Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220461

Usability of OMOP Common Data Model for Detailed Lab Microbiology Results

Emmanuelle THERON[a], Jean-François GORSE[a] and Xavier GANSEL[a,1]
[a] *bioMérieux, Data Science, Grenoble, France*

Abstract. Anti-microbial resistance surveillance systems in Europe are limited by the inability to link laboratory data and patient data. The OMOP Common Data Model (OMOP CDM) is an option to store and use patient data in an international context supporting observational research. Detailed medical microbiology laboratory data are usually not stored in OMOP CDM. We propose here a solution to deal with the inherent complexity of microbiology data and store those in the OMOP CDM v5.4. We demonstrate the feasibility of our approach by capturing data from a microbiology in vitro diagnostic middleware, modeling in OMOP CDM 5.4 and querying for visualization.

Keywords. Interoperability, OMOP CDM, data model, microbiology

1. Introduction

VALUE-Dx [1] is a European « Innovative Medicines Initiative » project, that aims to combat antimicrobial resistance (AMR) and improve patient outcome[2]. One of the objectives is to "assess and establish a proof-of-concept data interoperability network to allow connections between laboratory information systems and VALUE-Dx partners". The proliferation of AMR, with the loss of action of antibiotics, has already caused millions of deaths in 2019 [2]. It is estimated that, left unchecked, AMR will cause ~10 million deaths yearly by 2050 [3, 4].

The limits of existing AMR surveillance systems in Europe were highlighted by Tacconelli et al. [5]. This work shows that, among others, purely lab-based systems experience a lack of patient clinical information. Indeed, data management systems and networks, looking to mobilize AMR data, are challenged by the complexity of the data relationships, the practicalities of handling patient identifiable health information, the diversity of data sources, data sovereignty and security issues.

The capabilities of Observational Health Data Science and Informatics (OHDSI) federated network to analyze clinical data from a variety of sources, harmonized in the patient centric interoperable Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), are well documented [6]. The need for interoperable laboratory data to support digital medicine and real word data analytics is also documented [7], as well as the mapping of microbiology data across openEHR, Fast

[1] Corresponding Author, Xavier Gansel, bioMerieux, Centre C. Mérieux, 5 rue des Berges, 38000 Grenoble, France; E-mail : xavier.gansel@biomerieux.com

[2] This work has received support from the EU/EFPIA/bioMérieux SA, Janssen Pharmaceutica NV, Accelerate Diagnostics S.L., Abbott, Bio-Rad Laboratories, BD Switzerland Sàrl, and The Wellcome Trust Limited. Innovative Medicines Initiative 2 Joint Undertaking VALUE-Dx grant n° 820755

Healthcare Interoperability Resources (FHIR) and OMOP CDM data standards [8]. Nevertheless, to the best of our knowledge, detailed microbiology laboratory data are not stored in the OMOP CDM for analysis.

To answer some of the unmet needs described by Tacconelli et al. [5], we investigate how detailed clinical microbiology laboratory data from *in vitro* diagnostic systems, can be modeled, incorporated in the OMOP CDM 5.4 [9] and visualized for data analytics needs. This is the first step towards building an AMR data management system able to associate laboratory and patient data.

2. Methods

The OMOP CDM is a relational data model where specific tables are dedicated to the storage of clinical data (e.g. measurement or observation), to the description of care structures (e.g. care_site), or to manage OMOP vocabularies [10]. The main differences between OMOP v5.3 and v6.0 versus v5.4 is the way they handle explicit links between measurements or observations entries and other clinical tables. Where OMOP v5.3 or v6.0 require explicit bidirectional "fact-relationship" association table, the newly published OMOP CDM v5.4 [9] brings, among other features, foreign key usage through modifier_of_event field in measurement table / observation_event_id in the observation table. The original aim is to better support observational cancer research that requires tumors to be characterized by a large set of attributes such as grade, site of origin, biomarkers, etc. [11].

The data used in this paper come from an anonymized dataset containing 996 distinct patients and 4000 microbiology results obtained from bioMérieux® VITEK-MS® and VITEK-2® instruments through the microbiology middleware MYLA®. Alignment between data fields has been carried out using the OHDSI tools WhiteRabbit and RabbitInAHat. We re-used test description mapped to LOINC® and test results (organisms name) mapped to SNOMED CT®[12]. The association between LOINC®, SNOMED CT® codes and OMOP code were realized using ATHENA vocabularies defined by OHDSI. A local OMOP instance was created in a PostgreSQL database and subsequently queried with R, and OHDSI packages SqlRender and DatabaseConnector. Completeness, plausibility and conformance of data loaded in OMOP v5.3 have been assessed using the DataQualityDashboard (DQD) open source tool from OHDSI. Currently, DQD doesn't support OMOP v5.4.

