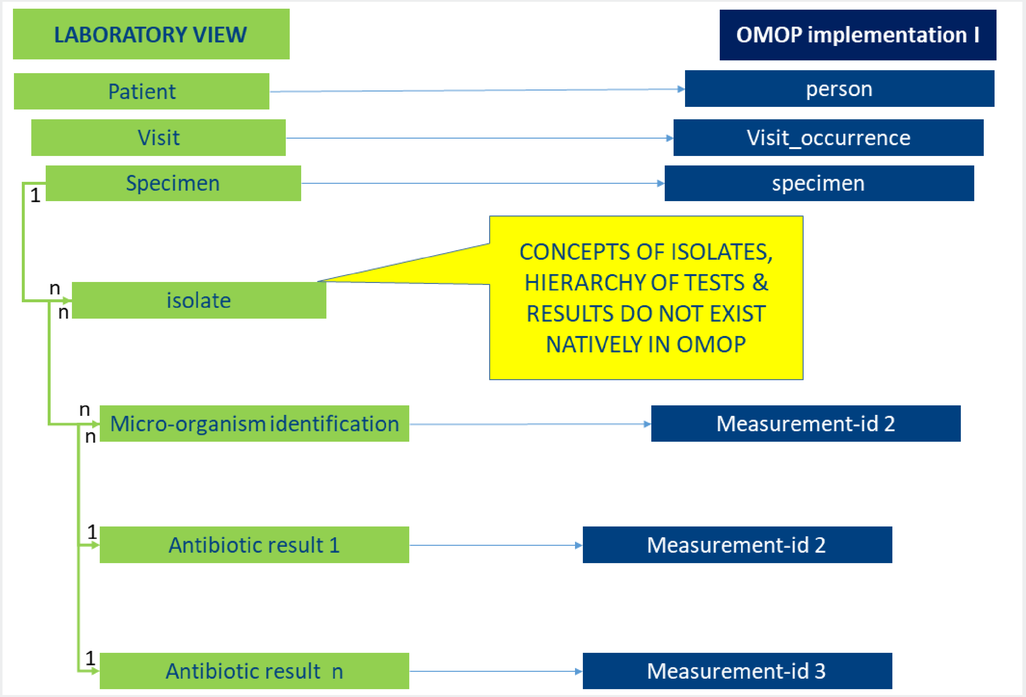

Figure 3-

This figure shows a logical representation of laboratory microbiology data (in green) compared to the OMOP CDM (in dark blue).

Model analysis & data mapping

Data model analysis was performed using all OMOP CDM v6 available documentation and tools from the OHDSI sites & forum. It also reuses laboratory workflow analysis as in [4]. Terminology mapping uses our previous work [3]. Mapping strategy is described in fig. 4.

