

EPISODE/EPISODE_EVENT-Based Representation of Microbiology Results and Antimicrobial Susceptibility Testing in OMOP CDM v5.4

Jared Houghtaling¹, Polina Talapova^{1,2}, Gilles Clermont³, Andrew E Williams¹

¹ Tufts Clinical and Translational Science Institute, Boston, MA, USA

² SciForce, Ukraine/Estonia

³ University of Pittsburgh, Pittsburgh, PA, USA

Background

Microbiology results are difficult to represent in OMOP CDM because they are generated from specimens and organisms, not only from person-level events. A single specimen may have multiple analyses, multiple organisms, and organism-specific antimicrobial susceptibility testing (AST) results. The modeling problem is therefore not simply where to store a susceptibility value, but how to preserve which organism from which specimen was tested against which in vitro antimicrobial agent, by which method, and with what quantitative or categorical result. Prior work has shown that microbiology data can be mapped to OMOP, but detailed organism-to-test relationships remain difficult to query consistently across sites.^{1,2}

More recent OHDSI work on hierarchical clustering of microbial resistance profiles and ventilation protocols demonstrated that EPISODE and EPISODE_EVENT can be used to organize sample-level microbiology data, microorganism attributes, resistance profiles, quantitative culture values, and ventilator-associated pneumonia(VAP)-related clinical criteria.³ However, that work identified important limitations: the oncology extension was not designed for VAP or antimicrobial resistance (AMR) data, and single-parent hierarchies may be insufficient when microbiology facts have multiple clinically meaningful ancestors. This abstract builds on that prior episode-based approach and proposes a microbiology-specific OMOP CDM v5.4 representation pattern for specimen-to-organism-to-AST context. The central requirement is recoverability: an analyst should be able to reconstruct, from OMOP rows alone, which AST result belongs to which organism, analysis, and specimen without inferring relationships from timing or row proximity.

Methods

We described a proposed OMOP CDM v5.4 representation. The design extends prior hierarchical-cluster logic by separating three levels of representation: specimen cataloging, workflow grouping, and row-level semantic qualification. SPECIMEN records catalog the collected biological sample. EPISODE provides a specimen-level parent anchor and child analysis anchors, allowing microbiology workflows to be grouped without forcing all interpretation through VISIT_OCCURRENCE. Unlike the earlier VAP/AMR implementation that reused oncology episode structures, this proposal treats microbiology episodes as domain-specific anchors that require finalized microbiology EPISODE concepts and conventions. The current draft vocabulary proposal defines MICRO_SPECIMEN_EPISODE (“Microbiology specimen episode”) and MICRO_ANALYSIS_EPISODE (“Microbiology analysis episode”) as required Episode-domain

anchors, with optional MICRO_AST_EPISODE, MICRO_PANEL_EPISODE, and MICRO_WORKFLOW_EPISODE concepts for AST panels, multiplex panels, or higher-level workflows where analytically useful. EPISODE_EVENT links the relevant domain rows to the appropriate episode. Organism identification and organism-level attributes are represented as OBSERVATION rows when they describe the organism or its properties. Standardized tests, organism-identification methods, AST methods or panels, and AST outputs are represented as MEASUREMENT rows when generated by laboratory testing; numeric results such as MIC or zone diameter should be retained as values with units when available, and categorical interpretations such as susceptible, intermediate, resistant, non-susceptible, or susceptible-dose dependent should be retained as coded result values where supported by vocabulary mapping. For example, MIC values are retained quantitatively (e.g., µg/mL) and zone diameters in their native measurement units (e.g., mm), while S/I/R/NS/SDD interpretations remain coded categorical outputs. The antimicrobial, organism, testing method, quantitative value, unit, and interpretation must remain jointly recoverable. Breakpoint standard/version harmonization (e.g., CLSI versus EUCAST) is not yet prescribed by this blueprint and remains an explicit ETL/vocabulary convention to finalize.