3. Results

3.1. Specific challenges of AMR representation in CDM

Clinical microbiology data is intrinsically complex [13] making modelling in OMOP CDM challenging. One issue is the numerous one to many relationships, indeed one to several pathogens may be identified per specimen and each pathogen may be tested for dozens of antibiotics. Moreover, explicit links between specimen, pathogen identification and AST results need to be retained.

Our analysis revealed that OMOP CDM v5.3 & v6.0 are poorly equipped to model detailed clinical microbiology data. Linking measurements through fact_relationships

(see methods, & figure 1-A) negatively affect the criteria of model interoperability & scalability, query performances and query complexity (due to multiple tables joins).

Leveraging new features of OMOP CDM v5.4 allows explicit links between microbial identification and AST results (see figure 1-B). Our guideline to represent AMR data is presented in figure 1-B. In this representation, a "measurement" is used to capture the lab test and its results (identification or detection). An "observation" captures the name of the identified organism and as many "measurements" as AST tests are linked to the observation (each host both the MIC result & the corresponding category).

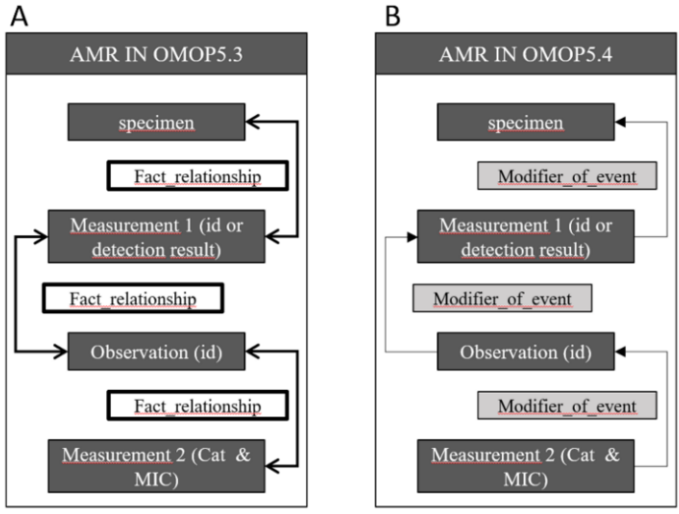


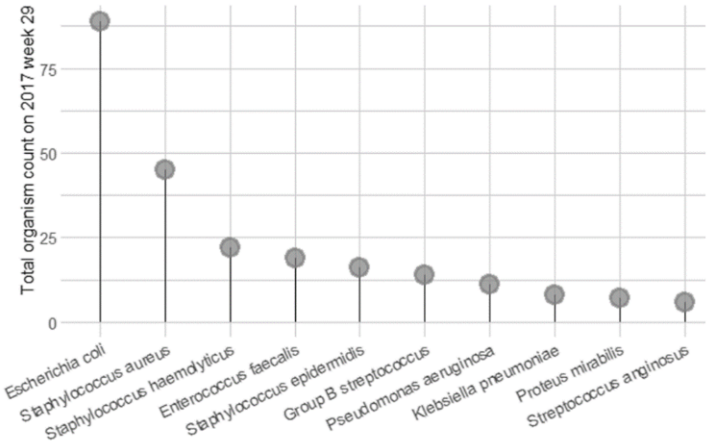
Figure 1: Part A shows modeling in OMOP CDM V5.3 / 6.0 and Part B in OMOP CDM V5.4

3.2. Model evaluation

To assess our model usability and compatibility with OHDSI tools, an implementation of this model has been carried out in a PostgreSQL database with an anonymized dataset (see Methods). Using OHDSI R packages we successfully queried the database and produced visualizations. The example presented in figure 2-A represents the top 10 pathogens identified over a given week in our data set. Data Quality was investigated through 2621 tests carried out with DQD (see figure 2-B). 99% were successful, thus validating the alignment of our model on OMOP standards as well as the plausibility and completeness of our implementation.

The query efficiency formicrobiology data modeled in OMOP CDM 5.4 versus OMOP CDM 5.3 have been compared. A query to obtain the date of a result, the name of the pathogen identified, the name of the antibiotic tested, and the AST status of the organism is five times quicker in our OMOP CDM 5.4 implementation. However, this measure has been performed on a local non optimized setting and will need to be confirmed on a controlled federated architecture.

A



B

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	1893	6	1899	100%	82	0	82	100%	1975	6	1981	100%
Conformance	354	17	371	95%	45	1	46	98%	399	18	417	96%
Completeness	211	6	217	97%	5	1	6	83%	216	7	223	97%
Total	2458	29	2487	99%	132	2	134	99%	2590	31	2621	**99%**

Figure 2: A TOP10 Organisms identified in an example database using OMOP model and OHDSI tools B Summary results of database evaluation using OHDSI DQD tool

4. Discussion

This work not only supports our VALUE-Dx work package objective, but also supports future AMR surveillance networks. Indeed, AMR surveillance is key to informed therapy recommendations, public health interventions, identification of emerging antimicrobial resistance. However, today very few surveillance systems can access data in real-time [5]. Common representation of AMR data across Europe could help surveillance to get up to speed with epidemic process.