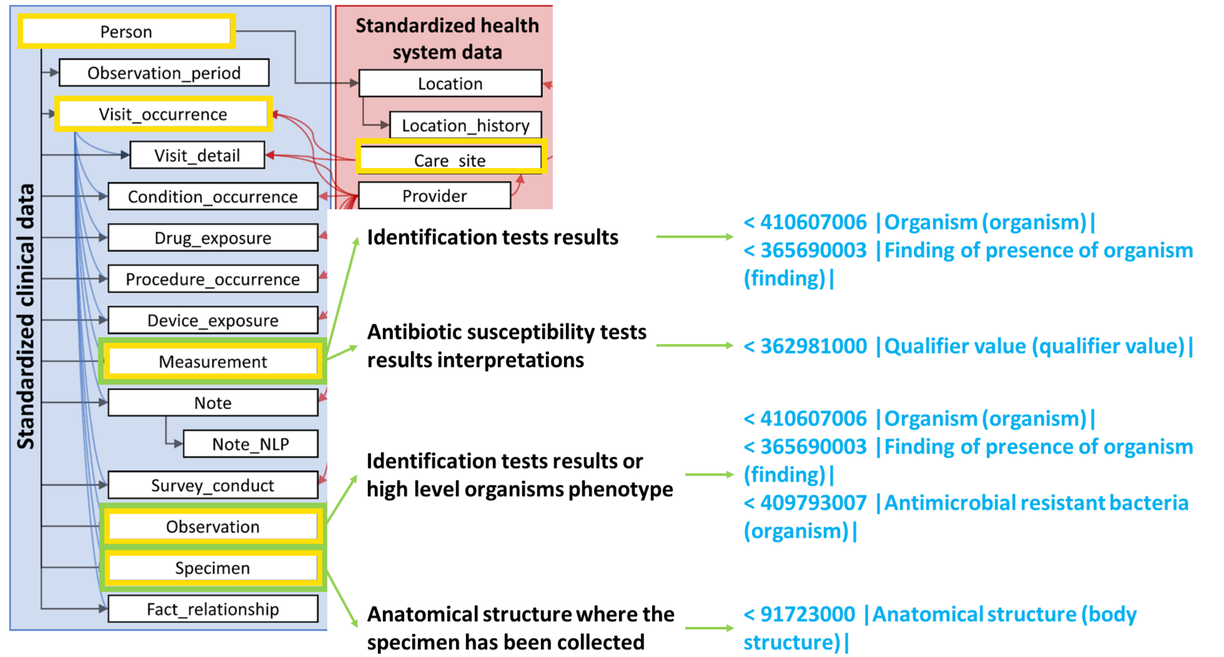

Figure 4- This figure presents the SNOM²ED CT® concepts mapped to data present in OMOP CDM table PERSON, VISIT_OCCURRENCE, CARE_SITE, SPECIMEN, MEASUREMENT, OBSERVATION

Analyzing microbiology *In Vitro* Diagnostics data using the OMOP CDM and SNOMED CT encoding

M. Le Gall, J-F. Gorse, X. Gansel | bioMérieux

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Results

We designed two options to represent laboratory microbiology data in OMOP v6. Both involve a root MEASUREMENT with the intention to (i) mimic the isolates ; (ii) anchor the specimen and all subsequent tests. Links are implemented through FACT_RELATIONSHIP.

The first one (fig. 5) is only base on MEASUREMENTs and stores susceptibility results in one single record. Limitations are that organism as SNOMED CT concept is not a permitted value to MEASUREMENT in CDM v6. The model only permits LOINC answers, that are not a sustainable option to describe identification results. Storing quantitative value (i.e. MIC) and qualitative interpretation (i.e. S/I/R) for drug susceptibility results in a single MEASUREMENT is not clearly allowed / prohibited in CDM v6.

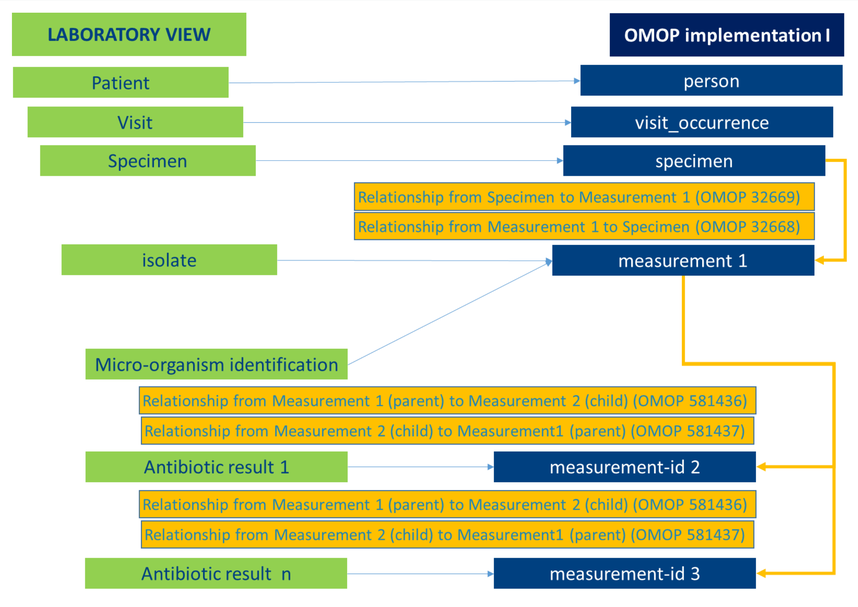

Figure 5- Model is based only on MEASUREMENTs. The root MEASUREMENT may be any lab test. In case of identification test, organism as SNOMED CT concept is not a permitted value to MEASUREMENT . Child MEASUREMENTs of identification are drug susceptibility tests.

The second model (fig. 6) is based on both MEASUREMENTs and OBSERVATIONs. It allows to store the chain of tests along a lab process and susceptibility results as two separate MEASUREMENTs, one for the MIC and second one for the corresponding category (using dedicated LOINC codes). The root MEASUREMENT host root identification test and corresponding identified organism is hosted in the anchored OBSERVATION.