Procedure events, when present in the source, remain in PROCEDURE_OCCURRENCE and can be grouped through EPISODE_EVENT. More granular MEASUREMENT and OBSERVATION rows use the v5.4 event-id and event-field concept fields to point upward to the organism, analysis, or resistance/susceptibility attribute they qualify. This creates a two-layer linkage strategy: EPISODE_EVENT preserves workflow membership, while event-id/event-field-concept fields preserve the immediate semantic target of a result or attribute. Operationally, EPISODE_EVENT answers “what belongs to this microbiology workflow?”, whereas event-id/event-field links answer “what specific organism or event does this result qualify?”. Antimicrobial agents tested in vitro are not automatically DRUG_EXPOSURE records; DRUG_EXPOSURE is reserved for drugs prescribed, dispensed, administered, ingested, or otherwise introduced into the patient.⁴ Thus, an antimicrobial can appear as part of an in vitro AST laboratory result without implying patient exposure to that drug.

Results

The proposed hierarchy supports negative cultures, single-organism cultures with AST, polymicrobial cultures, multiple procedures on one specimen, organism toxin or genetic attributes, and analytic groupings such as policy-period or visit-related questions. Worked examples were used to test the representation logic. A negative sputum culture produces a SPECIMEN, specimen/analysis anchors, and a negative MEASUREMENT without creating unsupported organism or AST rows. A single-organism blood culture retains an organism OBSERVATION together with organism-linked AST results. In a polymicrobial urine culture, *E. coli* and *E. faecalis* are represented separately, and their nitrofurantoin and ampicillin MIC/interpretation pairs remain linked to the correct organism rather than being flattened into specimen-level results. For instance, the polymicrobial test case contains an *E. coli* nitrofurantoin MIC of 32 µg/mL (S) and ampicillin MIC of 16 µg/mL (R), alongside an *E. faecalis* nitrofurantoin MIC of 64 µg/mL (I) and ampicillin MIC of 2 µg/mL (S); the proposed linkage pattern preserves all four organism-antimicrobial associations independently within one urine-culture workflow. It also addresses a key limitation of prior hierarchical clustering work by allowing microbiology rows to retain both

workflow-level episode membership and finer-grained organism/test/attribute qualification, rather than relying on a single parent-child path alone. The model preserves specimen-analysis-organism-AST context while avoiding a flattened record that obscures whether an antimicrobial was tested, interpreted, or administered. It also creates explicit targets for quality assurance: a source microbiology row containing specimen, organism, test, method, antimicrobial, and result can be checked against the expected OMOP domain rows and linkages (Figure 1).