In order for our proposed modeling to store every data required for correct AMR surveillance it is paramount that deduplication of the data is achievable. Depending on the organization, the deduplication methods and the data required vary: WHO uses the patient ID, the nature of the sample and the identified pathogen ; whereas the Japanese surveillance system (JANIS) uses the patient ID, the identified pathogen, the AST results and the date of the sample [14]. All this information can be represented in the work presented.

It is also important to be able to differentiate between community-acquired and hospital-acquired infection. This requires the date of admission of the patient to the hospital and more precisely the time between the admission and the collection of the sample. However, this may be poorly represented in a laboratory based system and call for EHR based data, see also below.

Limits of actual surveillance systems have been highlighted by Tacconelli et al. [5] The tendency is for the data to be purely lab-based leading to a lack of patient clinical information. By design OMOP CDM is well equipped to represent clinical outcome, feeding OMOP CDM 5.4 with data originating from both the laboratory and the EHR circumventing these limitations. The modeling proposed here is an option to reach this goal.

5. Conclusions

We propose a model able to represent the complexity of clinical microbiology laboratory data using the newly released OMOP CDM 5.4 [9]. Our work shows that changes in OMOP CDM 5.4, historically pushed by clinical oncology needs [11], benefit clinical microbiology. Compared to OMOP CDM 5.3 it allows for a better data representation and gains in performance.

However additional investigations are needed to confirm those preliminary findings. This should involve a more realistic and larger dataset, additional types of laboratory *in vitro* diagnostics devices, and be carried out in a federated architecture closer to those deployed in clinical settings.

References

[1] VALUE-Dx. VALUE-Dx https://www.value-dx.eu/ [Available from: https://www.value-dx.eu/.
[2] Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022;399(10325):629-55.
[3] Emberger J, Tassone D, Stevens MP, Markley JD. The Current State of Antimicrobial Stewardship: Challenges, Successes, and Future Directions. Current Infectious Disease Reports. 2018;20(9):31.
[4] WHO. Antimicrobial resistance: global report on surveillance. World Health Organization; 2014.
[5] Tacconelli E, Sifakis F, Harbarth S, Schrijver R, van Mourik M, Voss A, et al. Surveillance for control of antimicrobial resistance. Lancet Infect Dis. 2018;18(3):e99-e106.
[6] Maier C, Kapsner LA, Mate S, Prokosch H-U, Kraus S. Patient Cohort Identification on Time Series Data Using the OMOP Common Data Model. Appl Clin Inform. 2021;12(01):057-64.
[7] Gansel X, Mary M, van Belkum A. Semantic data interoperability, digital medicine, and e-health in infectious disease management: a review. Eur J Clin Microbiol Infect Dis. 2019;38(6):1023-34.
[8] Rinaldi E, Thun S. From OpenEHR to FHIR and OMOP Data Model for Microbiology Findings. Stud Health Technol Inform. 2021;281:402-6.
[9] OHDSI. CDM v5.4 https://github.com/OHDSI/CommonDataModel/releases/tag/v5.4.0: OHDSI; 2021 [Available from: https://github.com/OHDSI/CommonDataModel/releases/tag/v5.4.0.
[10] OHDSI. The Book of OHDSI: Observational Health Data Sciences and Informatics; 2021.
[11] Belenkaya R, Gurley MJ, Golozar A, Dymshyts D, Miller RT, Williams AE, et al. Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research. JCO Clin Cancer Inform. 2021;5:12-20.
[12] Le Gall M, Vachon R, Cellière B, Dante M, Holder F, Mary M, et al. Semantic laboratory data interoperability: Encoding microbiology tests and test results using LOINC ® and SNOMED-CT ® from an IVD manufacturer perspective. In: Microbiology ASf, editor. ASM microbe 2018; 2018 June 10; Atlanta: American Society for Microbiology; 2018.
[13] Fournier P-E, Drancourt M, Colson P, Rolain J-M, Scola BL, Raoult D. Modern clinical microbiology: new challenges and solutions. Nature Reviews Microbiology. 2013;11:574-85.
[14] Kajihara T, Yahara K, Stelling J, Eremin SR, Tornimbene B, Thamlikitkul V, et al. Comparison of de-duplication methods used by WHO Global Antimicrobial Resistance Surveillance System (GLASS) and Japan Nosocomial Infections Surveillance (JANIS) in the surveillance of antimicrobial resistance. PLoS One. 2020;15(6):e0228234.