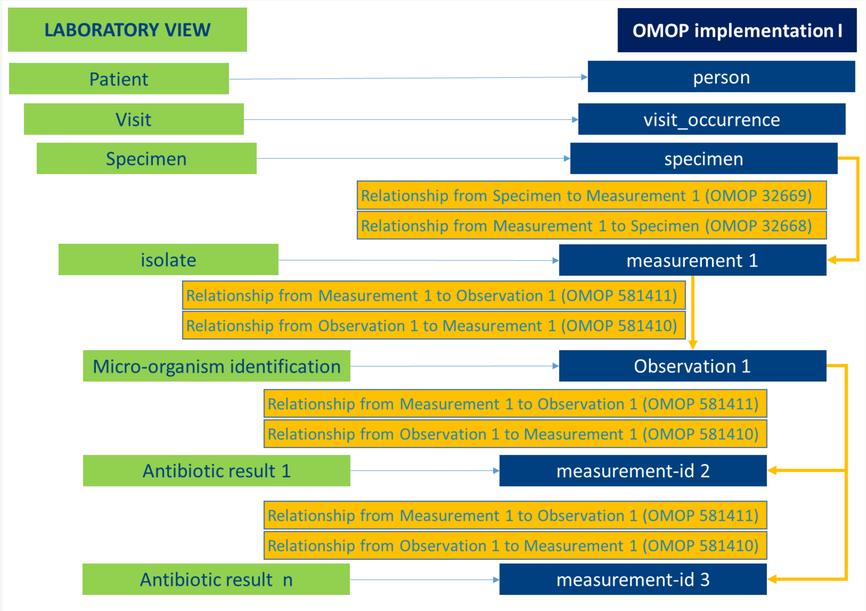

Figure 6- Model is based on both MEASUREMENTs and OBSERVATIONs. One MEASUREMENT as a root anchoring both (i) OBSERVATION carrying the organism concept and (ii) MEASUREMENTs for drug susceptibility tests.

Drug susceptibility tests are stored as 2 individual records for the quantitative MIC and its interpretation (S/I/R)

The second model was implemented in a data end-point, populated with data extracted from a middleware and an IVD device. Ongoing analysis show that OHDSI Athena tool does not support FACT_RELATIONSHIP.

Analyzing microbiology *In Vitro* Diagnostics data using the OMOP CDM and SNOMED CT encoding

M. Le Gall, J-F. Gorse, X. Gansel | bioMérieux

DISCUSSION

SNOMED International
SNOMED CT EXPO 2020
Virtual Conference October 8-9

Discussion

Representing laboratory data into OMOP is challenging

- MEASUREMENT do not allow using SNOMED CT concepts to represent microbial or viral identification results. The model allows using LOINC answers that is not a sustainable solution for identification results.
- Combination of measurements and observations allow to represent lab tests / results at the cost of clarity and heavy usage of FACT_RELATIONSHIP.
- A shared usage of FACT_RELATIONSHIP across all OMOP end-points is a blocking issue. If this is achievable in a given project it is very challenging across independent projects thus causing interoperability issues.

Under the OMOP CDM v6, a merge between the two models may give good results provided the implementation is shared across all end-points. Ideally an evolution of OMOP CDM is needed to accurately represent laboratory data and prevent usage of FACT_RELATIONSHIP.

Future Directions

In the close future, we will deepen our ongoing analysis of model implemented (fig. 6), pursuing definition of a better representation of lab. data under OMOP CDM v6 & upper, and implement a live Proof of Concept.

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 820755. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and bioMérieux SA, Janssen Pharmaceutica NV, Accelerate Diagnostics S.L., Abbott, Bio-Rad Laboratories, BD Switzerland Sàrl, and The Wellcome Trust Limited.

www.imi.europa.eu // www.value-dx.eu

Conclusions

OMOP CDM supports the representation of lab. microbiology data. with limitations. Loaded into a series of nodes within an OHDSI-ARACHNE federated architecture, it also proved to be usable with some limits that we are investigating.

Usage of FACT_RELATIONSHIP, may jeopardize implementations across projects and the wanted interoperability.

In line with some active discussions on the OHDSI forum, laboratory data need

- Clarity on interpretation of the so call "convention" notably using SNOMED CT concepts to describe identification results.
- Evolution(s) of OMOP CDM to better represent lab data, get rid as much as possible of FACT_RELATIONSHIP.

Finally we foresee the lack of concept model in the SNOMED CT® Organism hierarchy as a future limitation if we are to use of the ontological nature of SNOMED® CT to support data analytics.

References

1. https://value-dx.eu/
2. Hripcsak, G. *et al.* (2015). "Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers." Stud Health Technol Inform. 2015; 216: 574–578
3. Le Gall, M. et al (2019). "SNOMED CT Coding and Analytics of in vitro Diagnostics Observations." Stud Health Technol Inform 264: 1460-1461
4. Fournier, P.-E. et al. (2013). "Modern clinical microbiology: new challenges and solutions." Nature Reviews Microbiology 11: 574-585