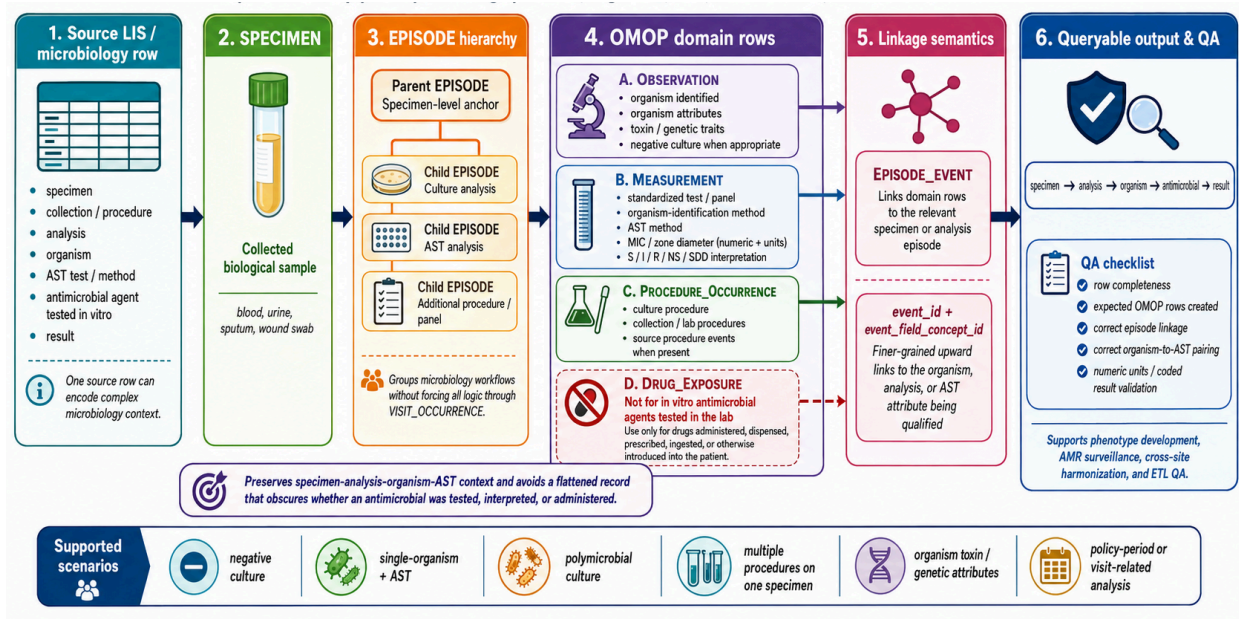


Figure 1. Microbiology and AST representation in OMOP CDM v5.4 via EPISODE_EVENT linkage

Expected QA targets include completeness of SPECIMEN, EPISODE, EPISODE_EVENT, MEASUREMENT, OBSERVATION, and PROCEDURE_OCCURRENCE rows; preservation of organism-to-AST pairing; valid distinction between in vitro antimicrobial testing and DRUG_EXPOSURE; retention of numeric units and coded interpretations; and reproducible reconstruction of the chain specimen → analysis → organism → antimicrobial → result.

Limitations

The current work is a representation blueprint supported by worked synthetic examples, not yet a cross-site ETL validation study. It does not provide performance benchmarks, network-scale row counts, or measured reconstruction accuracy against production LIS data. Draft microbiology EPISODE concepts have not yet been finalized through the OHDSI Standardized Vocabularies governance process, and coverage of organisms, AST methods, antimicrobial concepts, MIC/zone units, categorical interpretations, and breakpoint-standard/version metadata remains to be assessed systematically. The proposal also does not yet define one universal convention for reconciling CLSI, EUCAST, local, or version-specific susceptibility interpretations. Until such conventions are agreed, source method,

quantitative value, unit, categorical interpretation, and available breakpoint provenance should be preserved rather than collapsed into an apparently harmonized result.

Conclusion

An episode-event microbiology pattern in OMOP CDM v5.4 can represent AST as linked specimen, organism, test, antimicrobial-agent, and result facts while preserving standard table semantics. By building on prior OHDSI hierarchical episode work and generalizing it beyond a VAP-specific use case, the approach reframes microbiology representation as a reusable specimen-centered pattern for complex laboratory data. The approach is designed to support phenotype development, antimicrobial-resistance surveillance, quality assurance, and cross-site harmonization. Further work is required to finalize microbiology EPISODE concepts, row-level ETL conventions, vocabulary coverage, and validation queries before claiming network-scale implementation. Future validation should compare source LIS rows with reconstructed OMOP query outputs to test whether organism-specific AST relationships remain recoverable across sites. The next implementation step is therefore not simply to load more microbiology rows, but to demonstrate lossless reconstruction of negative, single-organism, and polymicrobial source cases, and then test whether the same queries recover organism-antimicrobial pairs consistently across institutions.

References

1. Theron E, Gorse JF, Gansel X. Usability of OMOP Common Data Model for Detailed Lab Microbiology Results. *Stud Health Technol Inform*. 2022 May 25;294:292-296. doi: 10.3233/SHTI220461.
2. https://ohdsi-europe.org/images/symposium-2023/posters/houghtaling_hierarchicalposter_2023_V2_-_Jared_Houghtaling.pdf
3. Houghtaling J, Jung F, Krishnan A, Padros Goossens M, Leus F, Maxwell L, et al. Hierarchical clustering of microbial resistance profiles and ventilation protocols using the oncology extension [poster]. Presented at: OHDSI Europe Symposium; 2023 Jul 1-3; Rotterdam, Netherlands. Available from:
4. Rinaldi E, Thun S. From OpenEHR to FHIR and OMOP Data Model for Microbiology Findings. *Stud Health Technol Inform*. 2021 May 27;281:402-406. doi: 10.3233/SHTI210189.